

Realtime
publishers

"Leading the Conversation"

The Essentials Series

VoIP

by Ken Camp


VoIP Primer: Introduction to Voice over IP

In order to understand and appreciate some of the business and technical issues driving the popularity of carrying voice over the Internet Protocol (IP), you need a basic framework of how telephony works. This article provides a brief overview of the basics so that you can explore why Voice over IP (VoIP) is a smart solution in today's network environment.

As communications has evolved over time, we've reached the point at which we can pick up the phone and call someone on the other side of the world. The delivery of information, or an idea, is nearly instantaneous. Our culture has shaped us to expect things to happen very quickly, and we have become impatient about delays in delivering information in any form.

Alexander Graham Bell made the first telephone call on March 10, 1876, but at that point, nobody fathomed the impact this technology would have on us all. AT&T put the first transcontinental phone lines in service in 1915, and communications again changed forever.

To transmit a telephone conversation, you must send sound over long distances. In a phone call, this sound is voice, but it could just as well be music or some other sound. The public switched telephone network (PSTN) has been optimized to carry voice traffic for more than 100 years. Today, separate voice and data networks may be too costly. VoIP allows integration of voice and data onto a single network.

 *IP Telephony Demystified* (McGraw Hill) contains more detailed explanations of many of the concepts touched on in this article.

Telephony Basics: Transmitting Voice

The PSTN is a circuit-switched network. Every time you make a phone call, telephone switches set up temporary connections between parties involved and establish a circuit that is used for a voice conversation. The Internet is a packet-switched network. Information, in digital form, is carried in packets over many diverse paths, or routes. The information in these packets can be anything from email messages to voice telephone calls. All that is required is that the message be encoded in some binary form for transmission.

Our vocal chords are used to produce sounds. Sound is a series of vibrations in air pressure. These sound waves travel through the air, much like ripples on a lake travel through water. Just like ripples in water, the farther they get from the source, the weaker they become.

Early telephones converted the physical energy of voice sound waves into electrical impulses that could be sent over copper wires. As electrical energy can use repeaters or amplifiers to regenerate or boost the signal, telephone calls can be carried over great distances. At the other end, the electrical signal is converted back to sound waves at the telephone handset. This function in the telephone set is referred to as a transducer. For a basic telephone call, the network establishes a connection between parties, encodes the signals, then transmits them. It decodes the signal at the receiver. When the call is finished, the network then disconnects and updates usage and billing records to bill for the call (see Figure 1.)



Figure 1: Phone calls on the PSTN.

Signals that are continuously variable are commonly thought of as being analog, whereas discrete signals are considered digital. For example, water flowing in a river is a form of analog signal. Flow meters can measure the variable rate at which the river flows. A typical light switch, in contrast, controls a digital signal. The light has only two states—on or off. Digital signals typically have predefined levels that represent specified conditions.

Analog signals are represented by a waveform or sine wave as Figure 2 shows. A digital signal can only represent predefined values. You frequently deal with a binary system in networking. There are two predefined levels. These are represented in binary math as a 0 or 1. A digital system can have a large set of predefined values. The alphabet is made up of 26 predefined values represented by the letters. A digital system doesn't necessarily have to be a binary system.



Analog signals are variable across an infinite of values



Digital signals represent a predefined set of discrete values

Figure 2: Analog and digital signals.


The square wave is not a sine wave but has enough similarities to suggest a relationship between the two. The relationship was completely described by mathematician Jeanne-Baptiste Fourier.

Packetizing Voice

To transmit voice over a data network, the voice conversation must be packetized. To transmit the conversation in packets, you need to digitize it. Voice digitization is not a new technology. In the 1960s, the phone companies began using a digital carrier system called T-carrier. Initially used as a trunking technology between phone company central offices, it's what is commonly called a T1 circuit today.

As your voice represents an analog signal, you must first convert it into digital format. This conversion is commonly accomplished through a technique known as pulse code modulation (PCM), using a coder and decoder (or codec). Using PCM, analog voice conversations are sampled 8000 times per second. A technique called pulse amplitude modulation (PAM) is used to convert each sample into one of 255 possible 8-bit words. Through a process of compressing and expanding (called companding), noise is reduced to help eliminate background hiss and changes in volume.

There are several voice processing standards that were developed by the International Telecommunication Union (ITU, formerly the CCITT) to standardize encoding of digital speech. In the PSTN, the G.711 standard is used worldwide.

 More information about encoding standards is available at the ITU Web site at <http://www.itu.int/>.

Understanding these encoding schemes isn't crucial to the process of packetizing voice for VoIP service, but provides useful background to understand the telephony processes that let you merge voice and data onto a single network infrastructure.

Connection Oriented vs. Connectionless

The PSTN is a circuit-switched network and is oriented to connections. Each telephone call represents a connection. Packet networks can be connection oriented, but a connectionless network requires no setup. Connectionless networks can't necessarily guarantee that packets will be delivered in the order they were transmitted. As packets can take different routes through the network, they can arrive at different times. In a VoIP network, this means the network nodes and end devices need resources to store packets until enough have been delivered that they can be reassembled into a message for delivery. This action all happens in milliseconds so that people don't notice a difference in call quality.

Packet Switching—Store-and-Forward

Although the telephone network uses circuit switching, there are other switching technologies that are quite effective. One example of store-and-forward switching is airline travel. Passengers travel between cities that have airports and can switch en route to reach their final destination. This network topology is referred to as "hub and spoke." The system has switching points, or nodes, around the world. To get from the starting point to a final destination, a passenger may need to switch routes from one plane to another at a node site along the way. In some cases, passengers are delayed (or buffered) while waiting for the next flight. With luck, all passengers are forwarded to their final destinations.

In data networks, links between switches are shared on demand. Routing calculations determine the best path for packets to take toward their destinations. Data networks are described as “bursty” in nature. That means that traffic patterns and volumes are sporadic and often unpredictable. This unpredictability doesn’t pose a problem to Web or email traffic. These types of traffic are not real-time, and slight delays don’t cause problems. Voice traffic is real-time in nature. Delays of more than 250 milliseconds (ms) or variations in timing degrade voice quality and can make the network incapable of carrying voice traffic.

In packet networks, you overcome this circumstance by breaking large chunks of information, such as a telephone call, into smaller IP packets that can take different paths through the network. In a store-and-forward network, each packet must carry the source and destination address of the information contained in the packet payload. Store-and-forward networks use statistical multiplexing techniques, such as first in, first out (or FIFO), so they’ve always been well suited for traditional, bursty LAN data traffic.

Today, packet networks carry huge volumes of data with far greater reliability and performance guarantees than in the past. Quality of service (QoS) to support call quality can be implemented through assessing the network requirements before implementing VoIP, adding bandwidth capacity where necessary, and implementing widely accepted methods for quality assurance such as Multi-Protocol Label Switching (MPLS) across a wide area network (WAN).

Replacing the Enterprise PBX with VoIP

Companies have long realized that maintaining separate voice and data networks is an expensive proposition. The telephone system, or PBX, is a large financial investment. As LAN and WAN technologies have become simpler to manage and VoIP solutions more widely available, it makes good business sense to integrate the two networks and consolidate resources. The benefits of this consolidation are numerous:

- A single wiring scheme can eliminate separate voice and data wires. Although most companies long ago standardized on Category-5 cabling, voice and data still often compete for cable resources.
- The PBX can be a very expensive investment. Integrating voice services into multi-function routers and VoIP gateway servers can reduce investment cost significantly.
- Administration skills can be consolidated into a single work group. Large organizations can benefit from consolidating internal support organizations.
- Large companies with many locations can simplify billing by integrating voice and data into a single network. This integration leads to simplified billing from providers, which means fewer company resources are needed to validate bills for payment. For a large enterprise, the impact can be significant.

A VoIP solution for PBX replacement might be as simple as implementing VoIP phones and services internally on the corporate network, with a single gateway to the PSTN to provide a connection for calls outside the company. One advantage to this approach is that simple IP phones can plug directly into the Ethernet LAN to provide telephone service within the corporate network.

This solution provides basic telephone service. Telephone calls are routed over the LAN for internal calls or directed to the gateway to reach parties on the PSTN. Figure 3 shows this simple implementation of a PBX replacement.

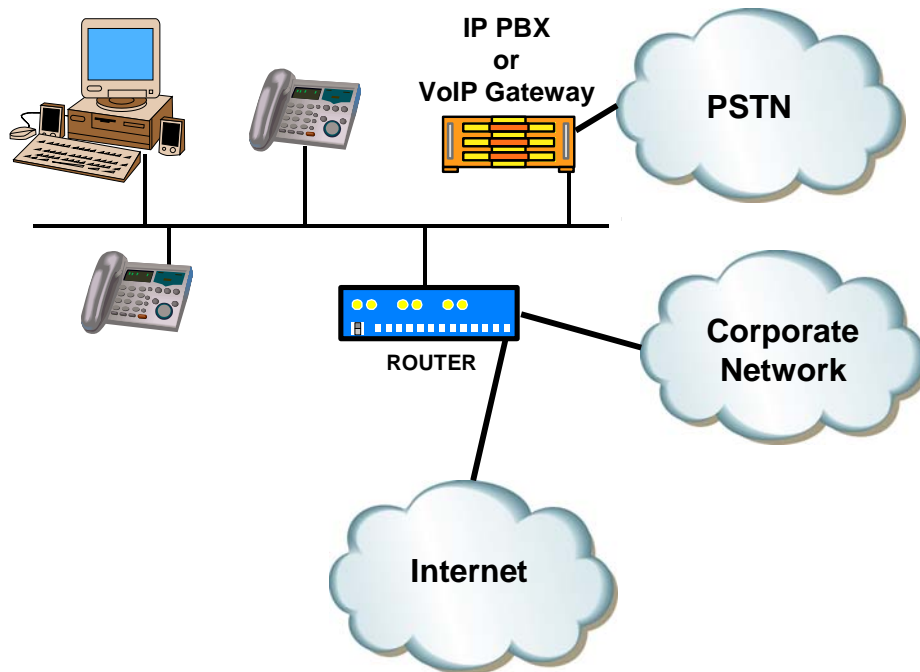


Figure 3: A simple PBX replacement example.

As VoIP systems have matured, standard features such as call waiting, placing calls on hold, conference calling, transferring calls, and music on hold are now common features in most business VoIP solutions. Voicemail systems are now commonly digital systems built on IP networked servers that integrate well with VoIP solutions.

The Converged Network

One of the greatest benefits a company can gain from VoIP is an evolution well beyond the simple consolidation of the network onto a single wiring infrastructure. When networks converge, integrated services can be realized using new approaches. In a fully converged network, the integration of computers and telephony leads to customer service opportunities.

Customer data, collected over both telephone and computer systems, can enable new approaches to customer service. Call centers can be geographically dispersed, even integrating telecommuters working from home. A single customer service center can handle both telephone calls and Web services.

Imagine customers clicking a Help button on a Web site that sets up a VoIP phone call to a customer service representative. With VoIP in a fully integrated environment, the customer service representative might actually get a copy of the Web screen that the customer last visited, complete with history of what this customer had been trying to do. How would you react as a customer if you clicked Help and the service representative came online and said “It looks like you’re trying to find information about our new product X. I’m sorry the Web information wasn’t more helpful. How can I help?” (see Figure 4).

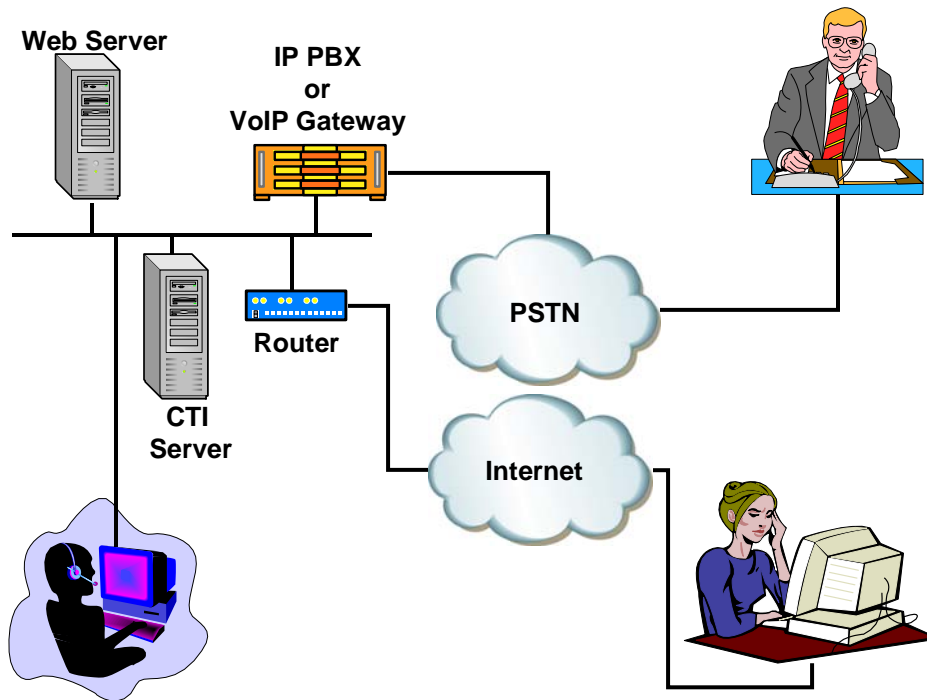


Figure 4: Computer telephony integrated into a converged network.

Computer telephone integration at its finest and most developed can provide a powerful competitive edge to customer service organizations.

Summary

Network evolution and corporate culture take time and effort to change. They evolve continually. Although no company can jump from legacy telephone systems to a fully integrated voice and data network overnight, VoIP provides a point of entry for your company to begin moving into a competitive converged network.

VoIP Management

By Ken Camp

A Network Management System (NMS) is a combination of hardware and software used to monitor and administer the addressable and manageable elements of the network. In IP networks, Voice over Internet Protocol (VoIP) services introduce a new set of manageable network elements that perform telecommunications service functions. These elements typically include gateways, call management servers, emergency responders, voice mail servers, and so on.

General network management involves functions such as network planning, traffic routing, user authorization, configuration management, fault management, security management, performance management, and accounting management. Many protocols exist to support network and network device monitoring and management. Common protocols include Simple Network Management Protocol (SNMP), Common Management Information Protocol (CMIP), Web-Based Enterprise Management (WBEM), Common Information Model (CIM), Transaction Language 1 (TL1), and Java Management Extensions (JMX).

When implementing VoIP systems, network management systems take on a new, crucial role in enterprise service delivery. Enterprises need to bolster their management capabilities to test and manage Quality of Service (QoS), performance, and availability VOIP metrics. To get started, companies should analyze their business requirements and determine key performance and QoS metrics.


A comprehensive, enterprise-wide data collection mechanism is required to provide effective service assurances. Collecting as much data about the network as possible will aid in the ability to ensure call quality and consistency of service.

Network Monitoring

An NMS constantly monitors and notifies the network administrator via email, pager, or other alarms in the event of outages or anomalies that exceed defined thresholds. Monitoring is vital to service assurance and VoIP management. An NMS continually monitors the network for problems that result from overloaded and/or crashed servers, network connections, or other devices. For example, to determine the status of a Web server, monitoring software may periodically send an HTTP request to fetch a page; for email servers, a test message might be sent through SMTP and retrieved by Internet Message Access Protocol (IMAP) or Post Office Protocol 3 (POP3).

Status request failures—such as when a connection cannot be established, it times-out, or the document or message cannot be retrieved—produce a pre-defined action from the monitoring system. These actions vary: an alarm may be sent out to the resident systems administrator, automatic failover systems may be activated to remove the troubled server from duty until it can be repaired, and so on.

Some of the most important characteristics of network elements monitored in the IP network include CPU utilization, physical memory, disk space usage, virtual memory, and fans and power supplies. Many systems monitor temperature to ensure a proper operating environment is maintained. Monitoring of system backups is incorporated to ensure positive confirmation that backup jobs run as scheduled. Many companies monitor Web server software (typically Apache or Internet Information Services—IIS), directory services systems, and Domain Name Service (DNS) servers. Security monitoring is often incorporated into the monitoring performed in this *command center* environment.

 Network monitoring tools rely heavily on standard network resources such as ping and SNMP. Basic tools such as SNMP servers, Cricket, IPCheck, and IPSwitch WhatsUp are widely used for basic network management in small to midsized networks. Larger, more sophisticated networks may employ simple tools such as SolarWinds or more advanced tools such as HP OpenView Network Node Manager (NNM) or IBM Tivoli Netview.

VoIP services introduce the need for specialized management tools. In a Cisco VoIP deployment, the AVVID architecture provides very specific tools oriented to Call Manager. Cisco's purchase of Digital Fairways, a VoIP software management company, in December of 2005 attests to the importance of management.

