
IP Telephony Design Guide

AN ALCATEL WHITE PAPER

April, 2003

ARCHITECTS OF AN INTERNET WORLD



IP Telephony Design Guide

Overview	3
IP Telephony vs. VoIP	3
Voice Bandwidth Requirements	3
Voice over IP Bandwidth Requirements	3
Voice Quality	4
Converting Voice into Data Packets	4
Buffering and Error Checking	6
IP Telephony/VoIP Audit	6
Bandwidth Management	6
What is QoS and why is QoS needed?	7
Availability	7
Throughput	7
Delay	7
Delay Variation	8
Packet Loss	8
Class of Service	8
Deploying IP Telephony in a Converged Alcatel Network	8
Deploying IP Telephony and VoIP in a Multi-Vendor Environment	9
Design Recommendations	9
VoIP Design Guide Check List	10
Acknowledgments	11
Bibliography	11

Overview

With the availability of today's new convergence technology more and more people are planning to deploy voice traffic over existing data networks. The practice of bandwidth sharing between voice and data traffic over a single network is not new. In the 1980's, time division multiplexing (TDM) made short work out of this task by carving up the bandwidth required to share a single, wide area network (WAN) connection between multiple locations. Although statistical multiplexing or packet-based networking was more effective for transporting voice, you still needed to maintain independent networks – one for data-/LAN-based traffic and one for voice.

With the Internet explosion and advanced PC applications that use more and more bandwidth, the data network volume has increased dramatically and is now the dominant bandwidth consumer. Therefore, it now makes sense to use the data network to transport voice instead of the voice network to transport data traffic.

This design guide covers the issues related to designing an IP telephony or voice over IP (VoIP) network for transporting voice and data over a common LAN or WAN infrastructure. Understanding the underlying technology used to transport voice traffic is important in designing an IP telephony network. Design principles used to deploy a successful LAN-based VoIP network will not necessarily work when you apply them to a WAN configuration. This document discusses the major hurdles that need to be addressed when designing either a LAN or WAN based VoIP network.

IP Telephony vs. VoIP

Let's define what IP telephony and VoIP mean in this document. IP telephony is the combination of voice, data, video, and wireless applications into an integrated enterprise infrastructure that offers the reliability, interoperability, and security of a voice network, the benefits of IP, and the efficiencies, mobility, and the manageability of a single network. IP telephony is based on circuit-switched and TCP/IP technologies and protocols, it removes the limitations of proprietary systems and provides increased productivity, scalability, mobility, and adaptability.

Voice over IP (VoIP) is the technology that is used to transmit voice over an IP network, which can be either a corporate network or the Internet.

Voice Bandwidth Requirements

In the traditional voice world a single T1 leased line is used to carry 24 toll quality telephone calls from the Public Switched Telephone Network (PSTN). Those with a private point-to-point T1 connection can compress the voice to less than 8 Kbps for more efficiency, but the quality may be sacrificed. The three most common modulation schemes for encoding voice are:

64 Kbps (PCM)/1.544 Mbps = 24 simultaneous calls on a T1

32 Kbps (ADPCM)/1.544 Mbps = 48 simultaneous calls on a T1

8 Kbps (CELP)/1.544 Mbps = 120 simultaneous calls on a T1

Efficiency is the primary WAN connection issue. Bandwidth use and voice compression play an important role in provisioning the WAN.

Voice over IP Bandwidth Requirements

What does it take to support traditional voice on data networks? The concept of combining voice on the data network is simple because voice traffic uses a lot less bandwidth than traditional LAN based computer networks. A single toll quality phone call over the public network uses 64 Kbps in each direction, that's only 0.0625% of a 100 Mbps full duplex link.

On a 100 Mbps Ethernet network each voice call takes up to 85.6 Kbps (64 Kbps + IP header + Ethernet header) in each direction supporting up to 1,160 calls over a full duplex link. On a Gigabit backbone up to 11,600 simultaneous calls can be handled.

If bandwidth were the only issue, LAN-based IP telephony networks would have been deployed years ago. But other elements, such as bandwidth hungry business applications, advancements in telephone technology, and network congestion have been the major stumbling blocks. Most of those issues have been resolved with newer VoIP technology, QoS, and the use of bandwidth managers or complex queuing schemes deployed on the LAN and WAN.

