UNIFIED COMMUNICATIONS STRATEGIES

Critical Success Factors in Design and Performance Management of UC Networks

March, 2008

Executive Summary

The purpose of white paper is to provide the Business Decision Maker (BDM) with a non-technical understanding of those factors that are critical to the successful performance of a Unified Communications (UC) solution with a focus on the voice application. Voice is the most widely implemented UC application to date.

Today the competitive tension between application vendors, network vendors and a new wave of UC vendors has introduced a polarization of opinion on whether - network management with attention to codec selection and Quality of Service (QoS), which concerns measurement of the treatment of the packets traversing a network including utilization, response time, latency (delays), delay variation, packet loss, jitter and availability, or application performance management with its focus on the unique VoIP Quality of Experience (QoE) requirements associated with differing business scenarios - is the key to reliable operational performance. The truth of course is that both these aspects must be appropriately managed. And there is no one size fits all.

The design and implementation of a UC solution must be targeted to each enterprise's unique needs if an optimal tradeoff of business requirements for enhancing individual and business process productivity, interoperability with legacy infrastructure and total cost of ownership is to be achieved. In achieving this balance the BDM must recognize that the productivity enhancing benefits of UC applications (e.g. UM, presence, integration with other business applications and simplicity of use) may be accompanied by performance considerations around bandwidth utilization and higher delay. However, with proper network design, attention to codec selection, QoS implementation in the network, and comprehensive performance management tools, it is possible to implement UC successfully.

UNIFIED COMMUNICATIONS STRATEGIES

Critical Success Factors in Design and Performance Management of UC Networks

1. Introduction

Interest in and adoption of Unified Communications (UC) by businesses of all sizes is on the rise as they opt for the unique opportunity afforded by UC to integrate multiple forms of communications, both real time and asynchronous - such as voice, video and conferencing, presence and messaging, as software applications - into their value chain processes; thereby deriving significant competitive advantage from more efficient operations both internally and across their supply chain, as well as enabling markedly improved customer intimacy.

Companies must address significant network design and communications application performance management concerns before the reality of UC meets the vision. Why? Real time applications such as VoIP drive more real-time traffic onto networks. These real time streams are intermixed with non-real-time traffic such as, email or other back office applications. Now, VoIP is extremely bandwidth and delay sensitive. From an IP-network perspective: for VoIP transmissions to be intelligible to the receiver, voice packets should not be dropped, re-ordered, excessively delayed, or suffer varying delay (otherwise known as jitter). To ensure voice quality, current best practice is to classify data and voice traffic into different categories and give voice traffic priority handling across a shared data network backbone. Giving voice traffic priority handling minimizes delays and drops, and whenever possible, gives voice traffic predictable transmission performance. In other words, the underlying converged IP network provides the foundation for implementing intelligent services, Quality of Service (QoS) and security as well as resilience and connectivity. From a voice application perspective the other factors determining voice performance, such as echo, delay, speech-level, noise-level and speech distortion, must also be maintained within required operational boundaries. From an end-user perspective both the network and voice application performance must be adequate to deliver satisfactory Quality of Experience (QoE).

The goal in managing a converged network is to tune it so that many types of application data traffic can coexist and perform within business tolerances specific to the scenarios in which the applications are used. QoS mechanisms are necessary, as well as visibility into the underlying factors that affect end-user QoE.

The purpose of this white paper is to provide the Business Decision Maker (BDM) with a non-technical understanding of those factors that are critical to the successful performance of a UC solution with a focus on the voice application.[1] Voice is the most widely implemented UC application to date. Today the competitive tension between application vendors, network vendors and a new wave of UC vendors has introduced a polarization of opinion on whether network or application management is the key to reliable operational performance. The truth of course is that both these aspects must be appropriately managed. And there is no one size fits all. However, as will be discussed below, with proper network design, attention to codec selection, QoS implementation in the network, and comprehensive performance management tools, it is certainly possible to implement UC successfully.

[1] See the glossary at the end of this white paper for useful definitions.

2. Operational Requirements

Voice performance is a critical issue for almost every business. Whether it is for internal communication or interaction with customers there is a minimum level of performance that must be achieved to ensure that productivity, business relationships, transaction rates and well-being are not affected. Business workers rely on the telephone many hours out of the day, from collaborating with business partners and co-workers to interacting with and helping customers and suppliers. Contact center agents spend the entire day on their telephones. Hence the first step to insuring an appropriately designed and managed UC system is to determine your business requirements. Key questions are: How will the system be used? How many calls per month (or day) are made out of your office? Are those calls to customers or internal employees? How many offices will you have on a system? Are there remote offices, mobile workers, or home workers to consider? How widely dispersed are they geographically?

Different business scenarios have different performance requirements. Consider the following examples in Table 1 where, in each case, there is a level of performance below which the call quality would be viewed as unsatisfactory. Note, however that high fidelity telephony with a minimum MOS[2] requirement of 4.3[3] is set to become common place with the introduction of UC. This will raise quality expectations in end users and reduce their tolerance for poor quality calls.

Table 1: Voice Performance Requirements by Business Scenario

Scenario	Voice performance requirement	Approximate min MOS requirement	Notes on requirement
A customer calling their bank	high	4.0	Voice quality partly defines the customer's perception of the interaction. Customers become extremely stressed if it requires high effort to understand important and often unfamiliar information
A sales person working toward deal closure	high	4.0	It is well known that sales people will deliberately choose a fixed line over a mobile connection to discuss transactions with customers. Better speech quality (often described as a clearer line) helps the parties to expend minimum effort hearing/understanding the dialogue
A call center agent (typical 8 hour shift)	high	4.0	High voice quality essential to maintain the well-being of the agent and control staff churn etc
An occasional user of making an internal call	medium	3.5	Occasional user will not be stressed by lower quality and interaction time is less critical
A mobile call	medium	3.25	The special utility provided by mobile phones is so high that people are prepared to accept incomplete coverage and low voice quality to have this utility. As mobile

[2] Mean Opinion Score (MOS) provides a numerical indication of the perceived quality of received media after compression and/or transmission. The MOS is expressed as a single number in the range 1 to 5, where 1 is lowest perceived quality, and 5 is the highest perceived quality.
[3] Based on wideband (7 kHz acoustic bandwidth) opinion scale.

			network performance continues to improve several major mobile service providers are taking great strides to improve quality and market themselves on this basis. However, Over time users may grow less tolerant of lower quality.
Long distance internet call	low	2.2	Almost complete cost avoidance makes users tolerant of almost unusable (poor) quality. This situation only remains true while the cost differentiation is sufficiently great and the operational/business risk is tolerable.

