Lucent NPS
The knowledge behind the network

Lucent Technologies
Bell Labs Innovations

Voice Over IP: An Overview for Enterprise Organizations and Carriers

By Craig J. LaCava, Senior Network Systems Consultant

Introduction

Convergence of voice and data over the same communications services has long been the goal of service providers and enterprise organizations. Telecommunications managers anticipate cost savings and enhanced control over their networks by combining voice traffic with data traffic.

A major driving force behind the desire for IP telephony in the corporate environment is the anticipated savings in long-distance toll calls, especially international voice calls where a substantial part of the cost derives from regulatory fees. In most cases, these surcharges don't apply to circuits carrying data traffic, making VoIP (Voice over IP) a much less expensive way to make voice calls. Because typical PBXs are also costly to operate, VoIP holds promise of significant savings in infrastructure costs even within a building or campus.

Enhanced Worker Productivity and Simplified Management

In addition to cost savings, combining voice and data offers enterprises the potential to increase worker productivity. Combining voice and data leads to entirely new, enhanced ways to communicate, such as integration of voice mail and email, teleconferencing, collaborative applications, and white boarding, resulting in more effective business processes.

In the corporate environment, convergence is also about enhancing the way companies manage their networks for more effective utilization and efficiency. A combined voice/data infrastructure holds forth the promise of simplified management of the network, elimination of redundant networks, and a single support staff (instead of one for voice and one for data).

Opportunity for New Services

As voice and data converge, carriers also stand to benefit by offering multiple services more efficiently on an improved infrastructure. In addition, they can begin to offer service bundles and more value-added features to increase revenue. Moreover, instead of buying switches and routers with different management structures, carriers increasingly are able to unify traffic management on compatible IP platforms.

Beyond collaborative applications, carriers will eventually be able to develop new revenue streams by offering many new kinds of services, such as universal messaging, integrated user-directed videoconferencing, and other advanced networking features for sophisticated users.

This paper examines some of the issues surrounding VoIP from the perspective of telecommunications managers responsible for specifying networks, as well as the network engineers charged with implementing those networks.

Differences Between Voice and Data Networking

Initially, running voice traffic over IP (VoIP) instead of a switched telephone network may seem like a peculiar concept because voice and data traffic have very different fundamental characteristics. Traditionally voice and data engineers have always been separated, along with their respective networks and expertise.

Circuit-Switched vs. Packet-Switched

Telephone calls are generally circuit-switched. This means that when a telephone call is initiated, a circuit is established between the callers that reserves a path, bandwidth, and processing time for the call. By contrast, nearly all packet switching networks do not reserve any bandwidth or a path for data streams.

There is an absolute limit to the number of telephone calls a voice switch or voice circuit can handle. In the United States, a T1 leased line can carry 24 simultaneous telephone calls. In Europe, an E1 circuit can accommodate 30 calls at once.

However, a data network engineer knows that a T1 circuit carries a maximum of 1.536MB of data while an E1 can provide 2 MB of bandwidth. How many user data connections do these circuits allow? How many e-mail, FTP, and web users can use a single T1 circuit to an Internet Service Provider (ISP)? These questions are difficult to answer due to the nature of data and packet-switched networks.

In packet-switched networks, a single data packet could be re-routed or dropped anywhere in the network due to congestion, link failure, or network noise. Though most of today's public data networks are very reliable, there are still few guarantees of packet delivery.

Constant Bandwidth Consumption vs. Bursts of Activity

While these conditions might sound discouraging, data networks actually thrive under these conditions due to the nature of data traffic, which can be described as burst-like in nature (meaning that constant flows of data traffic from customers are extremely unlikely). Data packets are often lost, retransmitted, or re-routed over greater distances without users even being aware of it.

Consider 100 web users in a single office using a single T1 circuit to an ISP. What utilization should be expected on the T1? This is very difficult to estimate. All web pages vary in file size and a user may take seconds to minutes to read the web page (thus won't request more data from the web server until they are finished).

Now consider 100 telephone users in one office with a single T1 circuit connection to the local carrier. If they all need to make a call, exactly 24 users will successfully receive dial tones and be given the opportunity to make their phone calls.

Voice Traffic Requires On-Time Delivery

In a circuit switched voice call, there are no bursts of voice traffic—each telephone call consumes a fixed amount of bandwidth. Voice traffic is always switched over the same network path and thus voice data is never received out of order at the remote end of the connection.