Voice Quality

Over the years voice quality has been very subjective: picking up the phone and listening to the quality of the voice. If you had two different users on the same call you may even receive reports of varying results. After years of research, human behavioral patterns have been recorded and scored, establishing an objective measurement of call quality.

The leading subjective measurement of voice quality is the Mean Opinion Score (MOS) as defined in the International Telecommunications Union (ITU) recommendation P.800. Mapping between network characteristics and quality score make MOS valuable for doing network assessments and tuning.

A MOS score can range from 5 (very satisfied) to 1 (not recommended), but keep in mind that each voice codec has a benchmark score based on several factors, including packetization delay and the inherent degradation that occurs when converting the voice to a digital signal. The highest MOS rating any codec could receive is 4.5. Each codec is given a MOS value based on any known impairments for the speed of the conversion, speech quality, and data loss characteristics. Below is a listing of the most common codecs used today for VoIP and their theoretical maximum MOS value.

Codec	Default data rate	Time between packets	Packetization delay	Default jitter buffer delay	Theoretical maximum MOS
G.711u	64 kbps	20 ms	1.5 ms	2 datagrams (40 ms)	4.4
G.711a	64 kbps	20 ms	1.5 ms	2 datagrams (40 ms)	4.4
G.729	8 kbps	20 ms	15.0 ms	2 datagrams (40 ms)	4.07
G.723.1 MPMLQ	6.3 kbps	30 ms	37.5 ms	2 datagrams (60 ms)	3.87
G.723.1 ACELP	5.3 kbps	30 ms	37.5 ms	2 datagrams (60 ms)	3.69

Source: *Voice Over IP 2nd edition*

Each network will have a different MOS value based on QoS, delay and codec that is deployed in the IP network. When deploying an IP telephony network the goal is to get the network to support the maximum MOS value and to achieve the best quality for voice traffic. All MOS values above 4.0 are considered to be toll quality speech.

Converting Voice into Data Packets

Digital signal processors (DSP) – the engines for voice coders – are making their way into IP telephony systems. The DSP is a specialized processor that has been in use for many years in other telephone applications such as mobile wireless networks. The DSP needs to be very fast due to the computation intensive operations required to process a typical telephone call. In essence, the DSP is what converts analog voice signal into data packets so they can be transported over an IP-based network.

In this document, DSP refers to the combined efforts of DSPs and codecs to perform the conversion of analog and digital signals into IP communication flows. DSP works by clarifying or standardizing the levels or states of a digital signal. A DSP circuit is able to differentiate between human-made signals, which are orderly, and noise, which is inherently chaotic.

Typically, the voice-coding algorithm used for an IP telephony or VoIP network in a LAN environment is G.711, which divides a voice stream up into 64 Kbps packet increments. It is regarded as toll quality. Some of the other more widely available voice coding algorithms/compressors on the market are the G.729a and G.723 codecs. The G.729a and G.723 codecs are normally used for WAN connections where bandwidth is at a premium and voice compression is a requirement. The majority of vendors who support IP telephony recommend the G.729a codec due to its superior quality over G.723, making it the de facto standard for WAN connections running IP telephony.

The chart below shows the bandwidth calculation for each codec.

Voice coder	Voice bandwidth Kbps	MOS	Codec delay	Packet size (bytes)	IP/UDP/RTP headers (bytes)	cRTP	L2 header (bytes)	Total BW	BW with silent suppression
Ethernet									
G.711	64	4.1	1.5	160	40		14	85.6	42.8
G.711	64	4.1	1.5	160		2	14	70.4	35.2
G.729	8	3.9	15	10	40		14	29.6	14.8
G.729	8	3.9	15	10		2	14	14.4	7.2
PPP									
G.711	64	4.1	1.5	160	40		6	82.4	41.2
G.711	64	4.1	1.5	160		2	6	67.2	33.6
G.729	8	3.9	15	10	40		6	26.4	13.2
G.729	8	3.9	15	10		2	6	11.2	5.6
G.723	6.3	3.9	37.5	30	40		6	16	8
G.723	6.3	3.9	37.5	30		2	6	8	4
Frame Relay									
G.711	64	4.1	1.5	160	40		4	81.6	40.8
G.711	64	4.1	1.5	160		2	4	66.4	33.2
G.729	8	3.9	15	10	40		4	19.7	9.9
G.729	8	3.9	15	10		2	4	9.6	4.8
G.723	6.3	3.9	37.5	30	40		4	15.5	7.8
G.723	6.3	3.9	37.5	30		2	4	7.6	3.8
ATM									
G.711	64	4.1	1.5	160	40		5 cells	106	53
G.711	64	4.1	1.5	160		2	4 cells	4	42.4
G.729	8	3.9	15	10	40		2 cells	2.3	14.1
G.729	8	3.9	15	10		2	1 cell	14.1	7.1
G.723	6.3	3.9	37.5	30	40		4	22.3	11.1
G.723	6.3	3.9	37.5	30		2	4	11.1	5.6

