

SIP, Security & Firewalls

The Trend Toward Unified IP Communications

by

Olle Westerberg
Chief Executive Officer
Ingate® Systems



Olle Westerberg has been active in the technology sector for more than 20 years. His experience with telecommunications, networks and product development include a range of roles from hands-on software development to corporate leadership. Westerberg has held senior management positions with Ericsson, DSC Communications and Alcatel with particular focus on international sales and marketing activities, including Vice President Private Operators at Ericsson Switzerland and Vice President Sales, EMEA at DSC. Westerberg boasts first-hand experience with small business needs as the CEO of Sweden's first dedicated e-service company, Buybridge, which he led from beginning to revenue generation within one year. Westerberg holds an MSc in Engineering from The Royal Institute of Technology (Stockholm) and an MBA from London Business School.

Contact details:

olle@ingate.com

tel.: +46 8 600 7750

mobile: +46 733 58 31 00



Index

1	<i>Executive Summary</i>	1
2	<i>Introduction</i>	2
2.1	Features	2
2.1.1	Presence.....	2
2.1.2	Instant Messaging.....	2
2.1.3	VoIP.....	2
2.1.4	Video.....	3
2.1.5	Unified Functionality.....	3
3	<i>Market Trends</i>	3
3.1	Geographical Trends	4
3.2	Adoptions	4
4	<i>The importance of SIP</i>	5
4.1	What is SIP	5
4.2	SIP vs. other protocols	5
4.3	Security features in the SIP protocol	5
5	<i>Security Issues</i>	6
5.1	Firewalls and NATs	6
5.1.1	Controlling the Firewall.....	6
5.1.2	The NAT Issue.....	7
5.1.3	Integrated Firewall Functionality.....	7
5.2	Identification of Users	7
5.2.1	Solutions.....	7
5.3	Media on an Open Network	8
5.3.1	Solutions.....	8
5.4	Quality	8
5.4.1	Solutions.....	8
5.5	Firewalls	8
5.6	SIParators®	9
5.7	Other Solutions	9



1 Executive Summary

Person-to-person realtime IP communications, like presence, VoIP and video applications, offer clear benefits for the enterprise as time- and money-savers. Unified functionality is now available, where all of the above are integrated into one streamlined application and interconnect with other networks. This creates a very important business tool and usage is increasing.

Session Initiation Protocol (SIP) has emerged as the protocol that will fuel this development. Its simplicity and the fact that it is based on Internet standards give SIP clear advantages over older protocols. Analysts predict that SIP will dominate and the evidence given by actual deployments among service providers and equipment vendors further support that prediction.

The beauty of the SIP protocol, besides its simplicity and standardization, is that SIP makes it possible for any enterprise to implement cost-effective solutions that offer the benefits of full communications capability while maintaining a high standard of network security. While SIP makes secure realtime person-to-person communications possible, it's critical to realize there are specific guidelines businesses must adhere to in order to maximize this security benefit.

For instance, SIP-based solutions reach maximum effectiveness only if the SIP capabilities are built into the enterprise firewall. By building the SIP capability into the firewall itself (SIP-capable firewalls), it is possible to keep communication ports open only as long as necessary to let communications pass. At all other times the ports remain closed. An onboard SIP proxy and SIP registrar also make it possible to handle situations where SIP users are sitting on private IP addresses behind a network address translator (NAT) without revealing private address information to the outside world.

There are additional security issues with the SIP protocol beyond the basic firewall functionality. Because of SIP's open nature, the systems involved need to provide functionality for authenticating users at registration. Special circumstances must also permit the possibility of encrypting messages before they are passed over the Internet. The SIP standard plays a part of such encryption, passing SIP over TLS. Any SIP implementation must contain this capability.

The final security issue for IT managers to address is how the heavy data usage that will likely accompany their SIP-based communications (VoIP, for example) may affect the quality of those communications (voice quality). A SIP-capable firewall should have the ability to limit bandwidth and set priority for different types of traffic. This will make it possible to ensure appropriate voice quality for the SIP traffic.