The most important characteristic of circuit switching is that these dedicated resources enable every bit of voice data to be delivered quickly and always on time. Packet loss or late delivery is unacceptable to voice customers.

However, when an Internet user sends e-mail, the mail messages might not get to the recipient's mailbox for seconds, minutes, or even hours. Similarly, when someone downloads a web page, packets containing the page's data might be dropped, retransmitted, delayed, or re-routed, but the page still appears at the user's workstation within an acceptable length of time. Likewise, FTP file transfers can take varying lengths of time depending on file size, distance from the user, and the bandwidth available between client and server—there is no relative time limit for each packet of the FTP session.

Voice Data (Circuit Switched)	Constant bandwidth and path provided by network. Data always arrives on time with very little loss.
TCP/IP Data (Packet Switched)	Data divided into packets and then sent over the network. Packets that are lost are retransmitted; late delivery is OK. Packets may take different paths and arrive out of order.

Quality of Service Is Critical

Despite these differences, it's good to remember that voice over IP is essentially just another application being run over a data network. What makes this application a challenge for network engineers is that a data network must also have a quality of service (QoS) mechanism in place to compensate for the differences between circuit-switched and packet-switched networking.

While a circuit switched network guarantees QoS by dedicating a circuit between the telephone callers, a packet switched network can cause a telephone call to degrade considerably when there are long delays between packets. However, by prioritizing voice packets over data packets, a packet switched network can deliver voice transmissions as if this traffic was sent over a circuit switched network. Success of a VoIP network comes down to delivering a quality voice transmission on time.

Why Consider VoIP for Business?

After examining the considerable differences between voice and data traffic and the networks that provide voice and data transport, it may seem somewhat ludicrous to attempt forcing voice traffic over a data network. Changes to the data network needed to accommodate voice traffic are significant and expensive. Why bother 'shoe-horning' voice traffic onto a data network when a dependable voice network already exists?

The simple answer is the same one that drives virtually all network development: VoIP will save money for many large business enterprises and give carriers new revenue opportunities. While it is obvious that a good deal of investment is needed initially to improve existing data networks so they can carry voice traffic, savings can be realized in the long term in several ways.

Infrastructure and Support Costs

Many large corporations that require voice and data networks to conduct their business must also invest in two separate infrastructures and support teams to support them. A company with both a voice network and a data network generally employs two separate groups of engineers and operators to maintain them. These two groups often have different skill sets and are usually mutually exclusive (except they may share wiring closets and patch panels).

If the data network carried both voice and data for the company, one of these groups could be disbanded or shrunk and absorbed by the other. However, even though the voice infrastructure can be reduced with VoIP, certain components will remain (e.g., the PBX). With a single networking group of engineers and operators and a reduced network infrastructure, a large business can reduce its support and management costs significantly.

Toll Bypass Savings

Consider a mid-sized company with offices in New York, San Francisco, and London where employees often call and fax each other to coordinate the company's business relations. Examination of the company's telephone bill shows that the majority of its long-distance telephone charges are between those three offices. Assuming this company already has a data network connecting these locations, implementing VoIP would allow calls between the three offices to go over the data network, greatly reducing the long-distance bill.

Additionally, the VoIP implementation could be engineered to save the company even more money. For example, suppose the London office needs to call suppliers in the U.S. that are in New York and California. Using the VoIP network to call between the London and San Francisco offices, then completing the call to Oakland using the local telephone services, is much less costly than a London-to-Oakland call.

Diverse Voice Call Routing

Although telephone service in the U.S. is extremely reliable, large companies often purchase multiple, diverse circuits to the local telephone company's exchange for redundancy. These reserve circuits, rarely if ever used, double the cost per month the company must pay for leased-line telephone access circuits.

With a VoIP implementation, if a failure occurs on the primary telephone circuit at one location, the data network could be used to route calls temporarily to PBXs at other company locations. This

functionality could eliminate the need for a redundant telephone circuit by leveraging the company's data network.

IP Phones Facilitate Adds/Moves/Changes

A typical desk or work area in a large company has one or more data and telephone jacks along with a telephone and a PC. Companies often must pay sizable sums of money each year to maintain these work areas – especially to move individuals from one desk to another.

With the implementation of DHCP (Dynamic Host Configuration Protocol, which enables dynamic assignment of IP addresses to devices on a network), moving a PC from one LAN to another has become a much simpler job due to auto-configuration functionality. However, moving a phone extension from one desk to another (or to another building) is not as simple, usually requiring reconfiguration of the office PBX systems.

