Sprint Enterprise Instant Messaging Security
Dale Bachman, PhD, CISSP

Instant Messaging is today’s killer application. Millions of consumer users are exploring the implications of real-time text messaging and finding that it is a perfect supplement to and, in some cases, substitute for voice communications. The enterprise market has only recently begun considering applications to business collaboration, and has found that consumer products are not generally suitable for enterprise use. Consumer applications were built without security in mind, so in general there is no protection for sensitive information that may be communicated through them. Furthermore, consumer applications were often designed specifically to circumvent security precautions. The time is ripe for an enterprise-class Instant Messaging solution.

Sprint has created two products to address the business need for Instant Messaging: Sprint Instant Messaging and Sprint Enterprise Instant Messaging. Sprint Instant Messaging is a desktop-based, completely hosted solution, offering a low-cost Instant Messaging service with security features, such as message encryption, without the management of an internal enterprise Instant Messaging server. Sprint Instant Messaging employs java clients that minimize installation issues. The clients interoperate with popular consumer Instant Messaging systems, so employees may interact with business contacts who are not on the Sprint Instant Messaging system.

Sprint Enterprise Instant Messaging is designed specifically for corporations. It is a closed enterprise Instant Messaging system, which can be hosted by Sprint or housed entirely inside a corporate network. Based on open standards, it offers authentication, encryption and interoperability with java-enabled devices, such as recently introduced PCS Vision phones as well as PDAs and PCs. In other words, it has clients for desktops and most wireless devices and networks, and all of the clients interoperate.

Since the security features of an enterprise Instant Messaging system are paramount, this document has been prepared to provide an overview of the security concerns of placing an Instant Messaging system in an enterprise setting and some of the technical measures and architectural options that have been incorporated in Sprint Enterprise Instant Messaging to address these concerns.

Motivation: Why Enterprise Instant Messaging?

Instant Messaging has gained immense popularity in the consumer space. But does this mean it should be adopted by the enterprise? What are the benefits we gain? What are the features that have made it popular, and which of these translate well to the enterprise? Here is a short list of Instant Messaging features.

- Presence information – Instant Messaging may be the first communications technology to give the user corner-of-the-eye information about the availability of important resources on a real-time basis. The effect of this property can’t be overemphasized, because it breaks down barriers to communication that we didn’t even know existed, effectively welcoming collaboration with the people who have the most to offer. It also encourages interaction by reminding us constantly that we can ask a colleague for information before we spend unnecessary time trying to find it on our own.
- Informality – Another way in which Instant Messaging breaks down communication barriers is by encouraging informal communication. This, along with the visibly open door represented by the presence information, means that even introverted personalities seek out collaboration.
Mild interruption model – A phone call demands response immediately, though voice mail mitigates this property. Instant Messaging, on the other hand, offers conversation without demanding immediate response. If the recipient of a message (or chat request) must take a short time to finish a task before responding, the conversation can start as soon as she is ready. And again, since the sender and the first sentence of the conversation are displayed, we obtain corner-of-the-eye information about the topic and probable length of the conversation.

Alternative communication channel – The fact that Instant Messaging is silent and may use alternate channels (especially in a multi-modal model) allows multitasking to improve efficiency. For example, small tasks can be accomplished during lulls in meetings, presentations or conference calls. This can, however, lead to breaches of courtesy. As in the early days of newsgroups, usage protocols will have to be established. Another benefit of the alternative channel property is a possible reduction in phone and email traffic, and therefore costs.

Improved collaboration – Secondary features such as whiteboard capability, voice chat, and file transfers may enhance the potential for collaboration. However, we should be careful to check each such added feature against our security requirements. For example, file transfer may open severe vulnerabilities in the system (to viruses, or data theft), and might be better left to a special purpose application such as an ftp server with integrated content analysis.

Natural path for application integration – The informality of the Instant Messaging experience makes it an ideal stepping-stone for application integration. First, applications which emulate human interaction can be used to give Instant Messaging users transparent access to information sources – in other words, a “buddy” doesn’t have to be human. Second, the APIs, protocols and behaviors developed for Instant Messaging applications can be leveraged for special-purpose services.

All of these features can result in increased productivity and communication, and should be considered as we design our architecture. However, there are also disadvantages to Instant Messaging applications. Four of these are

Improper usage – The ubiquity and informality of Instant Messaging may encourage non-business-related discussions that waste company time and distract employees from their jobs. Employees may use Instant Messaging to chat with their friends outside the enterprise, or to exchange gossip inside.