2.1 VoIP Performance Management Architecture

While the performance requirement may vary between operational situations, there is always a performance level that must be met from a business operational view point. Efficient delivery of this performance level requires the capability to monitor and manage ***both*** network and application performance.

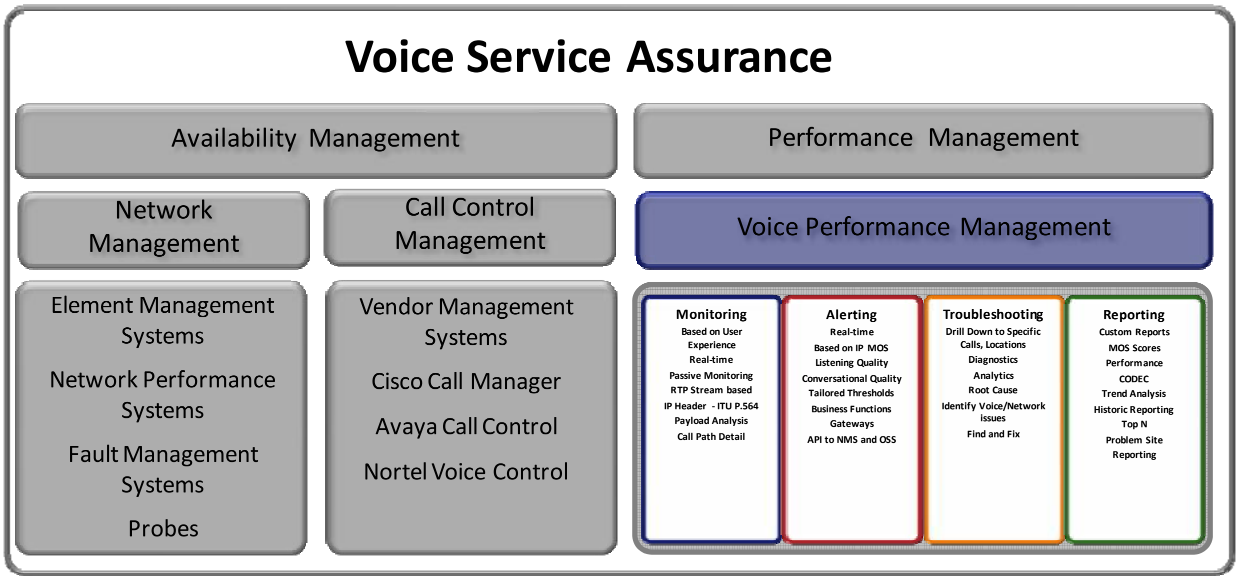
For the traditional packet transport network there's a whole ecosystem around data applications. On the Service Assurance side you have HP's OpenView, Tivoli's NetCool, Computer Associates' eHealth, etc. This operations support system infrastructure allows the network administrator an excellent vantage point from which to assess whether data applications are going to run adequately on the IP network. However, assessing voice application performance requires that you have additional information at your fingertips. You need to know, on a per session basis, whether the IP transport is service-affecting for that session, implying a need to know things like the packet loss and jitter. You also need to know how these packet behaviors are distributed in time and how much power the user device has to error correct in order to know whether that particular network connection is going to affect the quality or not.

You will also need to know several new things about the voice applications themselves: Was it noisy? Was there echo? Was the speech level sensible or was it too faint or too loud? Was the speech distorted? These are the things that you must know in order to ascertain whether the people involved were having a decent QoE or not. But these are things you can ***never*** learn by measuring just the packet transport. It's not that the IP data network tools out there today do a bad job of measuring those things. They don't measure them at all! And that leads to situations where you have unhappy employees and customers not able to converse successfully, while there is a solid row of green lights on the traditional IP network management tools.

Consider this VoIP real comment from a deployment in the financial services sector. "When the IPT system was only part deployed, we became very aware of the difference between quality of experience and quality of service. A small proportion of our call center staff were not satisfied with the call quality when using the IP telephony service, due to a number of factors including varying volume levels, or an echo or hiss on the line … We purchased numerous quality of service tools, all of which indicated that our IP telephony system was working well. None of the tools could explain the mystery behind the small number of call quality issues that we were experiencing. We needed a tool to identify phone call experiences of customers as well as check the network infrastructure was working."

The answer is an integrated voice service assurance as depicted in Figure 1, below.

Figure 1: Voice Service Assurance Architecture



One notable candidate for the missing voice performance management component is Experience Manager developed by Psytechnics[4]. Experience Manager is a performance management solution which can both measure the things you need to know about the IP network, but also can look inside the packets at the pair of wave forms for the two halves of the call. The measurements can be made either via embedded software at the endpoints or via a midpoint probe. The waveform analysis actually determines whether the call was noisy, had echo, whether speech levels were sensible, distorted and so on; providing that complete picture about whether you're achieving satisfactory performance with your real time voice communications applications. With Experience Manager you can set operationally relevant thresholds for different operational scenarios. If Experience Manager detects that you're not meeting your operational thresholds an alarm is raised. From that alarm you can go straight through to trouble shooting. The rest of product is all about drilling down to find specific faults, locations – detail around what's gone wrong to inform a very efficient find and fix process.