With the increased popularity of VoIP, a new type of PBX and IP telephone station has entered the marketplace. An IP Phone can automatically configure itself as a DHCP client. Now a desk requires only a single data jack, and the IP Phone can act as a hub or switch to provide a data port for the user's PC. One or more PCs, or perhaps an IP fax machine, can be directly connected to the IP Phone, and all of them can utilize the local IP network. Moving employees from desk to desk is now a simpler, less costly task.

IP Equipment Costs Rapidly Coming Down

Packetizing voice puts computing power for data networking equipment near the endpoints of the network, where the packet-switching equipment enjoys a much faster improvement in price/performance than switched equipment. With many companies feverishly developing new features for IP equipment, the price/performance of packet-switching equipment is likely to continue improving dramatically. Clearly, packet-switching equipment is the only technology that can possibly keep pace with increase in demand for IP traffic.

VoIP Technology and Components

The VoIP Gateway

The VoIP gateway is the demarcation point where the data packet-switched network meets the circuit switched voice network. VoIP data packets are collected, buffered, and then sent over the circuit switched network's voice channels at a constant rate. At the same time, voice transmissions are taken from the circuit switched voice network, compressed, converted to IP data packets, and then sent over the data network.

A gateway must be located wherever voice traffic will enter or exit the TCP/IP network. A typical VoIP TCP/IP network will have multiple gateways, and usually an enterprise that uses VoIP will have at least one gateway at each office.

The gateway will have at least one network connection (usually LAN-based such as 10/100BaseT, Token Ring, or FDDI) and one or more voice ports. These voice ports can be a single analog telephone port, an ISDN PRI, or a T1/E1 connection to a voice switch or PBX. Most of the voice compression, conversion, and jitter buffering occur at the VoIP gateway.

The VoIP Gatekeeper

The VoIP gatekeeper is where access on the VoIP network is controlled. The gatekeeper is especially important when a VoIP network allows IP phones or PC voice applications to make calls over the network (IP or public voice networks). Each VoIP user must log into a gatekeeper before being allowed to access the VoIP network. In addition to authenticating VoIP users, the gatekeeper also gets access policies for each user that can be configured to restrict calling to certain areas or countries.

A typical VoIP network will have one or two VoIP gatekeepers depending on its size and scale. An enterprise network that uses VoIP in two office locations could operate with a single VoIP gatekeeper. However, if access to this gatekeeper were lost, users would be unable to use the VoIP network in either location because they could not be authorized for service. Placing VoIP gateways in both office locations would help to prevent this situation.

Enterprise VoIP administrators may wish to restrict users from calling other countries, offices, or even certain people. Most VoIP implementations do require at least one gatekeeper component, even if the VoIP users will have no access restrictions.

The VoIP Billing System

When a VoIP network is operated by a carrier (not a private enterprise), VoIP call billing becomes extremely important. The billing system usually works closely with the gateways and gatekeeper to track VoIP users and their network usage.

The billing system is usually a large database of users, tariffs, calling plans, and access policies. Each user is assigned to a gateway by the billing system, typically the closest physical location. Sometimes each user's VoIP access rules are kept at the billing system and downloaded to the gatekeeper. The billing system and the gateway record each user's VoIP call, destination, and length. Based on the user's tariff and call destination, billing records are produced.

Many billing systems also provide accounting and invoicing functionality that can be used to bill and record payments from VoIP clients. A typical VoIP network will have only a single billing system server.

Modifying an Existing Data Network to Carry VoIP

In order for an existing TCP/IP network to successfully carry VoIP traffic, several modifications must be made to make the network infrastructure act like a circuit switched network. VoIP traffic is given priority and, when possible, network resources are reserved exclusively for the VoIP packets.

Compression

The data stream from an uncompressed telephone call approaches 64 Kbps (and uses a single DS0 channel). This amount of bandwidth is considered to be excessive, and is generally reduced for delivery over a typical data network. Several encoding and compression algorithms are available to reduce the bandwidth consumed by a telephone call. These compression mechanisms are usually found at the VoIP gateways and not within the data network routers or switches.