Products meeting the above criteria are available today. There are also solutions for enterprises that do not want to replace existing firewalls but still desire to make use of modern person-to-person IP communications. These solutions are cost-effective, eliminating investment as the principal limiting factor for the adoption of new services.



2 Introduction

Various forms of person-to-person IP communications are attracting increasing interest around the world. Several leading research institutes state that the Voice over IP (VoIP) is the next big trend in telecommunications. As VoIP brings about new opportunities it also raises some security concerns. Compared both to traditional telephony (PSTN) and other packet-based data communications like e-mail, VoIP installations demand certain prerequisites to work and to guarantee the same security level. The Internet architecture of today cannot in itself offer the same kind of physical security as with the PSTN. The main difference between VoIP (compared to traditional IP services like Web access or e-mail) is that VoIP takes place between two or more individuals, not between servers or between one person and a server. This puts entirely new requirements on network capabilities, and opens up new security issues as well. However, there are solutions to make VoIP as secure as traditional telephony and other data traffic.

This paper will examine the security concerns for businesses looking to implement VoIP and also present solutions to problems when creating one converged network to transport both data and voice.

2.1 Features

VoIP is the most well-known form of person-to-person IP communications. However, there are several other forms that complement VoIP which offer a rich palette of communication tools for businesses. We therefore start by looking at definitions of some of these other communication forms available today.

2.1.1 Presence

Presence functionality lets users know who is available to chat among their regular business contacts. Users add people to their contact lists, and receive immediate notification when one of these “subscribed” contacts comes online.

2.1.2 Instant Messaging

Instant messaging functions similarly to the intercom systems of offices in the past. Instant messaging is useful for short requests or questions; it is much less disruptive than a phone call but more immediate than e-mail.

Instant messaging enables users to send short text messages to another person. In combination with presence, instant messaging increases the chances to get simpler questions answered quickly, thereby enhancing productivity within the enterprise.

2.1.3 VoIP

Telephony over the IP network (VoIP) is perhaps the most well-known application of person-to-person communication. The VoIP trend is leading businesses down a path of using the Internet to the maximum extent possible and avoiding private networks, leased lines and other specialized



solutions for their telephony needs. VoIP also allows enterprises to minimize the use of traditional telecom networks (whether PSTN or mobile) as much as possible.

There are several definitions of VoIP. We define VoIP as the ability to call and be called by others over the Internet, whether the caller is using another VoIP system or traditional PSTN phone.

2.1.4 Video

With modern technology, it is quite possible to communicate *via* good quality video over the Internet even with relatively simple and cost-effective end equipment. The ability to share ideas over the Internet using video greatly enhances the productivity in a distributed enterprise environment.

2.1.5 Unified Functionality

Presence and instant messaging are naturally used through a software interface on a personal computer. Voice has traditionally been used with a separate handset (a telephone). While handsets will continue to exist, soft clients like Microsoft's Windows Messenger create added value by integrating all of the above functionalities into one package. This enables the user to seamlessly transfer between modes of communication with a mouse click.

An example: once you have established that the person with whom you want to communicate is available (presence), you send a message (instant messaging) asking a question. During this interchange, it becomes apparent that the issue at hand is too complex for a text message-type conversation. You then simply put on a headset and click "Start Talking" on your soft client to initiate Internet-based telephony (VoIP). When there is a need to show a practical example of what you are talking about, you can start a video session in the same way.

Another area of development is that of interconnectivity. Once the basic communications infrastructure is in place, it is possible to implement a variety of interconnecting services such as PSTN services, which can deliver the ability to make calls from an IP network to a traditional telephone. Another example is the ability to chat *via* instant messages over SMS or to an e-mail address.

3 Market Trends

Usage of VoIP and other person-to-person enterprise IP communications is growing. Info Tech research predicts that by 2008, 50% of all small and mid-sized business will implement VoIP. Further reports (Gartner Dataquest) predict that enterprise spending on IP phone systems in North America will more than double during the next few years to reach \$4.2 billion in 2007.

