Building Residential VoIP Gateways: A Tutorial
Part Three: Voice Quality Assurance For VoIP Networks

by David Jarrett and Keith Buchanan,
Senior Broadband Applications and VoIP Gateway Product Manager,
Texas Instruments Incorporated

One reason that VoIP technology is becoming more widely deployed, aside from the decreasing cost-per-channel of enabling technologies such as DSPs, is the ability of VoIP systems to match the overall service quality offered by traditional circuit-switched voice networks. Many incumbent local exchange and long-distance service providers use VoIP technology in the backhaul portion of their networks without the end user being aware that VoIP is involved.

These traditional service providers use techniques to manage service quality developed over the last 100+ years for circuit-switched networks -- namely, careful network design, and tracking of customer and network trouble reports. Service providers use well-understood rules to characterize service level in terms of voice quality (based on loss, delay, and echo), and in difficulty in establishing a call. Networks are pre-engineered to offer a certain level of service while taking into account these factors. Then, a service provider’s main tool to assess service quality while the network is in operation is based on trouble reports from users, as well as general network equipment failure notification through Network Management systems.

Voice quality is, in reality, the end user's perception of quality. Network performance characteristics will impact voice quality (as discussed below). Metrics such as Mean Opinion Scores (MOS) measure the subjective perception of voice quality, and analysis tools can be used to derive these metrics.

However, as VoIP technology gets pushed closer to the edge of the network with IP phones (wired and wireless) and residential voice gateways, VoIP service providers have a much more difficult time assuring the voice quality for their subscribers for two important reasons:

1. Lack of control over the underlying transport network -- e.g. when proving voice service from a residential voice gateway attached to another provider’s residential broadband Cable Modem or DSL service
2. Use of transport technology that can vary in quality -- e.g. using WLAN media to transport VoIP, especially when the subscriber is moving

Fortunately the increasing processing power in these edge devices, which has enabled them to support high-quality VoIP service in the first place, will also enable them to directly measure and troubleshoot issues with customer service quality. A service provider can make in-service measurements of the voice quality their end-users experience, and can also use this information to separate problems with their VoIP equipment from those with the underlying transport, and therefore help them more effectively address issues whether they own the entire network or not.

VoIP Network Issues

Factors that impact voice quality in a VoIP network are fairly well understood. While most of these can be mitigated with careful network design, good quality assurance tools both in the VoIP endpoint equipment and the network itself can allow these issues to be addressed with the best balance of effectiveness and cost. The level of control over these factors will vary from network to network. This is highlighted by the differences between a well-managed enterprise network vs. an unmanaged network such as the Internet.

Network operational issues impact network performance and will create conditions that affect voice quality. These issues include:

- Outages/failures of network switches, routers, bridges
- Outages/failures of VoIP elements -- call servers, gateways
- Traffic management during peak periods and virus/DOS attacks.

Scaling to very large networks increases exposure and places more importance on effective planning and implementation. The following are several factors that must be considered when planning, designing and deploying VoIP networks.

Delay -- Caused by processing in the endpoint equipment (and in the network), the collection of voice samples to implement voice compression, and the collection of voice (compressed or uncompressed) into network packets. One-way delays of 400 ms or more will impact the ability to carry on a normal conversation (ITU-T Specification G.114). Delay can be mitigated with efficient VoIP gateway and network design (e.g. prioritizing voice packets to minimize switching and routing delays), but also by selecting the appropriate packet length to lower packetization delay.

Jitter -- Caused by the variation in delay characteristic of packet transport networks. This is best mitigated by adaptive jitter buffer management in the packet receive path, to effectively remove the jitter before the voice samples are played out to the listener.

Packet Loss -- Caused by packet buffer or processor overload in the network or the receive VoIP endpoint, or by packet bit errors. Best mitigated using packet loss concealment techniques as part of the voice compression algorithm to replay previously received voice and/or comfort noise samples until new information can be received.

