White Paper

Implementing VoIP over Raptor Adaptive Switch Technology (RAST™)

Edwin Hoffman CSA

RAPTOR NETWORKS TECHNOLOGY INC

Scope and Purpose

The IP phone is coming—or has arrived—to a desk near you. The IP phone is not a personal computer (PC), but does have a several hardware and software features also resident in PCs. VoIP demands certain things before it will work correctly, and it is these demands that are the scope of this paper.

It is naïve to assume that because VoIP is over "IP" that it will work as well anything else on your network. In fact do you even know how well your applications are running over your existing network? It is not always the case that a major network issue will always cause the user to complain. If the network is not being monitored for bandwidth usage, latency, and jitter, you never know what is working well. This paper discusses VoIP operation and testing.

Basic Overview

What is VoIP?

Voice over Internet Protocol (VoIP), also called IP Telephony, is rapidly becoming a familiar term in technology that is inexorably entering enterprise, education, and government organizations. VoIP is designed to replace the legacy Time Division Multiplexing (TDM) technologies and networks with an IP-based data network. Digitized voice will be carried in IP data packets over a LAN and/or WAN network. Installing and testing the VoIP network of IP phones, gateways, and servers require new tools and expanded knowledge.

The legacy telephone network has provided reliable, high-quality voice communications for many years. It delivers voice and speech over a standardized 64-Kbps channel. The 64-Kbps bandwidth is guaranteed for each call and the speech path carries the voice as a continuous digital stream. Digital voice is NOT carried in packets. Enterprise and residential callers use DTMF (Touch Tone), TI channel, and ISDN D-channel signaling to set up and manage the call. Within ISDN, signaling is carried in packets on a separate signaling channel over the Basic Rate Interface (BRI) and Primary Rate Interface (PRI) connections to the carrier. The carrier then translates the signaling (all forms) to an internal signaling protocol called Signaling System #7. Signaling protocols are primarily active at the beginning and the end of a call.

