

White Paper

Voice Over IP 101

Understanding VoIP Networks

Stefan Brunner
Senior Network Security Consultant

Akhlaq A. Ali
Senior Marketing Engineer



Juniper Networks, Inc.
1194 North Mathilda Avenue
Sunnyvale, CA 94089 USA
408 745 2000 or 888 JUNIPER
www.Juniper.net

Part Number: 200087-001 Aug 2004

Contents

Introduction	3
Why VoIP?	3
VoIP Functions	4
Signaling	4
Database Services	4
Call Connect and Disconnect (Bearer Control)	4
CODEC Operations	5
VoIP Components	5
Call Processing Server / IP PBX	6
User End-Devices	7
Media/VOIP Gateways/Gatekeepers	8
IP Network	9
VoIP Signaling Protocols	9
H.323	10
Real-time Transport Protocol (RTP)	11
Real-time Transport Control Protocol (RTCP)	12
Media Gateway Control Protocol (MGCP)	13
Session Initiation Protocol (SIP)	13
Megaco/H.248	15
VoIP Service Considerations	15
Latency	15
Jitter	17
Bandwidth	17
Example	18
Packet Loss	18
Reliability	19
Security	19
Conclusion	20

Introduction

Although voice over IP (VoIP) has been in existence for many years, it has only recently begun to take off as a viable alternative to traditional public switched telephone networks (PSTN). Interest and acceptance has been driven by the attractive cost efficiencies that organizations can achieve by leveraging a single IP network to support both data and voice. But cost is not enough to complete the evolution; service and feature parity is a main requirement. Customers will not accept voice quality or service that is inferior to what they are used to with a PSTN and, until now, VoIP fell short in delivery.

Voice protocols have evolved to offer a richer set of features, scalability and standardization than what was available only a few years ago. The pace of service integration (convergence) with new and existing networks continues to increase as VoIP products and services develop. Critical to success is the ability to deploy value-added and high-margin services. For example, a service provider can deploy a unified messaging system that synthesizes voice and e-mails over a phone to the subscriber.

This paper will explain the fundamentals of VoIP, focusing on the functions and components that make up a VoIP solution. It will answer the following questions: What does it mean to an organization to deploy VoIP? What makes up a VoIP solution and how can they take advantage of it? Once a general understanding of VoIP is achieved, organizations are better prepared to tackle the more complex issues that go into deploying a secure, reliable and high-performance VoIP network.

Why VoIP?

The cost-effectiveness is initially attractive when looking into VoIP. It is evident that an organization can gain efficiencies by only having to support a single network infrastructure. By using a single packet-switched network, as opposed to having to manage both packet and circuit-switched networks, organizations can realize reduced maintenance and management costs. The same technical personnel are able to operate both voice and data systems instead of requiring resources with different expertise.

This convergence of voice and data networks onto a single IP network also provides some inherent flexibility, in terms of being able to easily add, change or remove nodes (e.g. phones) on the network. As a result, organizations can easily deploy and then redeploy equipment to maximize their investments, without having to do a truck roll or require special expertise on hand.

Finally, VoIP promises to deliver many nice new features, such as advanced call routing, computer integration, unified messaging, integrated information services, long-distance toll bypass, and encryption. Because of the common network infrastructure, it is also possible to integrate other media services, like video or even electronic white boards, to name a few. An example of such features would be the "follow-me" feature where a person is always reachable at the same extension, whether telecommuting from his Lake Tahoe cabin, staying in a hotel abroad, or sitting at his desk in the office. Another feature would be the integration of VoIP with customer relationship management (CRM) software. CallerID or dialed numbers could be linked to a customer's record, which automatically opens on the desktop when the Sales person receives or places a call.

Due to the cost-effectiveness, flexibility and promise that leveraging a single IP network offers, it is no wonder that organizations are looking hard at the VoIP technology and trying to figure out how best to use it to their advantage.

VoIP Functions

Before going into a discussion of the components that make up a VoIP solution, it is important to understand the basic functions of VoIP, particularly as they compare to current PSTNs. As mentioned above, in order to enable organizations to adopt VoIP as a viable solution, its components must be able to perform the same functions as the PSTN network. These are:

- Signaling
- Database services
- Call connect and disconnect (bearer control)
- CODEC operations

Signaling

Signaling is the way that devices communicate within the network, activating and coordinating the various components needed to complete a call.

In a PSTN network, phones communicate with a Class 5 switch (analog) or traditional private branch exchange (PBX) (digital) for call connection and call routing purposes.

In a VoIP network, signaling is accomplished by the exchange of IP datagram messages between the VoIP components. The format of these messages may be dictated by any number of standard protocols, which are covered later in this paper.

Database Services

Database services are a way to locate an endpoint and translate the addressing that two (usually heterogeneous) networks use. A call control database contains these mappings and translations. Another important feature is the generation of transaction reports for billing purposes. You can employ additional logic to provide network security, such as to deny a specific endpoint from making overseas calls. This functionality, coupled with call state control, coordinates the activities of the elements in the network.

A PSTN uses phone numbers to identify endpoints.

A VoIP network uses an IP address (address abstraction could be accomplished with DNS) and port number to identify an endpoint.

Call Connect and Disconnect (Bearer Control)

The connection of a call is made by two endpoints opening a communication session between one another. In the PSTN, the public (or private) switch connects logical (Digital Signal) DS-0 channels through the network to complete the calls.

In a VoIP implementation, this connection is a multimedia stream (audio, video, or both) transported in real time. This connection is the bearer channel and represents the voice or video content being delivered. When a communication is complete, the IP sessions are released and optionally network resources are freed.

CODEC Operations

Traditional voice communication is analog, while data networking is digital, as a result, the network needs a way to be able to convert the voice into a format that it can transport. Since the PSTN is often analog, this is not necessarily a major function, however, for VoIP, it is necessary for “packetiz-ing” the voice. The process of converting analog waveforms to digital information is done with a coder-decoder (CODEC, which is also known as a voice coder-decoder [VOCODER]). There are many ways an analog voice signal can be transformed, all of which are governed by various standards. The process of conversion is complex and beyond the scope of this paper. Suffice it to say that most of the conversions are based on pulse coded modulation (PCM) or variations. Each encoding scheme has its own history and merit, along with its particular bandwidth needs.

The output from the CODECs is a data stream that is put into IP packets and transported across the network to an endpoint. These endpoints must use the standards, as well as a common set of CODEC parameters. If two endpoints use different standards or parameters then the communication will be unintelligible. Table 1 lists some of the more important encoding standards covered by the International Telecommunications Union (ITU). Notice the tradeoff between encoding efficiency, reduced bandwidth consumption, and increased conversion delay.