Another key requirement met by Psytechnics' Experience Manager is the ability to provide performance management across hybrid (multi-vendor) networks. Hybrid networks will be the norm in UC deployments as businesses combine multiple levels of UC application, access networks and core providers. A typical hybrid network would be a combination of Microsoft Office Communications Server 2007 (OCS) and Cisco Unified Communications and Collaboration applications. Experience Manager is

[4] Psytechnics has supplied quality measurement software solutions to the communications industry worldwide. A spin-off company from BT in 2000, Psytechnics has intellectual property rights in 6 of the ITU standards in the voice performance area.

vendor independent integrating seamlessly with the left side of Figure 1. It can readily be applied across different network technologies (including wireless), is integrated with OCS (with OCS end-point data visible via the Experience Manager UI) and provides directly comparable standards based metrics for fault location and supplier management.

Other vendors also address these issues although, primarily, from a network level view of performance. For example, Cisco Unified Operations Manager (CUOM) provides visibility into the network infrastructure, its performance, the applications used across the network, and the end points. CUOM uses open interfaces and numerous types of diagnostic tests to continuously monitor and evaluate the current status of both the UC infrastructure and the underlying transport infrastructure of the network. Key service and voice quality metrics include utilization, response time, latency (delays), packet loss, jitter, availability and MOS (as achieved by the IP-Network). Cisco Unified Operations Manager does not deploy any agent software on the devices being monitored and thus is non-disruptive to system operations. Information presented by a series of 4 dashboards (Service Level View, Alerts and Events, Service Quality Alerts and IP Phone Status) provides the network manager with a comprehensive view of the UC infrastructure and its current operational status.

There are numerous practical causes of IP network and application performance issues in VoIP deployments. A few examples of each appear in Table 2.

Table 2: Practical Causes of IP Network and Application Performance Issues

Type of issue	Cause and Effect
IP network problems	
Packet loss and jitter	LAN congestion Leading to speech loss/distortion
Out of sequence packets, packet loss	Diverse routing Leading to speech loss/distortion
Delay, packet loss and jitter	WAN QoS misconfigured Leading to speech loss/distortion
Application level problems	
Volume Levels (Loudness)	Gateway pads, incorrect/faulty terminals, PBX faults, peering
Noise	Noise floor from automatic gain control, faulty terminals/PBX/DSP, transcoding
Echo	Poor edge devices (headsets, phones), faulty/under-provisioned echo cancellers
Speech distortion	Poor/faulty edge devices (PC sound cards), transcoding, gateway/PBX DSP
Delay	Low cost routing, failure of anti-tromboning[5], excessive coding stages (i.e., transcoding)

[5] Anti-tromboning is the term used to describe the local re-routing of media streams when a call is transferred over large geographical distances.

3. Hand-in-hand Partnership of QoS and QoE

A well engineered and managed converged network is necessary though not sufficient to achieve operational voice performance. The goal in managing a converged network is to tune it so that many types of application data traffic can coexist and perform well. Good QoS policies must be in place to give priority to VoIP traffic over the TCP applications, which aren't very delay-sensitive. The QoS for VoIP is mainly affected by latency, jitter (delay variation) and packet loss. It's bad business to have your voice traffic burdened by an excess of any of these effects.

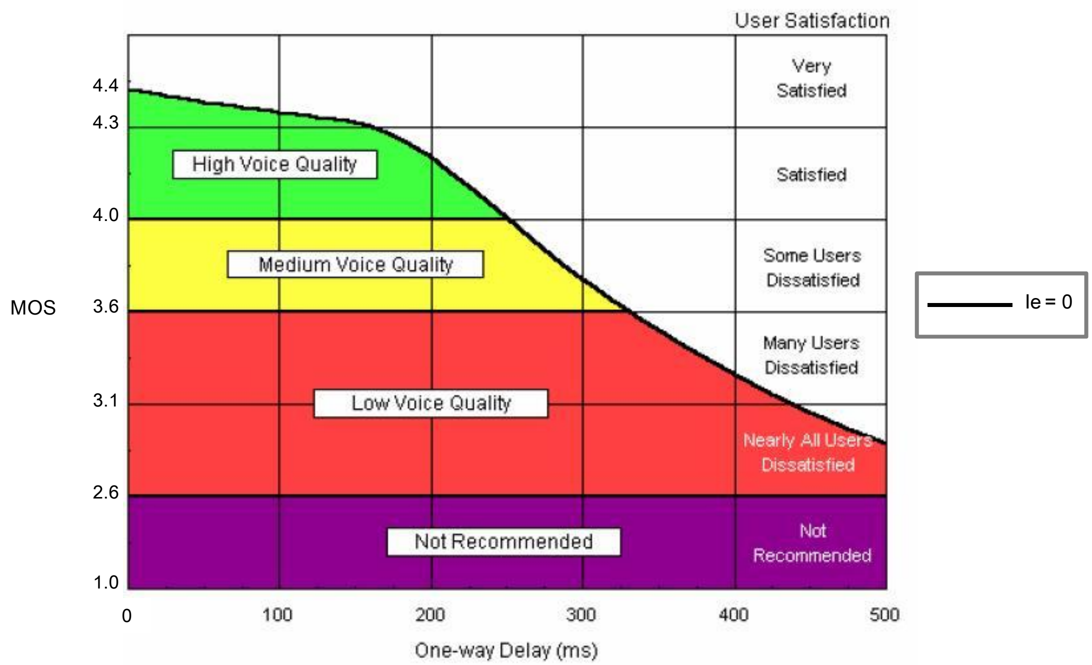
Latency, or delay, refers to the time it takes for a voice transmission to go from its source to its destination. As latency increases, it causes call participants to start interrupting each other because they believe the other person is finished speaking. A VoIP packet may be delayed for several reasons:

- Codec delays occur when the speech is encoded by the codec and the packet is created. Modern speech codecs operate on collections of speech samples known as frames. Each block of input speech samples is processed into a compressed frame. The coded speech frame is not generated until all speech samples in the input block have been collected by the encoder. In addition, many coders also look into the succeeding frame to improve compression efficiency. The length of this advance look is known as the look-ahead time of the coder.
- Transport delay includes time to transmit the packets, buffer and queue them if needed, and move the packets from hop to hop through the network. Corporate QoS-enabled IP networks use equipment with only about 25 to 100 microseconds of delay per hop. Without QoS, transport delay can be variable and high with congestion.
- Jitter buffers (used to compensate for varying delay) further add to the end-to-end delay, about 10 to 20 milliseconds (ms), and are usually only effective on delay variations less than 100 ms. Jitter must therefore be minimized.
- Transcoding delay occurs when one packet format is converted into another, or is formatted to cross the PSTN, like a current majority of inter-enterprise calls. Such delays may be significant (several ms or more) across a multi-hop call as multiple transcodings must be completed for end-to-end transmission.
- Propagation delay, the time taken for a signal to travel along its path from caller to called party can become relevant for off-shore call centers, cross-continent traffic, etc. For example, electrons travel through copper or fiber at approximately 125,000 miles per second implying that a fiber network stretching halfway around the world (13, 000 miles) induces a one-way delay of about 70 ms.

When these factors are combined, delay can easily become significant and there is a loss of synchronicity in the conversation and normal turn taking rules start to break down. Naturally the human participants interpret excessive delay as a hesitation to reply rather than the normal operation of codecs, jitter buffers and propagation delay. As a result, minimizing the overall delay is an important consideration for maximizing the experience of interactive VoIP calls. The International Telecommunications Union-Telecommunication standardization (ITU-T) recommendation G.114 establishes a number of time constraints on one-way latency. The upper bound for domestic calls is 150 ms for one-way traffic. However, this time constraint limits the amount of security that can be added to a VoIP network. Latency should not exceed 100 ms one way for toll-quality[6] voice and must not exceed 150 ms one way for acceptable quality voice. At 150 ms, delays are noticeable by the human ear, but callers can still carry on a normal, comfortable conversation.

The impact of latency on perceived voice quality can be visualized through use of the E-Model as demonstrated below in Figure 2[7]. Of course this simplified view does not represent the psychoacoustic interaction of delay with other performance factors such as echo, speech-level and noise-level.

Figure 2: VoIP Quality Categories



The current methods recommended by the ITU-T for operational monitoring VoIP voice quality are based on per-call measurements that reflect actual customer experience as described below in Table 3. Operationally useful voice performance management must be based on measurements that are accurate for individual calls. The E-Model is not a true psycho-physical model, and cannot be used to predict the opinion of an individual user. However, the E-Model is suitable for transmission planning purposes where it can model the average performance across large call volumes

[6] **Voice Quality Recommendations for IP Telephony,** Telecommunications Industry Association, TSB-116-A, p.21, March 2006. See also: M. Coluccio, **VoIP: Are the Networks Ready for Prime Time?** Connections Magazine, June 2005 http://www.connectionsmagazine.com/articles/5/049.html.

[7] Adapted from Figures 21-21 of **TIA Telecommunications Systems Bulletin TSB-116-A,** March 2006.

Table 3: ITU-T Operational Monitoring Standards for VoIP Voice Quality

Performance aspect	Description	Standard
IP-Network	Relationship between packet network statistics and the impact of the IP-Network on end-user perception of quality on a per call, per edge device basis	ITU-T P.564
	WB[8] IP-Network performance	ITU-T P.564/Annex B
Conversational performance	Combination of echo, delay, noise-level and speech-level to provide a per session measurement of conversational performance.	ITU-T P.562
	WB version of above	ITU-T P.562
	Underlying measurements for P.562 metric: echo, delay, noise-level, speech-level	ITU-T P.561
	WB	ITU-T P.561
Listening quality	NB[9]	ITU-T P.563
	WB	ITU-T P.563

Jitter can cause strange sound artifacts to contaminate the voice and users will complain of degraded voice quality. Jitter has many sources: network congestion, queuing methods used in routers and switches, or network routing policies such as traffic engineering or Multiprotocol Label Switching (MPLS) paths used by carriers. Introducing jitter in audio stream is more detrimental to QoE than the actual delays themselves. Jitter is often caused by low bandwidth.

Packet loss is caused by congestion, poor line quality and distance. If the delivery time of a packet exceeds the length of the receive buffer, then this packet "arrives too late" with respect to its intended play-out time, and will be discarded. Since IP telephony is a real-time voice service that uses the Real Time Protocol (RTP) running over User Datagram Protocol (UDP), there's no way to recover lost packets. Even with less than 150 ms of latency, a packet loss of 5% causes VoIP traffic encoded with G.711[10] to drop below the QoE levels of the Public Switched Telephone Network (PSTN), even with a packet loss concealment scheme. If even one or two percent of VoIP packets drop, voice quality degrades. Consequently, the speech carried in this packet is lost for the decoding process. This "packet loss" impairs speech transmission quality.[11]

As a reminder of the importance of both sides of the QoS/QoE coin, we point out that Table 2 provided examples of practical causes of both application level and IP network performance issues. Network QoS metrics cannot not fully address the application level issues in the second half of the Table 2. Effective QoE assessment and voice performance management requires additional application information on echo, delay, noise-level and speech-level/speech distortion.

[8] WB = Wideband (acoustic band width ~100 Hz to ~8 kHz)
[9] NB = Narrowband (acoustic band width ~300 Hz to ~3.5 kHz)
[10] G.711, also known as Pulse Code Modulation (PCM), is a very commonly used waveform codec.
[11] See ITU-T Rec. G.113, Series **G: Transmission Systems and Media, Digital Systems and Networks**

3.1 Packet Classification Overview

Resources in an IP network are shared by many different users. Therefore, in order to guarantee bandwidth for VoIP packets, a QoS scheme that provides prioritization for VoIP traffic is required for every VoIP deployment. Network devices use the source and destination IP address in the IP header or the source and destination User Datagram Protocol (UDP) port numbers in the UDP header to identify VoIP packets. This identification and grouping process is called *classification* and it is the basis for providing Network QoS.