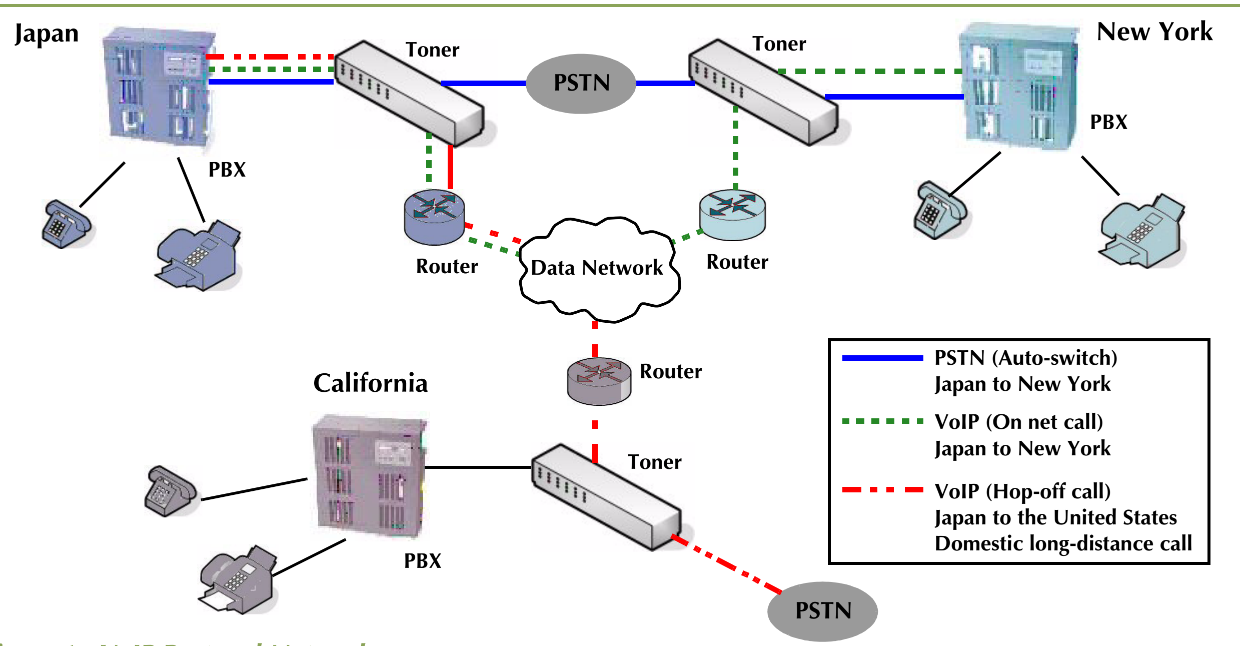


Figure 1: VoIP Protocol Network

In a VoIP network like that of Figure 1, there is a signaling protocol and a speech transmission protocol. Both protocols require that all information be carried in IP packets. Several standards-based choices are available for signaling protocols, including H.323, SIP, MGCP, and H.248 (MEGACO).

Most IP PBX vendors have developed their own proprietary signaling protocol, the most common of which is Cisco's SCCP (Skinny) protocol. Real Time Protocol (RTP) is the standard speech transmission protocol used with VoIP networks. The speech is digitized, placed in packets, and transmitted through the IP network. Multiple packets are required to carry a single spoken word. The voice is digitized using one of the G.7xx standards.

Why, then, do organizations move from a TDM-based telephone to an IP-based telephone? This is a major change in the underlying technology, yet people are still making phone calls as they have for years.

One or more of the following reasons may justify the move to VoIP:

- Reducing long-distance charges, especially international long distance
- Reducing staff by combining voice-network and data-network management and eliminating redundant functions
- Adding expanded applications that are not offered by TDM-based systems
- Having one common network for different forms of communication
- TDM vendors not offering new systems, thus forcing customers to eventually adopt IP-based telephone systems

Standards for VoIP

"*Standards are great, I have so many to choose from*" is a quote that amply describes VoIP. There are multiple signaling standards.

- H.323, the ITU standard that was published in 1995, started the development of VoIP products and services. There are four versions available. V.1 is obsolete and has been discontinued in virtually all products. Versions 2, 3, and 4 are used in today's products. These three versions are similar in design and are upwardly compatible. This is the dominant installed signaling protocol for use with hard and softphones.
- The Session Initiation Protocol (SIP) was produced by the IETF as an IP standard. Although SIP is gaining considerable attention, it will not become the dominant installed protocol for a few years. The attractions of SIP are better interoperability among vendors, easier application development, operation that is close to other existing IP protocols, and easier operation through firewalls. It is usually part of hard and softphones, but it may also be used with gateways. SIP is a completely different design when compared to H.323.
- MGCP is a protocol used primarily with gateways, although an occasional hard phone may support MGCP.
- MEGACO/H.248, another standard protocol, is a combined effort of the ITU and IETF. It can be used with gateways and server-to-server communications. It is not found in hard or softphones.

In addition to the standards, nearly every IP PBX vendor has produced a proprietary signaling protocol. The most commonly found protocol is previously mentioned Cisco's SCCP (Skinny) protocol. These proprietary protocols may be variations of the standards or may be uniquely designed. They each provide the call control found in the standard protocols. An IP PBX vendor usually supports one or more of the standard signaling protocols plus their proprietary protocol. All the signaling protocols follow the same path for control as shown in Figure 1 on page 1. The H.323 and most proprietary signaling are carried over TCP, while SIP operates over UDP.

Speech is carried in packets that use the Real Time Protocol (RTP) standard. Each RTP packet contains a piece of a digitized word.

Multiple RTP packets, when combined at the receiving IP phone, produce a spoken word. The RTP IP PBX vendors commonly implement RTP. Proprietary protocols that operate like RTP are uncommon. The speech paths connect directly between phones and gateways; speech does not pass through the server, nor is speech carried by the signaling protocols.