ITU Standard	Description	Bandwidth (Kbps)	Conversion Delay (ms)
G.711	PCM	64	< 1.00
G.721	ADPCM	32, 16, 24, 40	< 1.00
G.728	LD-CELP	16	~ 2.50
G.729	CS-ACELP	8	~ 15.00
G.723.1	Multirate CELP	6.3, 5.3	~ 30.00

VoIP Components

The major components of a VoIP network, while different in approach, deliver very similar functionality to that of a PSTN and enable VoIP networks to perform all of the same tasks that the PSTN does. The one additional requirement is that VoIP networks must contain a gateway component that enables VoIP calls to be sent to a PSTN, and visa versa. There are four major components to a VoIP network.

- Call Processing Server/IP PBX
- User End-Devices
- Media/VOIP Gateways
- IP network

Call Processing Server / IP PBX

The call processing server, otherwise known as an IP PBX, is the heart of a VoIP phone system, managing all VoIP control connections. Call processing servers are usually software-based and can be deployed as a single server, cluster of servers, or a server farm with distributed functionality. Call processors may also be based on a router platform or developed as a dedicated appliance.

VoIP communications require a signaling mechanism for call establishment, known as control traffic, and actual voice traffic, known as voice stream or VoIP payload. VoIP control traffic follows the client-server model, with VoIP terminals, including messaging servers that hold voice-mail messages representing the clients that communicate to the call processing servers.

With the exception of routed voice traffic to another call processing server, conferencing functionality and music-on-hold, call processing servers do not handle VoIP payload (which is the RTP stream carrying voice itself) traffic. VoIP payload flows in a peer-to-peer fashion – from every VoIP terminal to every other VoIP terminal. In this case, the VoIP terminals determine traffic flows and the call processing servers negotiate those flows within the control messages. A typical VoIP setup with Call Processing Server is shown in Figure 1.

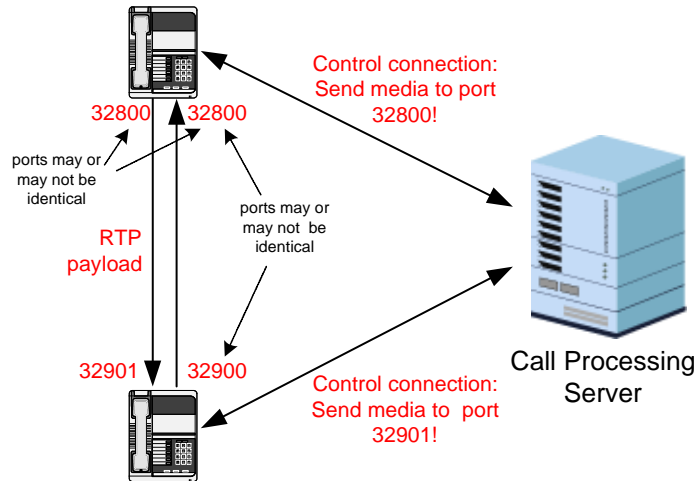


Figure1: Call Processing Server

Figure 2 shows how different signaling protocols can be used by these Call Processing Servers to communicate with IP Phones, Gateways/Gatekeeper, which will be discussed in a following section. Signaling protocols and their functions are also described later, in the Signaling Protocols section.

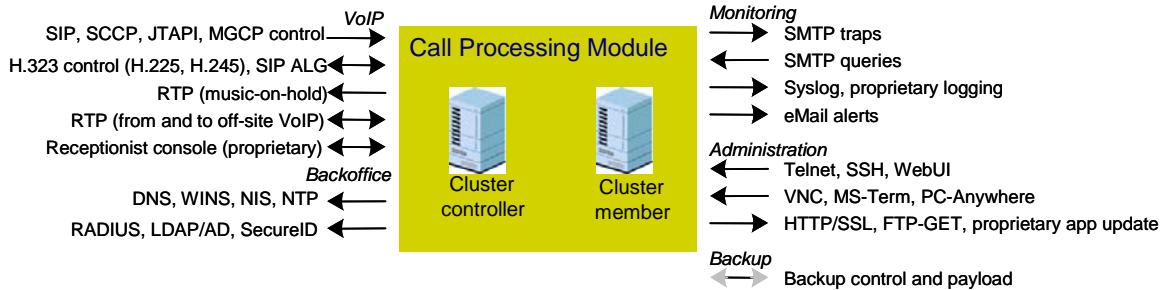


Figure 2: Call Processing Server Signaling

User End-Devices

The user end-devices consist of VoIP phones and desktop-based devices. VoIP phones maybe software based (“softphones”) or hardware based (“hard phones” or “handsets”, like traditional phones).

- VoIP phones use the TCP/IP stack to communicate with the IP network, as such, they are allocated an IP address for the subnet on which they are installed. VoIP phones may also use additional protocols to support VoIP-enabled features, such as built-in IM applications or directory search functions. Typically, VoIP phones use DHCP to auto-configure themselves, with the DHCP server telling the phone about the location of the configuration server, which most of the time is identical to the call processing server.
- Softphones are software application running on notebook computers, usually targeted towards mobile users. They have the same base features as VoIP phones.
- Consoles, on the other hand, are applications with certain control characteristics. Consoles usually include a Softphone, but may also interact with a legacy phone, via a voice gateway or a VoIP phone. Consoles are special-purpose applications to control call distribution. This includes receptionist consoles with the ability to connect calls, executive consoles with the ability to see call states of special groups of phones, and customer relations consoles with the ability to support call distribution. The distinction among the different types of consoles is not too clear. All VoIP consoles have in common the use of proprietary protocol extensions. Proprietary protocol extensions can be problematic for all stateful firewalls, unless the firewall can understand the non-standard signaling. Consoles should be installed on dedicated desktop computers, with no access to the Internet and only controlled access to data network services, in order not to expose the voice network. Consoles are usually static and should be confined to their own network within the module.

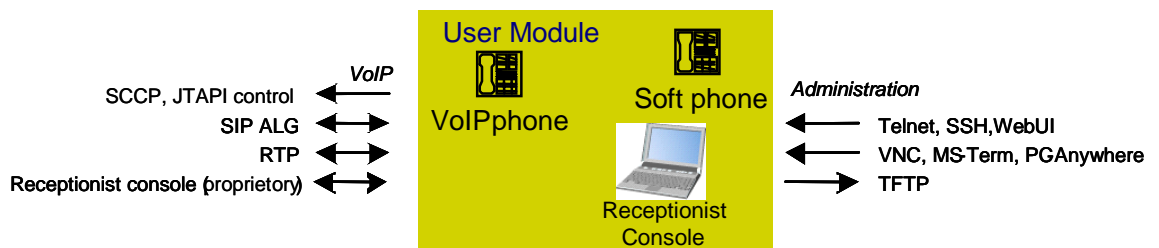


Figure 3: VoIP end user devices

Media/VOIP Gateways/Gatekeepers

The terms gateway and gatekeeper are sometimes used interchangeably. Traditionally gatekeepers have been mainly used for Call Admission and control and bandwidth management. But this has changed recently, as technology has allowed this functionality to co-exist within traditional gateways (described below).