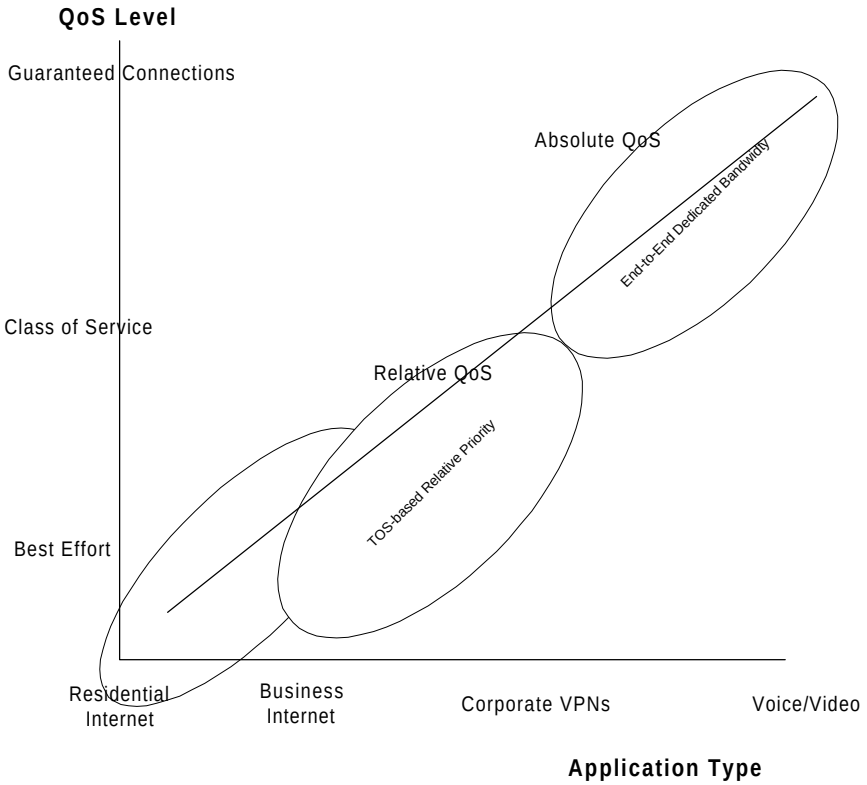
Encoding & Compression Method	Reduced Bit Rate	MOS
G.711 PCM	64 Kbps	4.4
G.726 ADPCM	16, 24, 32*, 40 Kbps	3.85*
G.729 CS-ACELP	8 Kbps	3.92
G.728 LD-CELP	16 Kbps	3.61
G.723.1 CELP	5.3 to 6.3 Kbps (variable)	3.65

The Mean Opinion Score (MOS) is a widely used measure of voice quality. In this completely subjective measurement, average people are asked to rate voice transmission on a scale of 1 to 5. Scores of 4 to 5 are deemed toll quality, 3 to 4 communication quality, and less than 3, poor voice quality.

Presently, the G.729 CS-ACELP compression is the most attractive option for VoIP gateway designers. This voice encoding technique will produce a 20-byte packet every 20 ms (for 50 packets per second). However, this is for the voice payload only—once transport and network layer headers are added to each packet, the resulting packet size will be larger.

The more compression that is applied to a voice transmission, the lower the quality of the voice call will be. Typical quality of a VoIP call after it has been compressed beyond 8 Kbps is about on par with a cellular telephone call.

Network Quality of Service



A TCP/IP network must have mechanisms in place to prioritize VoIP traffic above all other traffic on the network (except other real-time application traffic such as video). A protocol called Resource Reservation Protocol (RSVP) has been designed to reserve resources across the network for real-time transmissions. Quality of Service (QoS) mechanisms within TCP/IP have also recently been implemented by a number of TCP/IP router and switch vendors. ATM networks and, to a lesser degree, Frame Relay networks have QoS functionality already built into them.

Generally, TCP/IP routers and switches will use a priority queuing system to buffer non-VoIP packets and send them only after all of the VoIP packets have been transferred to the next network element. Large IP packets (non-VoIP) are buffered to the side so that the smaller VoIP packets can be sent on time. Other mechanisms will predict times of congestion over the wide area link and throttle back bandwidth demands from non-real-time applications.

Cisco Systems and other vendors offer several tools within their router IOS that deal with QoS. Some of the most popular QoS tools include IP packet precedence, weighted fair queuing (WFQ), resource reservation protocol (RSVP), and weighted random early detection (WRED).