The foremost reason for the increase: VoIP saves both time and money for the enterprise.

SIP communications offer so much more than just VoIP. As described above, there are several other communication tools that can be used and integrated into one unified system that will increase a business' efficiency. The traditional customer service department is one area with significant opportunity to benefit from these new communications tools, as they can provide a varied means of customer interface with the company. For example, platinum clients could access customer care



representatives *via* presence, giving them a far more personalized experience than, say, an interaction by email.

Another potential area for growth: the changes in working patterns. Companies are seeing the benefits of affording employees all over the globe the ability to work together seamlessly, regardless of location or time zone. They also acknowledge the benefits of home-based workers, or hiring specialists who work not from the home office, but rather from a satellite location. The result means that where you used to walk through the office to chat with a fellow co-worker, today instant messaging serves as the communications bridge between employees.

3.1 Geographical Trends

As always, these technology trends started in North America and have for a substantial period of time been growing significantly there. Recently, however, there have been significant developments in the Asian market, especially regarding VoIP. The underlying reason for this boom is likely due to the less extensive traditional telecom networks there, making the economics more favorable to VoIP solutions. Although Europe is still in its infancy when it comes to these services, several early deployments are currently underway.

3.2 Adoptions

More and more industry participants are now embracing SIP for the development of person-to-person communications services and products. Examples range from service providers and operators to equipment and software suppliers.

Among service providers, dominant players like MSN, AOL and Yahoo! have all announced support for the SIP protocol. Other telecom operators like AT&T, MCI and Telia of Sweden are also active in this field.

The number of equipment suppliers announcing products with SIP capabilities is increasing by the day. While many early adopters were new niche players, large, traditional industry stalwarts are now jumping on the bandwagon. As VoIP is going mainstream the demand for interoperability between vendors increases.

Perhaps the single most significant endorsement of SIP being not only the standard protocol for VoIP, but also for other forms of real time communications came from Microsoft, when they announced the inclusion of a SIP-based communications client with its Windows operating system. The Microsoft Office Live Communications Server (LCS) effort practically ensures the vision a SIP-based client on almost every enterprise desktop in the world.

Clearly, the SIP protocol is gaining momentum and there is no longer any doubt that it is the dominating protocol for the type of person-to-person communication discussed in this paper. Older solutions will exist for some time due to previous investments in technology, but it is quite likely they will eventually be phased out. Most new investments will be made with SIP as the enabling technology.



4 The importance of SIP

Compared with existing solutions, SIP's advantages can be summarized as follows: simplicity, scalability and global connectivity.

The SIP protocol has a simpler structure than its predecessors. This means that it is easier to implement and that there is less room for companies to create their own flavor of SIP. SIP can therefore be implemented more cost effectively.

Services based on SIP are scalable. Because the protocol makes use of the inherent capabilities of the Internet (routing, for instance) there need not be one single focal point for traffic, even in a very large network. Architecture can be kept distributed.

The most important advantage of SIP over other protocols, however, is the ability to achieve global connectivity. Anyone who publishes his or her SIP address (presence) on the Internet can be reached by anyone else on the Internet. That means that the only thing needed for true global connectivity is that an enterprise user can reach across the enterprise firewall to publish its presence. This is different from server-based presence and instant messaging services like MSN or ICQ in that you need not register on a central server and you can be reached by anyone, regardless where they are in the world and which service provider they are using. In fact, SIP users don't need a service provider at all except to provide their Internet access.

4.1 What is SIP

SIP is a text-based communications protocol for initiating two-way communication sessions over Internet. It is standardized by the Internet Engineering Task Force (IETF). This means fewer interoperability issues. Also, the SIP protocol supports all forms of realtime communication described above since it can be used for all the services and features mentioned. Finally, this Internet-based protocol is designed to make use of the features already existing in the large public IP network, the Internet. These functions include the routing mechanisms inherent in the Internet.