The major function of media gateways is analog-to-digital conversion of voice and creation of voice IP packets (CODEC functions). In addition, media gateways have optional features, such as voice (analog and/or digital) compression, echo cancellation, silence suppression, and statistics gathering.

The media gateway forms the interface that the voice content uses so it can be transported over the IP network. Media gateways are the sources of bearer traffic. Typically, each conversation (call) is a single IP session transported by a Real-time Transport Protocol (RTP) that runs over UDP or TCP.

Media gateways exist in several forms. For example, media gateways could be a dedicated telecommunication equipment chassis, or even a generic PC running VoIP software. Their features and services can include some or all of the following:

- Trunking gateways that interface between the telephone network and a VoIP network. Such gateways typically manage a large number of digital circuits.
- Residential gateways that provide a traditional analog interface to a VoIP network. Examples of residential gateways include cable modem/cable set-top boxes, xDSL devices and broadband wireless devices.
- Access media gateways that provide a traditional analog or digital PBX interface to a VoIP network. Examples include small-scale (enterprise) VoIP gateways.
- Business media gateways that provide a traditional digital PBX interface or an integrated soft PBX interface to a VoIP network.
- Network access servers that can attach a modem to a telephone circuit and provide data access to the Internet.

Figure 4 shows a VoIP gateway and the signaling protocols it uses to communicate with VoIP Call processing servers and other VOIP devices, such as IP Phones, messaging systems, etc.

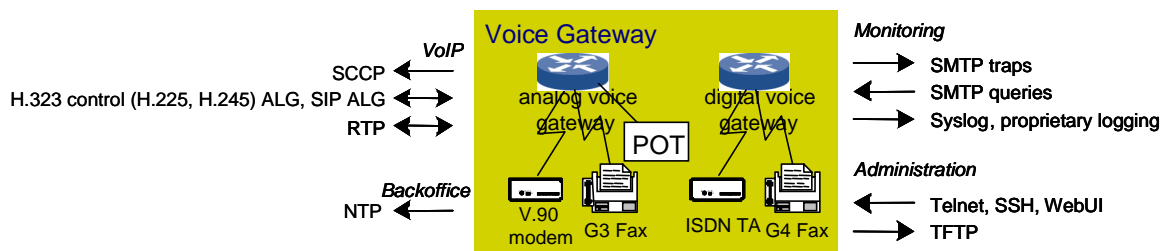


Figure 4: VoIP/Media Gateway

IP Network

You can view the VoIP network as one logical switch. However, this logical switch is a distributed system, rather than that of a single switch entity; the IP backbone provides the connectivity among the distributed elements. Depending on the VoIP protocols used, this system as a whole is sometimes referred to as a *softswitch architecture*.

The IP infrastructure must ensure smooth delivery of the voice and signaling packets to the VoIP elements. Due to their dissimilarities, the IP network must treat voice and data traffic differently. If an IP network is to carry both voice and data traffic, it must be able to prioritize the different traffic types, as VoIP traffic is extremely sensitive to latency.

While there are several similarities between VoIP and circuit-switching components, there are also several differences. One is in the transport of the resulting voice traffic. Circuit-switching telecommunications can be best classified as a TDM network that dedicates channels, reserving bandwidth as it is needed out of the trunk links interconnecting the switches. For example, a phone conversation reserves a single DS-0 channel, and that end-to-end connection is used only for the single conversation. This is not an efficient method of resource utilization.

IP networks are quite different from the circuit-switch infrastructure in that it is a packet-network, and it is based on the idea of statistical availability. Thus network resources are not completely tied up for the duration of the call, unlike in a circuit-switched environment. Class of service (CoS) ensures that packets of a specific application are given priority. This prioritization is required for real-time VoIP applications to ensure that the voice service is unaffected by other traffic flows.

VoIP Signaling Protocols

VoIP signaling protocols are the enablers of the VoIP network. The protocols determine what types of features and functionality are available, as well as how all of the VoIP components interact with one another.

There are a variety of VoIP protocols and implementations, with a wide range of features that are currently deployed. Two major standards bodies govern multimedia delivery (voice being one type) over packet-based networks: International Telecommunications Union (ITU) and Internet Engineering Task Force (IETF). H.323 is the ITU's standard for establishing VOIP connections, while IETF uses Session Initiation Protocol (SIP) as its standard. More implementations tend to be focused on the ITU specifications than those of the IETF, primarily because H.323 is more widely deployed today than SIP. This is expected to change, however. It should be noted, many of the standards in both bodies are based on solving the same problems. The result is some overlap of functionality, as well as differences in approach and nomenclature. To further confuse the issue, some vendors are implementing proprietary schemes that fill apparent gaps in the standards or add functionality that is product dependent. Also, not all of the protocols are used in one specific product group. Instead, the product vendor will code its offerings with what is most applicable for its scope, services, and market.

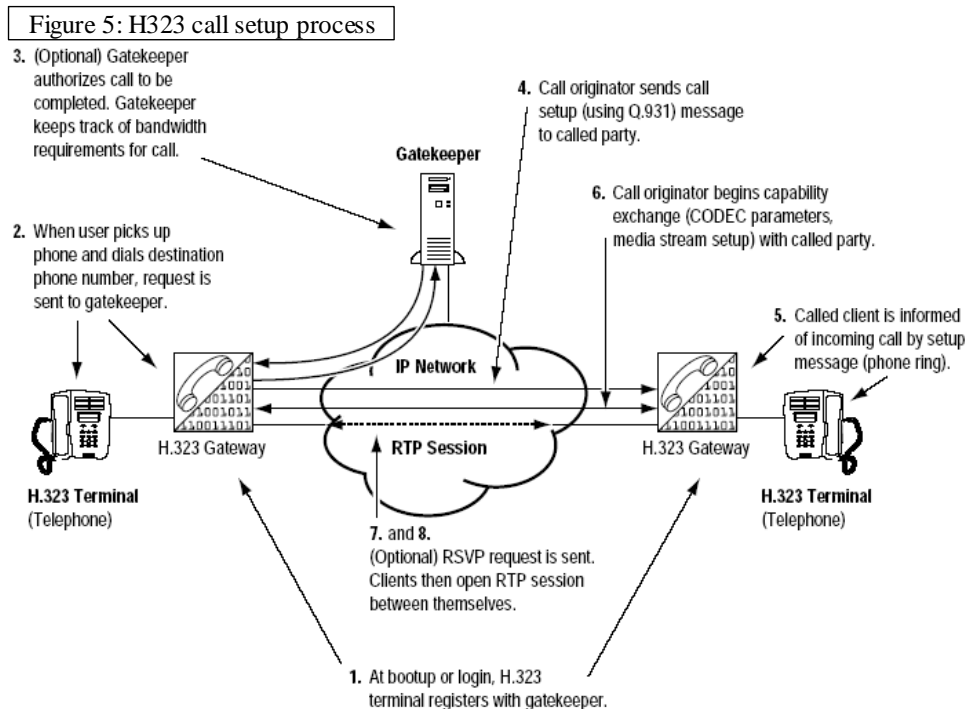
Each of the voice protocols has its own strengths and weaknesses, and each takes a different approach to service delivery. Each of these protocols is successful in different products having a specific market focus. The protocols listed in this section are the most prevalent, but do not constitute an exhaustive list; there are a few other protocol options available.