IP Packet Precedence

IP precedence bits should be set at the edge of the network, with VoIP traffic given the highest possible precedence. Cisco's IOS allows this to be done in a number of ways: using route-map statements, with Cisco's Committed Access Rate (CAR), and the IP Precedence interface command. Cisco allows for the following IP precedence levels:

Precedence Title	Precedence IP Setting
Network	7
Internet	6
Critical	5
Flash-override	4
Flash	3
Immediate	2
Priority	1
Routine	0

VoIP traffic and network protocol traffic should be set to IP precedence level 7. Cisco's IOS will automatically set TCP/IP routing protocol traffic, such as BGP routing information, to IP precedence level 7.

Data networks with protocols other than TCP/IP running on them are not as well suited for VoIP as a purely TCP/IP network because it's more difficult to give traffic priority when it is not a TCP/IP packet. Whenever possible, all bridged traffic, Novell, AppleTalk, DecNet, and Vines traffic should all be segregated from the TCP/IP WAN links where VoIP will flow. Bridging of any sort over the wide area will hinder TCP/IP QoS implementations.

Weighted Fair Queuing

Weighted fair queuing (WFQ) is a buffering mechanism that will buffer TCP/IP packets, classify them based on a number of different criteria, and then de-buffer the packets based on IP precedence or traffic flow. The classifications available are: source and destination address, protocol, and session identifier. During the de-queuing procedure, packets are given privilege based on the three IP precedence bits in the packet's IP header.

Resource Reservation Protocol

The purpose of resource reservation protocol (RSVP), a signaling protocol developed by IETF, is to reserve bandwidth and router-processing resources for a specific TCP/IP traffic flow. Cisco has implemented RSVP so that it will work in conjunction with WFQ. When a router sees an RSVP signal from another router or VoIP host, it will dynamically reserve the highest priority queue within multiple WFQ processes for packets from that traffic flow. In theory, this dynamic queuing policy will reserve enough bandwidth across the network for each VoIP flow. The amount of bandwidth reserved per flow is set within the configuration of each RSVP router interface.

Weighted Random Early Detection

In order to avoid congestion over a data network's core WAN links, Cisco has implemented a mechanism called weighted random early detection (WRED). When a router sees an increase in link utilization that could be a precursor to congestion, WRED actually drops TCP/IP packets at random. Due to the sliding windows flow control device built into TCP/IP, a TCP/IP device (host, client, or router) will notice the packet loss and actually throttle back on the amount of data it sends across the network so that a fully congested WAN link is avoided.

The weighted part of WRED implies that properly tagged (with IP precedence) VoIP packets will never be dropped and only the non-real-time applications will be throttled back. Please note that the WRED function is independent on each router running WRED over a wide area—there is no network-wide synchronization between routers.

Dealing with Delay

One of the most important design considerations in implementing voice is minimizing one-way, end-to-end delay. With VoIP traffic, a packet that arrives late might as well have not been sent at all. Retransmission is not an option with VoIP traffic flows. Voice traffic is real-time; if there is too long a delay in voice packet delivery, speech is unrecognizable.

Delay is inherent in VoIP networking and is caused by a number of different factors. Although an acceptable delay is less than 200ms, a better target for a carrier is 150ms, since the packet may have further to travel along the customer's network at either end of the IP network path.

Two kinds of delay are inherent in today's VoIP networks: propagation delay and handling delay. *Propagation* delay is caused by the characteristics of the speed of light traveling through a fiber-based or copper-based medium. *Handling* delay is caused by the VoIP gateway and has a significant impact on voice quality in a VoIP network.

One element of handling delay is the delay involved with compressing a VoIP packet.

Encoding & Compression Method	**Reduced Bit Rate**	**MOS**	**Compression Delay**
G.711 PCM	64 Kbps	4.4	0.75 ms
G.726 ADPCM	16, 24, 32*, 40 Kbps	3.85*	1 ms
G.729 CS-ACELP	8 Kbps	3.92	10 ms
G.728 LD-CELP	16 Kbps	3.61	3 to 5 ms
G.723.1 CELP	5.3 to 6.3 Kbps (variable)	3.65	30 ms

As seen in the table above, there is a trade off between the bit rate associated with each CODEC and how long it takes to compress and then uncompress the packet. While a PCM data stream has

the smallest compression delay and best MOS score, it also takes up the most bandwidth. The CELP CODEC uses the smallest amount of bandwidth and has an acceptable MOS, but the compression delay is considerably more than the other CODECs.