4.2 SIP vs. other protocols

SIP offers some significant benefits over other protocols like H.323 and those used specifically for VoIP. SIP is an application-level protocol that can be carried by several different transport protocols like UDP (User Datagram Protocol) or TCP (Transmission Control Protocol).

The UDP gets a data unit from one computer to another. However, UDP does not divide a message into packets and reassemble it at the other end. This means that the application program using UDP must be able to make sure that the entire message has arrived and is in the right order.

TCP takes care of keeping track of the individual packets of data that a message is divided into for efficient routing through the Internet.

4.3 Security features in the SIP protocol

Several security features for the SIP protocol are described in RFC 3261. These include:



- HTTP digest authentication for data signalling authentication
- S/MIME
- RTP encryption for confidentiality of media data
- TLS for the protection of SIP signalling messages against loss of integrity and confidentiality

IETF is currently working on several drafts regarding other security issues.

5 Security Issues

The SIP protocol boasts several security features in itself, but most companies adopting SIP incorporate an enterprise-class firewall to protect the network from attacks and unwanted traffic. However, most firewalls cannot differentiate between incoming SIP traffic and unwanted incoming traffic. The result: traditional enterprise firewalls treat all SIP communications (including VoIP) as unwanted traffic, and stop the communications from reaching users behind the firewall.

5.1 Firewalls and NATs

Reaching the Internet with SIP-based applications is one of the most important challenges for the growth of this industry. However, the Internet must be reached without compromising the security of the private enterprise LAN.

Unlike other Internet applications, realtime applications need to connect to a user on the private LAN, not to a server. The problem is that most companies use NATs (Network Address Translation devices), making private addresses on the LAN not addressable from the outside. Also, when a communication session is initiated the media streams will take a separate, dynamically allocated path through the firewall.

5.1.1 Controlling the Firewall

Several solutions exist whereby the enterprise firewall is either bypassed or “tricked” into letting communications through. The latter is sometimes referred to as “opening pinholes” through the firewall. Other old solutions include building an enterprise-class PSTN gateway that translates the traffic to the PSTN already within the enterprise, a solution that does not make use of the infrastructure available on the Internet.

The only long-term viable solution is one that makes the firewall aware of the SIP signalling and checks which dynamic UDP ports should be opened for just long enough to allow communication to pass and then be closed again. It presents the most secure scenario, as ports stay open for the minimum amount of time possible.

However, this solution presents unexpectedly large challenges on firewall vendors. A firewall is traditionally designed specifically not to allow dynamic control.



5.1.2 The NAT Issue

Network Address Translation (NAT) is used to provide the network behind a firewall with multiple private IP addresses even though there may be a limited number of public IP addresses available to the enterprise. The reason for using NAT has been the lack of public IP addresses in IP v4, but it has also been considered somewhat of a security feature as the NAT hides the private addresses used on the private network from users on the outside of the firewall.

The use of NAT poses additional problems for person-to-person IP communications applications. As the private addresses are hidden from the outside, there must be a way to translate incoming requests to the private address space and route them accordingly.

5.1.3 Integrated Firewall Functionality

The solution to the above issues is to use an enterprise firewall with integrated SIP functionality. Such a firewall must meet all traditional security standards for enterprise firewalls including stateful inspection, packet filtering, *etc.* At the same time, it needs to include a SIP proxy to interpret the SIP messages. The SIP proxy will do the following:

- inspect the SIP signalling to ensure that it is valid
- rewrite the headers to enable the SIP signalling to reach the intended recipients at their private IP addresses
- dynamically open the media ports to admit realtime communications on either UDP or TCP ports
- rewrite the headers to deliver the media to the appropriate destination
- rewrite the headers of outbound traffic to keep the private addresses hidden
- close the media ports when the session is finished

An on-board SIP registrar that enables the registration of SIP clients on the private network and allows the proxy to translate between the private and public address spaces is also necessary to integrate in the firewall.

