

The Border Patrol: Firewalls For VOIP

Gary Audin

Can your data-oriented firewall handle packet voice traffic? What if it can't?

Gary Audin is president of Delphi, Inc., an independent consulting and training firm. He can be reached at delphi-inc@worldnet.att.net.

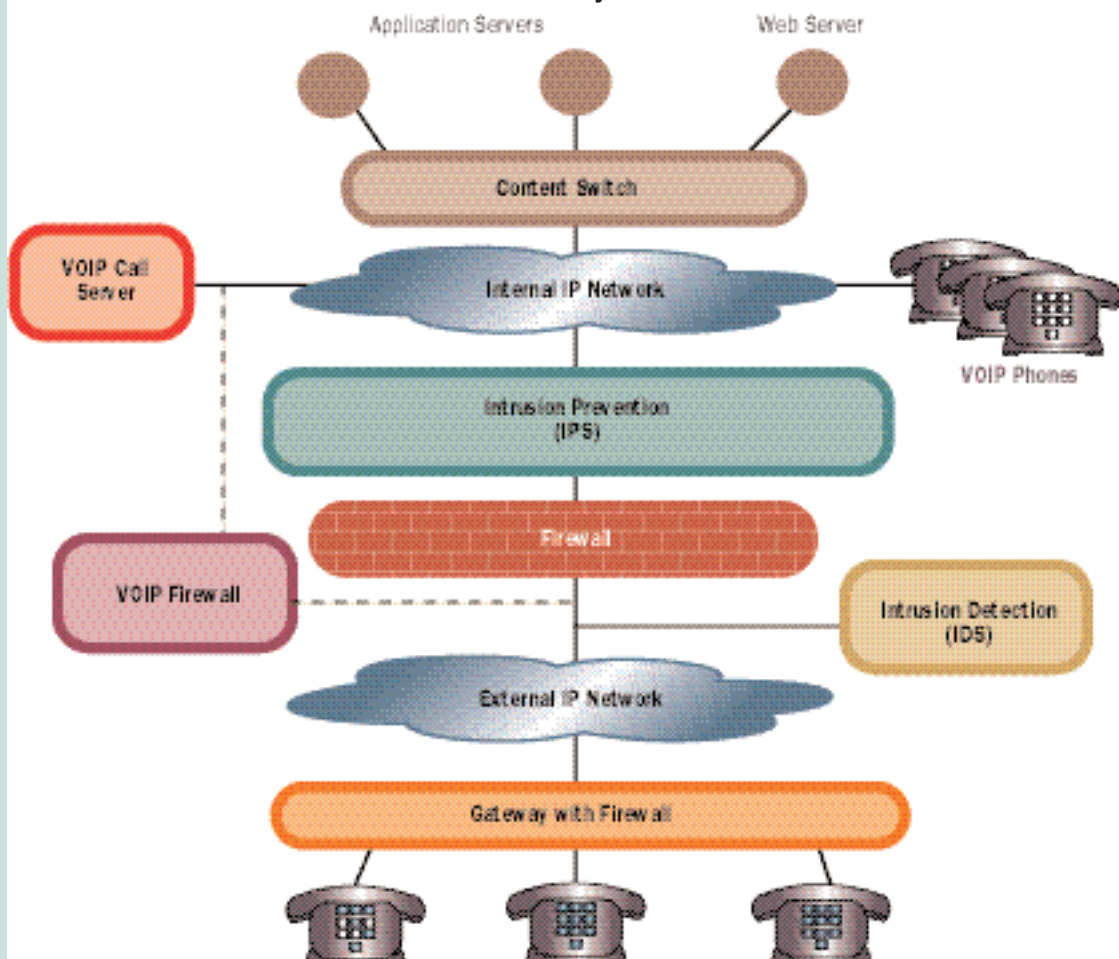
Firewalls provide security by blocking intrusions into an enterprise network. By allowing certain traffic in while blocking other kinds, they represent the physical implementation of an enterprise's security policies.

But firewalls also produce performance problems and cause delay. Most firewalls are designed

for data applications and are *not* application specific, though some firewall vendors (such as Checkpoint, Jasomi, Datapower, F5 and Sarvega) are moving toward packet content analysis (called deep packet inspection). This is a move to more application-specific security, though even it does not yet cover voice over IP (VOIP) packet analysis.