Nortel's Business Communications Manager strategy incorporates the company's Proactive Voice Quality Management (PVQM) to continuously measure the quality of the users' experience. The PVQM strategy integrates tightly with NetIQ AppManager (which also integrates nicely with Cisco solutions).

Managing VoIP services raises the need to monitor both the VoIP service elements and QoS facets of network performance to ensure acceptable call quality.

VoIP Service Elements to Monitor

Voice traffic carries a set of performance expectations users have come to expect through years of using the telephone. VoIP services introduce a new range of network elements to monitor.

Whenever a device (for example, phone, gateway, and gatekeeper) registers with the network, there will be an auditing entry to review. Problems with device registration, for any reason, can impact service availability. You'll want to be alerted when the number of registration attempts or failures exceeds pre-defined thresholds. If the number of registered telephones changes dramatically, it could be a signal that there is a problem with the VoIP network. Gateway registration monitoring will help identify new or missing gateway servers.

Call monitoring isn't eavesdropping on individual calls. It's really call traffic monitoring. It involves monitoring incoming and outgoing call volumes to identify failures. If your VoIP system supports fax calling, attempted fax calls also need to be monitored. Call monitoring looks at four specific areas:

- Calls in progress—When a VoIP phone goes “off hook,” a call is deemed in progress until it goes back “on hook.” If every call in progress connects successfully, the number of calls in progress will equal the number of active calls. When designing the VoIP network, you'll need to establish an upper-limit threshold for the number of calls that can be in progress at any given time.
- Active calls—Active calls have successfully connected a voice path. Again, when designing the VoIP network, you'll need to establish an upper-limit threshold for the number of active calls that can be handled at any point in time.
- Attempted calls—Designers strive to ensure that all calls attempted will be completed successfully, but such isn't the case in the real world. Monitoring calls attempted over time yields data that aids in identifying peak periods and the busy hour call attempt (BHCA) value.
- Completed calls—A completed call is any successful active call that completes without an abnormal termination code. Monitoring completed calls over time is also useful in identifying peak periods and the BHCA value.

VoIP services need to interconnect to the Public Switched Telephone Network (PSTN) through gateways. In addition to gateway monitoring, it is vital to monitor the PSTN side of the VoIP service network. PSTN connections are frequently established using ISDN Primary Rate Interface (PRI) channels over T-1 circuits. Monitoring active PRI channels, especially over time, can help identify call patterns and busy hour peak call volumes. Baseline data can also be used to identify underutilization of circuits. Data trending helps in capacity planning and the growth and maturation of the VoIP service.

One benefit in deploying VoIP services is the conference bridging capabilities. If your deployment supports conferencing, you must configure the maximum number of audio streams that will be supported. Monitoring will ensure that the number of available audio streams meets acceptable service levels.

IP phone functionality requires continual monitoring for service assurance. You should monitor IP phones for their registration status, the validity of their dial tones, jitter, latency, and lost packet count. These QoS parameters directly affect service delivery.

Monitoring Bandwidth and QoS

Voice traffic requires specific bandwidth based on the codec used in the VoIP design. G.711 requires about 64Kbps for each direction of a bi-directional call. G.723 and G.729 require significantly less bandwidth due to compression, but congestion can severely impact call quality.

When you add new applications to your network, there is always a risk of oversubscribing links. Oversubscription leads to congestion, and congestion can have a deleterious affect on call quality. Packet loss and increased latency are common side effects of congestion and can render VoIP services unusable.

In order for VoIP users to receive an acceptable level of voice quality, VoIP traffic must be given priority over other kinds of network traffic, such as data. The main goal of QoS is to ensure that VoIP traffic receives the preferential treatment it deserves, thereby reducing or eliminating the delay of voice packets that travel across a network. You should monitor the following metrics that affect VoIP call quality:

- Delay or latency is an estimate of the network delivery time expressed in milliseconds. It's measured as an average value of the difference between the time stamps noted by the senders and the receivers of messages. It is measured when the messages are received. The end-to-end delay, or latency, as measured between endpoints is a key factor in determining VoIP call quality.
- Jitter is also called delay variation. It indicates the variance of the arrival rate of packets. Jitter points directly to the consistency or predictability of the network. It is a call quality factor known to adversely affect call quality. Networks can compensate for jitter by implementing *jitter buffers* to normalize the timing of the traffic flow. Jitter buffer loss occurs when jitter exceeds that which the jitter buffer can hold. Jitter and jitter buffer loss affect call clarity, which affects the overall call quality.
- Packet loss indicates a packet lost during transmission. In VoIP, this could mean the loss of an entire syllable or word during the course of a conversation. Obviously, data loss can severely impair call quality. Monitoring systems measure the number of packets that were expected against the number actually received.
- Mean Opinion Score (MOS) is a subjective measure used in voice telephony, especially when codecs are used to compress the bandwidth requirement of a digitized voice connection from the standard 64Kbps PCM modulation. MOS is generated by averaging the results of a set of standard, subjective tests. In the past, a number of listeners rate the heard audio quality of test sentences read aloud by both male and female speakers then rate each as follows: 1-bad, 2-poor, 3-fair, 4-good, 5-excellent. The MOS is the arithmetic mean of all the individual scores. In current systems, MOS is often determined through software algorithms.

ITIL Information and Network Management

The Information Technology Infrastructure Library (ITIL) is a framework of best practices that promotes quality computing services in the information technology (IT) sector. ITIL presents a comprehensive set of management procedures with which an enterprise can manage its IT operations. Since the mid-1990s, ITIL has been treated as a standard for IT service management.

ITIL is published in a series of books, each of which covers one topic. ITIL Service Support describes best practices in service assurance for the following areas:

- Change Management
- Release Management
- Problem Management
- Incident Management
- Configuration Management
- Service Desk

Summary

Without NMS focused on the VoIP service, companies will fly blind as to service assurance. VOIP services that lack management are likely to deliver poor quality that cannot be tracked to any service-level metrics. It's vital that the appropriate QoS metrics be monitored continually. These should include some combination of MOS, jitter, latency, call completion and quality, and voice quality.

Needs Assessment and Network Readiness

When companies explore VoIP solutions, they focus on key business areas. Generally, the cost savings of a converged network is a business driver. When considering business needs, it's crucial to the success of a VoIP deployment that you do not set unreasonable expectations, and it's important to understand the business drivers behind the VoIP implementation. If reducing costs is the primary driver, there will be different factors to consider than if the primary driver is deploying new, converged applications. Understanding the impetus for the project helps maintain focus on factors that can ensure success. Communicating the objectives clearly with everyone involved helps maintain a clear view of the expected results.

Understanding Existing Voice Needs

Once the decision is made to explore VoIP as a network service, the next step is to gather data and assess the needs of the business. It is critical to understand the calling patterns in existing telecommunications services. Start with call usage reports from the current telephone provider or system. Whether the current telephone service is delivered through a company-managed PBX or a telco-provided Centrex-like service, this information is crucial and should be readily available. Key elements to identify and understand include the busy day of the week and month, the busiest hour of the day, and the upper call volume that must be supported. Some remote offices may have different calling patterns and requirements than either the main corporate site or other remote locations. It's important to be flexible and incorporate the needs of all business units.

Because many businesses are cyclical in nature, it's wise to review these calling patterns for the past several months or even for the past year. Retail companies may experience much higher call volumes as Christmas nears, whereas financial services companies' busy call period might fall during tax preparation times. Remember that voice services need to be designed to support peak activity periods, not just everyday call volumes.

The key in identifying business needs lies in understanding your business flows and volumes. Call patterns for inbound and outbound calls may be very different. It's important to know what is required to support your particular business.

Remember that the telephone network is a mission-critical facet of business operations. The phone network is built to support 99.999 percent uptime, which equates to roughly 5 minutes of downtime per year. Corporate networks rarely provide this level of reliability today. Thus, some redesign effort may be required in order to provide suitable reliability in the data network.

Network Design Considerations

Existing networks may have evolved from departmental local area networks (LANs) and grown over time. Many have been methodically redesigned every few years. Just as the PSTN has been designed and optimized to support voice calls, these IP-networks have been designed and tuned to support known and understood data applications. Applications in use today may have very different network performance requirements than VoIP services.

Many VoIP projects begin with the implicit understanding that network upgrades will be required. Network elements such as routers may already be running at high CPU utilization or unable to support VoIP services. WAN links may be overburdened supporting existing data applications. Network capacity in terms of bandwidth and processing power are vital factors in deploying any new service.

The pitfall of upgrading prior to VoIP is the need for multiple upgrades. A smart approach is to assess and understand calling requirements in conjunction with existing data services. Evaluate planned new data services as well as voice requirements. Rather than upgrading prematurely, focus early on data gathering and information analysis. This methodical approach will yield a better network design with the capacity to support a successful pilot deployment and expand to provide for all the needs of the business.

At this point, re-evaluate the scope of VoIP deployment. If VoIP will only be used in a single LAN environment, for internal calls only, design preparation will be much different than a multi-location network with gateways connected to the PSTN in several remote offices.

Voice Compression Methods

Different VoIP encoding mechanisms use different levels of compression. Although G.711 encoding is widely used, it generates a 64Kbps voice stream. For some companies, quality considerations might make a different encoding scheme more appropriate. Greater compression of the voice reduces the bandwidth requirement, but voice quality may suffer as a result.

Table 1 provides a comparison of several encoding schemes that are all widely used. It lists the Codec types and algorithms used, the bit rate and sample size, and the algorithmic encoding delay, then compares the Mean Opinion Score (MOS) for these approaches. It is important to remember that factors such as encoding delays are cumulative. If the total delay exceeds 250 milliseconds (ms), call quality will suffer. These algorithms are described in depth in ITU-T standards (<http://www.itu.int/>) and a wide variety of papers. Vendors and consulting partners will also be a valuable resource in identifying the best approach for your business needs.

ITU Codec	Coding Scheme	Bit Rate	Sample Size	Encoding Delay	Mean Opinion Score
G.711	Pulse Code Modulation (PCM)	64Kbps	8 bits	<1 ms	4.4
G.726	Adaptive Differential Pulse Code Modulation (ADPCM)	32Kbps	4 bits	1ms	4.2
G.728	Low-Delay Code Excited Linear Predictive (LC-CELP)	16Kbps	40 bits	2ms	4.2
G.729	Conjugated Structure Algebraic Code-Excited Linear Predictive (CS-ACELP)	6Kbps	80 bits	15ms	4.2
G.723.1	Algebraic Code-Excited Linear Predictive (ACELP)	5.3Kbps	160 bits	37.5ms	3.5

Table 1: Voice encoding schemes.

Pulse Code Modulation (PCM), or G.711, is the approach used in the PSTN today and is widely used in VoIP systems. Virtually every VoIP equipment vendor supports G.711. G.726, known as Adaptive Differential Pulse Code Modulation (ADPCM) could reduce the bandwidth requirements by half while only sacrificing .2 of a point on perceived quality in the MOS. G.728, or Low-Delay Code Excited Linear Predictive (LC-CELP) coding is widely used in voicemail systems for digitizing stored voice messages. G.729 can deliver an 8-kilobit sample with less than 16ms of processing time. This G.729 codec standard is often used in Voice over Frame Relay (VoFR) and is supported by many frame relay equipment vendors.

G.723.1 or Algebraic Code-Excited Linear Predictive (ACELP) actually creates models of the human voice, then predicts what the next sound will be. It encodes the difference between the actual sound and the predicted sound, and the difference is transmitted to the receiving end. In some implementations of this codec, there have been complaints that women and children's voices were not represented accurately, probably due to their higher pitch.

G.711 (PCM) is the most widely used voice digitization technology, but other encoding schemes merit consideration. PCM creates an 8-kilobit sample, which requires 64Kbps of network bandwidth and is often viewed as being the closest VoIP approach to traditional telephone call quality.

For decades, quality in the telephone network was rated by something called an MOS. This score was assigned by putting a group of people in a room and having them listen to sound in headphones. Evaluators would rate the sound quality from 1 to 5, 5 being the best quality. In real-world telecommunications, the human ear can clearly distinguish between a 4 MOS and a 4.5 MOS. Today, MOS is often determined using ITU Recommendation P.800, which details methods for subjective determination of transmission quality.

Security: Firewalls, VPNs, and Network Address Translation

When implementing VoIP solutions, network security is as great a concern as reliability and call quality. Corporate networks may include firewalls and multiple connection points. These security devices may impact VoIP service delivery. The more complex a rule set in a firewall, the more latency it induces to the data flow. Remember that latency is cumulative and counts toward the 250ms maximum tolerable delay.

VPN services are deployed in two typical fashions. Point-to-point VPN solutions may be used to connect remote offices over the Internet. Many companies also use VPN services allowing employees to connect to network resources while telecommuting or away from the office. Encryption algorithms consume processor power. A VPN device running DES or Triple DES encryption as a VPN end point will add further latency as packets are encrypted and decrypted. If the VPN end point is a firewall, this CPU load problem may be further compounded.

Network address translation (NAT) is often used within the corporate network as a segmentation mechanism. It can also provide an added layer of security, as the private addresses are not routed outside the company to the Internet or extranet partners. In some cases, NAT usage that works just fine for data applications has adverse impact on voice. It can result in failure to complete call setup, leaving users unable to make phone calls.

Many VoIP applications use Java applets to support user features. Java controls for speed dial list, feature programs, and attendant consoles are quite common. Java applets may not work as expected in an environment using NAT. It's important to work closely with your VoIP vendor in designing and testing any VoIP solution.

Security concerns reinforce the importance of pulling the appropriate technical team together for network assessment and readiness testing. It's crucial that the telecommunications, IT, and network security teams collaborate to be successful.

Making Sure Your Network is Ready for VoIP

Making sure the network can fully support VoIP is a significant task and not something to be overlooked or pushed aside. Remember that your data network has been tuned, honed, and optimized to support data applications. Voice traffic places new, sometimes-inflexible demands on your network. Call quality requires low latency and jitter, low packet loss, and adequate bandwidth.

It's important to test the characteristics of the network to ensure VoIP service requirements can be met. Understanding traffic aspects such as traffic type, frame size, prioritization schemes in use, utilization (both normal and peak), network latency, jitter, and loss will all help in benchmarking application performance.

Gateways to the PSTN may have specific performance requirements on both the data network side and the PSTN side of the gateway. Both need to be considered. Data network bandwidth must complement the trunking requirements when connecting to the traditional telephone network. Overlooking details such as this can result in performance bottlenecks between networks. Use a combination of network analyzers and monitoring tools to evaluate network throughput and performance. There are many tools available through highly capable partners.

Simple VoIP calculators like those available at <http://www.voip-calculator.com/calculator/> can help you identify key factors:

- Estimating the bandwidth needed to support a known number of phone lines
- Calculating the bandwidth required to support busy-hour call volumes
- Determining how many paths might be needed across a wide area network (WAN)

Network analyzers such as SolarWinds (<http://www.solarwinds.net/>) provide extensive performance measurement tools to help understand the current levels of performance your network provides.

There are a variety of products and tools to assist in measuring network utilization and classifying types of traffic that are VoIP-specific. Traffic-generation tools and VoIP simulators can thoroughly test the readiness of your network and help identify problem points. These tools help ensure that network upgrades identify and eliminate potential service problems before you integrate voice into the network. VoIP-specific tools such as NetIQ Vivinet Assessor (<http://www.netiq.com/products/va/default.asp>), Viola Networks NetAlly (<http://www.violanetworks.com/interior.php/sid/1/aid/2>), and ClearSight Networks Analyzer (<http://www.clearsightnet.com/products-analyzer.jsp>) can provide specific help with simulating VoIP traffic on your network to ensure your network is ready and able to support VoIP before any pilot efforts are begun.

Planning for Success

There are some basic steps every company must take to achieve a successful VoIP implementation. IP telephony is one of many new technologies being deployed today, but the basic considerations remain the same for any new technology:

- **Analysis**—Don't make the decision to move to IP telephony before identifying why you're making the change. Understand your expectations of what benefits will follow. The success of a VoIP deployment hinges on the data gathering and analysis. The more you learn ahead of time about your requirements and your network's ability to meet the needs, the smoother the entire process will be.
- **Planning**—How many times have we all heard "People don't plan to fail, they fail to plan." Do your homework and prepare. Be methodical and document every facet of the entire planning process.
- **Testing**—Once you understand the network requirements, test to make sure it can support your VoIP requirements. Thorough analysis on the front-end ensures that you know what to test. Integrate your testing to support a fully converged network. Don't just test VoIP services. Existing applications on the WAN and LAN need to be tested while voice is being simulated too.

-
- **Acquiring the Right Resources**—For many enterprises, resource acquisition could mean sending your technical staff to training beforehand. For other companies, it may mean bringing in outside help from either a consultant or a trusted vendor partner. There is no shame in admitting you need expertise from outside and working with a consulting partner. Leverage expertise where you can.
 - **Looking Ahead**—Don't look at VoIP deployment as the final destination. Give appropriate consideration to scalability for the future. Can your network continue to grow to support business? Once you can deliver voice traffic, can you continue to grow at an acceptable pace? How will anticipated new application services or bandwidth demands impact voice? This point is a good time to review the overall future capacity of your network.