H.323

H.323 is the ITU recommendation. It is a packet-based multimedia communication system that is a set of specifications. These specifications define various signaling functions, as well as media formats related to “packetized” audio and video services.

H.323 standards were generally the first to classify and solve multimedia delivery issues over LAN technologies. However, as IP networking and the Internet became prevalent, many Internet RFC standard protocols and technologies were developed and based on some of the previous H.323 ideas. Today there is cooperation between the ITU and IETF in solving existing problems, but it is fair to say that the RFC process of furthering the standards has had greater success than the H.323 counterparts.

H.323 networks consist of Call Processing Servers, (media) gateways and gatekeepers. Call Processing Servers provide call routing, and communication to VOIP gateways and end devices. Gateways serve as both the H.323 termination endpoint and interface with non-H.323 networks, such as the PSTN. Gatekeepers function as a central unit for call admission control, bandwidth management and call signaling. Although the gatekeeper is not a required element in H.323, it can help H.323 networks to scale to a larger size, by separating call control and management functions from the gateways.



H.323 specifications tend to be heavier (due to chattiness, in terms of control signaling) and with an initial focus in LAN networking. These standards have some shortcomings in scalability, especially in large-scale deployments. Primarily, limitations are due to chattiness or the heavy signaling required to establish H.323 sessions. H.323 is dependant on TCP-based (connection-oriented) signaling. There is a challenge in maintaining large numbers of TCP sessions because of the substantial overhead involved. However, most H.323 scalability limitations are based on the prevalent version two of the specification. Subsequent versions of H.323 have a focus on solving some of these problems.

Let's look at the main H323 process:

- With each call that is initiated, a TCP session (H.225.0 protocol) is created, using an encapsulation of a subset of Q.931 messages. This TCP connection is maintained for the duration of the call. Complete call setup process is shown in figure 5.
- A second session is established using the H.245 protocol. This TCP-based process is for capabilities exchange, master-slave determination, and the establishment and release of media streams. This group of procedures is in addition to the H.225.0 processes.
- The H.323 quality of service (QoS) delivery mechanism of choice is the Resource Reservation Protocol (RSVP). This protocol is not considered to have good scaling properties due to its focus and management of individual application traffic flows.
- Although H.323 many not be well suited in service provider spaces, it is well positioned to deploy enterprise VoIP applications. As a service provider, it might be necessary to bridge, transport or interface H.323 services and applications to the PSTN.

Real-time Transport Protocol (RTP)

RFC 1889 and RFC 1890 cover the Real-time Transport Protocol (RTP), which provides end-to-end delivery services for data with real-time characteristics, such as interactive audio and video. Services include payload type identification, sequence numbering, time stamping and delivery monitoring. The media gateways that digitize voice use the RTP protocol to deliver the voice (bearer) traffic.

The RTP protocol (Figure 6) provides features for real-time applications, with the ability to reconstruct timing, loss detection, security, content delivery and identification of encoding schemes. For each participant, a particular pair of destination IP addresses defines the session between the two endpoints, which translates into a single RTP session for each phone call in progress. RTP is an application service built on UDP, so it is connectionless, with best-effort delivery. Although RTP is connectionless, it does have a sequencing system that allows for the detection of missing packets.

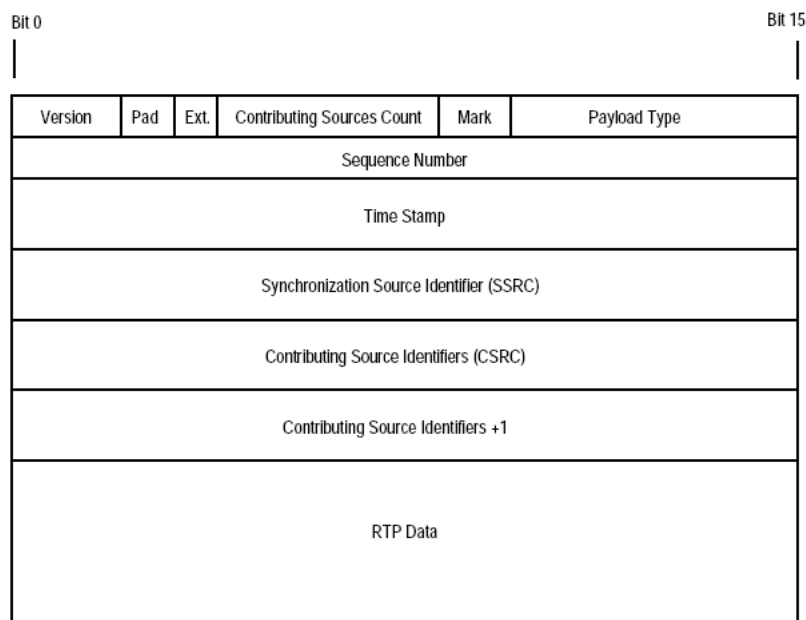


Figure 6: Upper Layer of RTP Protocol

As part of its specification, the RTP Payload Type field includes the encoding scheme that the media gateway uses to digitize the voice content. This field identifies the RTP payload format and determines its interpretation by the CODEC in the media gateway. A profile specifies a default static mapping of payload type codes to payload formats. These mappings represent the ITU G series of encoding schemes.

With the different types of encoding schemes and packet creation rates, RTP packets can vary in size and interval. Administrators must take RTP parameters into account when planning voice services. All the combined parameters of the RTP sessions dictate how much bandwidth is consumed by the voice bearer traffic. RTP traffic that carries voice traffic is the single greatest contributor to the VoIP network load.