Echo -- Caused by voice energy "bouncing" off the circuit at an analog PSTN interface (i.e. line to a telephone). Echo that is sufficiently attenuated and/or that is delayed by less than 15 ms will not be noticed. Echo between 15 ms to 35 ms will give the speech a "hollow" sound, while echo delayed more than 50 ms will be distinctly heard and should be cancelled (ITU-T Recommendation G.131). Echo is exacerbated by the additional delay caused by VoIP, typically in the range of 50 ms to 100 ms. Mitigation requires robust echo cancellation solutions in the gateways between VoIP and the PSTN.

Vocoder -- Voice quality is partially affected by the voice vocoder used. While the PSTN uses pulse code modulation (PCM - G.711), VoIP systems widely use low bit-rate vocoders such as G.729. The most commonly-used vocoders have acceptable MOS scores. Wide-band vocoders, such as G.722, can actually support voice quality on an all-IP voice network greater than that of a traditional circuit-switched voice networks.

Voice Activity Detection -- VAD is a popular extension to voice coding schemes that further reduce bandwidth by eliminating packets that contain silence. This sometimes affects call quality by clipping the beginning of a talk burst. This effect can be mitigated by careful tuning of the voice detection algorithm.

Other factors that may affect voice quality include:

- Signal loss and dropouts
- Background noise
- Signal attenuation/gain changes
- Level clipping
- Physical interface (e.g. analog vs. digital T1/E1)

Example Deployment-Related VoIP Quality Issues

New networking technologies and deployment models are causing additional challenges that impact the ability of VoIP service providers to guarantee the highest levels of service quality in their deployments. Two such examples are where the VoIP service provider does not control the underlying packet transport network, and use of packet networks with potentially high delay and loss, such as 802.11 (WLAN) technology.

Example 1:
A number of independent VoIP service providers are entering the market, offering consumer residential voice services at extremely low prices. These providers will provide a home gateway designed to be connected to a broadband internet connection (i.e. DSL or cable modem service), and will operate the infrastructure gateway equipment to connect subscribers to each other and to the PSTN.

These VoIP service providers are typically completely independent of the broadband access providers, so that the gateways will have no interworking with the transport network to allow support for end-to-end QoS. Indeed, since the transport network includes the Internet, there is no way to guarantee any level of packet jitter, loss, or delay. Therefore, more aggressive measures must be taken in the home and infrastructure gateways to mitigate possible degradation due to these effects. In addition, it is critical that these devices also provide robust measurement and troubleshooting tools to allow the service providers to know about, and hopefully localize, quality issues.

Example 2:
Significant progress has been made in the Wi-Fi Alliance and the IEEE 802.11 working group to add QoS-aware features to the WLAN MAC, such as access categories to handle the QoS requirements of voice (and streaming video) applications, and admission control policies to ensure WLAN channels are not oversubscribed. However, the fact remains that WLAN media will have relatively high loss and delay compared to wired Ethernet,

due to:

1. RF interference -- interference from other devices using the WLAN frequency bands (2.4 GHz for 802.11b and g), such as cordless phones and microwaves
2. Changes in the RF path -- e.g. due to moving objects reflecting RF energy, or motion from the end-station itself (e.g. because the user is walking or driving)

Again, it is necessary to have robust diagnostic solutions in the VoWLAN handset and in the overall network to identify voice problems per-call, to enable service providers and network operators to identify and most effectively address problems as they arise.

Voice-Quality Measurement Tools

The ability to capture and report events is critical for managing network performance. These tools must be extended to managing voice quality, allowing operations to identify and correct network problems that impact voice quality. In some cases the cause of the problem may not be determined in real time, requiring off-line analysis. Captured information can be reviewed to determine the root cause.

The oldest and most reliable voice-quality tool is the listening opinion tests where human listeners rate call quality in a controlled setting (ITU-T Specification P.800). Overall results are compiled to produce a mean opinion score (MOS), which is based on a panel of listeners ranking the quality of a series of call samples on a scale of 1 to 5 ("Bad" to "Excellent", respectively). An aggregate score of 4 or more is considered toll (i.e. PSTN) quality. While this test has the disadvantage of being subjective, expensive, and time-consuming to produce, it is recognized as the most consistent measure of voice quality available.