There are several voice digitization standards and some proprietary techniques in use. Most vendors support one or more of the following ITU standards and avoid proprietary solutions:

- G.711 is the default standard for all vendors, as well as for the PSTN. This standard digitizes voice into 64 Kbps and does not compress the voice.
- G.729 is supported by many vendors for compressed voice operating at 8 Kbps. With quality just below that of G.711, it is the second most commonly implemented standard.
- G.723.1 was once the recommended compression standard. It operates at 6.3 Kbps and 5.3 Kbps. Although this standard reduces bandwidth consumption, voice is noticeably poorer than with G.729 and is not very popular for VoIP.
- G.722 operates at 64 Kbps but offers high-fidelity speech. The three previously described standards deliver an analog sound range of 3.4 kHz, while G.722 delivers 7 kHz. This version of digitized speech will become common in the future In all cases, the IP phones and gateways collect about 10 to 30 ms of digital speech and place it in the RTP packet for transmission.

VoIP networks

Local (LAN) and wide area networks (WAN) can both support VoIP operation. However, there are significant differences between LAN and WAN performance that affect signaling execution speed and voice quality. The LAN implementation uses Ethernet as the transport network. VoIP devices operate following the IEEE 802.3 standards. No proprietary changes have been made to the Ethernet protocols and architecture. The Ethernet LAN operates at 10 and 100 Mbps on networks with very short delay, no jitter (delay variance), few errors, and no packet loss. Although VoIP traffic could share the LAN with data users, it is recommended (by many vendors) that the voice and data devices run on separate VLANs (IEEE 802.1q) on the LAN switches for both performance and security reasons. Voice quality and signaling execution speeds are as good as the TDM PBX's.

Latency (definition)

1. In general, the period of time that one component in a system is spinning its wheels waiting for another component. Latency, therefore, is wasted time. For example, in accessing data on a disk, latency is defined as the time it takes to position the proper sector under the read/write head.
2. In networking, the amount of time it takes a packet to travel from source to destination. Together, latency and bandwidth define the speed and capacity of a network.
3. In VoIP terminology, latency refers to a delay in packet delivery. VoIP latency is a service issue that is usually based on physical distance, hops, or voice-to-data conversion.

The WAN presents a number of performance-hindering issues. Bandwidth is limited, end-to-end delays are much longer, packet jitter occurs and, finally, packets are lost. The transmission error rate is low enough that signaling packets do not often require retransmission, and voice quality is not impaired by errors. The end-to-end delay goal between IP phones is 150 ms. The receiving IP phone must compensate for jitter by waiting for all the packets of a word to arrive before the word can be converted back into analog sound. The receiving IP phone must also insert simulated voice packets to eliminate holes in a word, which occur when packets are lost. Correcting jitter and packet loss causes extra delay between IP phones, so both LAN and WAN connections need to reduce these problems to an absolute minimum.

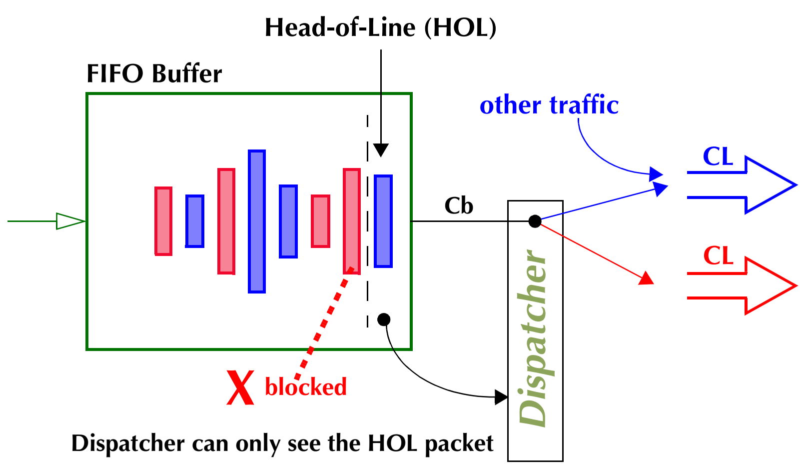
Extra bandwidth and Quality of Service (QoS) techniques can solve these problems. Voice calls consume bandwidth: about 80 Kbps when no voice compression is used (G.711) and about 25 Kbps (G.729) when the voice is compressed. The actual bandwidth consumption varies based on compression type and packet size. Although it should not be a problem, bandwidth can be an issue on the LAN because even though most LANs operate at low utilization (less than 10% to 30% utilization), many of the current network devices are reacting to congestion by losing packets or creating latency hikes. QoS can be supported on LAN switches using the IEEE 802.1p standard. This standard requires that IP phones and gateways also support the 802.1p standard. Routers produce QoS through the implementation of DiffServ, which must be supported by the IP phones and gateways, as well as MultiProtocol Label Switching (MPLS) in the routers. RSVP was an early technique for VoIP QoS, but is less frequently used in today's products and differential services (DIFFserv) are adequate for most QoS requirements.