Lack of security features – Instant Messaging applications were not designed with security in mind. The peer-to-peer communications model of many applications is notoriously hard to integrate with security, and most common Instant Messaging applications send messages in plaintext and are relatively easy to intercept.

Inappropriate transfer of information – Regardless of whether security features are in place or not, Instant Messaging encourages on-the-spot information exchange. This is great for collaboration, but it may eliminate the time between concept and transfer during which review can occur. Thus, inappropriate distribution of sensitive information is likely.

Anti-Security Tunneling – Instant Messaging applications have traditionally been designed to be very configurable in their protocols, specifically to circumvent security controls. For example, many public Instant Messaging clients and servers can be configured to use any TCP port, including well-known ports like 80 (http) or 443 (SSL), so that firewalls may pass traffic through without inspection.

Security features in any enterprise Instant Messaging system must address these issues.

The goal of Sprint Enterprise Instant Messaging is to offer multi-modal, enterprise-class Instant Messaging that retains as many as possible of the advantages of Instant Messaging, addresses the issues, and is robust and scalable. Probably the most important point here is that the system is multi-modal: it brings the benefits not only to the enterprise, but to mobile clients as well as desktop PCs, and ties these clients together in an innovative way.
Sprint Enterprise Instant Messaging Architecture

The basic components of the system are the Connection Manager (CM) and the Sprint Enterprise Instant Messaging server. The CM serves as an endpoint for client TCP sessions, and passes messages to the server. The CM also passes authentication requests to the server, which communicates with a RADIUS server. The Sprint Enterprise Instant Messaging server communicates with an enterprise directory for profile information, and with a mail relay to send notifications to some clients about waiting messages.

Security Concerns

There are three major concerns that we address in this document:

- Confidentiality and integrity of the traffic, as it relates to unauthorized entities – that is, entities outside the enterprise gaining unauthorized access to the traffic
- Confidentiality and integrity of the traffic, as it relates to authorized entities – that is, breakdown of internal segregation
- Use of the system as an element in wider security compromise – that is, an attacker gaining access to the Instant Messaging system and using it as a springboard for attacking other enterprise systems

Note that in this paper we do not address availability of the system (the third element of the information security triangle: confidentiality-integrity-availability), nor privacy concerns (intentional distribution of personal information or correspondence), except as the confidentiality and integrity measures have secondary effects.

The major point to remember is that Sprint Enterprise Instant Messaging is intended for use by an enterprise, and as such there may be confidential and/or proprietary information passing through it. For this reason, a server-based architecture (as opposed to peer-to-peer) is desirable, as it provides more central control over both the flow of information and the configuration of the system. Furthermore, our security recommendation is to configure the system not to be interoperable with consumer Instant Messaging products, so that employees are not encouraged to share sensitive information with outside friends or colleagues.

The multi-modal nature of Sprint Enterprise Instant Messaging suggests consideration of security issues on wireless networks. A number of wireless networks may be used by Instant Messaging clients: for example, AMPS, TDMA, CDMA, 802.11a or b. Several such networks have security (or “privacy”) features – some, such as WEP for 802.11b, are known to be weak, while others, such as CDMA privacy features, have few known issues. In any case, though, the security features of the medium will not be sufficient to protect sensitive information, since the Sprint Enterprise Instant Messaging system crosses public networks (specifically, the Internet) as well as private (such as the cellular network and the enterprise intranet). This is particularly true when mobile devices are used, since they are designed to travel outside the enterprise. The only solution to this problem is to layer security, that is, to specifically protect the Instant Messaging traffic, notwithstanding any protection afforded by one or more of the networks it traverses. This is precisely the philosophy adopted in the design of Sprint Enterprise Instant Messaging. We will see below a number of measures designed for this purpose.

Two problematic features of consumer Instant Messaging products are file transfer and file sharing. File transfer allows users to transfer files directly to each other using the Instant Messaging protocol, and file sharing allows creation of an informal file server that other users may access to store and retrieve files. These functions are the vectors for most malicious code transferred via Instant Messaging applications. Virus scanners, while beginning to examine this conduit, are not yet completely effective in controlling it. Furthermore, the file sharing capability was not designed with privacy or confidentiality in mind (though integrity-checking features are included in some applications), so it can be extremely difficult to control access to files shared in this way. It is strongly recommended that users exchange files using more robust and established mechanisms – for example, email, Unix or Windows file servers, ftp or web servers.
Because of the wide range of security issues associated with peer-to-peer file transfer and file sharing, we are balancing the required levels of security with ease of use for the next release.