Debunking a Myth

There was a widespread myth associated with early VoIP services that said “All you have to do is add IP telephony to your network and it will work just fine.” Don't believe it. Voice service is mission critical to the operation of any business. Anything you add to your network, whether it's an application, new software, or new devices will have an impact. If your equipment vendor tells you to add their product and your “voice will ride free,” be very skeptical. You will get what you pay for, and it's crucial that you protect your business network services, both voice and data.

Using Vendors, Partners and Tools

As this article has touched on the idea of using outside resources, it's worth reviewing how outside consultants will work for you. They will help look out for your best interests. It's what they get paid to do, and there are some VoIP consulting firms who specialize in VoIP technology. Their reputation is built on how well they take care of client needs. They build business through word of mouth referral and will be your advocate for success. They can ensure that you are being well served by your vendors.

Equipment vendors and service providers will gladly offer consulting services, and their sales engineers will graciously come in and help you design a solution. It's important to keep in mind that these people are specialists in their specific products, working for their company. They'll do a good job based on their own experience, but they often don't have comprehensive training on the workings of competitive solutions. Use your trusted vendor partners, but don't give them free rein to design your network. You need to remain in control of the project overall.

Summary

The key to success in any network implementation begins with the first steps. Needs analysis and readiness assessment represent two of the early stages in evaluating your network requirements and the capacity of your network to support VoIP services. Don't sidestep these important steps in the process and your VoIP deployment will succeed and fulfill your business needs.

Delivering Call Quality with VoIP

Over the past hundred years, the public switched telephone network (PSTN) has evolved into a finely tuned mechanism for delivering voice traffic. In business, for years the phrase “toll quality voice” has been used to describe a service level suitable for business—good enough to hear a pin drop. This network has been built and conditioned to do one thing really well—deliver voice conversations.

Data networks use Internet Protocol (IP) for delivery. IP is described as being unreliable and having no guarantees. It is a “best efforts” protocol. It was designed to carry sporadic, unpredictable traffic loads that burst to peak volumes at times. At that transport layer, above IP, there is also Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). TCP offers the ability to guarantee delivery through a process of synchronization and acknowledgement messages, but this delivery guarantee also adds overhead to the data flow. Table 1 provides a comparison of voice and IP traffic.

Voice Traffic on the PSTN	IP Network Traffic
Connection-oriented—A dedicated path is established for each telephone call; calls are long duration (4 minutes on average)	Connectionless—Conversations are packetized and transmitted over the best route based on routing protocols; packets are small, so conversations are cut up into many packets
Delivery is guaranteed once the call path is established	Best efforts are made to deliver traffic but there are no guarantees
Designed to use specific bandwidth; the PSTN use a 64Kbps voice channel	Uses the bandwidth that is available
Real-time voice traffic is very sensitive to delay	IP data traffic is delay insensitive

Table 1: Voice and data traffic comparison.

IP data networks have very different basic characteristics than the PSTN:

- Unlike the dedicated circuit connection used in the PSTN, data traffic is short in duration and *bursty* in nature. As packets are short, they can be routed over different paths and carry delivery information in each packet. It is connectionless.
- IP packets aren't delay sensitive. Email messages can be delayed 60 seconds or 60 minutes without problems. IP traffic is historically non-real-time traffic. Delays are expected in an IP network. The PSTN is designed to minimize delay and provide near-instantaneous delivery of voice conversations.
- IP uses the available bandwidth when it has data to deliver. IP doesn't require dedicated bandwidth.

IP networks today are designed, redesigned, and modified to support carrying any kind of traffic. With digitization, any media can be carried inside IP packets. The most common are data, voice, and video. The VoIP challenge faced today is delivering business-quality voice conversations with all the characteristics users have come to expect, such as fidelity, clarity, and near-instantaneous delivery. Conversations must be clear and intelligible.

When a user perceives unacceptable quality, that user is likely to terminate the call. If this happens frequently, dissatisfaction with the service lingers and is difficult to overcome. You need tangible mechanisms for measuring the quality of service (QoS) and ensuring that network resources support the traffic load without degeneration of call quality.

One approach to call quality has been to overprovision the network. In many cases, this means adding bandwidth. Although over-provisioning might work initially, in the long run, it's a path to problems. First, it requires investment. Upgrading the capacity of connections, switches, and routers can be an expensive proposition. This approach might work for a time in a small local area network (LAN), but in a metropolitan or wide area network (WAN), it's often too expensive to be a practical solution. If the bandwidth is available, users will fill it. Data applications can quickly consume the additional bandwidth leaving you with the same congestion and problems delivering voice service.

Quality of Service

You know that voice and data are different applications and place different demands on the network. The requirements of applications are described as *class of service*. The question becomes how many classes of service are necessary to meet the overall QoS needs of all the applications on the network.

To meet performance level requirements, the network needs to take into account all of the following factors:

- Availability is usually represented as the uptime percentage. In commercial networks, the standard is referred to as “5 nines reliability” or 99.999 percent uptime. This number equates to roughly 5 minutes of downtime per year.



Reliability is another aspect of availability that is a design consideration. Does the network need redundant paths and fault-tolerant or high-availability equipment to guarantee availability in the event of a failure?

- Throughput is typically measured in bandwidth. It's a simple measure to determine how much of the data inserted into the network makes it through to the destination.
- Error rate is a measurement of data loss in some form. IP suite relies on high-layer protocols, such as TCP, to overcome errors and guarantee delivery. Legacy mainframe applications are generally intolerant of lost data.
- Delay is a fact of life in IP networks because routers use statistical multiplexing to process traffic and user data has a *bursty* characteristic. There will be some delay in any transmission network. Naturally, a lightly loaded or over-engineered network will have lower delay.

-
- Jitter is variation in the delay. Because packets can take different routes across the network, delay variation in IP networks is common. Jitter in VoIP conversations results in unintelligible conversations that sound “jerky.”
 - Scalability is also a consideration. As companies grow and businesses change, the ability of the network to grow with increased needs is an important factor.
 - Manageability is a necessary requirement. There have been many studies that demonstrate the most expensive component of network services to be management and administration of the network.

Today, the network is designed to meet specific performance criteria required by applications. VoIP is one more application you need to design for. The danger is that every new application potentially adds a new class of service. Network complexity can spin out of control. To maintain manageability, most network engineers favor using only a few critical service classes:

- Quick delivery supports real-time traffic such as VoIP
- Guaranteed delivery supports mission-critical traffic
- Best efforts delivery is standard IP to support everything else

In a network that is correctly designed, QoS becomes a prioritization scheme and is a factor in aggregating similar traffic types to ensure that each type takes a suitable path through the network. There are several approaches to providing QoS for VoIP. Each adds some form of overhead to provide this prioritization “or traffic cop” function.

As a counterpoint to VoIP call quality, it’s crucial that the prioritization method doesn’t allocate all available bandwidth to voice services. Mission-critical data applications also need the proper prioritization so that they don’t become starved for network resources, primarily bandwidth. You must design for balance between higher priority voice traffic and lower priority bursty email and Web traffic.

IP QoS Approaches

There are three distinctly different approaches to delivering QoS in IP networks. The following sections touch on each approach briefly.

Signaled QoS

Integrated Services (IntServ) introduces a signaling protocol to IP. Using IntServ, the user’s application sends some sort of call setup signal to the network. This signal is basically a request for a set of service delivery parameters required to complete the call. Using the approach, the network can check resource availability and deliver traffic accordingly. This approach is very similar in process to the circuit setup and teardown associated with a telephone call on the PSTN.

Provisioned QoS

Differentiated Services (DiffServ—also called DiffServ Code Point or DSCP) takes a completely different approach, requiring specific routes through the network that are predefined and available for each type or class of traffic. Paths could be pre-existing as part of the network design or set up on demand in some manner. DiffServ is often used as an aggregation technique to direct similar traffic along the same paths.

Bypass or “Shim” QoS

Multi-Protocol Label Switching (MPLS) is an approach that eliminates the normal hop-by-hop routing IP uses. It adds a “tag” to each packet that essentially shortcuts delivery to the best available path for the type of traffic. MPLS is compatible with frame relay and ATM networks. It has become a widely used approach to delivering QoS.

There are many other approaches used in providing QoS for VoIP call quality. 8021p/Q tagging is frequently implemented with VLANs at Layer 2. IP Type of Service (TOS) markings and IP precedence approaches are also popular. There are a variety of Weighted Fair Queuing techniques used by vendors to specifically address delay and jitter concerns.

Each of these methods boils down to prioritization, aggregation, or some combination of the two. They either request network resources or rely on another protocol or method to ensure the necessary paths through the network. These QoS approaches also provide traffic management and, in some cases, traffic shaping capability.

QoS implementation presents a hurdle because there are many options and approaches from which to choose. Different vendors use different approaches and techniques, sometimes making comparison of solutions challenging. Because QoS is really an approach to resource management in the network, distinguishing traffic types is a key part of the process.

These issues need to be dealt with in your LAN as well as the WAN. WAN QoS raises the complexity. If you maintain your own private WAN, you have some control over traffic handling. If your WAN consists of circuits purchased through a network provider, you will need to review, and perhaps revise, the service level agreement (SLA) to ensure prioritized traffic is handled properly.

Network providers widely embrace MPLS technology. Negotiate how prioritization tagging applied in your network will be treated once it traverses the provider network to ensure suitable WAN performance. If the provider remaps MPLS tags to their scheme for backbone QoS, understand their approach and ensure you’re getting the service delivery guarantees you need for your business.

When testing network performance, especially across a WAN, test the end-to-end connection performance not just node-to-node. Delay is cumulative. So is jitter.

Traffic engineering and testing must consider peak traffic loads. Testing in a controlled lab environment will yield very different results than an implementation in your production network. When evaluating your QoS approach, use a combination of network testing tools. Traffic generators need to simulate all traffic types in conditions that match your busiest times. Analyzers and monitors will help verify that bandwidth, throughput, error rate, delay, and jitter meet the requirements for VoIP call quality.

The right balance of network performance and QoS can be blended with MPLS and 802.1p/Q and VLANs to enhance QoS. These techniques, when designed methodically, can also improve security by isolating voice service from other data traffic while enhancing call quality.

Summary

IP telephony is viewed as the next-generation technology for communications network infrastructure. The idea of a full multi-service network supporting data, voice, and video over a single IP infrastructure can provide a streamlined communications system. This converged network can generate cost savings, encourage new applications, and enable new ways of collaborating within a company.

VoIP is a strategic enabler. It can reduce the cost of network management and lower the total cost of ownership (TCO). It can support new, advanced applications fostering innovation and creativity. Integrating tools such as unified messaging, intelligent voice recognition (IVR) systems, and customer relationship management (CRM) systems can provide new ways to enhance revenue and increase value.

VoIP Security

By Ken Camp

With the dramatic increase in VoIP deployments during 2005, there have been numerous surveys about security concerns. A few major threats to VoIP appear consistently in every survey. Falling prey to a Denial of Service (DoS) attack is most often the greatest fear. Worms and viruses are another major concern. Eavesdropping on calls raises anxieties for many companies deploying VoIP. In addition, VoIP spam gets a lot of mention in the press, but so far hasn't become a real problem.

VoIP security is a large topic, and the subject of many books and papers. This article provides a simple overview into the primary areas of focus and concern when addressing VoIP security. To keep things focused, we'll explore threats and dangers in the context of three fundamental security concerns—confidentiality, integrity, and availability. We'll cover VoIP network threats from the outside and the inside. In addressing how to deal with the threats, we'll stick to three fundamental precepts—prevention, detection, and reaction.

Characteristics of Security: Confidentiality, Integrity, and Availability

Confidentiality most often relates to ownership or control of the data. Data confidentiality is always at risk to malicious users who might discover, disclose, inappropriately monitor, or copy propriety corporate data. Protecting information privacy in VoIP networks is important. Keeping telephone conversations private is expected, just like in the public telephone network. In the public phone network, eavesdropping is difficult and requires a court order. In VoIP networks, eavesdropping through means such as packet sniffing may be technically easier to accomplish than public phone network eavesdropping. A fairly inexperienced attacker can potentially capture traffic. Free programs from the Internet, such as Audacity, may be able to reassemble these packets into an audio stream.

Integrity

Eavesdropping isn't the only threat. Some malicious users try to disrupt business traffic by degrading the integrity of systems. Integrity relates to preventing unauthorized or erroneous changes to network information. VoIP system integrity can be damaged by the insertion or use of false data. Integrity attacks frequently lead to modification, removal, repudiation, or misuse of system data.

Protecting integrity is frequently tied to user authentication procedures. System integrity requires that unauthorized users cannot make changes. Passwords should only be reset by authorized administrators or end users—only authorized administrators make configuration changes to the VoIP service infrastructure. The ability to make configuration changes is the attacker’s playground. If an attacker can configure the network, malicious activity can continue completely unnoticed. If an attacker gains configuration access, it’s no longer your network...it’s the attacker’s network.

Any system disruption or data corruption threatens system integrity. Malicious intruders aren’t the only threat to integrity. Legitimate employees make errors, sometimes taking unauthorized actions that cause problems. Disgruntled employees might purposely take harmful action. It’s important that user access levels provide permission to access only the resources needed to perform their jobs.

Availability

Availability is simply the loss of access, either to the network itself or resources on the network. Data might be unavailable because of corruption or destruction (an integrity breach). Congestion on the network may cause delays in accessing information. Any problem that makes the network resource difficult to use or access degrades availability.

Threats and Attacks

Attacks may come from outside or they may come from within. Attacks from the outside tend to be focused and malicious in nature. Attacks from within, although sometimes malicious, are often caused by user error or improper system configuration.

From the outside, pings, probes, and scans hit the network constantly. These are viewed as reconnaissance attempts, as outside users try to learn about your network. Worms and viruses are a constant and growing threat to business networks. As these problems commonly arrive via email, they are often admitted to the network, then cleaned, purged, or quarantined in an antivirus system inside the network. VoIP deployment creates a new potential attack vector in the gateways, VoIP servers, and unified messaging systems. In a fully converged network, attackers may attack the data network only in an attempt to get at the VoIP services within. Once VoIP services have been deployed, scanning from outside may increase as intruders probe for vulnerable services. VoIP services may open new TCP/IP ports for calling or signaling traffic.

Internal attacks are dangerous because they originate from a trusted segment of the network. Although they may come from malicious users, these attacks often come from inappropriate use of the network or some unexpected condition. Internal attacks are often the simple result of some new software program being installed somewhere on the network. A newly added, but misconfigured, server often creates unexpected conditions that may look like a scan or attack against the network.

What Is Being Attacked and Where Are the Targets?

To protect the network, you need to understand the targets. Some targets are logical. Others are physical targets. Logical targets are frequently user accounts. User accounts carry an associated set of accessible resources. Guest accounts typically have very limited permissions, but a domain administration account might have permissions to completely reconfigure the network. Running processes—whether in servers, routers, or other network elements—represent another logical target. Often the target is a program being run—its data, stack pointers, registers, and so on. Malicious attacks attempt to overflow memory buffers, creating unexpected results.

Physical targets are most often the network infrastructure—servers, workstations, routers, and so on. Domain controllers, VoIP service elements, and management platforms are also crucial physical elements of service delivery, making them key targets.

Who Is Attacking and Why?

It's important to understand that people perpetrate direct attacks. These types of attacks are very different than the perceived attacks caused by a misconfigured system. Some attackers may attack for the challenge. Professional criminals seeking financial gain are becoming far more common. The rise in identity theft problems online provides evidence of this increase in criminal involvement. In some cases, corporate raiders may be attempting to find a competitive edge through malicious attacks or hacking. Some malicious intruders want to cause damage by corrupting data. For some businesses or organizations, attackers may be spies or terrorists looking for information that can be leveraged for political gain.

Attackers don't play by any defined set of rules. There is no honor among thieves. They have a single objective—gaining access, control, or information by any means. Some attacks may be incredibly sophisticated technical efforts. Many attacks employ subtle, non-technical techniques:

Technical Attacks

Technical attacks often focus on the data leakage problem. Their goal is to intercept. Wiretapping is a technical attack but requires physical access to the wiring. Packet capture, in contrast, requires little technical skill. It can easily be accomplished with a free program. Intruders will attempt to breach access controls and circumvent security measures in order to access network resources.

Other forms of technical attack might include password guessing, perhaps using a dictionary attack. Theft of electronic media or *dumpster diving* can yield a wealth of proprietary information.

Worms and viruses aren't the only malware delivered in email. Spyware and keyloggers are often inserted without the end user's awareness, creating another data stream the malicious intruder can use to learn about the network, to learn passwords, or to steal proprietary information.