The bulk of subsequent activity in voice quality measurement has been on producing algorithms and tools that can objectively measure voice quality -- i.e. based on direct mathematical calculation on sound samples, rather than listening tests. Such tests can be roughly classified as active (or intrusive) and passive (or non-intrusive). In general, active tests perform calculations on test or simulated calls and thus intrude on normal network usage (or are conducted in lab environments), while passive tests can perform calculations on active calls in live networks without any interruption of service. The following will explore the relevant tests in the categories further.

Active/Intrusive Tests

As described in various white papers [References 1, 2], a wide range of research into automated, objected voice quality testing led to the development of a number of algorithms based on perceptual modeling. The most widely used of these are:

- PSQM – ITU-T P.861 http://www.itu.int -- perceptual speech quality measurement; automated scoring system, design for circuit switched network

- PAMS http://www.psytechnics.com/downloads/pams/PAMS_white_paper.pdf -- perceptual analysis and measurement system; Intrusive speech quality assessment tool; end-to-end degradation analysis of injected signal
- PESQ- ITU–T P.862 http://www.pesq.org/ -- international standard for measuring end-to-end voice quality according to models of the human preception -- recent standard for assessing voice quality; leverages the best of PSQM and PAMS algorithms; supports voice encoding, jitter, packet loss, time-clipping and channel errors

From Ref. [2]: "Techniques are based on psycho-acoustic science, and use a common approach in which a sample of voice is input into a network, and the subsequent output is recorded. The output sample is then compared to the input sample to produce a score that represents how well (or poorly) the network reproduced at the output the original speech. The two key features of these techniques are that the input and output signal are both modeled in a "perceptual" domain first, and then the comparison determines audio-perceptual distances or disturbances as a human would perceive them. The objective of each technique is to produce scores, like MOS, that reliably predict the results of subjective tests."

Much analysis has been done on the relative merits of these (and other) techniques. Suffice it to say that while these algorithms have evolved over time to better model more situations that may arise in packet-based networks (e.g. packet loss, variable delay), PESQ was designed to combine the best aspects of the previous ones, and is recognized as providing the highest degree of correlation to subjective MOS testing.