Another handling delay is the time it takes to generate a voice packet. Within VoIP, the gateway generally generates a frame every 10ms. Two of these frames are then placed within one voice packet; the packet delay is therefore 20ms.

The actual delay to get into and out of the output queue is another source of handling delay, and should be kept to under 10ms whenever possible by using whatever queuing methods are optimal for the VoIP network. Usually network engineers have little control over these delays since they are part of the network router's IOS. (One hopes that the router vendors are doing their best to reduce this delay.) However, some precautions can be taken to make sure the output queue delays are kept to a minimum.

When QoS is enabled on an existing data network, the demands on the network router's CPU and memory will increase. If there are not enough computing resources available for each router, VoIP packets can be delayed or lost. Carriers must pay close attention to the demands placed on their routers and plan for expansion where routers are taxed the most by customer traffic. Diligent capacity planning should include network bandwidth as well as router memory and CPU resources.

Carriers should consider adding or even maximizing the memory and the number of CPUs on their backbone routers to ensure good QoS queuing performance. Using distributed processing and memory resources (such as with Cisco's VIP2-50 cards within the 7500 router line) is the typical choice for many carriers. Using distributed hardware platforms will improve router performance under heavy loads by offloading processes from the central CPU and memory pool whenever possible.

Jitter

Due to the nature of packet switched networks, it is difficult to get the VoIP packets to arrive at their destination at a constant rate. Jitter buffers are used at the receiving end of a VoIP transmission to deliver the VoIP packets to the receiving gateway at a consistent rate. VoIP packets are sent at a constant rate (usually once per 20 ms), and will arrive at their destination at the end of the TCP/IP network at slightly different intervals. A jitter buffer collects VoIP packets at the gateway itself and compensates for missing or late packets using certain algorithms. This functionality is usually found at the VoIP gateways instead of within the TCP/IP network components themselves.

Voice Over IP Implementation: Special Considerations for Carriers

Carriers would seem to have the inside track for obtaining the enterprise business, given their established infrastructures, customer base, and knowledge of data networking developed over the years. This section looks at some of the key considerations for carriers who are moving forward with integrating VoIP services into their infrastructures.

Ease of Use

In the ideal situation, VoIP users should not know they are utilizing aVoIP network as opposed to a circuit-switched voice network. The service provided to users should look and feel the same as the

public telephone services that everyone is familiar with. IP Phones look and feel exactly like telephones at first glance. Some VoIP PC software packages do a decent job emulating a telephone.

If a business or carrier adopts VoIP as a technology alternative, most users should never notice a difference in service (except maybe a very slight drop in the quality of the voice transmission). The VoIP equipment is hidden in the network away from the user population. A PBX can be configured to use the VoIP network and the public voice network, depending on the call destination and the availability and service level on the VoIP network.

The high cost of a typical IP telephone prevents these devices from being common among general ISP users. For many ISPs or IP carriers, VoIP service usually must deal with a PC at one or both ends of the voice call instead of a telephone. If the PC voice client is not easy to use and if the quality of the voice transmission is not high, then PC VoIP clients will tend not to use the VoIP services from their ISPs. If the PCs involved with a VoIP call are not equipped with quality, full-duplex sound cards, PC users will perceive the service as being poor and once again opt for the public voice network instead. These issues tend to severely shrink the VoIP market for most ISPs.

Scalability Concerns

Once VoIP is no longer confined to a single enterprise with slow and predictable growth, a number of scalability issues arise. IP carriers must be able to provide VoIP services to a dynamic, rapidly growing customer base. While the calling patterns found within an enterprise network can usually be anticipated, carriers could be faced with rapid user growth or decline at any one of their points of presence (POPs).

Typical VoIP carriers will offer TCP/IP and voice services before they bring to market a VoIP service. The majority of their POPs will contain TCP/IP routers, and their major POP or POPs will contain a voice switch. In order to offer a VoIP service, carriers will need to install gateways at each of their voice switches—at least one gatekeeper and one billing system. To make the service more robust and scaleable, a second or third gatekeeper could be installed at different POPs within the network.

If the carrier does not have its own voice switch equipment, voice services can be brought into the carrier's POP via T1, E1, or ISDN PRI. However, cost effectiveness will be more difficult to achieve if the carrier is not itself a voice service provider.