Table 1 - Bandwidth calculation by voice code

Buffering and Error Checking

Due to the bursty nature of business applications, data networks have large buffers built into them to sustain large bursts of traffic over a short period of time.

Large buffers in a voice network will only increase the delay of time sensitive traffic and cause poor call quality. Voice is very similar to constant bit rate (CBR) traffic – it requires a predictable, reliable throughput.

The majority of the LAN protocols used to transport data traffic include end-to-end error checking. So, if a packet is delayed or lost, the originating station will retransmit a copy of the frame. The end station will wait for the acknowledgement, then reassemble the packet stream and pass it on to the application. This is usually transparent to the user.

Voice transmissions on the other hand are very time sensitive. The originating station does not copy the transmitted frame into a buffer, since it would only increase the delay and degrade quality. With voice, if you lose a frame, it is lost. Both error and frame sequence checking is done at the upper level of the Real Time Protocol (RTP), but due to the time sensitive nature of the voice stream, if the frame is out of sequence it will be discarded and the next frame will be processed, thus affecting the quality of the call.

The majority of voice codecs can support minor frame loss, but the conversation will be choppy and of poor quality. Some of the IP telephony equipment manufacturers have tried to compensate for poor line quality by playing the preceding voice frame a second time, but this does not resolve the issue, it only makes it tolerable. This is why it is so important to understand the inherent behavior of voice running on a data network and the additional requirements like QoS and predictive delay that a network must meet.

IP Telephony/VoIP Audit

An IP telephony/VoIP audit should be performed for every proposed LAN/WAN segment prior to the addition of IP telephony traffic. The key to designing an IP telephony network is an understanding of the underlying technology used to transport the IP telephony traffic. The design principles used to deploy a successful LAN based VoIP network will not necessarily work when you apply them to a WAN configuration, due to a number of factors including limited bandwidth. QoS and traffic isolation are the key factors for the LAN, but bandwidth, priority and delay are important to the WAN. This can make a significant impact on the installation.

The most common cause for poor voice quality during a VoIP installation is inadequate WAN bandwidth to support both voice and data traffic. If an audit was performed prior to the installation, corrective action could have been taken to resolve the issue prior to deployment.

In some cases, a poorly designed WAN can be fixed by lowering the delay with fewer router hops, setting up QoS on the routers or increasing the amount of available bandwidth prior to the installation of voice. In other cases the solution may be too expensive or too complex and other products like bandwidth managers must be deployed prior to the addition of voice.

Bandwidth Management

If the MOS value is not in an acceptable range after completing the IP audit and tweaking the installed vendor's suggested parameters, a bandwidth manager may be needed for a successful installation. Bandwidth managers allow the end user to define how much bandwidth is going to be used by each application and guarantee what percentage of the WAN bandwidth is going to be used by voice applications.

What is QoS and why is QoS needed?

Voice quality is directly affected by many factors that can be divided into five QoS dimensions that impact the end user experience:

- 1) **Availability**
- 2) **Throughput (both committed and burst)**
- 3) **Delay or latency**
- 4) **Delay variation, including jitter and wander**
- 5) **Packet loss**

Availability

Availability is the percentage of time that the network is up. The traditional benchmark for a voice network is 99.999% ("five 9s"), or about 5.25 minutes of downtime per year. Availability is achieved through a combination of equipment reliability and network survivability. Availability is a probability calculation, so it is not simply calculated by summing the MTBF figures.