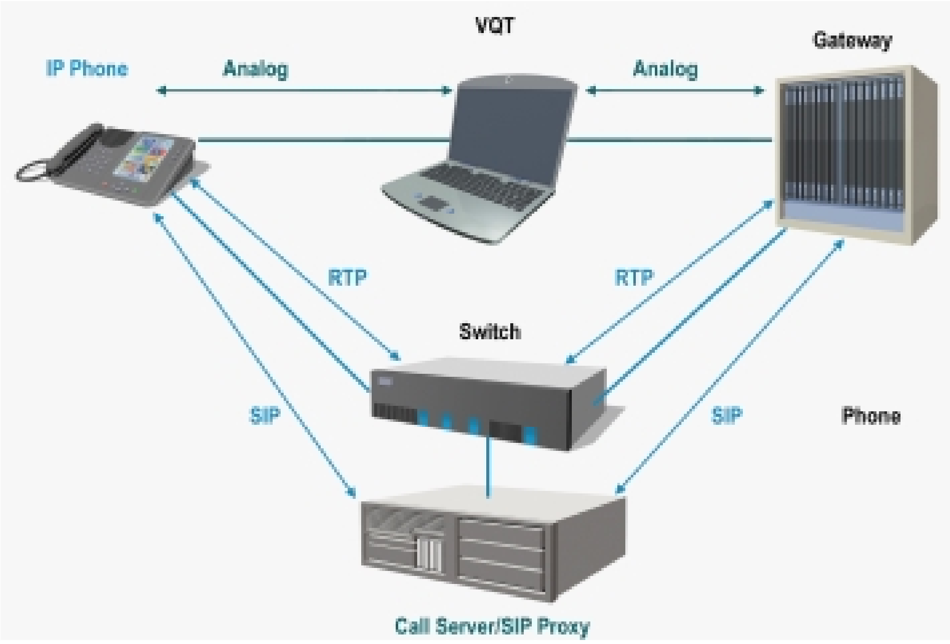

Fig. 1: Active/Intrusive Monitoring; Passive/Non-Intrusive Tests

Equipment from various manufacturers implementing these algorithms is widely used to test the quality of VoIP implementations at the component level and at the system level (Fig. 1). However, it is costly to use such equipment to measure the performance of active networks, since revenue-producing traffic must be interrupted to use it. In addition, while these algorithms can quantify deficiencies in speech quality, they do not produce information to help localize and identify the root causes of the situations causing the deficiency.

Passive tests, on the other hand, run in live networks without interrupting active calls and often use statistics gathered on active calls; they are therefore actually embedded into the VoIP equipment at the use site and in the VoIP service provider's network. As such, passive testing can therefore be used at lower cost as it eliminates both interruption to revenue-producing traffic and additional dedicated test equipment equipment.

Many tools used here are based on the E-model, as described in ITU-T Recommendations G.107 and G.108. The E-model is a transmission planning tool meant to account for a number of real-world factors to predict the performance of a network. The E-model calculates a transmission rating factor, R, calculated as:

$$R = Ro - Is - Id - Ie + A$$

where, *Ro* is SNR, including circuit and room noise;
Is is an impairment combination that occurs simultaneous to speech, including too-low send/receive loudness, non-optimal sidetone, and quantization distortion;
Id is a combination of impairments from delay, including talker echo, listener echo, and absolute delay;
Ie is an equipment impairment factor due to low bit-rate vocoders;
A is an advantage factor that accounts for the added convenience of different types of access. For example, mobile telephony has a higher advantage factor than wired telephony.

The ITU specs also describe how the R factor can be related to MOS.

Various vendors have adapted aspects of the E-model to support real-time calculation of call quality based on information about a call (e.g. jitter, packet loss, vocoder used). These calculations take minimal processing resources and can be combined into the overall VoIP DSP software load. In this way, they can perform the measurements on active calls through the gateways, on a per-channel basis (Fig. 2). Also, these measurements will be single-ended such that they don't depend on collecting end-to-end information about each call.