5.2 Identification of Users

As the servers enabling SIP communications must, by definition, be open to the public, issues may arise compromising the true identity of users. In order to prevent an unauthorized user from “kidnapping” the user name of someone else, there must be functionality in the firewall unit that serves as the SIP registrar for user authentication. In addition, the administrative capabilities of such a unit should be strictly protected.

5.2.1 Solutions

Digest and RADIUS offer common models illustrating how this should be done. There are also several organizations that issue certifications for this type of security.



5.3 Media on an Open Network

The trend toward unified IP communications means that more and more traffic will pass over open networks like the Internet. As with traditional telephone conversations, such communications can potentially be intercepted and interpreted by someone other than the intended recipient. The first solution to address this security issue is to use restraint in what is communicated over an open network. For more advanced applications there are, however, additional options available.

5.3.1 Solutions

Virtual Private Network (VPN) is frequently used for communications within an enterprise or between an enterprise and other organizations close to it. VPN uses encryption algorithms that are secure enough for any normal enterprise application.

In addition, SIP includes encryption methods for SIP signaling. By sending SIP over TLS, all signalling related to setting up a SIP session is transferred using an SSL-type encryption. As presence information and instant messages are sent solely over the SIP signalling path, this information will also be encrypted when TLS is used, providing increased security. This is a hop-by-hop encryption, which means that global connectivity can still be maintained even if all clients do not support TLS. Encryption can still be maintained between, for example, the firewalls in each location.

5.4 Quality

The final security issue discussed in this paper is the issue of media quality. The capacity of the backbone IP network today is such that there are normally no major issues with voice or other media quality in a SIP session. With Internet access, however, the behavior of a few individual users can severely affect quality. For example, downloading high bandwidth video could potentially consume the entire bandwidth of the firewall.

5.4.1 Solutions

The firewall must contain basic Quality-of-Service (QoS) functionalities that allow it to both limit the maximum bandwidth available to certain types of services and to set priorities based on service type. Ideally, the firewall should be able to distinguish between SIP traffic and other traffic, therefore being able to give priority to the SIP based VoIP, IM, *etc.*

5.5 Firewalls

More firewall vendors than ever before are recognizing the need for SIP capabilities in their products. Many solutions available today use the bypassing or pinholing strategies described in section 5.1.1. Currently, the only firewalls available with fully integrated SIP capabilities are the IX67 small office firewall from Intertex Data AB and the full line of InGate Firewall® products from InGate® Systems. Various service providers and solutions providers – including Avaya, Telia, 3Com, Pingtel – have certified these as part of their respective solutions.



5.6 SIParators®

There is a large installed base of (non SIP-capable) firewalls in the world. These are often key installations where IT managers are very reluctant to replace or touch existing security infrastructure. For these installations, there is a specific need for a product that provides the SIP capabilities while maintaining the existing security configuration. The Ingate SIParator® from Ingate Systems serves exactly this purpose.

Firewall and NAT issues are resolved by letting the SIParator take responsibility for the opening and closing of SIP media ports and for translation between private and public addresses. All the existing firewall has to do is to statically allow the SIP signalling (usually on port 5060) and a range of UDP ports to flow freely to and from the SIParator.

The authentication of SIP users and encryption of SIP signalling can also be handled directly by the SIParator, with the existing firewall transparently passing on the information.

In this configuration, the SIParator cannot itself take responsibility for the entire QoS issue. Here the capabilities of the existing firewall will have an important role in determining exactly what can be achieved.

5.7 Other Solutions

Other solutions for introducing SIP capabilities are generally not integrated within the firewall itself. This makes implementation more cumbersome and costly because of the need for several different units. For larger enterprises with greater, confirmed needs, these solutions can be extremely attractive. For very large solutions, the split between functionalities can serve to make administration more manageable and overcome capacity restrictions.

However, if usage of unified person-to-person IP communication is to take off, it is imperative that solutions be easy to implement and cost-effective for the small-to-medium enterprise to realistically consider an enterprise adoption. This is best achieved with an integrated approach.