**General Security Measures**

Several architectural steps have been taken to protect the traffic in the Sprint Enterprise Instant Messaging design, and will be applicable in all of the scenarios we outline below. Two of these measures are:

1. Secure Sockets Layer (SSL) connections between the client and the Connection Manager, for those clients capable of SSL. These clients include PCs, phones that are J2ME enabled, PocketPC devices, and devices running PalmOS v5. SSL is an open standard that has been thoroughly tested and adjusted by the Internet community until it represents one of the premier security protocols in use today. The Sprint Enterprise Instant Messaging implementation of SSL requires a server certificate on the CM, and requires this certificate to be verifiable by the client – which implies either an enterprise root certificate loaded on the client, or an enterprise certificate signed by one of the root certificates loaded by default on the client. No client certificate is required, since client authentication is handled separately. (See item 2 below.) Note that this does not address protection of transmissions to/from non-SSL-capable client devices.

2. Client authentication through a RADIUS server. This exchange is initiated after the SSL connection is established, and is thus protected by encryption. The CM passes the client through to the Sprint Enterprise Instant Messaging Server for authentication, which in turn cedes responsibility to an enterprise RADIUS server. Note that the RADIUS transactions between the Sprint Enterprise Instant Messaging server and the RADIUS server are protected by DES encryption using a long-term shared key that is loaded on both servers. Note, too, that communications between the CM and the Sprint Enterprise Instant Messaging server must be protected, either by encryption or by using a dedicated connection.

Furthermore, it is noted that the information obtained from the directory is not considered sensitive, since it is limited to user names and public information. The SMTP server passes only notification messages (i.e., that the server has an instant message waiting for a client), so neither is this traffic considered sensitive – though in an environment with specific security requirements, traffic analysis concerns could change this judgment.

Elements that will need to be protected therefore include:

1. The Sprint Enterprise Instant Messaging server itself, since sensitive messages pass through it in plaintext, and the RADIUS key may reside upon it. Further, it would be an ideal springboard for attack of other enterprise systems.
2. The CM, since it is the end of the SSL tunnels.
3. The data flow between the CM and the Sprint Enterprise Instant Messaging server.
4. The data flow between Sprint Enterprise Instant Messaging servers, if there are a number of distributed systems. This will probably be the case in large enterprise installations, both because of need for additional processing power and for geographic considerations.
5. Client devices, both while sensitive information is present and at other times to prevent installation of monitoring software and other malicious acts.

Of course there will be security processes needed, such as patch maintenance, log monitoring and activity tracking, and procedures such as incident response plans and termination activities. These will not be addressed in this document because they should be subsumed in the enterprise security policy.

**Scenarios**

Organizations may have a variety of security architecture strategies and drivers, so there will be a corresponding variety of Sprint Enterprise Instant Messaging architectures. In this section we outline three such architectures, depending upon whether the customer wishes to keep all systems strictly inside the corporate firewall, wishes to isolate the Instant Messaging components in a DMZ, or prefers to have the
whole system hosted. Each of these scenarios presents security advantages and disadvantages, and must be weighed against the goals of the organization.

Possibly the most important point to note is the flexibility of the Sprint Enterprise Instant Messaging system. This flexibility is essential to meeting the security needs of a variety of customers.

Scenario 1: DMZ

- CM and Sprint Enterprise Instant Messaging server in DMZ.
- Connection between CM and Sprint Enterprise Instant Messaging server is a dedicated line. If geographic distribution is required, links between CMs and Sprint Enterprise Instant Messaging servers are encrypted through a link encryptor or point-to-point VPN.
- RADIUS, LDAP, and SMTP servers in intranet.
- Connections from the Sprint Enterprise Instant Messaging server are allowed through the (internal) firewall to the RADIUS, LDAP, and SMTP servers.
- Clients access servers through SSL – firewalls (internal and external) pass SSL connections from any intranet or internet address to the CM.

Advantages

- Clients have direct access only to servers in the DMZ. In case of compromise, these servers have limited access to internal resources.