Today, a typical gateway can provide from two to four PRI ports for voice calls. In Europe, a gateway with four PRI connections can handle up to 120 VoIP calls at once (96 calls in the U.S.). When a gateway reaches its maximum utilization and VoIP customer calls are blocked due to congestion at the gateway, the carrier must add another gateway to the POP and connect additional PRI lines.

A typical gatekeeper can authenticate and authorize more than 1000 users per second. However, as new VoIP customers appear at different POPs within the carrier's network, most VoIP vendors will recommend adding a gatekeeper at the POPs where the customers connect to the carrier's network.

As additional VoIP customers are added to the carrier's network, the costs associated with expansion are considerable. Additional gateways, gatekeepers, PRI connections, and voice and data capacity will be needed.

Fault Tolerance

As with almost all data carriers, a service level agreement (SLA) usually exists between the carrier and its customers. Because the availability of a typical voice network in a moderately populated area is 100% (or at least this is the perception of most voice network users), a VoIP customer will expect an extremely high SLA from its carrier. Not only will the network need to be available 100% of the time, but also all VoIP packets must be delivered on time. Not only is it difficult to provide 100% VoIP network availability, it is also difficult to measure such an SLA.

When designing a VoIP network for a carrier that will offer a competitive SLA, there must be a great deal of fault tolerance at the physical and logical network. If a backbone link, router, or switch fails, the network should have a fully redundant path to which it can instantly fail over. Network equipment should be protected from power loss or power surges either by a UPS or a battery backup.

Having redundant routers in a network is not enough to guarantee a high SLA when it comes to VoIP traffic. Routing is critical to on-time delivery of packets. If a router failure or noisy backbone link causes the reconvergence of a carrier's internal routing tables, many VoIP packets could be dropped or delivered too late. The logical and network layers of a carrier's VoIP network must be as solid as the physical infrastructure. Internal routing reconvergence should happen as little a possible, and when reconvergence is unavoidable, it should take place as quickly as possible.

While static and floating static routing provide the quickest convergence times, such routing designs do not scale well and are difficult to manage. Most carriers must choose a more practical, dynamic routing protocol such as OSPF, RIPv2, EIGRP, or IS-IS. Through the use of a routing hierarchy, shortened keep-alive timers, and route aggregation (for smaller routing tables), reconvergence times can be minimized as much as possible.

Service Level Agreements

A carrier's VoIP customers will demand a very high SLA. High levels of network availability, low levels of packet loss, and end-to-end latency will often all make up a carrier's VoIP SLA. A VoIP SLA will often be more demanding than a typical ISP SLA due to the high availability and dependability of a typical, circuit-switched, voice network.

Network Availability

A carrier's VoIP network should be available for transmitting data all of the time. However, this SLA is usually not practical for most of today's data networks. A typical VoIP carrier will offer an SLA of 99.99%, which does not include scheduled maintenance windows. Maintenance windows are time periods where the carrier may take down the network to upgrade equipment or IOS software, clean or switch fibers, or perform any other work that could lead to network downtime. These windows are usually during the network's quiet period – typically in the early morning hours on weekends.

Network Latency

The round-trip time for a 64K packet to get from one end of the carrier's network to the other should be about 150ms. This time should be measured from the points of the network where the customer connects to the carrier. Network latency times can be higher for wide area links between continents. For example, 70ms from London to New York City is very reasonable given the speed

of light through fiber. WAN links that connect cities on opposite sides of the Earth will have response time greater than 150ms.

Packet Loss

When more than 0.2% of the packets on a VoIP connection are lost, the users will notice a degradation of voice service. VoIP network capacity planners must make sure there is enough capacity available for all VoIP customers in the network's backbone. All links in the network should be noise and error free so that packets are not dropped due to transmission errors. Network routing reconvergence can lead to unacceptable levels of packet loss even when redundant links exist to compensate for wide area link failures.

Measuring Service Level Agreements

Measuring SLAs is as important as setting SLAs for VoIP carriers. Since it is very difficult to accurately measure and report on SLA targets for VoIP networks, usually more than one network management tool will be required. SLA agreements will need to specify how network SLAs will be measured.

Sending pings from one end point of a network to another will measure network availability and latency. Questions arise concerning the interval in which these measurements are taken and how they are used.

For example, a carrier measures latency across the network at five-minute intervals and then averages all of these measurements per hour. This is a poor measurement for VoIP customers. During any given hour, the network latency could go well over 200ms, and this measuring method would never detect the slip in SLA.