Reconnaissance attacks, penetration testing, and the exploitation of known vulnerabilities are becoming more common every day. Many tools to accomplish these attacks are downloadable from the Internet and freely available to anyone.

Non-Technical Attacks

Non-technical attacks focus not on technology, but on people. These types of attacks are frequently described as *social engineering*. It's human nature to be helpful. This truth can make it easy for attackers to misrepresent themselves to gain unauthorized access.

An intruder doesn't need much information to call the enterprise Help desk posing as an employee and requesting help in resetting a password. One very common technique is for intruders to masquerade as third-party vendors providing system support. In either of these cases, it's natural for support staff to want to be helpful, perhaps easing unauthorized access to the network. Don't underestimate human targets. They're always the weakest link in network security.

DoS Attacks

A DoS attack is always among the top fears. DoS attacks exhaust network resources. Because there are so many different resources available, the attack vectors can vary widely. IP addresses, network bandwidth, and processor memory are the most common attack vectors. Although DoS attacks take on many different forms, they all focus on starving out those resources critical to network operations.

These attacks can take on several forms. Buffer overflows exhaust system memory or CPU capacity, creating unexpected conditions. Some attacks simply attempt to consume all available bandwidth, degrading the ability of the network to deliver traffic. Routing and DNS attacks can lose or misdirect packets and disrupt information.

Distributed DoS Attacks

Distributed DoS (DDoS) attacks represent an especially troublesome variation. DDoS attacks involve several elements and are made of up *botnets*. Usually an intruder or worm delivers some form of malicious program, the *bot*, to the network. This bot program is often called a *zombie*. It's a software agent that gives outside control of the user's system to a *master* or *handler*, controlling the actions of the botnet. The bots will often log in to some chat channel and wait for a command from the handler. When an attack is launched, the users who have been infected with *bots* often don't even know that their systems are being used to attack some other victim.

A VoIP network presents many attack points for a DoS or DDoS attack. Every VoIP endpoint or phone contains some form of call agent software. Trunking gateways, signaling gateways, access gateways to other networks, and media and application servers are all server-based resources on the network. They all represent potential attack points. These VoIP services are often installed on servers running general-purpose operating systems (OSs) such as Windows servers. As with any Windows server, these systems are vulnerable to many Windows exploits.

A DoS attack against a VoIP network will disrupt the delivery of calls, presenting several manifestations. VoIP phones may not be able to register with the network. Internal calls might work but calls to the Public Switched Telephone Network (PSTN) might fail or vice versa. Users might not get a dial tone when picking up the phone, calls might be blocked due to lack of available resources, the caller might hear nothing but silence after dialing, or users might be able to dial, but calls may just fail to complete. Even if the VoIP network allows successful calling, conversations may be disconnected midstream. Calls could simply disconnect when the answering party picks up the phone. Network congestion could increase delay and jitter, making conversations unintelligible.

An overloaded network, under DoS attack might add 2500 milliseconds (ms) delay to every packet. This network would still work fine for email and Web services, but voice service would be unusable.

Viruses and Worms

Worms are self-replicating programs and are similar to viruses. A virus usually attaches itself to some other program; a worm is self-contained. Worms tend to propagate by exploiting file transmission capabilities of computers, most often email or network file sharing. Worms often attempt other undesirable actions. They may delete files or send documents via email. Some recent worms have carried other malicious code in their payload.

Worms tend to consume all available network resources, creating a DoS situation. Domain controllers and Active Directory (AD) servers are often targets, but VoIP resources bring a new rich target set to the network. You can expect new worms and viruses to exploit potential vulnerabilities in VoIP servers, endpoints (softphones and hardware phones), and the VoIP protocols (SIP, H.323, SCCP, and so on).

Eavesdropping and the Man in the Middle

There is a fear with VoIP that an intruder might be able to insert into the network between authorized endpoints, allowing a man-in-the-middle attack. Eavesdropping is simplified because the intruder either appears to be on the network or is listening in transparently. In network terminology, this security breach is called promiscuous listening. The intruder might be able to redirect traffic through spoofing or broadcasting alternate addresses.

Intruders eavesdropping on the network might also be able to alter call detail records (CDRs), modifying call setup information or corrupting billing data. An attacker might spoof SIP responses and redirect the caller to a rogue SIP address that intercepts the call. In practice, these exploits are achievable, but as yet, not widely seen in real-world VoIP implementations.

Spam Over Internet Telephony


Spam is a troublesome problem for anyone who uses email today. Users are bombarded with a stack of annoying, unsolicited junk mail daily. Imagine those messages filling not only your email inbox but also your voicemail inbox as spam moves to VoIP systems. Spam over Internet Telephony (SPIT) resembles spam in that it's essentially unsolicited junk phone calls or messages. SPIT has the potential to fill voice mailboxes with junk messages, just as spam fills your email inbox.

For now, documented cases of VoIP spam remain isolated and few, but SPIT could clearly pose a major headache. Although real-world examples remain slim in this area, it's certainly an area that gets a lot of press coverage. The largest real concern in the future is perhaps seeing SPIT used as a form of DoS attack.

Defense Mechanisms: Prevention, Detection, and Reaction

Preventive measures are the first line of network defense and can range from locking the server room door to setting up high-level security policies. Some preventative measures are simple and seem quite obvious. Other techniques are quite sophisticated. The best place to start is by identifying what needs to be protected.

1. First, protect the network. There are several problems to which networks are vulnerable. The classic problem is a DoS attack, but address spoofing is another common problem in networks under attack. To protect the network against increasingly sophisticated attacks, you need to build layers of defense in the network, each adding to the overall security and defense of VoIP services and other critical resources.
2. Second, protect the services. Every network service will present its own set of security requirements based on the intended use of the service. An internal service will have a much different set of security requirements than a service designed for external use. If VoIP is used only for internal calls within the company, simply protecting the internal servers from external access with access control lists (ACLs) might be adequate.

 Don't locate both internal and external servers on the same server hardware if possible. External traffic should be isolated using an outside-facing demilitarized zone (DMZ) segment of the network. Use firewalls and ACLs between networks to allow only authorized communications to pass.

3. Third, protect the protection. Be sure to protect the security and management platforms. It may be smart to create a special management segment of the network to house these critical services. As the management segment oversees the entire network, systems on this segment may need access to the entire network. It's important to ensure that the reverse is not left as a default access configuration. Management servers should not be accessible to anyone outside the network administration staff. Log servers should only permit authorized users to view log contents. Intrusion detection systems (IDSs) shouldn't be seen or touched by anyone outside the network security staff.

Firewall rules and ACLs can help determine where users can gain access, but only strong user authentication mechanisms will keep unauthorized users away from management systems. Discovery protocols such as Simple Network Management Protocol (SNMP) and Internet Control Message Protocol (ICMP) should be disabled whenever possible. Disable all services that aren't necessary on management systems. For example, don't run IIS, FTP, Telnet, or SMTP on a Windows server simply because that is the default configuration. Evaluate system requirements and only enable necessary services.


Effective network security is a combination of policies, processes, and technology. Firewalls can monitor the state of traffic and handle anomalies. Firewalls can monitor the state of VoIP sessions as well. Voice and data services should be treated as different zones of trust in the network. The goal is to control traffic flow between differing trust zones. Security policies define the rules that describe what types of traffic are allowed and what types are denied.

IDSs can monitor the flow of network packets and alerts from firewalls, routers, switches, and other systems in the network. They can identify traffic patterns either based on a digital signature or a newer heuristic approach that compares network traffic patterns with known *normal* baseline traffic patterns.

An intrusion prevention system (IPS) is a more advanced variation of this type of solution that has some ability to change network configurations in an automated manner. Instead of passively monitoring and alerting administrators, the IPS becomes an active management component. In one sense, the network becomes *self-defending*, with the ability to protect itself from some types of malicious traffic automatically.

Detection

In SANS security classes, the need for detection is often stressed as being even more important than the need for protection. Regardless of whether prevention is successful, detection systems will help you know about every event that occurs in your network. Sound detection tools minimize response time and speed mitigation. Incident management techniques can help you be proactive rather than reactive.

 For more information about SANS security classes, see <http://www.sans.org>.

Audit procedures and regular review of system logs help detect when information has been damaged, altered, or stolen; how it has been damaged, altered, or stolen; and who has caused the damage. As with prevention, detection is a combination of policies, processes, and technology. In addition to firewalls and intrusion detection and prevention technologies, syslog is a critical tool at your disposal that can be used in a variety of ways.

Syslog

The term *syslog* is used to describe both the syslog protocol and the application that sends syslog messages. The syslog protocol is very simple. The syslog sender transmits a small text message to the syslog receiver or server.

Syslog is used for network management and security auditing. Although syslog is very simple and may not be the very best tool for auditing, it has one huge advantage—it is supported in virtually every element of the network, allowing a corporate syslog server to become a central repository for audit log data. Syslog data is in plain-text format and is easily manipulated using simple tools and scripts. Many small to midsized organizations use a combination of scripts and spreadsheets to analyze syslog data when first starting out. Large networks produce much larger log files and will require more tools and processes that can scale to handle the larger data files.

Technology alone can't solve the detection problem. What you monitor, how you use log data, and how you react to incidents at the time of detection are all a critical part of the cycle of network defense. You employ detection mechanisms so that you'll know as soon as possible when an intrusion or other malicious event occurs. Network threats mutate quickly. Worms spread almost instantaneously. The threat of *zero-day* attacks will not allow for weak incident management prevention and detection processes. Effective incident management tools and processes ensure quick reaction and recovery when an event does occur.

Reaction

Layered defense is a term that is used in many ways. Just as you build security in layers, you build your incident management process layers. You take steps to prevent security breaches with firewalls and other preventative measures. You know that no matter how good your defense is, breaches will occur. You put detection tools and processes in place so that you can speed your reaction time. Your operations processes can be either reactive or proactive. How an organization responds to an incident is driven by how well prepared everyone is.

The proactive strategy is a pre-attack strategy. It requires steps to minimize existing security policy vulnerabilities and develop contingency plans. Identifying the damage that an attack will cause ahead of time requires careful analysis of weaknesses and vulnerabilities.

The reactive strategy is a post-attack strategy. It focuses on assessing and repairing any damage caused by the attack, then implementing any contingency plans developed in the proactive strategy. It's vital to document and learn from every incident, and restore business functions as quickly as possible.

Summary

Business networks are dynamic and complex systems. VoIP services add another layer of complexity to corporate networks. Security threats are real and continually on the rise. Security countermeasures come, not as a simple product to be purchased, but as a balance between policies, procedures, and technology solutions. Complexity and sophistication of both attacks and defensive strategies will continue to evolve. The malicious intruders have the upper hand in that they can use any means of attack. The best defense comes through comprehensive policies, diligent monitoring, and a cycle of ongoing review and improvement.

A Look at H.323

By Ken Camp

In 1995, the ITU-T began work on a series of standardized signaling protocols. One outgrowth of this work was a product called The Internet Phone from VocalTec. Initially, this was a proprietary solution and did not focus on interworking between the PSTN and Internet. Transmission of audio signals and the idea of video conferencing over the Internet were key issues. These standards fell within the H.323 family of protocols for multimedia transmission over packet networks and were to some extent an outgrowth and extension of H.320 ISDN videoconferencing standards.

Interworking with the PSTN quickly became a major focal point of this technological development work. Designers recognized the need to incorporate some method for calls to cross over from the IP network to the PSTN. This led to efforts in gateway protocols and connectivity to the SS7 network, which provides extensive signaling (command and control) functionality in the PSTN. H.323 products began to appear from vendors in 1996.

H.323 embraces a set of goals that are quite simple and straightforward in principle; however, implementation, over the years, proved far more complex.

- Internetworking with the PSTN became a central theme.
- H.323 had to handle the conversion of signaling from whatever packet protocols are used to the PSTN signaling format used by SS7.
- H.323 had to have a call control mechanism for call setup and teardown.
- H.323 had to encode the media—that is, digitize and packetize the audio voice transmission for the IP network.

The last three functions were ideally performed in a single device referred to as a gateway; however, overall H.323 encompasses a complex suite of protocols and approaches to signaling, media conversion, registration, and admission to the network as Figure 1 shows.

A Broad View of H.323

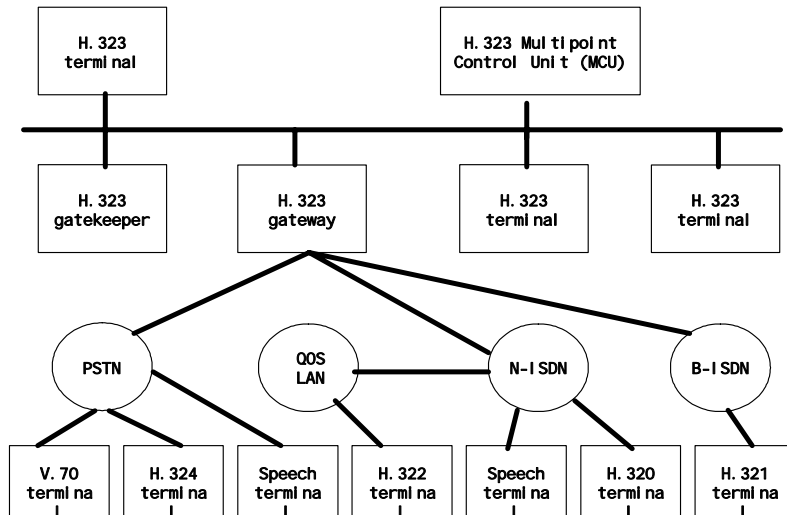


Figure 1: A look at the overall scope of the H.323 standards.

H.323 has always been viewed as an “umbrella standard” because it includes multiple individual standards, each addressing an individual requirement. Codecs are defined for the transport of both audio and video signals. Audio codecs are included for compression as low as 5.3Kbps for voice. It is important to note that “voice” can mean audio, modem data, fax messages, or touch tone signals (DTMF). Encoding schemes don’t necessarily work equally well with each of these. Real Time Transport Protocol (RTP) is defined for use with both audio and video. It is used for delay-sensitive information and includes time-stamping of packets for sequencing and timing. Whenever RTP is used, Real Time Control Protocol (RTCP) is also used. This control protocol establishes and monitors RTP sessions.

It is important to note that RTP uses UDP for transmission of the audio and video packets. Because this media is providing for delivery of real-time information, UDP is used for the quickest delivery. TCP can provide guaranteed delivery of information, but the overhead associated with TCP coupled with the retransmission of any lost data is too intrusive to support real-time data delivery. What this means is that a voice packet that is lost during transmission is simply lost. Testing has shown that in real-world applications, the human ear is far more tolerant of lost packets containing a fraction of a syllable than to the call quality delays introduced by using TCP.

The H.225 standard is used for Registration, Admission and Status (RAS). An H.323-compliant terminal upon connection with the network will register with a Gatekeeper in order to participate as a member of the voice network. Stations need to request network resources, make calls, and resolve the IP address of the called station. A Gatekeeper often performs these functions, although the device is optional under the H.323 standards.

Q.931 signaling is used between devices for call setup and teardown. This is the identical signaling protocol used in ISDN services, complete with all the features used there. This signaling can be supported by a gateway, but can also be supported by a telco central office switch. This signaling has provided a crucial interoperability function in allowing signaling to move between IP networks and the PSTN.

H.245 standards are used to provide an exchange of media capabilities between end stations. The H.323 family of protocols supports multimedia including video. H.245 might be used when a video call is requested to negotiate for voice-only connectivity if the called party does not have a video-capable terminal.

Per the standards, reliable transport is used for signaling. In an IP network, this means that the overhead and performance of TCP are necessary. A good way to view this is to correlate this signaling to the support provided by the SS7 network in the PSTN. Without guaranteed delivery of call control and signaling message, call processing would halt and the system could not function.

The T.120 standard was implemented in 1996-97 and contains protocols and services that support real-time, multipoint data communications. This has often been implemented in the form of a “whiteboard” application that both users can share, but is also used for sharing files or multiplayer gaming. One example is two users collaborating on a spreadsheet while talking about the changes being made. Many vendors—including Apple, AT&T, Cisco Systems, Intel, MCI, and Microsoft—have implemented and support T.120-based products or services. Figure 2 shows the relationships between these protocols.

Protocols Encompassed by H. 323

Video		Audio		Control			Data
H. 261 H. 263 (video codec)		G. 711 G. 722 G. 723 G. 728 G. 729 (audio codec)		H. 225 Terminal to gatekeeper signaling	Q. 931 Call signaling	H. 245 Control Channel	T. 120 (Data terminal sharing)
RTP	RTP	RTP	RTP				

Figure 2: Under the umbrella of H.323 protocols.

Call Setup Using H.323

Many people have suggested that H.323 is cumbersome. Products supporting it have often been accused of *bloat* because of the comprehensive functionality included and supported. Yet, H.323 is widely deployed and supported still today. Given that, this section explores the steps it might require to establish a telephone call between two workstations using H.323 protocols.