Disadvantages

- Sensitive information resides on Sprint Enterprise Instant Messaging server in DMZ.
- J2ME clients have port 443 hard-coded for SSL connections. This means that either the firewall (or some other network device in the path) must translate to port 5223 or the CM must have port 443 open. Since port 443 is privileged, this means that the application must run as root, at least momentarily. The server software is based on the Apache web server, and though this can start as root and then change effective UID to a non-privileged user, there have been reported issues with some threads continuing to run as root. The effect: if port translation is not done, unauthorized users may have direct access to a root-level process on a server in the DMZ. This is not considered a good security practice.

Scenario 2: Internal Servers

All servers are placed in the enterprise network. SSL is enabled from external clients through the firewall to the CM. SSL should still be used by internal clients as well, to help enforce internal segregation.

Advantages

- No sensitive information outside enterprise (except on clients!)
- If additional CMs or Sprint Enterprise Instant Messaging servers are necessary to scale to a large user base, it would be feasible to provide unencrypted channels between servers on the enterprise network. However, this does not support internal segregation.

Disadvantages

- SSL access by unauthorized users to internal server. At the very least, the firewall should be configured to allow such access to only the CM. Note that this is the classic setup for an IP fragmentation attack, and the firewall and routers should be carefully checked for their behavior in the face of such an attack. In general, direct access from the Internet to an enterprise address through the firewall is not a recommended practice.
Scenario 3: Hosted

- CM and Sprint Enterprise Instant Messaging servers in Sprint Internet Center.
- Connection between CM and Sprint Enterprise Instant Messaging server is a dedicated line. If geographic distribution is required, links between CMs and Sprint Enterprise Instant Messaging servers are encrypted through a link encryptor or point-to-point VPN.
- RADIUS, LDAP and SMTP servers in enterprise network.
- Connections from the Sprint Enterprise Instant Messaging server are allowed through the (internal) firewall to the RADIUS, LDAP and SMTP servers.
- Clients access servers through SSL – firewalls (internal and external) pass SSL connections from any intranet or Internet address to the CM.

Advantages

- Clients have direct access to only servers in the Sprint Internet Center. In case of compromise, these servers have limited access to internal resources. Alternatively, if no access to internal servers is allowed, directory services and authentication can be replicated in the Internet Center, and a Sprint mail relay can be used for notification.
- Sprint provides support and maintenance for major components, while customer retains responsibility for standard authentication and directory services.
- The SMTP server can be hosted by Sprint or be in the enterprise.
- Sensitive information resides on Sprint Enterprise Instant Messaging server in the Sprint Internet Center. The security of these servers is carefully controlled, so that even though they are directly accessible from the Internet they provide excellent protection. Port translation (see the discussion under Scenario 1) can be handled by Sprint systems.

Disadvantages

- Customer security policy may require modification to allow protection of sensitive information by a third party (namely, Sprint).

Legacy Clients

The above architecture and discussion focused on SSL-capable client devices. However, there is a large installed user base of 2G phones and older PalmOS devices that may not have SSL capability. There are three natural solutions for extending the architecture to these devices.

First, Short Message System (SMS) capability could be leveraged by providing a gateway device between the Sprint Enterprise Instant Messaging server (or a proxy) and SMS. In this case, security would depend upon a secure link between the proxy and SMS gateway, protection of either of these servers that resides on a public or semi-public (for example, DMZ) network, and the security of SMS. This solution does not accommodate older devices using PalmOS releases earlier than version 5.

The second possibility is to provide a Wireless Application Protocol (WAP) interface to the system, with the proxy becoming a WAP server. In this case legacy clients with browsers would establish a session with a WAP gateway (not pictured), which would in turn establish an SSL connection with the proxy/WAP server. The session between the client and the WAP gateway is protected by Wireless Transport Layer Security (WTLS), a lightweight version of SSL that cuts down on computational and bandwidth overhead, but does not compromise on security. Note that traffic is decrypted and reencrypted on the WAP gateway (this is called the “WAP gap”), so placement and control over this server is critical – if an attacker could gain control of this server, s/he could either passively monitor traffic or perform a “man-in-the-middle” attack to modify transmissions. Furthermore, the proxy/WAP server also decrypts and reencrypts, so it must also be protected. It is recommended that the proxy be placed either on the same machine or in the
same physical and logical location as the CM. This solution may be relatively expensive for the end user, as a WAP session must be maintained while messaging is in progress.