A recent Gartner Research Note, titled "Four Paths to True Network Security," describes both present and next-gen devices this way: "The underlying 'secret sauce' is a generic engine that

FIGURE 1 Security Positions



Next-gen firewalls must understand the concept of a voice “call”

performs packet assembly and compares the contents to a list of rules and lots of memory to cache the packet stream.” The note defines four approaches (Figure 1):

1. Intrusion detection vs. prevention (IDS/IPS).
2. Content switching (also called “application switch”) products.
3. Application-specific firewalls.
4. Traditional firewalls.

IDS devices passively report probable security intrusions, but they have proved frustrating to users; IDSs tend to produce many false positives, prompting users to either turn off rules or ignore the reports.

IPSs, the next-gen evolution of IDS, are more active devices that sit on the boundary between the internal and external networks and use an extensive set of rules to stop attacks that pass through the firewall. The rules must be implemented at the very beginning of the attack, and the subsequently received packets must be blocked. The IPS requires considerable processing power if it is to relay real-time voice and video traffic and avoid performance problems.

Content switching devices perform deep packet inspection to load-balance across multiple servers. These products could evolve into security systems, but are not yet designed to block traffic for security purposes.

Application-specific firewalls are targeted toward data and Web server environments, not real-time voice or video traffic. Gartner believes that “software running on hardened Unix platforms can handle application defense, but the demands on this technology quickly will escalate to the point where specialized hardware acceleration is required.”

Traditional firewalls are those that have been on the market for some time; they are typically software-based. A few firewalls are hardware based, and these likely will require hardware changes to support VOIP traffic. In general, traditional firewalls will have a difficult time supporting the short delay, jitter and throughput that application-specific firewalls will demand.

VOIP And Firewalls

VOIP traffic requires real-time delivery, short delay, low jitter and low packet loss across networks. Data firewalls are not designed for real-time applications. Among other issues, they have difficulty dealing with Network Address Translation (NAT) and VOIP signaling (see *BCR*, April 2003, pp. 55–58).

Besides these challenges, other performance and control issues arise when voice passes through a firewall. Next-generation firewalls will have to understand the concept of a “call” in order to do voice traffic analysis.

These complexities point toward the central question: What is the best way for enterprises to deploy firewall capabilities in converged voice/data networks?

Traditional Data Firewalls

Put up a firewall and it solves security problems. That’s a nice image, but it’s not reality. The firewall is necessary for security, but so are other devices such as intrusion detection and prevention systems (IDS/IPS).

Firewalls can provide a false sense of security. They do not protect the internal private network from rogue authorized users (employees), who are more likely than outsiders to be the source of security problems. Firewalls also introduce performance issues for real-time traffic (Table 1). What’s more, they can be cumbersome to install and use, they add a bottleneck between the private and public networks, sometimes deny access when they should allow it and can impair network performance.

The goal of firewalls, IDSs and IPSs is to detect and stop attacks, which come in many forms, including:

- Denial of service.
- Virus/Worms.
- Unauthorized access (hackers and crackers).
- Unauthorized data transfer.

Another major factor when dealing with security systems is their management and reporting capabilities. Next-generation security systems must detect attacks and intrusions and report on these while they’re occurring, rather than after the fact. The system should then automatically block future traffic based on its knowledge of past improper access.

Finally, the security system should constantly review temporary security policy changes, such as

TABLE 1 Firewall Performance Issues

	Data-Only Firewall	Real Time Firewall
Delay	More	Less
Jitter	More	Less
Rapid Dynamic UDP Port Assignment	Rare	Fast
Packet Rate	Moderate	High
Data Throughput	High	Moderate
Simultaneous Dynamic Port Assignment	Slow	Fast
NAT	Fast	Very Fast
Simultaneous Dynamic UDP Port Assignment	Rare	Fast
Signaling Delay	Longer	Shorter
Application Aware	No	Yes
Packet Inspection for Attack	HTTP and TCP	SIP and RTP

opening a port for a one-time transfer, because a temporarily opened port frequently winds up being left open permanently.