The call setup scenario example is simple and straightforward. Two users, Bob and Alice—in the same network, on the same LAN, and in the same building—want to communicate. Bob needs to call Alice to discuss a project they are working on. This example assumes that there is a Gatekeeper on the network for administration purposes and will step through the entire process of establishing the necessary connections and sessions to conduct a telephone call. It will also assume that they are both using their computer as the H.323 softphone for simplicity.

4. When Bob and Alice boot up their computers in the morning, each station must send a Discovery message to locate the Gatekeeper.
5. The Gatekeeper must reply and provide its IP address
6. Each workstation must now transmit a Registration Request to the Gatekeeper, and receive an Acknowledgement in return.
7. Bob wants to call Alice, so his workstation sends a Locate Request, and receives an Acknowledgement in return. This is to provide the IP address of Alice's workstation.
8. Bob's station now transmits an Access Request seeking resources and permission to make the call, and receives an Acknowledgement in return.
9. At this point, a TCP session is established for H.225 setup.
10. Alice's workstation has received an incoming call request and now must transmit an Access Request for resource and receive Acknowledgement.
11. An H.245 Connect message is exchanged between the two workstations.
12. A second TCP session is established for the H.245 session.
13. Bob's workstation must open a Logical Channel for the media stream in the forward direction and receive an Acknowledgement.
14. An RTP media forward channel is opened, immediately followed by an RTCP control stream in the reverse direction.
15. Since the stations negotiated a full duplex, or two-way call, at the capabilities exchange phase, Alice's workstation must open a Logical Channel for the media stream in the reverse direction and receive an Acknowledgement.
16. An RTP media reverse channel is opened, immediately followed by an RTCP control stream in the forward direction.
17. Bob and Alice have established a full duplex, two-way path for voice audio.

These steps are shown in Figure 3.

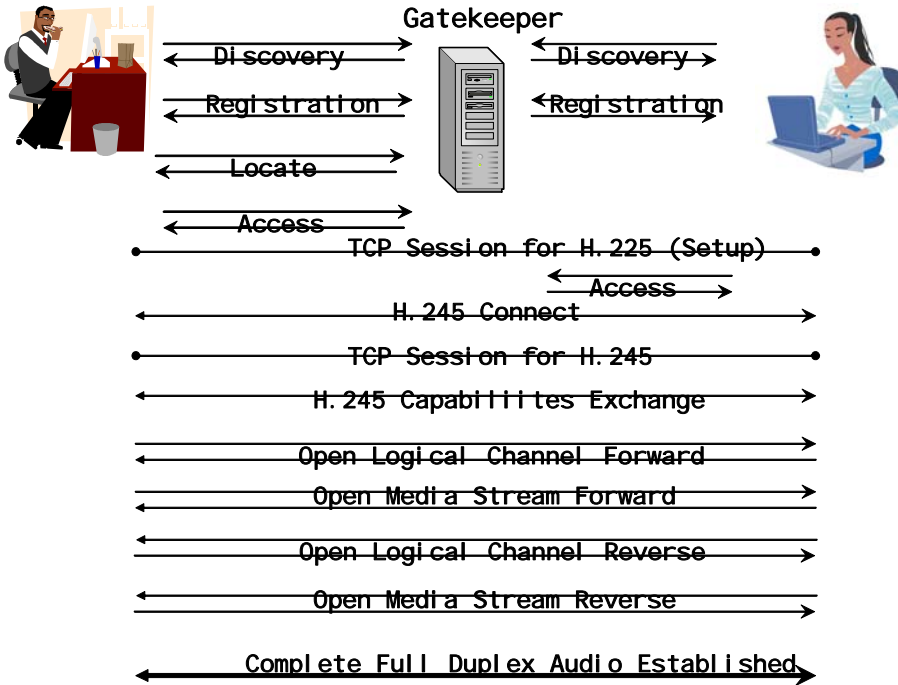


Figure 3: Making a phone call with H.323.

If this seems complex, keep in mind that H.323 supports multimedia beyond pure voice. If full-duplex video were negotiated, steps 10 through 14 would be necessary to set up those media streams. If data sharing were also required, steps 10 through 14 will have to be repeated again to establish that stream, and another TCP session will need to be established for T.120 data.

It's easy to see that establishing a phone call in H.323 can be very overhead intensive. Although this process works very quickly on a local LAN segment where bandwidth is not a real problem, over a WAN or the Internet, the delay in call setup can pose a serious problem. Call quality just in setup can render the service unusable. In some cases, setting up an H.323 call can take several seconds. Because TCP has timers for retransmission of packets, the delays can be much worse if the network is suffering from congestion and packet loss.

Summary

H.323 is widely used and supported by every major equipment vendor. This support remains to maintain full compatibility with the PSTN and ITU-T standards, but designers quickly observed that the complete feature set of H.323 might not be necessary for IP telephony. And the inclusion of the complete set of protocols led to products that were large in size, and often inefficient in their coding and use of system resources.

H.323 version 2 allows for fast connections and supports opening of media stream simultaneously, but the issue of overhead remains a concern to many designers of voice services and software. It's also important to note that nothing described here provides a method for call transfer, hold, parking a call, call pickup, call waiting, or message waiting services. These are all common services in business systems. Users expect these common features in any working telephony solution, whether it's based on traditional telephony or VoIP. There's is also no alternative routing mechanism if a node is congested or doesn't have the resources necessary to support the call.

Many designers, the author included, argue that although H.323 provides the necessary functionality, there are far too many scalability and performance-related issues to treat it as an acceptable solution for VoIP service delivery.

A Look at Session Initiation Protocol

By Ken Camp

As a follow-up to recent online conversations, this article begins an exploration of common VoIP protocols. This first installment of an ongoing exploration of the technical workings of VoIP protocols will explore Session Initiation Protocol (SIP). As this is the first paper in the series, it will begin with an overview of how the open standards widely used in the Internet are developed.

Standards used in the worldwide Public Switched Telephone Network (PSTN) are products of the International Telecommunications Union-Telephony (ITU-T) sector, formerly known as the CCITT. This group operates under the auspices of the United Nations (UN), which is important to note because this body acts in many ways as an administrative unit focused more on international interoperability than other areas. Although this group has responsibility for telephony standards, they have not historically been known for being quick or nimble at responding to technology needs. ISDN standards took 10 years to get through this group. There are many different international political agendas that come into play when dealing with the UN, and change takes time.

Internet standards are a much different story. TCP/IP standards used throughout the Internet are primarily the work of the Internet Engineering Task Force (IETF). This group is also global—primarily an organization of volunteers. Any interested party can join the IETF and participate in standards development, and many do every year. The IETF working groups are made up of technology specialists from colleges and universities, hardware and software vendors, telecommunications providers and competitive local exchange carriers (CLECs), Internet service providers (ISPs) and Internet telephony service providers (ITSPs), government organizations, and a variety of other interested parties. Because the organization is voluntary and created by technologists focused on progress and efficiency, the structure of developing IETF standards is much different than that for the PSTN.

Internet standards are developed through the Request for Comments (RFC) process, which is quite efficient. In many cases, interested individuals or organizations collaborate to jointly present a draft for a new open standard to improve the network or enhance a capability. A proposal for a new standard that is jointly presented by team collaboration can carry broad support at introduction. Vendors will often work closely together to “grease the wheels” and move standards proposals forward. Thus, in the real world of standards development, change can often occur very quickly in Internet standards, as they move through the draft process.

In telecommunications, between the PSTN and the Internet, there are two different networks, with technical standards developed and approved by completely separate standards bodies. Although this idea might seem straightforward, it often isn’t—whether it’s the PSTN or the Internet, people often draw the network as a cloud. Although this is done for the sake of simplicity, the cloud concept breaks down when complex problems related to open standards and vendor interoperability are concerns. When you look at VoIP technologies, interoperability with the incumbent voice network, the PSTN, is absolutely necessary for global success and widespread deployment of IP telephony solutions.

Figure 1 shows three iterations of the convergence between the PSTN and the Internet. At the top is the beginning stage. The PSTN and Internet began as distinctly separate networks, each with a specific purpose. There was no connection or interoperability required between the two. This

model represents the relationship between the two networks throughout the 1970s and most of the 1980s. They operated independently and many people thought they would never meet or join.

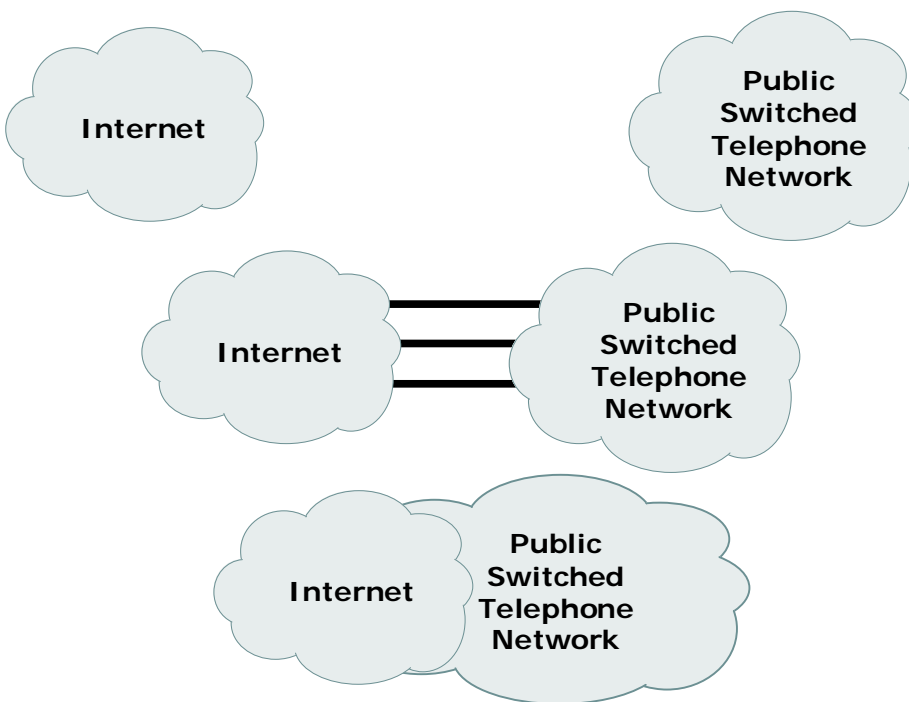



Figure 1: Network clouds converge—the Internet meets the PSTN.

In the center are connections between the two networks—separate but connected. Although there are only three lines shown, there are thousands of links. When users began dialing into the Internet using modems, the two networks began to interweave in a limited fashion. As VoIP technologies began to evolve, it became obvious that an *Internet phone call* would have to maintain some form of compatibility and interoperability with the PSTN in order to provide ubiquitous service. Connectivity alone cannot support telephony. Interoperability is a crucial factor. This conclusion presented the IETF with an often-unspoken mandate to work in cooperation with the ITU-T and assure that Internet standards supported and remained compatible with PSTN standards. The goal for many has been seamless interoperability between the two networks. The phrase *transparent to the end user* has never been used more than in discussions of interoperability between these two environments. This collaborative effort toward a ubiquitous, interoperable technology was the dominant philosophy of many, but not all, developers of the early IP telephony standards.

Today, convergence is talked of as if it's behind us, but the efforts continue. We view convergence as the migration of voice, data, and video services to a single consolidated network infrastructure. Most often, we envision this infrastructure to be a TCP/IP network, like the Internet. In the past, people said that the Internet was growing so quickly it would absorb the PSTN. Others claimed that telephony would migrate to the Internet and that one day the old legacy PSTN would simply be turned off.

After several years of evolving VoIP technologies, we see that both networks are vital, growing, and critical, although the moderate pace of improvement in VoIP technologies is creating the potential for far deeper penetration of a single infrastructure. One common view in the past was

that the two networks would become linked only at high capacity access points, passing traffic through gateways to each other. Today, they are far more inextricably connected than anyone imagined, with thousands and thousands of gateways. The two networks haven't become one, but they are so tightly coupled delivering business services that they are beginning to represent a single cloud as shown in the bottom of Figure 1. This cloud is nothing more than raw service capacity, with services being defined through agreements between users (as service level agreements—SLAs) and providers (as peering agreements). The network of tomorrow is partly here today. It's a high-performance cloud of capacity that provides whatever service the end user requests of it.

 For a more detailed review of this concept and the evolution of the networked world, see Steve Shepard's book *Telecommunications Convergence* (McGraw-Hill).

Session Initiated Protocol

The original work on SIP was performed by the IETF as one of several efforts. The Multiparty Multimedia Session Control (MMUSIC) working group took much of the lead in very early efforts. Since 1999, the IETF-SIP working group has led this work. Their specific charter states that SIP is a text-based protocol, similar to HTTP and SMTP, for initiating interactive communication sessions between users. Such sessions include voice, video, chat, interactive games, and virtual reality. This group has worked to bring SIP from proposals to drafts and standards in addition to specifying and developing proposed extensions that have arisen from very aggressive protocol and feature requirements. The SIP working group will concentrate on the specification of SIP and its extensions, and will not explore the use of SIP for specific environments or applications.

Throughout its work, the group will strive to maintain the basic model and architecture defined by SIP:

- Services and features are provided end-to-end whenever possible
- Extensions and new features must be generally applicable rather than applicable to only a specific set of session types
- Simplicity is key
- Reuse of existing IP protocols and architectures, and integrating with other IP applications, is crucial

SIP provides protocols and mechanisms that allow both end systems and proxy servers to provide services including:

- Call forwarding under a variety of scenarios (no answer, busy, and so on)
- Calling party and called party number identification using any naming scheme
- Personal mobility allowing a single address that is location and terminal independent
- Capabilities negotiation between terminals
- Call transfer
- Instant messaging

-
- Event notification
 - Control of networked devices

There are also extensions to SIP that provide for fully meshed conferences and connections to multipoint control units (MCUs).

SIP uses an addressing structure very much like email addressing. Given that users might log on from any location and receive an IP address dynamically, there must be some way to resolve some common convention to the active and current IP address. As people are familiar and comfortable with email addresses, this structure seems most appropriate and remains a popular choice.

SIP is a text-based protocol like HTTP or SMTP, so the addresses, which are SIP Uniform Resource Locaters (URLs), can be imbedded in email messages or Web pages. Also, as a text protocol, the addresses are network neutral. Thus, the URL might point to an email-like address, using SIP, an H.323 address, or it might point to a PSTN telephone number. The ITU-T E.164 standard defines the telephone numbering structure.

SIP provides a comprehensive set of building blocks that can be extended to allow for E911 or Advanced Intelligent Network Services. Because it can support forking to multiple destinations, SIP can support call forwarding, Automatic Call Distribution (ACD) techniques for call centers, and redirecting a call to multiple alternative locations.

SIP operates independently of the network layer and requires only unreliable datagram or packet delivery. It provides its own reliability mechanism. Although in the IP environment of the Internet, SIP is used over UDP or TCP, it could run over IPX, Frame Relay, ATM AAL5, or X.25 without modification. Generally, UDP is used to avoid the overhead associated with TCP.

The protocol model for SIP as originally described in RFC 2543 is shown in Figure 2. It provides four functional components:

- User Agents either initiate call requests or are the destination of those requests. A User Agent might be IP telephony software running in a computer or an IP telephone.
- The Registrar keeps track of users within the network or domain. User Agents register with the registrar as members of the network
- The Proxy Server is an application layer routing process that directs SIP requests and replies within the network.
- The Redirect Server receives requests for users (UAs) and provides the location of other SIP User Agents or servers where the called party can be reached.

Within the SIP Server, the Registrar, Proxy Server, and Redirect Server can be implemented in the same software package.

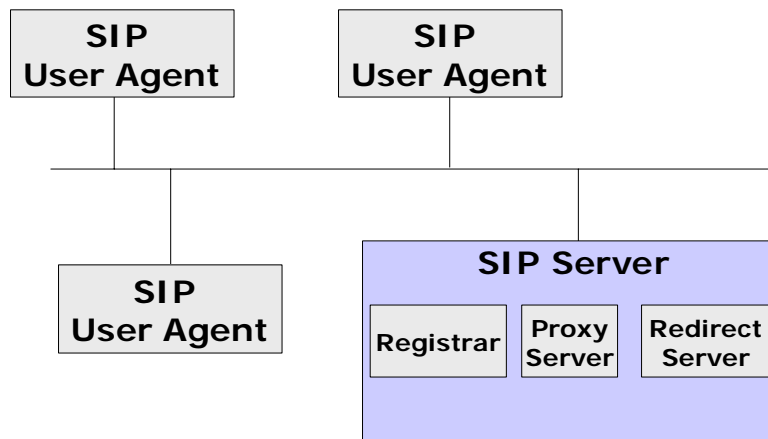



Figure 2: The SIP model.