Real-time Transport Control Protocol (RTCP)

Real-time Transport Control Protocol (RTCP) is the optional companion protocol to RTP; it is not needed for RTP to work. The primary function of RTCP is to provide feedback on the quality of the data distribution being accomplished by RTP. This function is an integral part of RTP's role as a transport protocol and is related to the flow and congestion control functions of the network. Although the feedback reports from RTCP do not describe where problems are occurring (only that they are), they can be used as a tool to locate problems. With the information generated from different media gateways in the network, RTCP feedback reports enable an administrator to evaluate where network performance might be degrading.

RTCP enables administrators to monitor the quality of a call session by tracking packet loss, latency (delay), jitter, and other key VoIP concerns. This information is provided on a periodic basis to both ends and is processed per call by the media gateways.

Some gateway devices might not employ RTCP because the facility to report such information is not applicable to the end user. For example, a single residential user (with an analog phone) might not have access to the gateway providing the service. Also, the media gateway vendor can use a more scalable approach of tracking call quality statistics. In this case, the storage, transport and presentation of statistical info are device dependent.

If using RTCP (or a vendor-specific implementation) in the network, the organization needs to take into account bandwidth calculations for the protocol. Administrators need to limit the control traffic of RTCP to a small and known fraction of the session bandwidth. It should be small so as not to impair the ability of the transport protocol to carry data. An organization should investigate the amount of bandwidth needed so that they can include the control traffic in the bandwidth specification. RFC specifications recommend that the fraction of the session bandwidth allocated to RTCP be fixed at five percent of RTP traffic.

Media Gateway Control Protocol (MGCP)

The Media Gateway Control Protocol (MGCP, RFC 2705) is along the lines of a softswitch architecture philosophy. It breaks up the role of traditional voice switches into the components of media gateway, media gateway controller and signaling gateway functional units. This facilitates the independent managing of each VoIP gateway as a separate entity.

MGCP is a master-slave control protocol that coordinates the actions of media gateways (Figure 7). The media gateway controller in MGCP nomenclature is sometimes referred to as a call agent. The call agent manages the call-related signaling control intelligence, while the media gateway informs the call agent of service events. The call agent instructs the media gateway to create and tear down connections when the calls are generated. In most cases, the call agent informs the media gateways to start an RTP session between two endpoints.

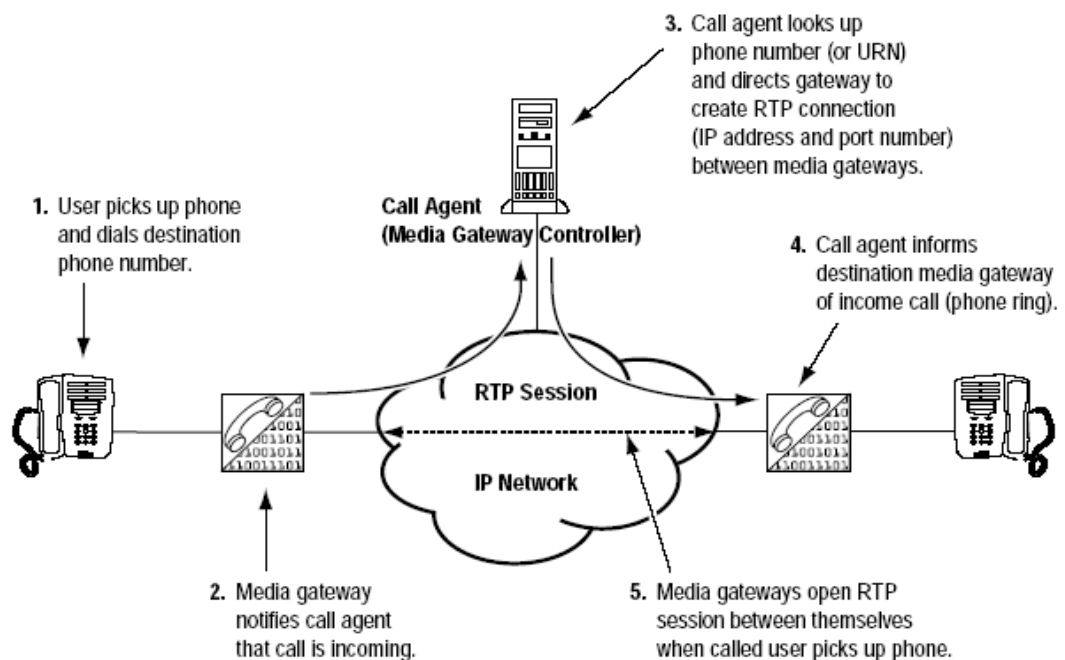


Figure 7: How MGCP Coordinates the Media Gateways

Session Initiation Protocol (SIP)

The Session Initiation Protocol (SIP, RFC 2543) is part of IETF's multimedia data and control protocol framework. SIP is a powerful client-server signaling protocol used in VoIP networks. SIP handles the setup and tear down of multimedia sessions between speakers; these sessions can include multimedia conferences, telephone calls, and multimedia distribution.

SIP is a text-based signaling protocol transported over either TCP or UDP, and is designed to be lightweight. It inherited some design philosophy and architecture from the Hypertext Transfer Protocol (HTTP) and Simple Mail Transfer Protocol (SMTP) to ensure its simplicity, efficiency and extensibility.

SIP uses *invitations* to create Session Description Protocol (SDP) messages to carry out capability exchange and to setup call control channel use. These invitations allow participants to agree on a set of compatible media types.

SIP supports user mobility by proxying and redirecting requests to the user's current location. Users can inform the server of their current location (IP address or URL) by sending a

registration message to a *registrar*. This function is powerful and often needed for a highly mobile voice user base. The SIP client-server application has two modes of operation; SIP clients can either signal through a *proxy* or *redirect* server.

- Using proxy mode (Figure 8), SIP clients send requests to the proxy and the proxy either handles requests or forwards them on to other SIP servers. Proxy servers can insulate and hide SIP users by proxying the signaling messages; to the other users on the VoIP network, the signaling invitations look as if they are coming from the proxy SIP server.