Throughput

Throughput is the amount of traffic – or bandwidth – delivered over a given period of time. Generally speaking in the LAN environment, more throughput is better.

For the majority of WAN users, throughput depends on the amount of money paid to lease carrier facilities. So efficiency, compression, and bandwidth management play key roles in designing an IP telephony network.

Delay

Delay or latency is the average transit time of a service from the ingress to the egress point of the network. Many services – especially real-time services such as voice communications – are highly intolerant of excessive or unnecessary delay. Interactive conversation becomes very cumbersome when delay exceeds 100-150 ms, when it exceeds 200 ms users find it disturbing and describe the voice quality as poor. To provide high quality voice, the VoIP network must be capable of guaranteeing low latency. The ITU-T G.114 recommendation limits the maximum acceptable round trip delay time to 300 ms between the two VoIP gateways (150 ms one-way delay).

There are many components of delay in a network that must be understood, including packetization delay, queuing delay, and propagation delay.

- **Packetization Delay** is the amount of time it takes the codec to complete the analog to digital conversion. Realize that IP telephony/VoIP always creates some measure of delay, as the algorithm specifies to "listen" or sample the voice for a specified period, followed by packetization.
- **Propagation Delay** is the amount of time it takes information to traverse a copper, fiber, or wireless link. It is also a function of the speed of light, the universal constant, and the signaling speed of the physical medium. For example, if a call has to pass through a transit node more delay is introduced.
- **Queuing Delay** is imposed on a packet at congestion points when it waits for its turn to be processed while other packets are sent through a switch or wire. For example, as previously stated ATM mitigated queuing delay by chopping packets into small pieces, packing them into cells, and putting them into absolute priority queues. Because the cells are small, the highest priority queue can be serviced more often, reducing the wait time for packets in this queue to deterministic levels. At gigabit speeds, however, the waiting time for high-priority traffic is very small even under the worst conditions, due to the speed of the links and available processing power.

Delay Variation

Delay variation is the difference in delay exhibited by different packets that are part of the same traffic flow. High-frequency delay variation is known as jitter, while low-frequency delay variation is called wander. Jitter is caused primarily by differences in queue wait times for consecutive packets in a flow, and is the most significant issue for QoS. Certain traffic types—especially real-time traffic such as voice, are very intolerant of jitter. Differences in packet arrival times cause choppiness in the voice. All transport systems exhibit some jitter. As long as jitter falls within defined tolerances, it does not impact service quality.

Excessive jitter can be overcome by buffering, but this increases delay, which can cause other problems. With intelligent discard mechanisms, IP telephony/VoIP systems will try to synchronize a communication flow by selective packet discard, in an effort to avoid the "walkie-talkie" phenomenon caused when two sides of a conversation have significant latency. Jitter must be less than 60ms (60ms = average quality, 20ms = toll quality).

Packet Loss

Loss – either bit errors or packet drops – has a bigger impact on IP telephony/VoIP services than on data services. During a voice transmission, loss of multiple bits or packets of stream may cause an audible pop that will become annoying to the user. In a data transmission, loss of a single bit or multiple packets of information is almost never noticed by users. In contrast, during a video broadcast, consecutive packet loss may cause a momentary glitch on the screen, but the video then proceeds as before. However, if packet drops become epidemic, then the quality of all transmissions degrades. Packet loss rate must be less than 5% for minimum quality and less than 1% for toll quality.

Class of Service

The main objective of Resource Reservation Protocol (RSVP) is to guarantee end-to-end QoS throughout the network by reserving bandwidth unicast and multicast applications on an individual flow basis.

Differentiated Services (DiffServ) is designed to group all flows with the same service requirement into a single aggregate. For example: RSVP would reserve bandwidth for a single VoIP call, while DiffServ would group all VoIP traffic together in the same flow. This aggregated flow would then receive its class of service based on the application priority.

When a QoS mechanism like DiffServ is enabled, it will provide complete flexibility in defining service classes that can be provisioned in a converged voice and data network. This means that the network management system provides access to the mechanisms that allow the end user to create customized service classes for each application.