In a SIP session, a user initiates a call, which prompts the User Agent to transmit a SIP message. These messages will then traverse one or more SIP Servers. Once the destination User Agent information is obtained, actual message transfer takes place directly between the User Agents. If one end of the call is located in the PSTN, some gateway between the IP-based SIP network and PSTN is required to provide all the necessary protocol conversions between networks.

Session Description Protocol

The MMUSIC working group of the IETF also provided RFC 2327—the Session Description Protocol. SDP is intended for describing multimedia sessions for the purposes of session announcement, session invitation, and other forms of multimedia session initiation. SDP is used within both SIP and Megaco implementations. SDP is not intended to support negotiation of session content or media encoding but to act as a general-purpose tool. It also supports multicast media and can be used for broadcast environments such as Internet radio or television.

 This article doesn't explore SDP in depth, but it's beneficial to have a high-level grasp of how the information describing multimedia sessions is transmitted between user systems.

SDP provides Session Announcements as the mechanism used to convey session description information between devices or nodes and proactively delivered to users. These announcements might also be delivered via email or the Web, allowing for automatic launching of the appropriate application on the called parties workstation. SDP includes:

- The session's name and purpose
- Time the session is active

-
- The type of media used in the session; this might be voice, video, data, and so on
 - The format of the media (MPEG video, H,261 video, and so on)
 - The transport protocol used (UDP, TCP, IP, and so on)
 - Information necessary to receive the media (TCP/IP ports, addresses, and formats)

The actual syntax for the port and addressing information vary depending on the transport protocol in use. Following is an example of an SDP description:

```
v=0
o=kcamp 2890844526 2890842807 IN IP4 126.16.64.4
s=SDP Seminar
i=A Seminar on VoIP
u=http://www.realtime-voip.com/seminar/voip.52.ps
e=ken@ipadventures.com (Ken Camp)
c=IN IP4 63.215.128.129/127
t=7944393265 8746931596
a=recvonly
m=audio 49360 RTP/AVP 0
m=video 51782 RTP/AVP 31
m=application 32416 udp wb
a=orient:portrait
```

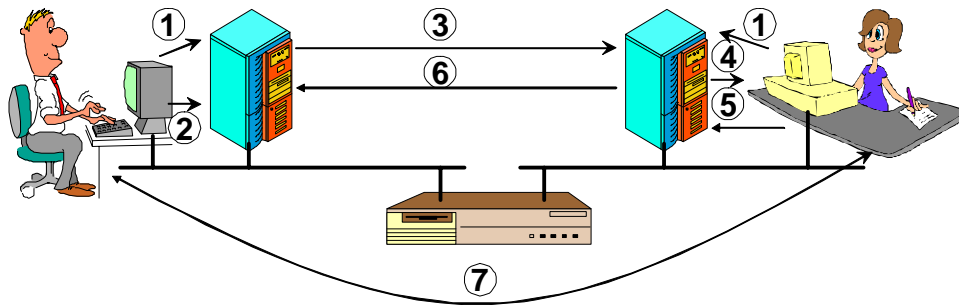
In general terms, SDP is used to convey enough information for a user to join the session or call. This information might not include encryption keys in the VPN environment. Those might be handled by another set of protocols such as IPsec.

Call Setup Using SIP

Whether it involves voice or data communications, we often use the same couple, Bob and Alice, to explain how a communication occurs, and we'll use them in Figure 3. Let's step through a call using SIP. We'll assume Bob and Alice are on different LANs and that the two networks are connected by a router. Each network has a SIP Server on the local LAN in this simulation.

18. When Bob and Alice turn on their computers, the User Agent software (part of the VoIP client), automatically registers them each with their local SIP Server.
19. Bob initiates the telephone call, and the User Agent on his computer transmits an *invitation* to the SIP Server on his local network. This *invitation* contains the session description information.
20. Because Alice registered with the SIP Server on her own local network, the SIP Server on Bob's network doesn't know how to reach her. Bob's SIP Server has to forward the *invitation* to every SIP Server it knows how to reach—Alice's SIP Server, in this case.
21. As Alice is on the same LAN and already registered with her SIP Server at startup, it knows how to reach her and forwards the *invitation* to her.

22. Alice also wants to talk to Bob, so she answers that call, which returns and acknowledgement (ACK) over the same path the *invitation* followed. Alice's session description information is included in the acknowledgement.
23. Now that both ends have exchanged session description information, they have the IP address and port information to directly contact the other party on the call and can now transmit RTP encapsulated media directly. The SIP Servers are no longer needed on the session.
24. The conversation takes place.



- ① Registration – performed by each station
- ② Bob initiate call, sending invitation and SDP
- ③ SIP Server forwards invitation to all known SIP Servers
- ④ Alice's SIP Server delivers invitation to called party
- ⑤ Alice accepts invitation and returns SDP
- ⑥ SIP Server returns ACK to calling party
- ⑦ End-to-end telephone conversation

Figure 3: Bob and Alice make a SIP phone call.

One advantage to this process is that it's much simpler than H.323, using fewer messages for call setup. H.323 is a protocol we'll explore in another article. SIP doesn't require the overhead of TCP's three-way handshake and guaranteed delivery. The registration process with SIP Servers can provide better extensive support for mobile users. And, as SIP is a text-oriented protocol, a simple BYE command is used to terminate the session.

📖 This article serves as a very brief technical overview of SIP. It presents one view of how SIP works. There are numerous references on the Internet for more information about SIP. For readers wishing to explore further, Wikipedia has an excellent starting point article at http://en.wikipedia.org/wiki/Session_Initiation_Protocol. There is also an excellent set of resources maintained by Columbia University at <http://www.cs.columbia.edu/sip/>.

Summary

SIP is one of several protocols being used to implement VoIP solutions today. It offers several competitive advantages as an efficient protocol and is very popular in some circles. In articles to follow, we'll explore other VoIP protocols.

Appendix

The following list highlights SIP RFC references:

- RFC 3261 SIP: Session Initiation Protocol - The core protocol specification; obsoletes RFC 2543.
- RFC 3524 Mapping of Media Streams to Resource Reservation Flows
- RFC 3515 The Session Initiation Protocol (SIP) Refer Method
- RFC 3487 Requirements for Resource Priority Mechanisms for the Session Initiation Protocol (SIP)
- RFC 3486 Compressing the Session Initiation Protocol (SIP)
- RFC 3485 The Session Initiation Protocol (SIP) and Session Description Protocol (SDP) Static Dictionary for Signaling Compression (SigComp)
- RFC 3428 Session Initiation Protocol (SIP) Extension for Instant Messaging
- RFC 3420 Internet Media Type message/sipfrag
- RFC 3388 Grouping of Media Lines in Session Description Protocol (SDP)
- RFC 3361 Dynamic Host Configuration Protocol (DHCP-for-IPv4) Option for Session Initiation Protocol (SIP) Servers
- RFC 3319 Dynamic Host Configuration Protocol (DHCPv6) Options for Session Initiation Protocol (SIP) Servers
- RFC 3327 Session Initiation Protocol (SIP) Extension Header Field for Registering Non-Adjacent Contacts
- RFC 3326 The Reason Header Field for the Session Initiation Protocol (SIP)
- RFC 3325 Private Extensions to the Session Initiation Protocol (SIP) for Asserted Identity within Trusted Networks Defines
- RFC 3324 Short Term Requirements for Network Asserted Identity
- RFC 3323 A Privacy Mechanism for the Session Initiation Protocol (SIP)
- RFC 3329 Security Mechanism Agreement for the Session Initiation Protocol (SIP)
- RFC 3313 Private Session Initiation Protocol (SIP) Extensions for Media Authorization
- RFC 3312 Integration of Resource Management and SIP Framework for preconditions
- RFC 3311 The Session Initiation Protocol (SIP) UPDATE Method
- RFC 3262 Reliability of Provisional Responses in the Session Initiation Protocol (SIP)
- RFC 3263 Session Initiation Protocol (SIP): Locating SIP Servers
- RFC 3264 An Offer/Answer Model with the Session Description Protocol (SDP)
- RFC 3265 Session Initiation Protocol (SIP)-Specific Event Notification
SIP event model; defines SUBSCRIBE and NOTIFY
- RFC 3087 Control of Service Context using SIP Request-URI

RFC 3050 Common Gateway Interface for SIP

RFC 2976 The SIP INFO Method

RFC 2848 The PINT Service Protocol: Extensions to SIP and SDP for IP Access to Telephone Call Service

VoIP and Firewalls

By Ken Camp


Firewalls are a critical element in protecting business networks. They provide protection against malicious traffic and are crucial monitoring points in the daily operation of many networks. Firewalls also present a potential performance bottleneck. Given the sensitivity of VoIP services to performance impairments, especially delay and jitter, it's important to incorporate thoughtful design methodology into VoIP services. How firewalls are used in the network will absolutely have an impact on voice services. Good design minimizes this impact, providing added security while allowing acceptable call quality.

Although firewalls have been around for many years, the creator of the firewall remains a topic of open debate. Marcus Ranum has been widely recognized for his work on the TIS Firewall Toolkit and Trusted Information Systems' Gauntlet product. Shlomo Kramer and other Checkpoint colleagues also made significant contributions to firewall technology. Regardless of who was there first, firewalls have become a crucial component in networking technology.

 For more information about Marcus Ranum, see <http://www.ranum.com/>, and for details about Shlomo Kramer, see <http://www.imperva.com/company/management.html>.

Different Types of Firewalls

There are many types of firewalls. Personal firewalls are typically software applications running on a personal computer. Network firewalls typically operate on a dedicated network device sitting at a boundary point, or demarcation, between two or more networks and are often located in demilitarized zones (DMZs). These firewalls filter all traffic entering or leaving the connected networks.

 More information about DMZs is provided later in this article.

There are several variations in firewall technologies:

- Firewalls that protect between a single node and the network; others protect between two or more networks
- Firewalls that operate at the network layer; some operate at the application layer
- Firewalls that track the state of the communications

Network traffic can be intercepted and monitored at different layers in the technology. Some firewall approaches—iptables, for example—operate at the network layer. Other firewall approaches, such as TCP Wrapper, work at the application layer. These network and application layer approaches might also overlap.

Network Layer Firewalls or Packet Filters

Network layer firewalls are essentially packet filters. Based on defined rules, they determine which packets are permitted to pass. These filtering rules operate on the source address, destination address, and TCP/UDP port, and sometimes whatever higher-level network protocols the packet contains. Network-layer firewalls tend to be quick in processing and transparent to end users.

The simplest network-layer firewalls are primarily stateless firewalls. They have packet-filtering capabilities, but they can't make more complex decisions based on the patterns of communication between hosts. Stateless firewalls somewhat resemble a router in their ability to filter packets. They are often implemented as a firewall rule-set feature in multi-function routers.

A computer running an operating system (OS) that supports packet filtering and routing can be used as a network-layer firewall. Linux, Solaris, BSDs, and Windows Server all include this capability.

Proxy Firewalls

A proxy server is a computer that offers indirect network connections from clients to other networks or network services. A client connects to the proxy server, which then requests a connection, file, or other resource on behalf of the client. The proxy might provide client access to the resource by connecting to the resource itself or by serving cached data to the client. The proxy may alter either the client's request or the server's response for a variety of reasons. A proxy server can also serve as a firewall.

Proxies hide internal resources, making it more difficult to tamper with internal systems. With a proxy in place, misuse of one internal system would not necessarily create a breach that could be exploited from outside.

Stateful Inspection

Several years ago, Checkpoint coined the phrase *stateful inspection* and invested a great deal of effort in promoting this firewall approach. Stateful firewalls monitor information about the state of sessions or connections. For example, they monitor whether connections have been established, session initiation attempts, the TCP three-way handshake, and proper session teardown.

Packet filters, or stateless firewalls, treat each packet as an individual unit. They operate on a packet-by-packet basis. Because of the way they work, stateless firewalls have no way of correlating an individual packet with a data stream, session, or connection. A packet might be part of an existing connection, trying to establish a new connection, or just some rogue packet—the firewall can't make that determination. Prior to the technique of stateful inspection in firewalls, this stateless method was how all firewalls worked. Current firewall solutions are aware of connections, or *states*. This awareness allows for increased granularity in how traffic can be controlled through the firewall.

Stateful firewalls maintain *state tables* in memory. These tables contain all the attributes of each connection through the firewall for the duration of every session. The state of the connection often includes data such as the IP addresses of the source and destination, the TCP ports involved in the connection, and the TCP sequence numbers of the packets being transmitted. CPU utilization is heaviest when a new session is established. Once a session is identified as being permitted, later packets are simply associated with a known session.

Stateful firewalls rely heavily on the TCP three-way handshake for state monitoring. The TCP SYN, ACK, and SYN ACK flags help ensure only sessions that complete the three-way handshake become ESTABLISHED as active sessions. Stateful firewalls can also monitor the state of connectionless protocols such as UDP. With these protocols, connections typically register in the ESTABLISHED state as soon as the first packet hits the firewall.

By monitoring state tables, these firewalls add efficiency, security, and granularity of control to the packet-inspection process. Instead of checking each packet against the entire firewall rule set, a simple comparison against the active state table can be performed.

Session Border Controllers

Session Border Controllers (SBCs) represent a new breed of VoIP-aware firewall technology. They're optimized for VoIP traffic and specifically support call controls for admission at the edge of the network. SBCs may also provide some call-control functions to offload them from other call agents within the network. There are two logical components of functions in SBCs:

- The signaling function in an SBC, SBC-SIG, provides access control over the VoIP signaling protocol messages sent to the VoIP services core of the network.
- The media function in an SBC, SBC-MEDIA, provides access controls over the actual media (VoIP) packets in the network. The function provides support for Quality of Service (QoS) mechanisms such as differentiated services (DiffServ) for different types of media (voice, video, shared applications, and so on).

SBCs can be used as a firewall or operate as a VoIP-specific adjunct to an existing firewall. SBCs eliminate bad VoIP signaling and media protocols at the network boundary. Doing so improves security at the VoIP perimeter and can prevent malformed or malicious signaling packets from entering the network causing potential problems. In addition, SBCs can enable VoIP signaling and media streams to pass to devices behind a firewall and NAT at the network border, thereby minimizing the need to upgrade existing network-element firewalls.

As part of a defense-in-depth approach to layered security, SBCs help hide the topology of the network. This functionality is particularly important for service providers who offer VoIP services to multiple customers.

SBCs often provide Network Address Translation (NAT) service. It's a common industry practice to use RFC-1918 private address space for carrying VoIP traffic internally.

In large VoIP deployments and service provider networks, SBCs can provide call admission control. This feature allows for gracefully rejecting unwanted calls, a measure of protection against Denial of Service (DoS) attacks and unexpected surges in traffic that lead to network congestion. Admission control can also support service guarantees and help police service level agreements (SLAs).

Additional functions provided by SBCs include:

- QoS—Resource reservation processing
- Media bridging—To support not only voice but also fax or modem traffic over VoIP
- Fault tolerance—Backup, failover, and redundancy
- Policy-based routing of VoIP calls
- Interworking between signaling protocols such as H.323, SIP, Megaco/MGCP, and Skinny
- Media transcoding when different codecs are used on each VoIP end point
- Call Detail Reporting (CDR) for monitoring call progress for billing purposes

SBCs are usually deployed in the DMZ of a network. The DMZ is the conceptual term for a small subnetwork (or individual device) that sits between a trusted private network, such as a corporate private LAN, and an untrusted public network, such as the public Internet. The purpose of the DMZ is to prevent hostile or unwanted traffic from entering (or, in some cases, leaving) the private network.

SBCs are somewhat controversial to proponents of end-to-end systems and peer-to-peer networking, as SBCs might impede those services. However, they are well received in enterprise networking environments that want to secure against either real or perceived security threats. Such is the case, in part, because SBCs break the end-to-end transparency of a peer-to-peer connection. SBCs must understand new VoIP features to be able to support them. For VoIP encryption to work, the SBC may need to have the encryption key.

An SBC might provide session media, generally Real-time Transport Protocol (RTP), and signaling, usually Session Initiation Protocol (SIP) wiretapping capability. This may be important for public VoIP service providers to comply with law enforcement requests to monitor or intercept conversations.

Which Firewall Is Right for VoIP?

There is no simple answer to the question of which firewall is the ideal VoIP firewall type. Simple packet filters may be adequate for small businesses but are rarely acceptable as the primary security method in larger enterprises. These simple access control lists (ACLs) are most widely used inside large networks to keep traffic flowing where it should.

The battle between proxy firewall and stateful inspection techniques has a long history, with outspoken proponents of each side. Proxies require awareness of every protocol they are proxying, which often means an add-on or plug-in module has to be installed and configured to support each and every protocol. Stateful inspection requires significant memory and CPU resources to maintain state tables. In a large network, state tables can contain information about many thousands of simultaneous connections.


SBCs have been gaining in popularity as VoIP has gained deeper penetration in the past 2 years. They provide enhanced security designed specifically to support VoIP traffic. As VoIP continues to gain acceptance and become a universal service in networks, SBCs will continue to grow in popularity and widespread use.