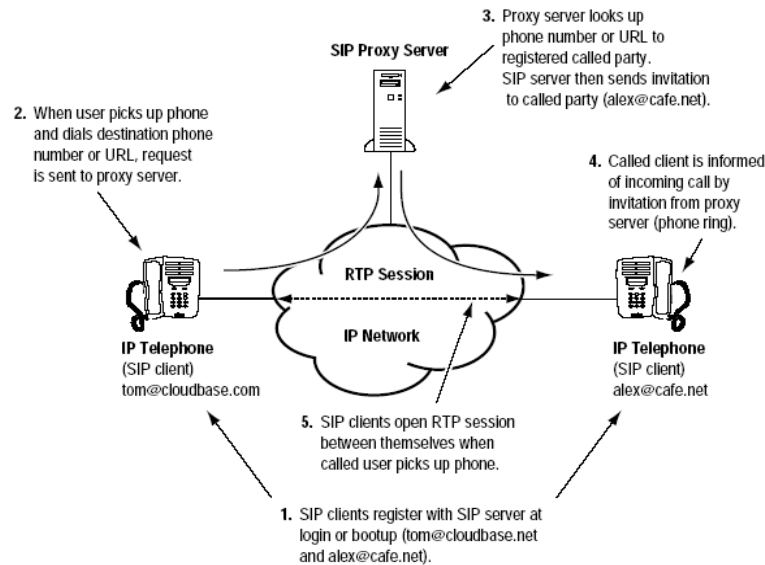


Figure 8: SIP Proxy Operation

- Under redirect operation (Figure 9), the signaling request is sent to a SIP server, which then looks up the destination address. The SIP server returns the destination address to the originator of the call, who then signals the SIP client

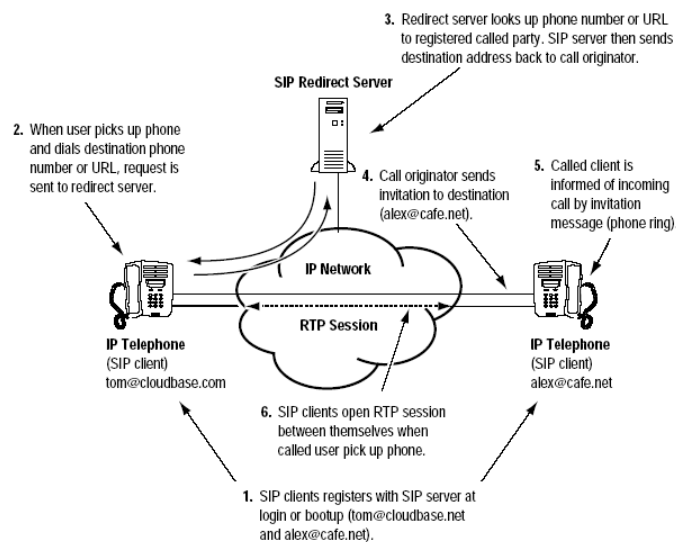


Figure 9: SIP Redirector Server

Megaco/H.248

Megaco/H.248 is a current draft standard and represents a cooperative proposal from the IETF and ITU standards bodies. Megaco has many similarities to MGCP and borrows the same naming conventions for the VoIP elements. The Megaco architecture defines media gateways that provide media conversion and sources of calls, while media gateway controllers provide call control.

Megaco addresses the same requirements as that of MGCP and, as a result, there is some effort to merge the protocols. It defines a series of transactions coordinated by a media gateway controller for the establishment of call sessions. The primary focus of Megaco is the promotion to standardize IP telephony equipment. Some of the design goals are as follows:

- Megaco IP phone meets the basic needs of the business user from day one.
- Provides a path for rapid expansion to support sophisticated business telephony features.
- Allows for a wide range of telephones and similar devices to be defined from very simple to very feature rich.
- Implements a simple, minimal design.
- Allows device cost to be appropriate to capabilities provided. Package and termination types have characteristics that enable reliability.
- IP phone meets the appropriate Megaco/H.248 protocol requirements, as provided in the Megaco requirements document, and are a straightforward application of the Megaco/H.248 protocol.

VoIP Service Considerations

Now that this paper has gone through the functions, components and protocols related to VoIP traffic, let's take a quick look at some of the issues an organization must carefully consider when deploying VoIP solutions, such as traffic parameters and network design. Without such due diligence, an organization could be faced with service that does not function reliably or is severely degraded. The important considerations are as follows:

- Latency
- Jitter
- Bandwidth
- Packet loss
- Reliability
- Security
- Interoperability

Latency

Latency (or delay) is the time that it takes a packet to make its way through a network end-to-end. In telephony terms, latency is the measure of time it takes the talker's voice to reach the listener's ear. Large latency values do not necessarily degrade the sound quality of a phone call, but the result can be a lack of synchronization between the speakers, such that there are hesitations in the speaker's interactions.

Generally, it is accepted that the end-to-end latency should be less than 150 ms for toll quality phone calls. To ensure that the latency budget remains below 150 ms, administrators need to take into account the following primary causes of latency. When designing a multiservice network, the total delay that a signal or packet exhibits is a summation of all the latency contributors.

- One source of latency is the time it takes for the endpoints to create the packets used in voice services. These “packetization” delays are caused by the amount of time it takes to fill a packet with data. Generally, the larger the packet size, the greater the amount of time it takes to fill it. Packetization delay is governed by the CODEC standard being used. This problem also exists on the receiving side because the media gateway must remove and further process the packet data. If the packets are kept small, this amount of delay, in both directions, is usually quite small, depending on the hardware / software implementation of the media gateways. All considerations being equal, nominal operation of any media gateway unit should not exceed 30 ms.
- Another source of latency is the delay it takes to serialize the digital data onto the physical links of the interconnecting equipment. This delay is inversely proportional to the link speed. In other words, the faster the media, the lower the latency. This value is somewhat dependent on the link technology used and its access method. For example, it takes 125 microseconds to place one byte on a 64-Kb circuit. The same byte placed on an OC-3/STM-1 circuit takes 0.05 microseconds. Although this delay is unavoidable (regardless of the bandwidth used), keeping the number of intervening links small and using high bandwidth interfaces reduces the overall latency.
- Propagation delay is the time it takes an electrical (or photonic) signal to traverse the length of a conductor. The speed of these signals is always slower than that of the speed of light. There is always propagation delay; however, it only becomes an issue when the signal (or packet) travels a great distance. The accepted formula for calculating propagation delay is as follows.
- $\text{Propagation delay} = \text{Circuit km} / (299,300 \text{ km} \times .6)$
- **Example:** Calculation of one-way propagation delay of a 6,000 km fiber run (discounting any signal repeaters in between)
- $0.0334 \text{ sec} = 6000 \text{ km} / (299,300 \text{ km} \times .6)$
- By this calculation, the latency contributed by just propagation delay would be 33.4 ms.
- A queuing delay, which is a large source of latency, is the amount of time that a packet remains buffered in a network element while it awaits transmission. Network traffic loads result in variable queuing delays. The amount of buffer that a queue uses is usually a configurable parameter, with a smaller number being better for latency values. However, this delay is also based on the amount of traffic the element is trying to pass through a given link, and therefore it increases with network load. Hence, you need to set aside adequate bandwidth and resources for voice traffic. If the queue used for voice traffic is not serviced fast enough and that queue is allowed to grow too large, the result is greater latency.
- Packet forwarding delay is the time it takes a network device (router, switch, firewall, etc.) to buffer a packet and make the forwarding decision. Included in that decision could be which interface to forward the packet to, whether to drop or forward the packet against an Access Control List (ACL) or security policy, etc. Packet forwarding delay is variable and depends on the function and architecture of the networking device. If a packet must be further buffered as a part of its processing, greater latency is incurred.