The third possibility is to use Virtual Private Network (VPN) software to encrypt transmissions between the client and the server (or the network the server resides on). Essentially, this establishes an encrypted tunnel that extends the domain of the private network that the server inhabits to include the client device. VPN software using the industry standard protocol IPSec is available for all PCs and PDAs, but not for legacy 2G phones. It is important to note in this case that if the VPN is terminated at a gateway device, as is most common, then this device should be carefully configured to allow only specific access to required servers in the enterprise network, because the clients that will be using it may be at greater risk of compromise than those protected by being physically inside corporate boundaries. Sprint supplies managed VPN solutions specifically designed for wired and wireless environments.

It is likely that at least two of these systems will be necessary to satisfy legacy users – for example, WAP for 2G phone users, and SMS because Instant Messaging is such a natural extension of the service that users will become confused if they are not compatible. Instant messaging and SMS should be seen as extensions of one another, not separate systems that don’t intercommunicate. This is the cornerstone of Sprint Enterprise Instant Messaging: providing interoperable communication between an unprecedented variety of devices, and allowing the user to communicate no matter where s/he is and how s/he is connected to the network.

Protecting Client Devices

As a culture, we are beginning to become more familiar with security procedures necessary for laptops: protect them from theft and unauthorized physical access by keeping them under supervision; protect the data by using BIOS or hard drive passwords and disk or file encryption; protect them from network attacks by installing personal firewall, IDS, and virus scanning software. But we are not used to thinking of mobile devices – PDAs and cell phones – as security risks.

For PDAs, the user has direct access to hardware – that is, there is no concept of separation of system, administrative, and user roles. Passwords are only marginally effective as access control, as are data encryption programs, since an attacker with physical access to the device (or its sync device) can install monitoring software and/or download memory contents. Furthermore, many devices respond to the insertion of external memory cards by running application(s) from the card. This is convenient for quick backup, but is also useful for an attacker who wants an image of the device in only a few seconds.

Still, installation of third party file encryption or firewall/IDS/Antivirus applications raise the security bar, so that physical access is required for compromise. If possible, these security applications should be centrally managed, so that security professionals are charged with setting parameters correctly. After this, protection of the device can be accomplished by good supervision by the user, which can in turn be improved with user awareness programs. Users need to be aware that sensitive data does reside on these devices, and that the information is easily stolen if someone obtains even brief (coffee-break) unsupervised access to the device.

Note that encryption software is more effective in mitigating damage in the stolen-device scenario: if an attacker steals the device, s/he does not have ongoing access to user interaction (for example, through an input monitoring program), so s/he may be forced into a head-on attack on the encryption. Such an attack may be facilitated by the strong tendency of mobile device users to employ short passwords, but many encryption programs add a data destruction feature for excessive incorrect login attempts. Thus, the attacker would have to find a stored version of the password hash and attack it offline. Encryption software should be evaluated against this scenario if the sensitivity of data stored on mobile devices is high, but in general software is available to protect and/or destroy information on stolen mobile devices.
Cell phones and PDAs are increasingly integrated, so for these devices the previous discussion applies. For Java-based phones, operating system and application vulnerabilities are not currently known. However, we may assume that the person in possession of the device will have essentially full access to the hardware, so that the same kind of physical protection of the devices will be necessary as for PDAs.

Note that there is nothing specific to Instant Messaging in this discussion – we already have these problems since users download email and transfer files to PDAs for use when away from the office. The above security measures should be adopted by any enterprise wishing to protect its information, and will become even more important as always-on devices become more prevalent and more vulnerable to network attacks.

Conclusion

We have presented here some options for tailoring the installation and security controls to specific customer needs – and this flexibility is a critical security component, since any new system must be configured to conform to the security profile of the owner. Furthermore, any new system will require modification, or at the very least review, of security procedures to ensure that adequate monitoring, maintenance, response capability, and awareness are provided.

Instant Messaging has become a de facto standard for collaboration, and an enterprise necessity for speeding up business, cutting costs, and improving customer relationships. But consumer Instant Messaging products suffer from a lack of interoperability and security. Sprint Enterprise Instant Messaging has been designed to meet this need by providing a secure, enterprise-class platform that is interoperable between traditional desktop clients and most mobile devices and networks. As we have seen, the system was designed with security in mind, with features such as message encryption and a server-based architecture. These features along with the ability to manage the system and control Instant Messaging traffic into and out of the enterprise make Sprint Enterprise Instant Messaging a welcome alternative to consumer-grade solutions.