A firewall that possessed such powerful reporting and management functions could evolve into a broader traffic-management tool. Since firewalls process all the traffic at the enterprise network edge, they could provide valuable, detailed traffic statistics and relate the traffic to the applications used, addresses in use, packet characteristics, diagnostic and troubleshooting transmissions and patterns of traffic.

At the low end of the spectrum, a newer class of product is software or desktop firewalls for teleworkers. These are more limited and, therefore, more vulnerable to attack than enterprise-class firewalls.

Like their larger-scale cousins, desktop firewalls will be called upon to support real time voice and video applications. Unfortunately, desktop firewalls may actually block VOIP traffic. In one case, the installation of a common home firewall/router translated the UDP port numbers assigned to voice traffic traversing the firewall. The call processing was successful, but the speech path was closed. (The “Changing Policies” section below explains just why this happens.)

The instructions included with the teleworker firewall product never mentioned the port number translation. The lesson: Deploying teleworker firewalls could be a real headache.

Protecting And Passing VOIP Traffic

As the previous example illustrates, VOIP creates a whole new set of firewall problems. To understand these problems, we first have to understand how VOIP traffic crosses the firewall perimeter.

A VOIP call uses either the TCP or UDP protocol with well-known application ports to set up a call. TCP port 1720 is used as the primary port for H.323, and UDP port 5060 is used for SIP (which rarely employs TCP—though the latest version of the standard recommends that TCP be used with SIP in the future).

VOIP also requires one or two additional UDP ports to be opened for each individual voice traffic stream. One port is used for the Real-Time Protocol (RTP) traffic that carries the voice packets, and a second optional port may be assigned to monitor the performance of the RTP call, using the Real-Time Control Protocol (RTCP). This means that three UDP ports are required for a SIP-based call (for call control, monitoring, and the voice payload itself). The early version of H.323 required two UDP ports for RTP and two UDP ports for RTCP.

The UDP ports *should* be opened only for the duration of the call. Static UDP port assignment—that is, keeping ports open permanently—essentially leaves the firewall open and not really secure. And not only does the firewall have to open the UDP ports dynamically, it must do it

rapidly, for multiple calls simultaneously, with short delay and without introducing jitter or packet loss. Cheaper and older firewall products lack this dynamic UDP port assignment capability.

One possible VOIP-specific solution is to embed security functions in VOIP gateways, as we saw in Avaya’s July announcement of its Security Gateway product line (see *BCR*, September 2003, p. 62). The Avaya gateway integrates VOIP firewall protection, virtual private network (VPN) functionality and IP-telephony support. It also includes a bandwidth manager to provide QOS for the voice traffic.

“The mobility and dynamic nature of [VOIP] technology introduces not only new security issues related to VOIP support, but just as importantly, significant configuration and management concerns that can become significant obstacles to deployment,” said Jorge Blanco, VP, products and solutions marketing at Avaya. “The logical step is to integrate security within the communications infrastructure and applications such that it becomes part of an overall trusted communications framework.

“In this model, security is...an integral feature that becomes part of a simplified communications deployment process, whether in the enterprise, within a branch office location, or distributed to a DSL-connected home telecommuter using an IP phone,” Blanco said.

Changing Policies

As noted earlier, the firewall implementation represents the physical manifestation of the company’s security policies. These policies will need to be revised as data firewalls are modified to accommodate VOIP traffic.

For a start, many firewall installers use the product’s default settings, but these aren’t adequate for handling VOIP, since the characteristics of packet voice traffic do not resemble those of data traffic. The following conditions must be considered when making policy changes in the data firewall to accommodate voice:

- VOIP packets are small (64–100 bytes), with near-constant length.

- The packet rate is about 25 to 200 packets per second, most commonly 50 packets per second. Firewalls *should* recognize different types of real-time media to determine if the packet rate is reasonable for a given application type. High packet rates could be a sign of a denial-of-service attack.

- If silence suppression is used, the packet streams will alternate: One direction will carry packetized voice while the other direction will carry no packets at all. Without silence suppression, the packet streams will be constant in both directions.