Another question arises concerning packet loss. If packet loss is detected and measured along a backbone link in the network, who can say which customers were affected by this loss? Were VoIP packets lost or data packets? It is impossible to say when the packet loss is measured from the backbone nodes.

The only way to truly measure network availability, latency, and packet loss is from the end-user perspective. Measuring agents must be placed on the end-user's PC or VoIP device and also on the VoIP gateways. There are very few products on the market today that can accomplish this sort of measurement and reporting.

Security Concerns

A carrier offering VoIP services to multiple enterprise clients must take steps to ensure the carrier network and all of the VoIP components are properly secured. VoIP customer calling and billing records must be kept confidential. Customer access codes and the VoIP gatekeepers must not be easily accessed by Internet hackers.

While Cisco Systems has implemented VoIP gateways and gatekeepers within its own line of routers, many other vendors offer solutions based on Windows NT or UNIX platforms. If an NT or UNIX administrator does not do a proper job of securing these hosts, confidential customer data will not be protected and theft of telephone service could occur.

Billing Hierarchy

The most difficult part of offering VoIP services as a carrier is the management and logistics concerned with billing and invoicing customers. A large number of users must be added or modified within the VoIP billing system on a continual basis. In addition, the telephone tariffs charges between countries will often change at least once per day. A significant amount of resources will be needed in order for a carrier to properly manage its VoIP billing system.

If a VoIP carrier is already offering voice services to customers, chances are that there is a group of people already in charge of setting telephone tariffs and examining least-cost-routing of telephone calls to other countries. These tariffs will vary in each location that the carrier has a telephone switch. This is also true for each location where the carrier has a VoIP gateway. All of this tariff information must be entered into the VoIP billing system so that customers that utilize VoIP at each of the carrier's POPs are charged the correct tariff. Many carriers will need to augment their least-cost-routing team so that they will have the resources to update the VoIP billing system as well.

If a carrier offers dial-up Internet access to customers and then includes a VoIP service for any interested members, the carrier will need to develop a business process to invoice and record payments from all customers for their VoIP usage.

If the carrier only offers its VoIP service to other, down-stream carriers, then a single bill can be produced for each of the smaller carrier customers. A majority of the billing logistics can be spread out to the carrier customers in this case. The VoIP billing system will keep a customer hierarchy and allow each carrier customer access to the billing database. These carrier customers can then create, modify, and invoice their own VoIP customers. The carrier customers then have the option of adding a premium to the charges billed by the parent VoIP carrier.

The Next Step

Since the use of the Internet and the Internet Service Provider has become so widespread and popular, the switched telephone networks around the world are being pressured to offer increasingly more capacity to customers. The duration of a typical modem call to an ISP is a lot longer than the typical voice call. Carrier class voice switches are very expensive, and the classical voice network was never designed to carry data traffic in the first place.

Many voice carriers have now built a second network for carrying data traffic and have become ISPs themselves. But as the Internet continues to gain in popularity, the carrier now has two networks it must continue to expand, maintain and support. Operation of both these networks is the biggest part of the carrier's cost associated with providing service.

The alternative is to build a single network that can carry both voice and data traffic. With voice and data convergence, a carrier could cut its infrastructure costs and then lower its prices for customers, enabling the carrier to gain market share and become more profitable.

Imagine the massive savings for a carrier once it can use a single network for both voice and data services. Now only half the support and maintenance efforts, half the infrastructure equipment, half the co-location space, half the hired engineering expertise, and half the management tools are needed to operate the network. The reduction in cost is very substantial.

Getting Help

Anticipated cost savings from VoIP will not be realized if the underlying network infrastructure is inadequate to support it. Many organizations need outside expertise to help them take full advantage of VoIP, often requiring support from an external service provider to analyze, design, implement, and deploy this new technology.

Lucent NetworkCare℠ Professional Services (Lucent NPS) is the world's leading network consulting and solutions provider. Our 5,500 engineers, architects, and consultants working in over 40 countries are equipped with knowledge derived from years of Bell Labs research and thousands of field engagements.

Most of the major network hardware and software vendors and service providers rely on Lucent NPS for help with their networks. As the leader in both voice and data technologies, Lucent NPS is uniquely positioned to help companies faced with the challenge of creating next-generation networks based on the convergence of voice, data, and video.

For further information, see the Lucent NPS website at http://www.lucent-networkcare.com, or call 1-877-369-1115 in the U.S. or 1-727-217-2303 outside the U.S.