It is be important to recognize VoIP as a technology leading toward ubiquitous multimedia networks. As video- and data-sharing tools improve and gain acceptance, the SBC used today for VoIP may provide the framework for the Multimedia Border Controller of the next-generation network.

Telephony, Regulation, and VoIP

By Ken Camp

Bringing new technologies such as VoIP into service presents a wide range of technical challenges. Given the highly regulated environment of telecommunications, VoIP presents a set of regulatory challenges. For the most part, these challenges present hurdles to VoIP service providers who want to deliver commercial services to consumers and businesses and don't directly impact business VoIP deployment. The intent of this article is to provide a brief glimpse into VoIP regulations and considerations for business users. Businesses that embrace managed VoIP services might want to review some of these regulatory issues, such as E-911 services, with the managed VoIP service provider.

 For readers truly interested in the depth and complexity of global telecommunications regulation, *The Practical Handbook for Telecommunications Regulators* is an excellent resource (<http://www.infodev.org/content/library/detail/842>).

In the United States, regulation began in 1866 with the signing of the Post Roads Act, which gave the U.S. Postmaster General control over the telegraph industry. Today, in the U.S., the interstate telecommunication industry is regulated by the Federal Communications Commission (FCC), which was formed with the Communications Act of 1934. Individual states' public utility commissions (PUCs) regulate communications within their jurisdictions. The FCC also regulates the use of wireless radio frequencies through a system of spectrum allocation and licensing.

Historically, the telecommunication industry around the world has been regulated by governments and international organizations, although some telecom regulations are being revamped and even removed to promote healthy competition. In the deregulated environment, many competitive telecom carriers appeared offering new and unique services.

The Internet, conversely, has historically been a highly unregulated environment. Requirements placed on traditional telecom providers often do not apply in the Internet services environment. This reality has placed the traditional telecommunications carriers and the emerging IP services providers, including the broadband cable companies, on opposing sides of regulatory battles.

IP-based service providers want to remain as unregulated as possible in order to explore every opportunity to create revenue streams. Traditional phone companies want VoIP providers to be treated just like phone companies—subjected to all the same rules and regulations. This difference in direction is further exacerbated by the fact that traditional telecommunications companies are also Internet providers and backbone carriers, so they also campaign for loosening of regulations that serve their interests in that regard. Regulation is a very complex issue.

Key Areas of U.S. Telephony Regulation

There are five key areas of regulatory focus in the U.S.: E-911, Common Carrier, Universal Service Fund (USF), Communications Assistance for Law Enforcement (CALEA), and Telephone Number Mapping (ENUM). Let's look briefly at each to get a basic understanding of what the focal area is and why it's important or, perhaps, not applicable to VoIP services.

E-911 Emergency Services

E-911 services represent perhaps the most visible concerns about VoIP. It also represents an issue everyone understands. VoIP providers are all working aggressively to find workable solutions for E-911 services. The market is driving this need just as much as the threat of regulation. VoIP service providers that deliver effective E-911 service will find success. Those without will quickly be left behind and viewed as second-rate. Many VoIP providers make it clear in their Terms of Service that they do not support "*lifeline service*" such as E-911, or that there are restrictions and special conditions that apply.

The general trend is that regulation of these services continues to be somewhat market-driven. There is some agreement that the basic elements of E-911 should be delivered as a public service and not morphed into some commercial revenue stream in addition to VoIP services.

Figure 1 shows one model for next-generation E-911 for mobile VoIP as presented by Tim Lorello, SVP, Chief Marketing Officer, TeleCommunication Systems, Inc., and William Clay, Vice President E-911 Product Development at Level 3 Communications at a recent CTIA tradeshow. Mr. Clay's slide addresses how a next-generation service might:

- Support end user mobility
- Support any Internet-enabled end points
- Allow medical information, video, and so on to be passed during session
- Enhanced PSAP capabilities (transfers, and so on)
- Support disaster recovery (remote PSAPs)
- Increase reliability
- Simplify carrier access to PSAPs

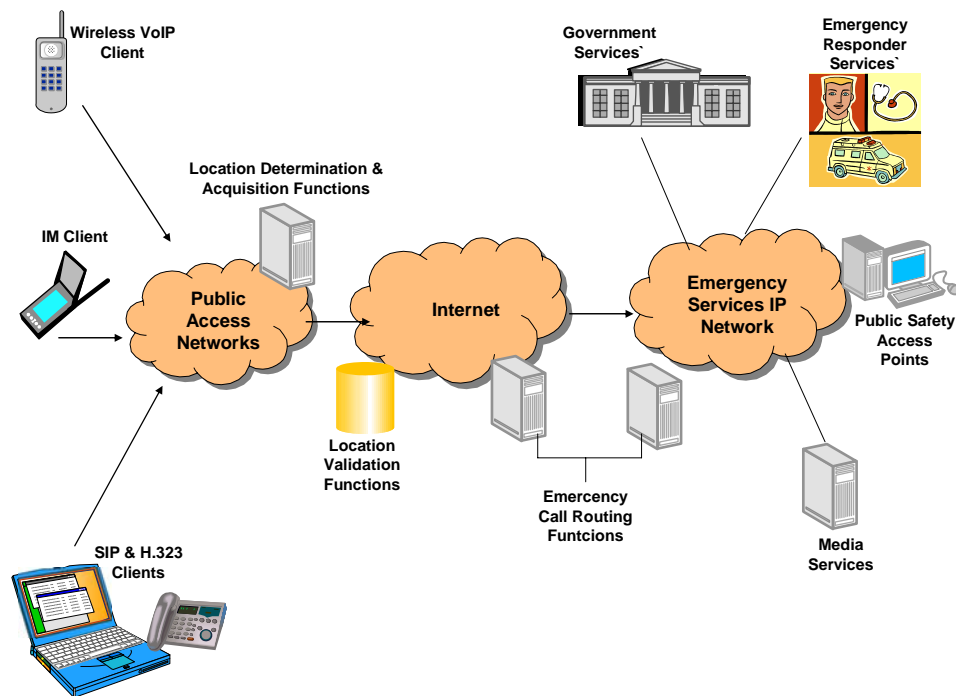


Figure 1: A model for future emergency service.

Future enhancements to the network—such as mobility detection, network/endpoint-generated location, and Location Information Server (LIS) integration—will all be required to make “nomadic” E-911 VoIP work effectively in providing for public safety.

Common Carrier

Carriers of layer-1 (wires) services are defined as common carriers that have traditionally been responsible for packet delivery for a major portion of the Internet at a reasonable price. The FCC has clearly ruled that Internet services do not fall into USF-funded services. It might be several years before the overhaul of layer-1 taxation will be considered again. There is much debate and lobbying about regulation and common carrier status for providers that possess some form of perceived monopoly with regard to serving data (packets, voice, whatever) to a geographic area. As an offset to the penalties accompanying common carrier status, government subsidies or rate controls often enter the picture. One option often promoted in this area of regulatory control is that common carrier status might be levied upon firms that deliver layer-1, 2/3 services (physical facilities and IP transit) in circumstances in which a monopoly or virtual monopoly exists. VoIP providers are not common carriers under that definition, but that definition is not widely accepted by everyone.

USF

There is a payment you see on your phone bill every month, labeled "*USF charge*" This Universal Service Fund was established to help pay for phone service to rural or otherwise disproportionately expensive endpoints of the traditional phone network. Some portions of the USF are taken for educational networking, called "e-rate" funding. Many rural states depend on USF to provide service to residents of remote parts of their states.

The USF represents billions of dollars in cash flow. Any threat to that revenue gets the attention of legislators. Those rural residents and farmers vote. The companies that benefit from USF funding happily contribute to the election campaigns of those legislators, further solidifying the status of USF. VoIP services and networks generally do not pay USF fees. This means that as the population of an area where VoIP is successful migrates off the traditional PSTN, revenues associated with those disconnected traditional phone lines will disappear. There are critics of the USF, who claim the phone companies are fleecing the government for millions of dollars. Certainly the USF is facing increased scrutiny. It's another area of legislation that impacts VoIP far less than traditional telephony, but creates a divisive gap between providers of each service.

CALEA

CALEA standards address support for law enforcement's authorized wiretapping or eavesdropping. In the end, CALEA may not directly apply to a number of VoIP services. This reality is sure to cause many governmental agencies to howl in rage. CALEA may just be impossible in some instances in which VoIP is deployed. Peer-to-peer services or SIP-native bridges may preclude VoIP from being able to offer any support at all. Even where interception is possible, the relevance of methods used by law enforcement make the usefulness of information or recordings gathered somewhat questionable for many people.

It remains difficult for Congress or law enforcement to leverage CALEA as a regulatory roadblock to VoIP. In some cases, CALEA access is just not possible with VoIP. Voice communications are no longer easily accessed through a central-office wiring plant. The technology has changed with VoIP. CALEA still must apply to gateways where VoIP and the PSTN meet, but the regulatory and technical challenges remain a very heated issue.

The Internet and Internet applications were not included in the regulatory mandates CALEA. The U.S. Senate is currently considering the VOIP Regulatory Freedom Act.

The Federal Bureau of Investigation (FBI), the U.S. Department of Justice (DoJ), and the U.S. Drug Enforcement Agency (DEA) filed a joint petition for rulemaking with the FCC arguing that legal wiretapping needs to be "technology neutral" and needs to apply to all telecommunications carriers equally. Specifically, the FBI is asking the FCC to rule that broadband access and broadband telephony are subject to CALEA and to adopt rules that provide for the easy and rapid identification of future CALEA-covered services. This petition envisions a new regulatory and enforcement scheme that imposes deadlines by which VoIP providers must implement CALEA intercept capabilities for their existing services. The FBI is also asking the FCC to confirm that carriers bear sole financial responsibility for CALEA implementation costs.

Two prominent industry groups, the Telecommunications Industry Association (TIA) and The VON Coalition, have each filed public comments with the FCC that are critical of the FBI's recent proposal to update the CALEA wiretapping law to cover VoIP and future packet-based communications.

ENUM

ENUM is a complex topic that could fill many articles or books. In a nutshell, it addresses the integration of Domain Name Services (DNS) and how DNS might be used to identify IP-capable phone switches designated to handle traffic for a dialed telephone number. ENUM is a suite of protocols to unify the telephone system with the Internet by using E.164 addresses with Dynamic Delegation Discovery System (DDDS—defined in RFCs 3401 through 3405) and DNS. ENUM also refers to “E.164 Number Mapping.”

Although it facilitates VoIP, ENUM isn't a VoIP requirement. It shouldn't be confused with VoIP routing based on SIP or H.323 protocols with a Uniform Resource Identifier (URI). VoIP service providers assign a URI to a customer in order to complete calls over the Internet. ENUM is a DNS-based protocol that is envisioned to facilitate the way calls over VoIP networks can be completed. It provides a user with a domain name on an E.164 server that associates a common international telephone number with a URI. The server is maintained by the service provider and is expected to become standard because it can successfully address locating URIs with nothing more than a common international telephone number.

This could be somewhat analogous to the MX record used in DNS to denote mail exchange server information. If a DNS query received a positive result, the phone call could be routed over IP to the destination. There would be no PSTN or phone companies involved. There would be no charge. That “no-charge” component gets a lot of attention.

The phone number could be under the control of a number of parties. If the phone number is under control of Verizon, Qwest, or SBC they won't be very happy about the loss of revenue.

ENUM is a critically important aspect of DNS as networks mature. It will replace, or augment, a system that currently has more users than the Internet does. Many view current ENUM approaches as a stopgap measure. Today, people use and remember phone numbers. We've become accustomed to them over time. In the future, URI-based dialing strings will certainly become more common (that is, ken@ipadventures.com easily becomes a SIP dialing string). Many see this as the future standard for dialing—many VoIP and Internet service providers (ISPs) that is. The SIP URI dialing doesn't translate well to the traditional telephone dialpad.

When we look at the concepts around “regulation,” the owner of the referral servers upstream from a caller may hold the key to the ability to receive calls to a phone number for some years to come. As long as the owner is a traditional telephone company with ties to the telephone instrument dialpad and revenues associated with telephone calls, there will always be impetus for legislative regulation.

Summary

The concepts of regulation have heretofore had little or no impact on Internet services, including VoIP. Congress even took the unprecedented step of declaring the Internet a “regulation-free” zone for a time. This approach led to incredible growth and the proliferation of new technologies. IP technology has now become an everyday staple of business markets. It’s also bringing about wide-scale replacement of legacy technologies. The industry has to accept that some regulation is part of that evolution.

As network technologies such as VoIP converge with highly regulated industries, such as telecommunications, developers and implementers must consider the implications of their designs with regard to current regulations. Otherwise, unpleasant results are sure to follow.

When regulations don’t fit, we need to incorporate alternative legal frameworks. Whatever path the U.S. takes to regulation, it will likely have significant impact on other countries’ regulatory stance. The European Union has been ahead of the U.S. in defining VoIP technology regulation, although there is still some inertia to overcome with state-owned telephone companies holding enormous power in the E.U. The rest of the world is likely to follow the U.S. and E.U in developing their own VoIP regulation guidelines.

VoIP is undergoing incredible growth. 2006 is forecasted on all fronts to be a huge growth year for VoIP deployment in every sector. The combination of growth in service and uncertainty in regulation is potentially volatile as regulatory stances are being established on both sides. Regulations need to mature. That often means removing outdated ideas of technology from regulatory wording.

For businesses implementing VoIP services, the regulatory environment bears watching. Any enterprise contracting for managed VoIP services needs to understand, at a minimum, what E-911 and CALEA services are being provided. Today, the requirement is on service providers, but as businesses create peering agreements and become providers in their own right, regulation that once governed the telecommunications industry could spill over into other business sectors.

Network Neutrality and Enterprise Business

by Ken Camp

There has been a great deal of conversation on the Internet about the issue of *Net Neutrality* and how it impacts the evolution of the Internet. What we now call *Network Neutrality* began as something called the *End-to-End Principle*. The nature of the *Net Neutrality* debate is driven by unified communications subjects, including the broad unified communications efforts including VoIP, IM Multimedia Subsystems (IMS), video collaboration, and fixed mobile convergence.

This article will introduce the concept of *Network Neutrality* for business and technical managers. It will survey some of the published viewpoints on *Net Neutrality*, both for and against, and will begin delving into the potential impact on enterprise business. Let's begin with background information and published opinions from the Web on the subject. Although not quoted in their entirety, the articles are extensively hyperlinked to ease further research into the discussion.

What Is Network Neutrality?

We'll begin with a definition from Wikipedia, an excellent online resource (http://en.wikipedia.org/wiki/Network_neutrality):

Network neutrality is synonymous with common carrier status for providers of the pipe to the network. It is a term devised by law professor Tim Wu to support an open access theory of network regulation that holds that packet data networks such as the Internet access as common carriage should be just like traditional railroads and telephone networks. According to this view, the Internet transport and service system, having received valuable public right of way should not be allowed to be privately owned without open access so that owners could provide services that they own on favorable terms and thus unfairly compete with third-party services.

Network neutrality has been expanded by others into a general theory of network operational architecture. It means that the network is operated under the three principles of neutrality: non-discrimination, interconnection, and access. The principles can apply to any network, but are generally ascribed to the Internet. They govern the operation of the network, not the content or business practices of the network operator. Inherent in the definition is that network operations are distinct from the content side. Network neutrality is one way to describe the operational architecture of the global Internet. Nearly every nation operating a portion of the Internet, often by default, has adopted some form of the neutrality principle, depending on its definition.

Voices Favoring Net Neutrality

David Isenberg

David's seminal work, "Rise of the Stupid Network," (<http://www.isen.com/stupid.html>) written in 1997, provides much of the framework and foundation for today's discussion of Net Neutrality. David's work in the years since has included "The Paradox of the Best Network" (<http://netparadox.com/>) with David Weinberger and "The End of the Middle" (<http://www.spectrum.ieee.org/WEBONLY/publicfeature/jan03/clude.html>).

The premise of David's work, in a nutshell, is that the traditional, legacy telephone network contained all the intelligence at the core, using unintelligent telephone devices at the edge. With the advent of faster and cheaper computing and advances in networking technology, the need for a "smart" network no longer exists. The intelligence of the network, the Internet, lies not within the core of the network itself, but at the periphery. Whether the end device is a computer, a mobile phone, a PDA, or a set top box, the edge device has become the key delivery mechanism for network services.

David's consulting, speaking, and writing drew him to coordinating a conference entitled Freedom2Connect (<http://www.pulvermedia.com/f2c/>). This conference was joined by many telecommunications and networking industry experts including former FCC Chairman Michael Powell.

The idea behind Freedom2Connect is as follows:

The need to communicate is primary, like the need to breathe, eat, sleep, reproduce, socialize and learn. Better connections make for better communication. Better connections drive economic growth through better access to suppliers, customers and ideas. Better connections provide for development and testing of ideas in science and the arts. Better connections improve the quality of everyday life. Better connections build stronger democracies. Strong democracies build strong networks."