Most networks are deployed with some level of QoS at layer 3 that supports the following classes of service:

- **Expedited forwarding (EF) for control frames like RTCP**
- **Assured forwarding (AF) for VoIP traffic**
- **Best Effort (BE) for all other data traffic**

It is possible to map different QoS parameters to one another (i.e., 802.1p to ToS or ToS to DiffServ) to enable the network designer to provision an "end-to-end" class of service for voice, video and data traffic.

Deploying IP Telephony in a Converged Alcatel Network

Today's business depends on scalable network communications that allow future expansion of business options and facilities. The groundbreaking OmniSwitch family (6600 series, 7000 series, and the 8800) and OmniPCX Enterprise voice products target that future networking and business solution. The OmniSwitch family series is a new line of data infrastructure switches that spans the core, edge, and desktop of networking. The design combines Alcatel's experience and expertise building carrier and enterprise network equipment with all of the company's cutting-edge convergence technologies.

e-Business solutions must provide availability, security, intelligence, and manageability. These values are both essential to successful modern business and fundamental to appropriate new technology.

The OmniSwitch family offers carrier-class availability throughout all networking components to deliver the infrastructure mandatory for IP telephony and mission-critical applications. A multi-layered approach to security is offered, securing traffic to, through and between switch nodes, preventing unauthorized access to business traffic and ensuring privacy. Intelligence mandates that all switching decisions are distributed and performed at wire-rate. Alcatel's implementation is wire-rate into, through the backplane, and out all network interfaces without performance bottlenecks. Manageability involves both networking and management system features. OneTouch QoS means that complex QoS policies are implemented consistently with a simple point-and-click interface.

Deploying IP Telephony and VoIP in a Multi-Vendor Environment

Even though IP telephony and VoIP technology have made some vast reliability and quality improvements over the past couple of years, customers and network designers still struggle with implementing the technology in a multi-vendor network. There are many reasons for this such as: inter-operability issues, proprietary protocols, and just plain old finger pointing. Please check with the manufacture of your installed equipment for their recommendations on how to design and deploy an IP telephony or VoIP network in a multi-vendor setting.

Design Recommendations

One of the most important recommendations that can be made is to pay close attention to the infrastructure that the VoIP network is built on. The foundation must be solid otherwise there will be ongoing quality issues until the network design issues are resolved. The more time spent upfront investigating and verifying the design of the LAN and/or WAN will make a more successful ending. Verification is critical, and although it may seem reasonable to believe that the "network is new and should support QoS" it's important to check. In some cases, like running VoIP over a WAN, an audit is a must. For example, the total end-to-end delay to support a quality voice conversation must not exceed 200 ms and can only be verified by an IP audit. Remember, the longer the delay the worse the quality.

After a VoIP audit is preformed the designer must engineer the network to support the worst-case scenario, even if it happens only 1 % of the time. Engineering the network for peaks not averages maintains the highest quality of voice traffic while the network is performing at its maximum potential.

When designing a VoIP WAN, the designer is required to calculate the amount of available bandwidth for all applications required to transit the link. In most cases the link traffic is miss-calculated or the IP audit is not performed prior to the installation and the quality of the VoIP calls suffer. As previously sated, a good rule of thumb for a WAN link is to keep at least 25 % of the bandwidth available for routing table and administrative updates.

As in most architecture's, the more redundancy and availability options designed into the network the better the odds are for a successful installation. The designer must also understand that engineering all of the redundancy options available into the system could adversely affect the performance of the network. For example, adding IP redundancy into the network could increase the jitter because the VoIP packets might take multiple paths to reach the end point. This is not a major concern, but it must be evaluated prior to deploying the VoIP network.

Redundancy features cost real money, so the main task of the design engineer is to make sure the product meets the customer's requirements and at the same time keeps the proposal price competitive. In some cases this could be the difference between winning and losing the opportunity.

The following is a list of questions, thoughts, and ideas that should be considered and reviewed with customers/prospects when designing a VoIP network. It is unlikely that a network configuration will implement every feature on this list, but it's a good checklist to review prior to completing the final design.

VoIP Design Guide Check List

Is the LAN equipment designed to support 99.999% availability?