Jitter

Jitter is the measure of time between when a packet is expected to arrive to when it actually arrives. In other words, with a constant packet transmission rate of every 20 ms, every packet would be expected to arrive at the destination exactly every 20 ms. This situation is not always the case. For example, Figure 9 shows packet one (P1) and packet three (P3) arriving when expected, but packet two (P2) arriving 12 ms later than expected and packet four (P4) arriving 5 ms late.

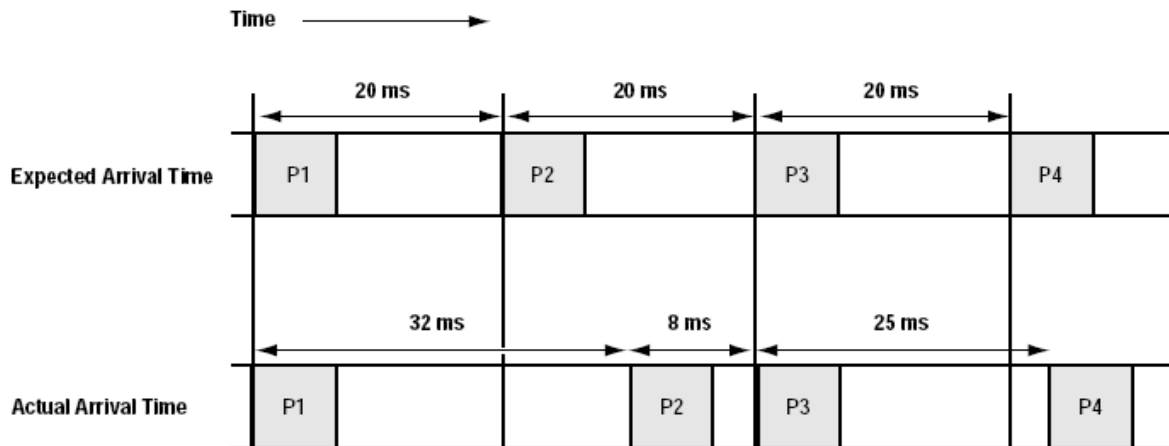


Figure 10: Jitter Example

The greatest culprit of jitter is queuing variations caused by dynamic changes in network traffic loads. Another cause is packets that might sometimes take a different equal-cost link that is not physically (or electrically) the same length as the other links.

Media gateways have *play-out buffers* that buffer a packet stream, so that the reconstructed voice waveform is not affected by packet jitter. Play-out buffers can minimize the effects of jitter, but cannot eliminate severe jitter.

Although some amount of jitter is to be expected, severe jitter can cause voice quality issues because the media gateway might discard packets arriving out of order. In this condition, the media gateway could starve its play-out buffer and cause gaps in the reconstructed waveform.

Bandwidth

An organization can determine how much bandwidth to set aside for voice traffic using simple math. However, in a converged voice and data network, administrators have to make decisions on how much bandwidth to give each service. These decisions are based on careful consideration of the organization's priorities and the available bandwidth that can be afforded. If an administrator allocates too little bandwidth for voice service, there might be unacceptable quality issues. Another consideration is that voice services are less tolerant to bandwidth depletion than that of Internet traffic. Therefore, bandwidth for voice services and associated signaling must take a priority over that of best-effort Internet traffic. If a network were to use the same prevailing encoding (CODEC) scheme as the current PSTN system, bandwidth requirements for VoIP networks would tend to be larger than that of a circuit-switched voice network of similar capacity. The reason is the overhead in the protocols used to deliver the voice service. Typically, an organization would need speeds of OC-12c/STM-4 and higher to support thousands of call sessions. However, VoIP networks that employ compression and silence suppression could actually use less bandwidth than a similar circuit-switched network.

The reason is because of the greater granularity in bandwidth usage that a packet-based network has in comparison to a fixed, channel size TDM network.

Allocations of network bandwidth are based on projected numbers of calls at peak hours. Any over-subscription of voice bandwidth can cause a reduction in voice quality. Also, you must set aside adequate bandwidth for signaling to ensure that calls are complete and to reduce service interruptions.

The formula for calculating total bandwidth needed for voice traffic is relatively straightforward. The formula to calculate RTP bearer voice bandwidth usage for a given number of phone calls is as follows:

$$\text{bits per sec} = \text{packet creation rates per sec} \times \text{packet size} \times \text{number of calls} \times 8 \text{ bits per sec}$$
$$\text{where samples per sec} = 1,000 \text{ ms} / \text{packet creation rate}$$

Example

2,000 full-duplex G.711 encoded voice channels that have a packet creation rate of 20 ms, with a packet size of 200 bytes (40 byte IP header + 160 byte payload)

$$50 \text{ samples per second} = 1,000 \text{ ms} / 20 \text{ ms}$$
$$160 \text{ Mbps} = 50 \times 200 \times 2,000 \times 8$$

Note that this number is a raw measure of IP traffic and does not take in account the overhead used by the transporting media (links between the routers) and data-link layer protocols. Add this raw IP value to that of the overhead to determine the link speeds needed to support this number of calls. Note this value represents only the bearer (voice) content.

Signaling bandwidth requirements vary depending on the rate at which the calls are generated and the signaling protocol used. If a large number of calls are initiated in a relatively short period, the peak bandwidth needs for the signaling could be quite high. A general guideline for the maximum bandwidth requirement that an IP signaling protocol needs is roughly three percent of all bearer traffic. Using the previous example, signaling bandwidth requirements, if all 2,000 calls were initiated in one second, would be approximately 4.8 Mbps (3 percent of 160-megabits).

With the calculation of bearer and signaling, the total bandwidth needed to support 2000 G.711 encoded calls would approximately be a maximum of 164.8 MB. This bandwidth requirement is a theoretical maximum for this specific case. If the parameters change, such as call initiation rate, voice encoding method, packet creation rate, employment of compression and silence suppression, the bandwidth requirements would change as well.

With large VoIP implementations requiring sizable bandwidth, it becomes imperative that the IP network delivers the needed service at predictably high performance.