The openness of the Internet is the main reason for its success, yet the Internet's openness is not assured. The most viable bill before the U.S. Congress would require even the smallest wireless hotspot provider to register with the U.S Government. In addition, the FCC is imposing requirements to provide Internet emergency dialing and law enforcement assistance in ways that only the biggest, least innovative connection providers can hope to meet.

Too often the discussion of telecommunications policy turns on phrases like "overregulation," and "investment incentives." These are critical issues, to be sure, but like the term "last mile," such phrases frame the issues in network-centric terms. As more and more intelligence migrates to the edge of the network, users of the network need to be part of the policy debate. Let's put the user back into the picture. F2C:Freedom to Connect provides the frame.

David has also posted numerous articles on the Web discussing the premise behind the end-to-end principle, or *Net Neutrality*.

In “Financial Sector Needs Net Neutrality” (<http://isen.com/blog/2006/05/financial-sector-needs-net-neutrality.html>), he writes:

If the telcos offer a high-security tier of service, then banks MUST use it or face liability if there’s a security incident, writes Philip Corwin in the April 21 issue of American Banker [link, two-week trial subscription required]. Corwin writes:

Those payments [for premium-tier service -- David I] may become both a competitive and legal necessity. The security of Internet access to financial institutions customer information is encompassed by the “[Safeguards Rule](#)“ of the Gramm-Leach-Bliley Act; and failing to provide the highest level of security could become a liability factor for online data breaches. Those that do pay may find they cannot pass these new costs directly along to customers for whom “free” is an essential element of online financial services. And those that decline to pay will face growing customer dissatisfaction as their transactions are shunted to the Internet’s increasingly congested, low-priority slow lanes.

Corwin is urging the banks to support Network Neutrality legislation.

Indeed, any entity that’s a customer of the telcos and depends on the Internet to do business should support non-discrimination! This includes hospitals, catalog sales companies, airlines, newspapers . . . just about any company that does business on the Internet.

Corwin continues:

A major ISP could designate a single diversified financial services company as its exclusive preferred partner and deny the same quality of connectivity to its competitors.

As David points out, network neutrality is difficult to define, and to implement, but the need for open, unfettered access to the Internet is a crucial facet of the information age we live in.

In “Net Neutrality: an international issue” (<http://isen.com/blog/2006/05/net-neutrality-international-issue.html>), David quotes Ethan Zuckerman (<http://www.ethanzuckerman.com/blog/>), putting an international spin on the network neutrality issue from the May 2006 issue of Inc. Magazine (<http://www.inc.com/magazine/20060501/index.html>):

[Network Neutrality] is under assault. Foreign countries have led the charge. Saudi Arabia blocks content that runs counter to the clerics’ interpretation of Islam. China bars its citizens’ access to sites created by, among others, practitioners of Falun Gong. [Apparently China also blocks pulver.com and pulvermedia.com -- David I] What results is the fragmentation of the Internet. The network we’ve grown accustomed to over the past decade is, in a very real sense, becoming multiple Internets, because the Internet you encounter from within China is different than the Internet you encounter in the United States . . .

David’s passion and focus on the issue has long been widely recognized in the telecommunications sector. David isn’t alone in his support for the notion of *Net Neutrality*.

 For more information about David Isenberg, see <http://www.isen.com/>.

Junaid Islam in Converged Digest

Junaid Islam's career in telecommunications and networking spans several generations of packet networking, including X.25, Frame Relay, ATM, and MPLS. Junaid has been active in the development of standards and holds multiple patents. He has a B.S. degree in computer science from the University of Toronto. In "Fair Use Networking: Preserving 'Net Neutrality' with Enhanced QoS" (<http://www.convergedigest.com/bp-ttp/bp1.asp?ID=354&ctgy=Intelligence>), his second article about fair use networking, he explains how Quality of Service (QoS) mechanisms might be used to preserve the idea of *Net Neutrality*:

"Net Neutrality" has become shorthand for a debate over the best ways to ensure that consumers of broadband services can access any and all Internet sites, services and applications, and to ensure that Internet-based content and service providers are not hindered in making their offerings available to the public. The industry has been debating this topic for years. In recent months, the debate has become more contentious, however, and has essentially condensed into two warring camps: those who argue that any management of broadband network traffic will severely harm the Internet and American technological innovation; and those who claim the uninhibited right to impose any traffic management rules and pricing structures that they deem necessary to achieve an acceptable return on investment.

In the meantime, congestion on broadband networks continues to grow as new, bandwidth-intensive services and multimedia applications proliferate. Indeed, without such traffic overload on the Internet, everyone would be content and there would be no anxiety over Net Neutrality. So the underlying issue here—often ignored in the debate—is how to manage congestion effectively and fairly. Developing an industry consensus regarding acceptable approaches to traffic management would help minimize or even eliminate regulatory uncertainty, and finally give network operators satisfactory returns on the investments needed to expand network capacity.

Junaid goes on to introduce a framework for fair use that addresses many concerns through the application of three increasingly granular traffic management policies:

- Proportioning available bandwidth equally among all active users
- Prioritizing available bandwidth for critical traffic needs, particularly those that involve public safety
- Provisioning available bandwidth into QoS tiers that may include different rate structures

Success of Junaid's technique argues fairness is provided by enforcing policy on every packet in the network. This approach accomplishes three objectives toward the goal of fair use of network resources:

- Proportion bandwidth equally among all active users
- Prioritize traffic for regulatory compliance
- Provision available bandwidth into nondiscriminatory QoS tiers

As Junaid states in his conclusions, this method may provide a versatile and non-intrusive means of managing network congestion that could be a moderate approach in the contentious debate over Net Neutrality. His suggested Fair Use Policy Framework offers a hierarchical approach that could mitigate congestion and the impact congestion has in a *best efforts* IP network.

Daniel Berninger, VP, Sr Analyst, Tier1 Research

Daniel Berninger offers thoughts on “Why Even Bells Need Net Neutrality” (<http://gigaom.com/2006/05/09/why-even-bells-need-net-neutrality/>) via Om Malik’s blog. Malik is with *Business 2.0* magazine. Daniel begins by defining Net Neutrality as “Internet access without discrimination by use or user except as required for network management purposes.”

In describing recent legislative action, he says:

The FCC’s decision to relieve AT&T and Verizon of net neutrality requirements in August 2005 definitively broke the chain of events the companies use to assert right-of-way privileges. The Bells claim privileges based on over 100 years of practice that may or may not coincide with the intent and limits of the original deals, but the resulting laws explicitly require a public purpose in exchange for the right-of-way concessions.


The obligations established on a state by state basis sometimes include build-out requirements or other compensation, but they all specify that access to state right-of-way at largely no cost or limit requires common carrier status (aka net neutrality.) The loss of common carrier status invalidates the contracts. The Bell companies have no access to state right-of-way for deployment of private, closed, non-neutral, non-common carrier network deployments.

He cites operational practices and historical behavior of the major telecommunications carriers and concludes saying:

The Bell companies need to stop the neutral Internet from erasing the legacy telephone network’s voice revenues. Price discrimination enables metering of Internet access by keeping per bit price of low bandwidth voice relatively high while offering relatively lower per bit prices to initiate a video revenue stream. Net neutrality stands in the way of their becoming digital economy toll collectors.

Daniel posted a follow up piece entitled “If It’s Not Neutral It’s Not Internet” (<http://www.danielberninger.com/dbessays/2006/05/if-its-not-neutral-its-not-internet.html>), and encapsulates his core argument in the opening paragraph:

The success of a proposal by AT&T and Verizon to end net neutrality does not threaten the Internet. The broadband customers of AT&T and Verizon will just no longer have access to the Internet. The development appropriately creates alarm among AT&T and Verizon’s customers, but the combined customer bases of these companies represent less than 2% of the billion or so users of the Internet. The fact access to the Internet requires net neutrality does not depend on laws passed by the US Congress or enforced by the FCC. Neutrality arises as a technical and business imperative facilitating the interconnection 250,000 independent networks that choose to participate in the Internet. Net neutrality will remain a requirement as long as it serves the interests of the global Internet community.

 Other Voices Supporting Net Neutrality

Vint Cerf, Chief Internet Evangelist, Google, Inc. Speaks Out on Net Neutrality
(<http://googleblog.blogspot.com/2005/11/vint-cerf-speaks-out-on-net-neutrality.html>)

Rep. Ed Markey - Rep. Markey (D-Mass.) has introduced a bill in the House to bring back Net Neutrality.

Sen. Olympia Snowe - Sen. Snowe (R-Maine) will cosponsor a bill with Sen. Byron Dorgan supporting Internet Freedom.

Voices Opposing Network Neutrality

Martin Geddes

Martin is a globally recognized telecommunications industry consultant. Formerly with Sprint, Martin lives in Edinburgh and speaks around the world about burning telecom issues. He spoke at the Freedom2Connect conference, and posted his notes in “OPINION://F2C: Network neutrality speech” (<http://www.telepocalypse.net/archives/000905.html>):

Within the current funding and construction approach to networks, I believe a network neutrality law is a tactical, practical, strategic and philosophical error. It takes us further away from Freedom to Connect.

A tactical error

As a tactical proposition, it supposedly is to solve a problem. So what’s the problem it might be solving?

Well, we certainly have a consumer protection and disclosure issue. Consumers are buying a product that has ‘broadband internet’ on the tin can, but they don’t really know what’s inside. The terms are often too obscure, hidden in obtuse language in the terms and conditions. That means we need disclosure along the lines that exist (in the US) for credit cards, where the key terms must be presented in a standard format. Furthermore, I’d suggest that there be regulated terms like “full internet access” and “partial internet access” to make it crystal clear.

Then there’s the issue of monopoly and unfair competition. But there already are laws to deal with this. Sherman Act, RICO, and so on. The problem is process and organization; getting the FTC, not the FCC, to enforce them.

Martin continues later in his talk:

Is there a “moral right” to neutrality and traditional common carriage rules, given that the networks were (supposedly) built using public money? Sorry, no. The elected representatives in the US made a bunch of terrible deals on behalf of the public. Telcos got great deals in exchange for empty, unenforceable promises. The capital has all been transferred into private hands. Gone!

There’s a frequent complaint that “the Net needs us”, and is under attack. But it’s never been healthier. We’ve never had so many people so well connected. It’s an emergent outcome of individual actors expressing their preferences via voluntary exchange. And they, by and large, demand an open connection! It’s not sacred, an object of worship. We can think of better Internetwork architectures.

A practical error

Network neutrality can’t be made to stick. Telcos will evade whatever definition you put up; it’s easier than fighting UNE-P unbundling rules. It’s easy to create tilted playing field.

Not just a tactical error and practical error, Martin goes on to articulate why the premise of Net Neutrality is also both a strategic and philosophical error. He explains his view of a new funding model that would encompass a broader base of beneficiaries. Martin believes this new model would give the traditional telecommunications carriers incentives to innovate in pricing, financing, and ownership of networks. Although his opinions are detailed and encompass vagaries of the telecom business that boggle the minds of business people, he summarizes his position quite handily:

An open, free net is an emergent outcome, not an a-priori input to be legislated into existence. We need to capture and accelerate the experiments in how networks are built, financed and sold; and protect those experiments from incumbent wrath until the results are in.

But most critically, don’t fossilize the network in 2006 by adopting network neutrality.

 For more information about Martin Geddes, see <http://www.telepocalypse.com/>.

John Earnhardt on the Cisco High Tech Policy Blog

Vendors and carriers' viewpoints are a bit more challenging to understand. Telcos speak through press releases and often don't address the issue directly. In "Net Neutrality Debate is a 'False Choice'" (http://blogs.cisco.com/gov/2006/05/net_neutrality_debate_is_a_fal.html), John articulates a viewpoint from Cisco Systems quite well:

Today, I was talking to Robert Pepper, former long-time FCC'er and current Cisco colleague, and he described Net Neutrality as a "false choice." The way the debate has currently been framed has it as an all or nothing scenario, i.e. you need to regulate or legislate or, the alternative, consumers get it in the shorts. This is just patently false.

It is also too bad that some in this debate are trying to make this consumers vs. businesses. Do you really think that the big businesses who are arguing to legislate net neutrality have the best interests of consumers in mind? Sure they do...as long as it also makes them a buck.

Our argument has been give to everybody access to legal applications on the internet but to allow providers to optimize the consumer experience, much like cell phone companies do. Let me 'splain: If you only want to talk a little bit, then you only pay a little bit. If you want to talk a lot, then you pay more. If you only use your phone at certain times during the day (i.e. nights or weekends) then you can get a package that provides for that as well. In my mind, this is all providers are asking for - flexibility to give the consumers the best experience.

It's not overly surprising that Cisco's position seems to support the views of its largest customers, the telecommunications industry. Here are two excerpts from online news articles presenting reaction from Verizon:

"Verizon: Net neutrality concerns are 'hypothetical'" (http://news.zdnet.com/2100-9588_22-6068848.html):

C. Lincoln "Link" Hoewing, an assistant vice president at Verizon Communications, said that the ability to charge for services such as high-quality video is crucial to being able to afford the multibillion-dollar price tag of upgrading its network-to-fiber links.

"We could put other services on those pipes--it's got a lot more capacity to do this," Hoewing told the Computers, Freedom and Privacy conference here. That would help "to make it more viable economically and financially and to help us compete."

Calling concerns about Net-favoritism entirely hypothetical, Hoewing said: "I'm getting tired of it...We've never done anything that I know to interfere with anyone's traffic."

“Verizon: Net Neutrality Limits Broadband’s Potential”

(<http://www.pcworld.com/resource/article/0,aid,125666,pg,1,RSS,RSS,00.asp>):

Current congressional attempts to write a so-called Net neutrality provision into law would stop broadband network operators from offering VPNs (virtual private networks) to online gaming vendors looking to improve connectivity or to hospitals launching home health-monitoring services over IP (Internet Protocol), said Tom Tauke, Verizon’s executive vice president of public affairs, policy, and communications.

Note: Tauke’s speech--at a broadband policy summit sponsored by Pike & Fischer, a research and publishing company--was a focused rebuttal to consumer groups and e-commerce firms calling for a Net neutrality provision to be included in telecommunications reform legislation now being debated in Congress.

Clear Voices of Reason

Adam D. Thierer is director of telecommunications studies at the Cato Institute

(<http://www.cato.org>) and coauthor, with Clyde Wayne Crews Jr., of *What’s Yours Is Mine:*

Open Access and the Rise of Infrastructure Socialism (Cato Institute, 2003). In “Digital

Discrimination or Regulatory Gamesmanship in Cyberspace?” (<http://www.cato.org/pubs/pas/pa-507es.html>), he writes:

...far from being something regulators should forbid, vertical integration of new features and services by broadband network operators is an essential part of the innovation strategy companies will need to use to compete and offer customers the services they demand. Network operators also have property rights in their systems that need to be acknowledged and honored. Net neutrality mandates would flout those property rights and reject freedom of contract in this marketplace.

The regulatory regime envisioned by Net neutrality mandates would also open the door to a great deal of potential “gaming” of the regulatory system and allow firms to use the regulatory system to hobble competitors. Worse yet, it would encourage more FCC regulation of the Internet and broadband markets in general.



Other Clear Voices

Ben Worthen, CIO Magazine “What is the net neutrality debate about?”

(<http://blogs.cio.com/node/200>).

Andy Oram, O’Reilly Emerging Telephony “Network neutrality and the false idol of innovation”

(http://www.oreillynet.com/etel/blog/2006/05/network_neutrality_and_the_fal.html)

Net Neutrality in Summary

Summarizing a complex issue such as *Net Neutrality* is an ongoing task. The debate rages on, and this topic is likely to remain in the public eye for some time to come. But the question as to business impact remains open.

Net Neutrality has direct implications on the service provider market. How telcos deliver services, what constraints they might place on the network, and how they charge other content providers will all be affected throughout this ongoing struggle. Traditional enterprise business will likely see little impact. Large enterprises typically have longstanding relationships with telecommunications providers. Industry segments that are not directly Internet-content-related (banking and finance, manufacturing, healthcare, and so on) will see little change at the large enterprise view. Enterprise engaged in content delivery (search engines, audio and video broadcast, and so on) and the emerging online media services companies will likely continue as strong advocates for *Net Neutrality*. It's an issue that directly affects their ability to compete.

Small and medium businesses may well be impacted across the board, regardless of industry focus. These companies are often in a position that leaves them feeling at the mercy of the large telcos and service providers. For the small business, sole proprietorship, and technology consulting segments, the issues surrounding Net Neutrality play directly to their ability to employ VoIP, IP video collaboration services, and other IMS solutions. Although the Net Neutrality debate rages on, it behooves us all to keep a close eye on the issues and take a stand where our business services may be affected.