The first architectural approach to providing end-to-end QoS required that the application signal its QoS resource requirements (such as bandwidth and guaranteed delay) to the network. In a VoIP scenario, this architectural approach meant that either the IP telephone or voice gateway needed to make QoS requests to every hop in the network so that end-to-end resources would be allocated. Every hop needed to maintain call state information to determine when to release the QoS resources for other calls and applications, and if enough resources were available, to accept calls with QoS guarantees. This method is called the Integrated Services QoS model. The most common implementation of Integrated Services uses Resource Reservation Protocol (RSVP). RSVP has some advantages, such as Call Admission Control (CAC), where a call can be rerouted by sending an appropriate signal to the originator if the network does not have the QoS resources available to support it. However, RSVP is relatively complex and therefore it is not widely implemented. Two key alternatives to RSVP include the differentiated services architecture and virtual LANs (VLANs).

Differentiated Services or DiffServ is a widely deployed and supported QoS model today. It can be used to provide low-latency, guaranteed service to critical network traffic such as voice or video while providing simple best effort traffic guarantees to non-critical services. A group of routers that implement common, administratively defined DiffServ policies are referred to as a DiffServ Domain. Network traffic entering a DiffServ domain is subjected to classification and conditioning. Traffic may be classified by many different parameters, such as source address, destination address or traffic type and assigned to a specific traffic class. Scalability comes from the fact that packets are classified at the edges of the DiffServ domain and marked appropriately so that the routers can provide QoS based simply on the DiffServ class. DiffServ requires no advance call setup, no reservation, and no time-consuming end-to-end negotiation for each flow, as with integrated services. This leads to relatively easy implementation.

Alternatively, voice may be assigned to a separate VLAN (or separate VLANs). This segregates traffic for improved performance and security. 802.1p prioritization is a specification that gives Layer 2 VLAN Ethernet switches the ability to identify and prioritize traffic by controlling the outbound port queue priority for traffic leaving the switch, and (if traffic exits through a VLAN-tagged port) sending the priority setting with the individual packets to the downstream devices. Eight classes are defined by 802.1p, which uses the priority fields within the packet's VLAN header to signal the switch of the priority-handling requirements.

3.2 VoIP Bandwidth Considerations

VoIP is a bandwidth- hungry, delay-sensitive application. There are two primary strategies for improving IP network performance for voice: allocate more VoIP bandwidth or implement QoS to ensure voice packets get through and deliver an IP-Network QoS commensurate with your business requirements, as discussed above. Of course application performance must also satisfy QoE needs. The amount of bandwidth to allocate (loaded bandwidth) depends on the packet size for voice (10 to 320 bytes of digital voice payload), codec used and header compression (RTP + UDP +IP). While the RTP header is required to support the real-time nature of the protocol, the accumulation of RTP headers can add a lot of overhead, especially considering the relatively small size of VoIP codec payloads. The protocol packet overhead in bytes is:

RTP[1] (12 bytes) + UDP (8 bytes) + IP (20 bytes) = 40 bytes of overhead per packet.

When you add in layer 2 headers dependent on the physical media, such as point-to-point protocol (PPP), Frame Relay and Ethernet, this overhead increases even more as shown in Table 4, below.

Table 4: VoIP – Per Call Bandwidth

Layer 2 Protocol	**Bytes** (1 byte = 8 bits)
Point-to-Point Protocol (PPP) or Frame Relay (FR) header	6
End-of-frame flag on PPP and FR frames	1
Ethernet L2 headers, including 4 bytes of Frame Check Sequence (FCS) or Cyclic Redundancy Check (CRC)	18

Loaded bandwidth is calculated as follows:

$$RTP + UDP + IP + Layer\ 2\ packet\ Overhead = Overhead\ bytes\ per\ packet$$

$$\frac{Overhead\ bytes\ per\ packet}{Payload\ bytes\ per\ packet} = Voice\ overhead\ ratio\ per\ packet$$

$$(1 + Voice\ overhead\ ratio) \times Payload\ data\ rate = Loaded\ bandwidth\ in\ Kbps$$

By way of example we calculate loaded bandwidth for codec G.711 for a voice payload of 20 ms.

Payload bytes per packet = 160 (64,000/8 x 0.02)
Overhead bytes per packet = 40 (RTP + UDP +IP)
Voice Overhead ratio = 0.25 (40/160)
Loaded bandwidth (Kbps) = 80 (1 + 0.25) x 64

It is imperative to minimize transcoding to maintain voice quality. This requirement can be addressed by standardizing on a particular codec within the VoIP network, or by deploying end-points that negotiate a consistent codec with the corresponding end-point. Voice quality as measured by MOS reduces rapidly with each time a voice conversation is re-processed by an additional codec such as can occur when

[1] Compressed Real-Time Protocol (cRTP) reduces the IP/UDP/RTP headers to 2 or 4bytes (unavailable over Ethernet).

calling across a third party WAN This direct digital-to-digital conversion from one codec to another without returning the signals to analog form generally introduces delay and compression artifacts (i.e., distortion) into the audio coding which typically show up as ringing, pre-echo, drop-outs, warbling, metallic/robot voice, an underwater feeling, hissing, or graininess. Transcoding occurs during the conversion between PSTN circuit-switched and LAN/WAN (Frame Relay, ATM, and IP) packet-switched networks.

Table 5 captures the bandwidth and MOS achievable by those codecs around which industry attention is currently focused.