- Is the LAN configured with the following redundancy options?
 - Management modules
 - Links
 - Protocols (i.e., Fast Spanning Tree)
 - Power supplies
 - UPS system (in the event of a power outage) in the wiring closet
- How are the IP phones going to be powered?
 - Does the LAN switch support in-line power (802.3af)?
 - Is it connected to a UPS system?
 - Does the IP phone model support in-line power?
 - Is an external power patch panel required?
 - Is it connected to a UPS system?
 - Are you using local power?
 - Is it connected to a UPS system?
 - What is the ratio of IP phones with UPS to IP phones without UPS?
 - Are digital/analog terminals intermixed with the IP phones in geographic layout to provide for "emergency dialing" in the event of power or network outages?
- Is the PBX configured with the following redundancy options?
 - Management modules
 - Redundant IP modules
 - Are the VoIP links connected to multiple LAN switches?
 - Is the switch configured to support battery back-up power?
 - Is there a back-up signaling path configured for all networked sites?

Does the installed LAN equipment support QoS?

- Do you know the speed and performance of the installed equipment?
 - Manufacture
 - Product type
 - Link speeds and WAN protocols
 - Routing Protocols
- What is the QoS design strategy?
 - 802.1p/Q
 - DiffServ
 - Is the priority set and respected on every LAN switch in the network?
- ToS (type of service) or CoS (class of service) for the WAN
- Do you have a current local area network diagram? This is a must.
 - When was the network diagram last updated? If it's older than 45 days, ask for an up to date diagram.
 - Has the cable plant been verified to support 100 Mbps Ethernet? (i.e., Cat 5 cable)

❑ Isolation

- Do you have an isolated VLAN configured just for VoIP phones?
- Has the excess broadcast traffic been removed from VoIP VLAN
 - Is IP multicast support enabled on the LAN?

❑ Does the installed WAN support QoS?

- Do you have a current wide area network diagram? This is a must.
- Has the packet forwarding latency and jitter been verified not exceed the maximum tolerance of the 200 ms. An IP audit is a requirement for all WAN connections.
- Is guaranteed bandwidth, packet forwarding rate and capacity specified for all WAN links? A good rule of thumb is to have a 25% available for overhead and routing table updates. Please refer to Table 1 for the bandwidth required for each codec.
- Let's look at a simple calculation using the 25% rule, using a T1 (1.536 Mbps) as the line speed.
 - $1.536 \text{ Mbps} - 25\% = 1.152 \text{ Mbps}$, so this means that both voice and data must share the available bandwidth.
 - Is a bandwidth manager required?

Acknowledgments

This document was authored by John Garrison, Director Systems Engineering, North America with technical input from Jim Martin, SE Manager, Ed Stouffer, Sr. Systems Engineer and Jerome Secher. Vicki Vaughn, Sr. Marketing Manager, provided technical editing. I would also like to extend a special thanks to the many who took a moment to review it and provide their thoughts.

Bibliography

Checklist of VoIP Network Design Tips. NetIQ Corporation, 2001

John Q. Walker. A Handbook for Successful VoIP Deployment: Network Testing, QoS, and John Q. Walker. Assessing VoIP Call Quality Using the E-model. NetIQ Corporation, 2001

Uyless Black. *Voice Over IP second edition*. Prentice Hall, 2002

Scott Hamilton & Charles Bruno. *What You Need to Know Before You Deploy VoIP*. Tolly Research, 2001

Vicki Vaughn. *IP Communications is in Your Future*. Alcatel, 2002

Voice & Data Convergence Solution Guidelines. Alcatel, 2001

Voice Over IP Management. R4.2 System Documentation. Alcatel 2002

www.alcatel.com/enterprise

Alcatel

26801 West Agoura Road
Calabasas, CA 91301 USA

Contact Center

(800) 995-2612 US/Canada
(818) 880-3500 Outside US
www.alcatel.com/enterprise

Product specifications contained in this document are subject to change without notice. Contact your local Alcatel representative for the most current information. Copyright © 2003 Alcatel Internetworking, Inc. All rights reserved. This document may not be reproduced in whole or in part without the expressed written permission of Alcatel Internetworking, Inc. Alcatel® and the Alcatel logo are registered trademarks of Alcatel. All other trademarks are the property of their respective owners.

P/N 031348-00. 4/03

ARCHITECTS OF AN INTERNET WORLD