Major areas for consideration when adding VoIP into an existing network are:

- Do you actually have the bandwidth you think is there?...Check and verify
- Are you experiencing excessive packet loss on any ports?.....Check (How)
- Use your existing switch port logs to see if there any issues you do not know about.
- Do your switches suffer from Head-of-Line blocking (HOL)?

WHAT IS HOL?

Head-of-Line blocking (HOL) is a phenomenon that appears in buffered telecommunication network switches. A switch is usually made of buffered input ports, a switch fabric and buffered output ports. HOL arises when packets arriving at the same input port are destined for different output ports. Because of the FIFO nature of the input buffers and switch design, the switch fabric can only switch the packets at the head of the buffer per cycle. If the HOL packet of a certain buffer at the input cannot be switched to an output port because of contention, the rest of the packets in that buffer are blocked by that Head-of-Line packet, even if there is no contention at the destination output ports for those packets. Head of line blocking is disastrous for VoIP, HOL causes latency to rise very quickly and cause wild latency variations, which cannot be allowed for in the VoIP receiver. **Ether-Raptor switches cannot cause HOL blocking.**



When a single FIFO queue feeds two or more links and Cl<Cb or the links are shared with other traffic, the head-of-line (HOL) packet can block packets behind it that are destined to other outputs.

Figure 2: FIFO Queueing: Head-of-Line Blocking

Are there any specific redundancy or failover implementations that might cause latency to increase?

- Some Layer 3 failover implementations cause a latency increase that prevents VoIP protocols from operating correctly.
- Typically speaking VoIP needs a round-trip delay of less than 150 ms, but because most of the delay perceived is inside the CODECs that convert from analogue-digital-analogue, this often leaves less than 30 ms for the network (30 ms round-trip delay).

Resiliency

Is the existing network resilient enough to allow for cable, line card, or switch failures without seriously compromising the ability for VoIP to traverse the network?

A POTS (Plain Old Telephone Service) telephone has traditionally been so reliable that using anything else is problematic, and for VoIP this is a problem. Predominantly LAN-based VoIP is traveling over an Ethernet network, which cannot guarantee delivery of any data. TCP (Transmission Control Protocol) protocol over IP (Internet Protocol) provides addressing and delivery guarantees. The problem with today's switched networks is that they can only provide high levels of resiliency at Layer 3 and 4 (IP/TCP), which very often causes greatly increased latencies between VoIP end stations.

Many end users find that when they implement VoIP, their network cannot support the requirements of a VoIP system, and they have to spend large sums to correct their existing networks or build completely new VoIP networks.

Defining the Problem

VoIP requires a very “clean” network with adequate bandwidth to cover VoIP calls (consider a 500-user site with all phones with calls in progress at 40 Mbps) being added to the backbone (not always easily done). Latency is fast becoming the main cause why VoIP implementations fail.

Trying to make VoIP access resilient requires complex Layer 3 and 4 configurations to be implemented, which may suddenly cause VoIP to stop working because of latency hikes. Installing a completely new VoIP network almost seems attractive when a company is faced with these problems.

Fixing the problem

Raptor Network's Ether-Raptor system creates a highly resilient (Layer 2) network adjunct for an existing network that provides very low latency switching and allows Layer 3 functions to become low-latency operations. When VoIP must be delivered to other buildings in the same campus, Ether-Raptor provides high resiliency connectivity for the entire campus, thereby allowing the VoIP system to gain a level of reliability that it would normally not have.