- The average length of a telephone call is three to five minutes, so the UDP ports will have to open and close at that rate. Voice mail calls average about 30 to 45 seconds. These are normal call



Multiple UDP ports must be opened on the firewall for each VOIP call

The next-gen firewall must do deep packet inspection if it's to carry out even basic tasks for VOIP

patterns that must be provided for in the firewall rules as implemented by the enterprise.

■ During a conference call, the packet flow will be predominantly one direction. If the conference call is supporting a Webcast or on-line lecture or presentation, the packet flow will be exclusively in one direction, behaving like software distribution *without* any acknowledgments, which is unusual in legacy applications.

■ You'll have to allocate a set number of UDP ports up-front for VOIP, and this will restrict the number of simultaneous calls that can be set up through the firewall. For example, if 512 UDP ports are set aside for VOIP, then at most only 256 calls can be carried, since each call requires a minimum of two UDP ports. Thus, the firewall can become a call-blocking point in the network.

■ The TCP (for H.323) and UDP (for SIP) signaling ports may be left open 24/7, otherwise they will have to be opened for each call. If left open, the signaling ports also will be idle for long periods when no calls are made, such as on nights and weekends. A firewall may normally be configured to shut down these ports if they remain idle for a specified time, to increase security. But that practice has a major drawback with VOIP, because employees who work or access voice mail on the weekends won't be able to get through if the signaling ports are closed.

■ Endpoints may not be aware of port number translation occurring at the firewall, which will cause the speech (RTP) path through UDP to fail even though the signaling operation succeeded. This has happened with some home firewalls, as described above. The fix is for the firewall to edit the text of the signaling packets and modify the RTP/UDP port numbers so that they translate properly, ensuring that the telephones connect to the translated UDP ports. This type of function is not commonly supported by data firewalls.

■ If the firewall can support quality of service (QOS), then both the signaling and voice packets must be given a high QOS level, with the voice packets provided the highest level.

Some of the above items concern identifying abnormal packet patterns. If your firewall can be configured to detect abnormal packet patterns, then these modifications will probably be required. If your firewall cannot detect these abnormalities, you should look for one that does before you implement VOIP through a firewall.

Real-Time Firewalls

To deal with these issues, a few vendors have created a new class of product, the real-time firewall (RTF), specifically designed to handle both data and real-time applications like voice and video over IP (Table 2). The significant difference between data and real-time firewalls is their performance for voice traffic.

Deep packet inspection looks into the content of a TCP or UDP packet for a thorough, all-encompassing view, including such crucial information as the IP address of the destination device. This inspection is performed by disassembling and reassembling the IP packets (called datagrams), as well as the TCP and UDP streams as they pass through the firewall. All of this extra processing requires significantly more computing power than most smaller and software-driven firewalls possess.

The key areas that must be addressed by all solutions, regardless of vendor, include:

- Compatibility issues with NAT.
- Control of dynamic voice sessions.
- Call admission control.
- Invalid signaling and voice stream challenges.
- Network latency and jitter.
- Visibility and control over legacy voice and resulting cross-network connections.

"A solution to the signaling and the UDP port-assignment problem is to provide a signaling filter resident in the RTF firewall," recommends Mark Collier, CTO of SecureLogix. "This filter would terminate and regenerate all signaling requests. The filter can then monitor for valid call requests and dynamically open and close the appropriate ports. It can ensure that the UDP ports are open by

the time the call has been established. It can also interact with other components in the network to provide authentication and encryption of signaling requests.

"The signaling filter can work with the NAT to rewrite the addresses in the signaling stream and provide NAT pathways for the voice stream," he added. "If these translations are incorrect, the voice stream will be blocked by the firewall."

TABLE 2 Firewall Security Features/Functions		
	Data	Real Time
Denial of Service	√	√
Intrusion Detection	√	√
Intrusion Prevention	Usually	√
Virus Scanning	√	√
VPN	√	Some
Dynamic Port Assignment	Usually	√
IP Security	√	√
Encryption	√	√
Malicious Behavior Blockage	Some	√
Network Address/Port Translation	√	√
Application Awareness	Few	√

A second, more complex approach is called a back-to-back user agent (B2BUA), which supports encryption as well as the other features of a signaling filter. A combination of both approaches provides the most comprehensive approach to security, according to Collier.