Table 5: Codec Bandwidth Requirements and Voice Quality as Measured by MOS

Codec	**Acoustic Bandwidth**	**Payload data-rate (kbps)**	**Loaded bandwidth[13] (20 ms pkts)**	**Max[14] MOS achievable (indicative)**	**Standards body**
G.711	NB	64	80	4.4	ITU-T
G.711.1	WB[15] (embedded)	64, 80, 96	80, 96, 110	4.1	ITU-T
G.729	NB	8	24	3.8	ITU-T
G.729.1	WB (embedded)	12-32	28-48	4.1	ITU-T
G.726	NB	32	48	4.0	ITU-T
G.722	WB	64	80	4.0	ITU-T
iLBC	NB	15.2	31.2	3.8	GIPS proprietary
GSM (AMR 12.2)	NB	12.2	29.2	4.1	3GPP
RT-Audio	WB	29	45	4.2	Microsoft proprietary
RT-Audio	NB	11.8	27.8	3.8	Microsoft proprietary

A lot of interest has surfaced in comparing Microsoft's OCS codec, RT-Audio, with a range of other codecs. RT-Audio utilizes advanced jitter-buffer/error-concealment techniques to provide high quality wideband speech (higher quality than a typical G.711 or G.729-based application) over relatively poor quality or lower bandwidth networks. RT-Audio's flexibility is useful, especially for mobile/remote workers (including in wireless hotspots), home-workers (including call center operatives) or anywhere a voice conversation will be carried over a connection with limited network management (for example,

[13] Taking account of IP, UDP and RTP headers but excluding layer 2 overhead.

[14] The practical MOS achieved operationally will be below this value depending on network performance, associated error resilience capabilities and the presence of disturbing signals such as noise. This makes codec comparison complex, since a better theoretical maximum MOS is of little practical relevance if the codec (and associated error correction) has low resilience to typical network losses. For example, iLBC and G.729 have very similar maximum quality performance, but with 10% packet loss iLBC will provide significantly better voice quality than G.729 due to better error resilience.

[15] Based on a wideband subjective scale and therefore off-set (favorably) from narrowband scale judgments.

one that does not prioritize voice). However, on a network with QoS implemented, the jitter-buffer and error concealment in RT-Audio is less relevant and does add undesirable extra delay not evident in standard codecs like G.711 and G.729.

Keep in mind that the edge-device (IP phone or soft client) is just one component in the overall transmission path with it and other network components all contributing to the overall delay. As we pointed out above, latency should not exceed 100 ms one way for toll-quality voice and must not exceed 150 ms one way for acceptable quality voice. And this time constraint, in and of itself, limits the amount of control/resilience that can be added to a VoIP network. Table 6, below, shows the impact on latency of various endpoint choices.

Table 6: Contribution of Endpoint Choice to One-way Delay

Endpoint	Codec	Error resilience	Typical delay (ms)
IP-phone (hard phone)	G.711	Med	56
IP-phone (hard phone)	G.729	Med	75
IP-phone (Tanjay)	RT-Audio	High	109
Soft-client (soft phone)	iLBC	High	110
Soft-client (soft phone)	RT-Audio[16]	High	145[17]
Wireless IP-phone	G.729	Med	90

The above table demonstrates that the compromise in utilizing the benefits of Microsoft's Office Communicator client with advanced jitter-buffer/error-concealment techniques comes at the price of higher delay. However, the counterbalancing operational benefits of integrated UC functionality, higher fidelity voice and robust operation over less well managed networks (thanks to the additional error resilience provided by advanced coding/jitter-buffer/error concealment) might be appropriate in some instances. Clearly, the BDM must target the design and implementation of a UC solution to his/her enterprise's unique needs if an optimal tradeoff of business requirements for enhancing both individual and business process productivity, interoperability with legacy infrastructure and total cost of ownership is to be achieved. And that solution may be a single or multi-vendor solution from any of the major suppliers (Microsoft, Cisco, IBM, Siemens, Mitel, Nortel, Avaya, etc.).

---

[16] Given the additional delay overhead of RT-Audio, data center placement of mediation servers for conversion of RT-Audio to G.711 may not have been the best decision for branch office QoE noted J. Lewis, Microsoft IT Manager Australia: "[I]f we were to do things over again I feel we would very likely decide to deploy the Mediation servers at the branch offices to provide the best user experience." http://download.microsoft.com/documents/australia/business/uc/track1/Session4_How_MS_IT_Deployed_UC.pptx

[17] Based on measurements directly taken between two soft-clients. See M. B. Hommer and R. J. Smithers, **Lab Test: Microsoft OCS 2007--Voice Communication for the Next Generation?** Jan. 15, 2008, http://www.nojitter.com/showArticle.jhtml?articleID=205602869.

4. VoIP Security

Security is a critical issue that cannot be ignored because security attacks of various kinds can have a direct impact on voice QoE. Though VoIP security is not the focus of this white paper, several summary recommendations should be mentioned:

1. If feasible assign voice and data on logically separate networks (VLANS) due to their different QoS and security requirements. Make sure your Ethernet switches are equipped with 802.1p prioritization so they can identify and prioritize traffic based on VLAN tags and support multiple queues.
2. Implement VoIP-ready firewalls capable of handling the latency sensitive needs of voice traffic. Such firewalls provide rich granular controls, protocol conformance checking, protocol state tracking, security checks, and NAT services. These are essential components in the VoIP network. In addition, state-of- the-art intrusion detection and prevention systems should also be installed.
3. Use VoIP network encryption. Transport layer security and IPSec are two main encryption methods. Make sure your firewall can provide for the inspection of encrypted voice traffic.
4. Perform requisite traffic analysis and implement network rate-limiting. This will help ensure that voice traffic receives the bandwidth it requires should a virus or worm infect your data traffic.
5. Ask your carrier how they can help you mitigate Distributed Denial of Service (DDoS) attacks. Recent technology advancements are available to protect against 'botnet' attacks.
6. Make sure your anti-virus system and all other network software is regularly updated.
7. Apply adequate physical security to restrict access to VoIP components.
8. Allow for sufficient power backup and the ability to roll-over your voice calls to the public switched telephone network (PSTN) should your IP WAN experience an outage.