The network shown in Figure 3 on page 6 was based on multiple large chassis switches that were incapable of working resiliently and passing VoIP traffic.

The introduction of the Ether-Raptor network allows the old core switches to be supplemented by the newer Raptor switches, which has the unique capacity to allow Layer 2, active/active resilient connections for any device that can support IEEE802.3ad, Cisco EtherChannel®, or Nortel’s Multi-Link Trunking (MLT) system.

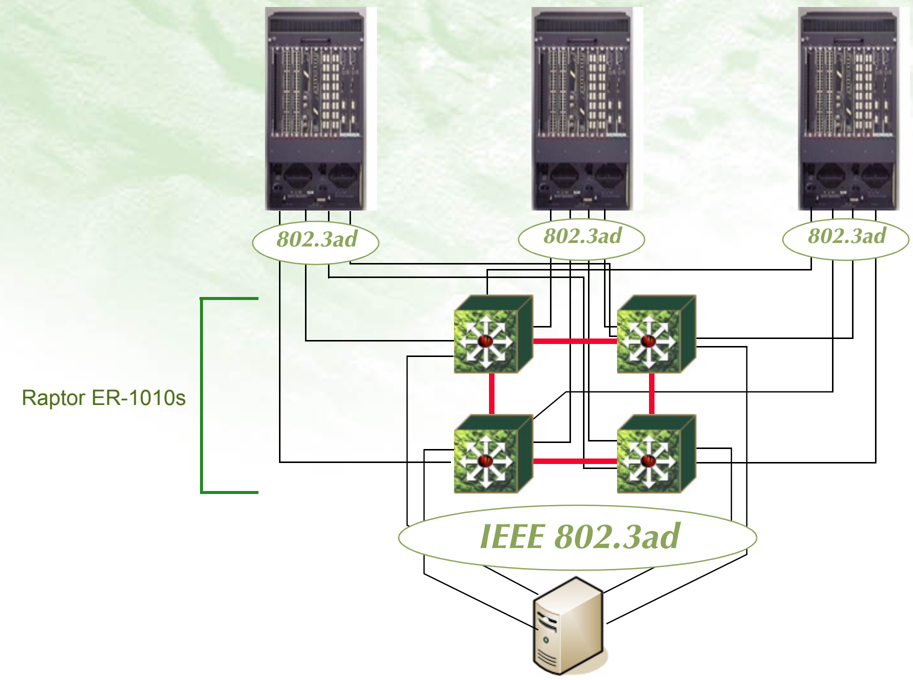


Figure 3: VoIP Server with Raptor Switches

In Figure 3, a VoIP server is connected to all four ER-1010 switches as an active/active failover backbone system, which will allow for 3 of the 4 Raptor switches to fail while still passing VoIP traffic. The existing chassis systems are also connected in a resilient fashion to the Raptor backbone. The Raptor backbone is a 640 Gbps total switching capability with 476 Mbps throughput as a highly resilient network that supports 144 ports of 10/100/1000 BaseX and 12 x 10-Gigabit ports. Assuming the quad connection profile shown in the diagram, this backbone can support 36 highly resilient (4 x 1-Gigabit LACP Groups) active/active Ethernet devices using the IEEE802.3ad (Link Aggregation/LACP) standard to create a massively resilient network over which VoIP, Video/IP, and Storage/IP can safely operate in a low latency environment.

Summary

The topics presented in this paper can be encapsulated in four areas:

- VoIP as a phone technology must have the same reliability as the POTS systems that users are presently operating.
- A resilient network is essential to deliver VoIP to every end user who requires access, but to also make sure that the service is reliable.
- Low latency is necessary to ensure good quality VoIP traffic and to prevent call loss.
- Ether-Raptor switches support all the requirements of VoIP, and can benefit any network by adding performance, low latency, high reliability at a cost point that is lower than other options.

Corporate Headquarters: 1241 E. Dyer Road, Suite 150 Santa Ana, CA 92705

Phone: 949-623-9300 /Fax: 949-623-9400 / Web: www.raptor-networks.com / E-mail: info@raptor-networks.com

CD-WP1400 10/09/2006