Packet Loss

Packet loss occurs for many reasons, and in some cases, is unavoidable. Often the amount of traffic a network is going to transport is underestimated. During network congestion, routers and switches can overflow their queue buffers and be forced to discard packets. Packet loss for non-real-time applications, such as Web browsers and file transfers, is undesirable, but not critical. The protocols used by non-real-time applications, usually TCP, are tolerant to some amount of packet loss because of their retransmission capabilities.

Real-time applications based on UDP are significantly less tolerant to packet loss. UDP does not have retransmission facilities, however, retransmissions would almost never help. In an RTP session, by the time a media gateway could receive a retransmission, it would no longer be relative to the reconstructed voice waveform; that part of the waveform in the retransmitted

packet would arrive too late.

It is important that bearer and signaling packets not be discarded, otherwise, voice quality or service disruptions might occur. In instances where service disruptions may occur, Class of Service (CoS) mechanisms offer a means of controlling packet delivery priority. , which is equivalent to DSCP in IP, but at the Ethernet layer, mechanisms become very important. By configuring CoS parameters, administrators can give packets of greater importance a higher priority in the network, thus ensuring packet delivery for critical applications, even during times of network congestion. Note that CoS is equivalent to DSCP in IP, but at the Ethernet layer.

Although packet loss of any kind is undesirable, some loss can be tolerated. Some amount of packet loss for voice services could be acceptable, as long as the loss is spread out over a large amount of users. As long as the amount of packet loss is less than five percent for the total number of calls, the quality generally is not adversely affected. It is best to drop a packet, versus increasing the latency of all delivered packets by further buffering them.

Reliability

Although network failures are rare, planning for them is essential. Failover strategies are desirable for cases when network devices malfunction or links are broken. An important strategy is to deploy redundant links between network devices and/or to deploy redundant equipment. To ensure continued service, organizations should plan carefully for how media gateways and media gateway controllers can make use of the redundant schemes.

IP networks use routing protocols to exchange routing information. As part of their operation, routing protocols monitor the status of interconnecting links. Routing protocols typically detect and reroute packets around a failure if an alternate path exists. Depending on the interconnecting media used for these links, the time taken to detect and recalculate an alternate path can vary. For example, the loss of signal for a SONET/SDH connection can be detected and subsequently rerouted very quickly. However, a connection through an intervening LAN switch might need to time out the keep-alive protocol before a failure is detected.

Having media gateways and media gateway controllers that can actively detect the status of their next-hop address (default gateway) as part of their failover mechanism decreases the likelihood of a large service disruption. Another possible option is that the media gateway and media gateway controller could be directly connected to the router. In this case, the possibility of a link failure (depending on the nature of the failure) could be immediately detected and the network devices would take appropriate action. Still another option for reducing long-term failure could be to employ a redundancy mechanism such as failover.

Security

Security, especially in a converged voice and data network, is a high priority. Organizations need to protect the voice communication devices from unauthorized access and malicious attack. While organizations can thwart unauthorized access by using security protocols (such as RADIUS and SSH), Denial-of-Service (DoS) attacks can be a real danger to voice services. It is conceivable that such attacks would either cripple or completely disable voice services.

One way to secure VoIP devices is to use private addressing to enumerate the media gateways and call processing servers. Private addressing is not advertised to the public Internet and, therefore, the devices are inaccessible to the outside world.

Additionally, all VoIP processing servers, gateway, and messaging systems should be placed behind firewalls to enforce access control policies and protect them from any DoS attacks. The firewall needs to understand the signaling protocols in use in the network to be able to dynamically open and close ports for the VoIP traffic only for the duration of a call, so that

these ports are not left open and cannot be usurped for unauthorized use. These servers are critical to VoIP communication; therefore, firewall policies should be in place to protect communications between these servers and VoIP end-devices. These policies should restrict VoIP communication, based on authorized end-devices or traffic sourced or destined for a particular IP address or interface. Firewalls can be used to segment the VoIP network, separating the voice traffic from other traffic to ensure appropriate priority and policies are applied. Firewalls may also be placed to mitigate DoS attacks and to create logs for forensics. Furthermore, intrusion prevention systems can be deployed to help detect and prevent certain attacks, such as manipulated DHCP messages or flooded FIB tables.

Conclusion

Organizations are increasingly looking to VoIP as an attractive alternative to traditional PSTN. However, deploying VoIP is not as easy as flipping a switch, so it is important an organization consider all of the functionality they are going to require from their VoIP network and are aware of the potential issues that go along with deploying a VoIP network. Just as companies choose various protocols for their data networks, they will choose various protocols for their VoIP requirements, depending on the business and technical requirements at hand. Although the variety in VoIP protocols has caused some confusion in the marketplace, it is precisely this protocol flexibility that makes VoIP-based voice systems so much more useful than legacy voice systems. Companies should choose vendors based on three very important requirements:

- Commitment to supporting open standards within their products. Any vendor should be actively developing voice strategies that consider interoperability with all VoIP protocols. Without this commitment, VoIP systems are in danger of becoming as proprietary as legacy voice systems.
- Support of multiple protocols. This way, if a company finds it needs to migrate its systems or add products that support a different protocol, it will not be required do a wholesale infrastructure redeployment or perform significant upgrades to the network.
- End-to-end support for all VoIP protocols, meaning vendors must provide solutions that work in both single-protocol and multi-protocol environments.

Organizations also want to ensure that the solutions they choose are able to address the issues brought up by this paper. VoIP solutions should be able to meet the latency, jitter, bandwidth, packet loss, reliability and security requirements necessary for their part in the VoIP infrastructure. By working with vendors that can provide this VoIP flexibility, companies can take advantage of the efficiencies of VoIP and focus on building scalable and reliable networks that support the requirements of next-generation networks.

Copyright © 2004 Juniper Networks, Inc. All rights reserved.

Juniper Networks, the Juniper Networks logo, NetScreen, NetScreen Technologies, GigaScreen, and the NetScreen logo are registered trademarks of Juniper Networks, Inc. NetScreen-5GT, NetScreen-5XP, NetScreen-5XT, NetScreen-25, NetScreen-50, NetScreen-100, NetScreen-204, NetScreen-208, NetScreen-500, NetScreen-5200, NetScreen-5400, NetScreen-Global PRO, NetScreen-Global PRO Express, NetScreen-Remote Security Client, NetScreen-Remote VPN Client, NetScreen-IDP 10, NetScreen-IDP 100, NetScreen-IDP 500, GigaScreen ASIC, GigaScreen-II ASIC, and NetScreen ScreenOS are trademarks of Juniper Networks, Inc. All other trademarks and registered trademarks are the property of their respective companies.

Information in this document is subject to change without notice.

No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without receiving written permission from:

Juniper Networks, Inc.

1194 N. Mathilda Ave. Sunnyvale, CA 95014 ATTN: General Counsel