5. Summary

VoIP deployments are proving complex to manage given the need to orchestrate (1) the successful integration of multiple components, technologies, network demarcation points and vendor equipment/relationships, and (2) the unique VoIP QoE performance requirements associated with differing business scenarios. Success integrating and managing these complexities requires that BDMs focus their attention on overall network design, QoS ***and*** QoE to ensure a successful UC implementation. Some summary recommendations are:

1. Use priority scheduling for voice-class traffic.
2. Actively minimize one-way delay. Latency should not exceed 100 ms one way for toll-quality voice and must not exceed 150 ms one way for acceptable quality voice.
3. Keep (random) packet loss well below 1%.
4. Rationalize the use of speech compression codecs for wireless networks and packet networks to minimize transcoding issues.
5. Avoid transcoding where possible. It has the potential to increase distortion and delay and, hence, a loss of quality.

6. Focus on both the network and the endpoints as the strategy for implementing QoE. Endpoints only have “knowledge” or “visibility” into conditions affecting themselves and cannot solve for overall network QoS requirements.
7. Structure the design and implementation of your UC solution for the best tradeoff of business requirements for enhancing both individual and business process productivity, interoperability with legacy infrastructure and total cost of ownership.
8. Recognize that the productivity enhancing benefits of UC applications (e.g. unified messaging, presence, higher fidelity/wideband voice, integration with other business applications and simplicity of use) may be accompanied by performance considerations around bandwidth utilization and higher delay.
9. It is wise to use a real-time voice performance management solution that can detect, alarm and diagnose both IP-Network and application (echo, delay, speech-level, noise, distortion) issues to ensure that operational quality requirements are met on a per session basis and that effective alarming, trouble-shooting and issue resolution address the full range of factors that determine end-user QoE.

About the Authors

Paul Robinson, PhD
David Yedwab
Founding Partners

www.mktstrategy-analytics.com

Market Strategy and Analytics Partners custom designs marketing and sales strategies that are consistent with client core competencies, market focus and competitive environment, and coupled with operationalized go-to-market plans across the value chain to ensure elimination of bottlenecks and complete consideration of end-to-end financials. Our clients include equipment and software providers, service providers and information intense enterprises.

Glossary of Key VoIP QoS and QoE Terms

Acronym	Term	Definition
CAC	Call Admission Control	**Call Admission Control** prevents oversubscription of VoIP networks. CAC is a concept that applies only to real time media traffic and not to data traffic. CAC mechanisms complement the capabilities of QoS tools to protect voice traffic from the negative effects of other voice traffic and to keep excess voice traffic off the network. CAC is used to prevent congestion in Voice Traffic. It is a *preventive* congestion control procedure that is used in the call setup phase.
CNG	Comfort Noise Generation	To counteract the negative perception of total silence on circuits, **comfort noise** is added, usually on the receiving end in wireless or VoIP systems, to fill in the silent portions of transmissions with artificial noise. The noise generated is at a low but audible volume level, and can vary based on the average volume level of received signals to minimize jarring transitions.
Codec	Coder Decoder	A **codec** is a device or program capable of performing encoding and decoding on a digital data stream or signal. The word *codec* may be a combination of any of the following: '**co**mpressor-**dec**ompressor', '**co**der-**dec**oder', or '**co**mpression/**dec**ompression algorithm'. The G.711, G.722, G.729, iLBC and RTAudio codecs discussed in this paper are used in VoIP networks.
DiffServ	Differentiated Services	**DiffServ** operates on the principle of *traffic classification*, where each data packet marked and placed into a limited number of traffic classes.. Each router on the network is configured to handle traffic based on its class. DiffServ-aware routers implement *Per-Hop Behaviors* (PHBs), which define the packet forwarding properties associated with a class of traffic. Each traffic class can be managed differently, ensuring preferential treatment for higher-priority traffic on the network such as, VoIP or video.
E-Model	E-Model	The E-Model (**ITU-T Rec. G.107**) is a transmission planning tool that provides a prediction of the expected voice quality, as perceived by a typical telephone user, for a complete end-to-end (i.e. mouth-to-ear) telephone connection under conversational conditions. The E-Model takes into account a wide range of telephony-band impairments, in particular the impairment due to low bit-rate coding devices and one-way delay, as well as the "classical" telephony impairments of loss, noise and echo. It can be applied to assess the voice quality of wireline and wireless scenarios, based on circuit-switched and packet-switched technology.
FEC	Forward Error Correction	In telecommunication, **forward error correction** (FEC) is a system of error control for data transmission, whereby the sender adds redundant data to its messages, also known as an **error correction code**. This allows the receiver to detect and correct errors (within

		some bound) without the need to ask the sender for additional data. The advantage of forward error correction is that a back-channel is not required, or that retransmission of data can often be avoided, at the cost of higher bandwidth requirements on average, and is therefore applied in situations where retransmissions are relatively costly or impossible.
Ie	Equipment Impairment Factor	The equipment impairment factor represents impairments caused by low bit-rate codecs. This factor should ideally cover all perceptively very diverse effects (distortion, sound degradation, degradation of voice quality, etc.) which can be associated with the codec, except those already covered in another way by the E-Model (e.g. overall attenuation, absolute delay).
IETF	Internet Engineering Task Force	The **Internet Engineering Task Force** (**IETF**) develops and promotes Internet standards, cooperating closely with the W3C and ISO/IEC standard bodies and dealing in particular with standards of the TCP/IP and Internet protocol suite. It is an open standards organization, with no formal membership or membership requirements. All participants and leaders are volunteers, though their work is usually funded by their employers or sponsors.
IP	Internet Protocol	The **Internet Protocol** (**IP**) is a data-oriented protocol used for communicating data across a packet-switched internetwork. IP is a network layer protocol in the Internet protocol suite and is encapsulated in a data link layer protocol (e.g., Ethernet). As a lower layer protocol, IP provides the service of *communicable* unique global addressing amongst computers.
ITU	International Telecommunication Union	**ITU** is the leading United Nations agency for information and communication technologies. As the global focal point for governments and the private sector, ITU's role in helping the world communicate spans 3 core sectors: radio communication, standardization and development. ITU also organizes TELECOM events and was the lead organizing agency of the World Summit on the Information Society. **ITU** is based in Geneva, Switzerland, and its membership includes 191 Member States and more than 700 Sector Members and Associates.
MOS	Mean Opinion Score	**Mean Opinion Score (MOS)** is the average of the opinions expressed by a group of subjects presented with a sample stimulus , e.g. a voice sample. Subjects express their opinion against a 5 point scale, e.g.: excellent (5), good (4), fair (3), poor (2), bad (1). Objective measurement methods attempt to predict human opinion to provide a numerical indication of the perceived quality of received media after compression and/or transmission.
PESQ	Perceptual Evaluation of Speech Quality – ITU-T	**PESQ**, *Perceptual Evaluation of Speech Quality*, is mechanism for automated assessment of the speech quality enjoyed by the user of a