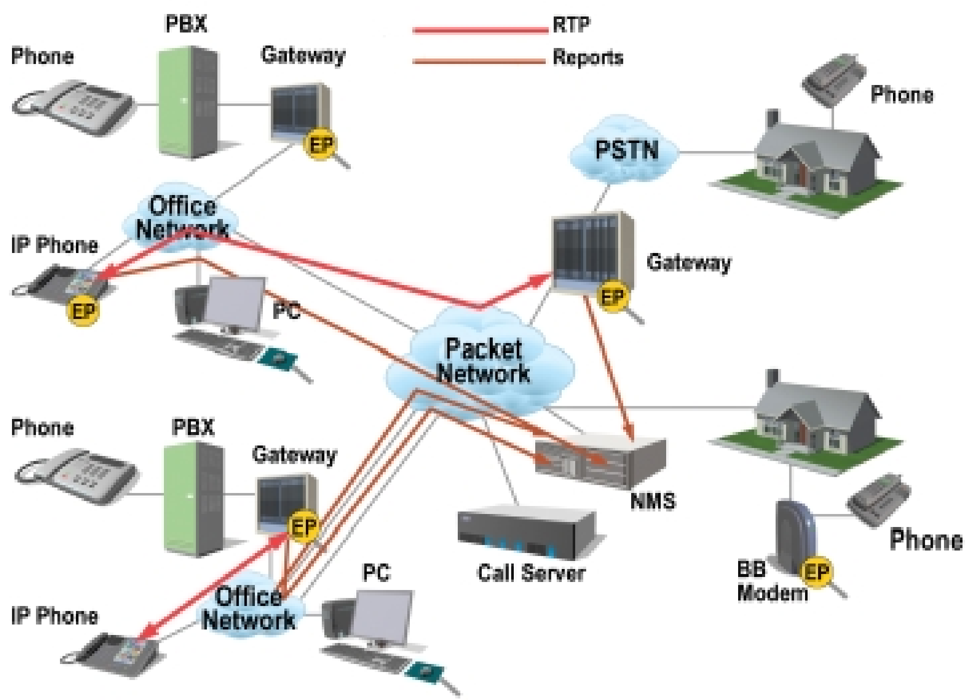

Fig. 2: Passive/Non-Intrusive Monitoring

While such uses of the E-model to perform passive measurements are not standardized, the ITU-T has recently standardized single-ended quality assessment measurements in ITU-T P.563.

Practical VQM Solution Aspects

Having passive measurements embedded in the VoIP equipment represents just the first piece of an overall solution for comprehensive voice quality monitoring. Once measurements are made in the VoIP gateway, this information must be reported to a Network Management system, where it can be used for problem detection and isolation. This and other information gathered from the network and the VoIP gateways can also be used for off-line analysis to diagnose the root cause of problems, so that they can be addressed in tweaks to the network or gateway configuration. The following will outline current work in these areas.

Reporting: RTCP XR

The Real Time Protocol (RTP; IETF RFC3550; http://www.ietf.org/rfc/rfc3550.txt) is the IETF standard for the transport of real-time data, including voice and video, in packet networks (including IP). It includes both a data part and a control part; the latter is Real Time Control Protocol (RTCP). RTCP offers general feedback of quality information in the context of a multicast group, as well as information to allow synchronization of multimedia streams. The Extended Report extension to RTCP (RFC3611;

http://www.ietf.org/rfc/rfc3611.txt) defines a format to transport information gathered in VoIP gateways in a standard and interoperable format.

Baseline RTCP includes sender and receiver reports that include some basic information (such as jitter, total packet count, and packets lost) about each call. RTCP XR defines a format to send report blocks that can be used for much more detailed information about RTP sessions. RFC3611 defines seven such report blocks; two are defined to carry summary metrics that are useful VoIP quality measurements:

1. Statistic Summary Report Block -- lost packets, duplicate packets, minimum, maximum, mean, and standard deviation jitter measurements, and packet TTL or Hop limit values by time interval (as defined by start and stop packet sequence number)
2. VoIP Metrics Block -- five categories of information including:
    - Packet loss/discard statistics -- loss rate, discard rate, burst metrics (burst density, gap density, burst duration, and gap duration)
    - Delay -- round-trip and end-system
    - Signal metrics -- signal level, noise level, residual echo return loss
    - Call quality metrics -- R Factor, listening quality MOS estimate, and conversational quality MOS estimate
    - Configuration parameters -- thresholds used, use of packet loss concealment, and use of adaptive jitter buffer

In addition, the specification defines a framework by which other implementation specific reports can be defined.

Having this information available, per VoIP endpoint, is clearly of great benefit in identifying potential trouble areas for entire networks, or even individual users.

Summary

Comprehensive QOS monitoring and management is required for VoIP services. Networks that are poorly implemented will adversely affect the end user experience and impact the broad acceptance of VoIP as a viable alternative to traditional telephony. Techniques and standards exist today for measuring and monitoring voice quality within the VoIP network elements. These tools must be included as the VoIP networks are deployed.

References

[1] Anderson, John, “Methods for Measuring Perceptual Speech Quality - White paper,” August, 2002
[2] Anderson, John, “Addressing VoIP Speech Quality with Non-intrusive Measurements,”

About The Authors

David Jarrett provides field applications support for customer definition and design activities using TI's VoIP products. He earned his BS in Electrical Engineering from Kansas State University, and MS in Electrical Engineering from the University of Arizona. djarrett@ti.com

Keith Buchanan is product manager for TI's VoIP gateway product line. Keith has more than 20 years of experience in information technology and telecommunication and held system engineering and marketing positions with IBM and Ericsson before joining TI. He earned his engineering degree from Virginia Polytechnic Institute and State University. kbuchanan@ti.com

as published in...
analogZONE