VOIP Performance And Firewalls

Five factors impair the performance of VOIP calls that pass through any network or device: delay, jitter (delay variance), errors, packet loss and packet out-of-sequence delivery. These factors may also affect call setup/teardown time for signaling.

Delay through a firewall is caused by the packet store-and-forward processing and analysis of each packet. VOIP packets are very short (tens of bytes), therefore the store-and-forward delay is short. Packet analysis delay, which is in addition to the store and forward delay, may be caused by factors including:

- Too much time spent in policy enforcement.
- Slow firewall processing speed.
- Packet headers/trailers that require excessive analysis.
- Limited bandwidth at the exiting network port.

The first two problems are affected by the firewall product design, while the second two can be problems for any firewall, regardless of design. Inadequate bandwidth causes congestion; heavy congestion adds delay, jitter and packet loss.

Jitter, or delay variance, will fluctuate depending on the number of packets ahead of the VOIP packets in the output queue. This may become an issue because data-only firewalls can have a large output queue and still satisfy user expectations—a few hundred more milliseconds' delay does not affect the operation of most data streams. In contrast, even 100 milliseconds of jitter is intolerable for VOIP.

Another factor that causes jitter is packet length. Data packets come in all sizes, but the largest ones—such as FTP and e-mail—can be 10 to 40 times longer than a VOIP packet. Longer packets cause longer queue delays, and variable-size data packets cause jitter.

Mixing VOIP and data packets through a common firewall penalizes the voice packets but *not* the data packets. The jitter introduced by a dedicated VOIP firewall would be far less than the mixed VOIP/data firewall, since the jitter would only be a factor of congestion, not packet size variations.

Packet loss occurs when the exiting network bandwidth is less than required for the incoming packet load. This is a congestion problem. Packet size also affects packet loss. When congestion is heavy, long packets take longer to deliver, thereby relieving the congestion problem more slowly and making packet loss more likely. Data firewalls often add buffering to mitigate this problem, but the added buffering causes delay, again affecting voice quality.

Short VOIP packets can be delivered faster because of their small size. Congestion is then relieved faster, producing less opportunity for packet loss. Again, mixing VOIP and data packets in a common firewall penalizes the VOIP packets, *not* the data packets.

It is very unlikely that a firewall will introduce *errors*. Errors are usually the result of a transmission problem. A firewall will discard faulty VOIP and data packets equally. The firewall sees no difference between the two.

Out-of-sequence packet delivery occurs when there are multiple router paths to a destination. As congestion over the first route increases beyond the limits set in the router, an alternate path with less congestion is chosen. Packets entering later may actually be delivered sooner, causing the out-of-sequence delivery problem. Firewalls do not cause this out-of-sequence delivery, nor can they fix the problem. They must be tolerant of this condition when they check sequence numbers for validity.

Finally, data firewalls were not designed for dynamic port assignment, because before VOIP, only static port assignments were required. Newer data-oriented firewalls can perform dynamic port assignment.

The way that the firewall handles dynamic port assignment will affect voice quality on a VOIP call. Configuring UDP ports dynamically takes time. Because signaling is done over statically configured ports, the call may be established between the end points (the telephones) and speech may start *before* the dynamically configured UDP ports are operational. This means packets are lost at the beginning of a call. This lost packet problem is worsened when the firewall has to set up many UDP port assignments simultaneously, as in a large office or call center, or during an emergency.

Conclusion

Where is all of this leading? For many enterprises, the solution may be a separate application-specific real-time firewall (RTF) running in parallel to the existing data firewall, a hardware- rather than software-based device. In this way, the VOIP traffic passes through a firewall specifically designed for its needs, while blocking data traffic. At the same time, the data firewall blocks VOIP signaling and traffic without penalizing the VOIP traffic. You get the best of both worlds□



One possible configuration: Voice and data firewalls running in parallel

Companies Mentioned In This Article

- Avaya (www.avaya.com)
- Checkpoint (www.checkpoint.com)
- Datapower (www.datapower.com)
- F5 Networks (www.f5networks.com)
- Jasomi (www.jasomi.com)
- Sarvega (www.sarvega.com)
- SecureLogix (www.securelogix.com)