	Recommendation P.862	telephony system. It is standardized as ITU-T recommendation P.862.1 (02/01) for Narrow Band signals and P.862.2 for Wide-band signals.
PSTN	Public Switched Telephone Network	The **public switched telephone network** (**PSTN**) is the network of the world's public circuit-switched telephone networks, in much the same way that the Internet is the network of the world's public IP-based packet-switched networks. Originally a network of fixed-line analog telephone systems, the PSTN is now almost entirely digital, and now includes mobile as well as fixed telephones. The PSTN is largely governed by technical standards created by the ITU-T, and uses E.163/E.164 addresses (known more commonly as telephone numbers) for addressing.
QoS	Quality of Service	In the fields of packet-switched networks and computer networking, the traffic engineering term **Quality of Service**, abbreviated QoS, refers to resource reservation control mechanisms rather than the achieved service quality. Quality of Service is the ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance to a data flow. QoS mechanisms implemented in the IP data network are key to providing high quality VoIP connections.
QoE	Quality of Experience	En –user Quality of Experience is determined by the performance of both the network and the communications application. In the case of VoIP QoE is determined by the performance of the IP-Network (to deliver the packets across the network) and application level factors such as; echo, speech level, delay, noise level, and speech distortion. Effective and performance management must account for both network and application performance.
RSVP	Resource Reservation Protocol	The **Resource ReSerVation Protocol** (**RSVP**), described in IETF RFC 2205, is a Transport layer protocol designed to reserve resources across a network for an integrated services Internet. RSVP provides receiver-initiated setup of resource reservations for multicast or unicast data flows with scaling and robustness. RSVP can be used by either hosts or routers to request or deliver specific levels of quality of service (QoS) for application data streams or flows. RSVP defines how applications place reservations and how they can relinquish the reserved resources once the need for them has ended. RSVP operation will generally result in resources being reserved in each node along a path.
RTP	Real-Time Transport Protocol	The **Real-time Transport Protocol** (or **RTP**) defines a standardized packet format for delivering audio and video over the Internet. It was developed by the Audio-Video Transport Working Group of the IETF and first published in 1996 as RFC 1889. RTP does not provide mechanisms to ensure timely delivery of packets. They also do not give any Quality of Service (QoS) guarantees so QoS needs to be

		provided by some other mechanism.
SIP	Session Initiation Protocol	The **Session Initiation Protocol** (**SIP**) is an **application-layer control** (signaling) protocol for creating, modifying, and terminating sessions with one or more participants. It can be used to create two-party, multiparty, or multicast sessions that include Internet telephone calls, multimedia distribution, and multimedia conferences. It is based on IETF RFC 3261. It is widely used as a signaling protocol for Voice over IP, along with H.323, MGCP and other protocols.
TIA	Telecommunications Industry Association	**TIA** is a leading trade association for the information, communications and entertainment technology industry. TIA serves industry suppliers to global markets through its leadership in standards development, domestic and international policy advocacy, and facilitating member business opportunities. TIA represents the communications sector of the Electronic Industries Alliance (EIA).
Transcoding	Transcoding	Transcoding is the direct digital-to-digital conversion from one codec to another without returning the signals to analog form. Transcoding has the potential to increase distortion and delay and, hence, a loss of quality. How much distortion depends on the codecs involved.
UDP	User Datagram Protocol	**User Datagram Protocol** (**UDP**) is one of the core protocols of the Internet protocol suite. Using UDP, programs on networked computers can send short messages sometimes known as *datagrams* to one another. UDP does not guarantee reliability or ordering in the way that TCP does. Datagrams may arrive out of order, appear duplicated, or go missing without notice. Avoiding the overhead of checking whether every packet actually arrived makes UDP faster and more efficient, at least for applications that do not need guaranteed delivery. Time-sensitive applications often use UDP because dropped packets are preferable to delayed packets. Common network applications that use UDP include: the Domain Name System (DNS), streaming media applications such as IPTV, Voice over IP (VoIP), Trivial File Transfer Protocol (TFTP) and online games.
VAD	Voice Activity Detection	**Voice Activity Detection** is an algorithm used in speech processing wherein the presence or absence of human speech is detected in regions of audio. While talking to someone, there will be silent periods when we are not talking. A VAD feature in VOIP can disable the silence packets and use the silent period to transmit some traffic other than voice thus conserving bandwidth.
VGW	Voice Gateway	A **Voice Gateway** is used as the connecting point between a VoIP system and the PSTN or other legacy equipment such as, analog phones. Thus it is used to convert from IP to traditional analog or digital formats to provides connections such as, FXS, FXO , PRI, T1, or other types of ports. Voice gateways can be implemented in dedicated devices or are often implemented in routers.

VoIP	Voice over Internet Protocol	**Voice over Internet Protocol** (**VoIP**) is a protocol optimized for the transmission of voice through the Internet or other packet switched networks, typically as an RTP stream. VoIP is often used abstractly to refer to the actual transmission of voice (rather than the protocol implementing it). VoIP is also known as **IP Telephony**, **Internet telephony.**
WAN	Wide Area Network	**Wide Area Network** (**WAN**) is a computer network that covers a broad area (i.e., any network whose communications links cross metropolitan, regional, or national boundaries. The largest and most well-known example of a WAN is the Internet. WANs are used to connect LANs and other types of networks together, so that users and computers in one location can communicate with users and computers in other locations. Many WANs are built for one particular organization and are private.