jabber inc.®
The Power of Presence™

WHITE PAPER

Jabber, Inc. :: 1899 Wynkoop Street, Suite 600 :: Denver, CO 80202 :: P: 303.308.3231 :: F: 303.308.3219 :: E: info@jabber.com :: W: www.jabber.com

SIP, RTP, AND XMPP IN THE EMERGING REAL-TIME INTERNET

Complementary Technologies for Next-Generation Solutions

INTRODUCTION

Since large-scale commercialization of the Internet began in the early 1990s, its dominant applications have been the World Wide Web and email. Although they have been augmented by some dynamic additions, the Web and email are essentially static, pull, or asychronous technologies: a user requests a Web page or his or her mail, a Web or mail server provides the relevant file or files, and a browser or email client presents the results to the user. By contrast, the emerging breed of next-generation applications are driven by more dynamic, push, or real-time technologies: Voice over Internet Protocol (VoIP), video conferencing, collaborative editing, shared whiteboarding, multiplayer gaming, instant messaging (IM), chat, and just-in-time content feeds are good examples.

The first generation of massive Internet adoption was built using a now-familiar alphabet soup of protocols developed by the Internet Engineering Task Force (IETF) and the World Wide Web Consortium (W3C): Hypertext Transfer Protocol (HTTP), Hypertext Markup Language (HTML), Simple Mail Transfer Protocol (SMTP), Post Office Protocol 3 (POP3), Internet Message Access Protocol (IMAP), and Multi-Purpose Internet Mail Extensions (MIME) foremost among them. What are the protocols powering the emerging real-time Internet, what are their strengths and weaknesses, how do they work together, and how can enterprises and service providers choose between them? This paper provides answers to those questions, focusing on Session Initiation Protocol (SIP), Real-time Transport Protocol (RTP), and the Extensible Messaging and Presence Protocol (XMPP).

THE INTERNET: A NEW GENERATION

What is the real-time Internet, and why do people feel they need it? Aren't the Web and email good enough? Apparently not. Adoption of next generation Internet technologies such as VoIP and IM has been intense, especially among the younger cohort ("Generation Y") that is coming of age in the era of the commercial Internet. Although the Web and email won't disappear anytime soon (after all, television never supplanted radio), the new technologies will rapidly expand to fill niches that the Web and email cannot address.

The focus of these applications is already becoming clear, and it centers on relatively rapid, continuous, high-attention activities between individuals or groups (voice, video, application sharing, IM, chat, gaming), as well as timely delivery of content that individuals or applications want to learn about in as close to real time as possible (news, sports, weather, blog updates, market information, incident alerts, emergency notifications, and the like). Usage of such applications is commonly driven by a special kind of data called "presence" — dynamic information about an entity's availability for communications over the network (e.g., a person's availability to talk on the phone or chat over an IM network).

There are many dimensions to categorizing these real-time communication modes and information sources. Perhaps the most helpful addresses the nature of the data being exchanged: Is it relatively *structured,* such as Extensible Markup Language (XML) data; relatively *unstructured,* such as binary voice or video data; or a combination of both? As we will see, different technologies and protocols are appropriate for exchanging structured data as opposed to unstructured data.

In addition, all of the new real-time applications emerging on the Internet today are framed in terms of stream-based sessions. For example, a VoIP call or video conference is a multimedia stream between two or more individuals, multiplayer games consist of many-to-many data streams among the players, IM conversations and chatrooms are one-to-one or many-to-many text sessions or sub-sessions, and even real-time content notifications happen in a streaming fashion as a user or application receives a steady flow of data from various information sources.

Thus, the next-generation Internet mainly consists of a wide variety of real-time applications in which structured or unstructured data (or both) is exchanged in the context of sessions or streams that are negotiated (explicitly or implicitly) between two or more users or applications, all driven by availability information in the form of "presence."

Today's Internet	Emerging Internet
▪ Static	▪ Dynamic
▪ Pull	▪ Push
▪ Intermittent	▪ Real time
▪ Latency	▪ Immediacy
▪ Smaller, text-based data	▪ Rich media
▪ Information silos	▪ Rich context for communication
▪ Person-to-machine or Person-to-person interaction	▪ Group interaction

REAL-TIME TECHNOLOGIES: SIP, RTP, AND XMPP

The best-of-breed technologies and protocols for building the real-time Internet are still being sorted out. Currently, market requirements are being addressed by a wide range of proprietary technologies (e.g., many VoIP providers, consumer IM services, gaming communities, and enterprise solutions for functionality such as application sharing) as well as products built on a mix of protocols approved by the IETF and other standards-development organizations.

Although proprietary technologies can be quite successful in the short and even medium term, over time market and regulatory pressures have tended to push towards convergence on standards-based solutions, since large enterprises, service providers, and governments do not want to be locked in to proprietary technologies. The remainder of this paper focuses on the standards efforts occurring within the industry to address the needs of real-time communication and data exchange.

SIP is a "rendezvous technology" since it enables network endpoints to negotiate and manage data streams but does not handle the data itself. SIP actually consists of two parts: various email-like wrappers and headers provided by SIP itself, and the underlying session semantics provided by the Session Description Protocol (SDP). Most commonly, SIP is used for user-oriented multimedia sessions such as VoIP calls and video conferences. Because of its multimedia focus to this point, SIP is often perceived as a multimedia technology. However, in fact SIP is a rendezvous protocol, and SIP negotiation results in a "handoff" to a data streaming protocol such as RTP for multimedia or, potentially, XMPP for structured data exchange.

For multimedia data exchange, the IETF's RTP is rapidly supplanting older technologies such as the International Telecommunications Union's (ITU) H.323 standard. Although less well-known than SIP, RTP is the actual data streaming protocol used to perform the "heavy lifting" in most modern, standards-based VoIP, video conferencing, and multimedia applications. The data of a VoIP call started by SIP is typically transferred using RTP. In particular, RTP is optimized for the transfer of large amounts of binary or unstructured data between two or more users or applications, typically where quality of service is not of utmost importance (e.g., a VoIP call in which it is acceptable for packets to be dropped, since both software and the human ear can compensate for small amounts of data loss).

In the realm of real-time exchange of structured data, the IETF's XMPP has been adopted and deployed by numerous government agencies, enterprises, and service providers worldwide. XMPP, which grew out of the XML streaming technology originally built by the Jabber open-source community, is a robust technology for transporting XML data between any users or applications in real time. Existing applications built on XMPP include IM and chat, workflow, financial trading systems, emergency incident handling, network management, remote instrument monitoring, gaming, content syndication, and generic publish-and-subscribe systems. Because XMPP is a pure XML streaming technology, it is ideal for transporting a wide variety of XML data, such as Remote Procedure Call (RPC), Simple Object Access Protocol (SOAP), Really Simple Syndication (RSS), Atom syndication data, Financial Information Exchange Markup Language (FIXML), and domain-specific formats defined by schemas developed in various industry consortia.

STRENGTHS AND WEAKNESSES

As is true of any protocol or technology, SIP, RTP, and XMPP were each developed to address a specific class of requirements. They are optimized for those requirements and can be extended to address a different class of problems only with some inevitable stretching and misalignment. Each technology's greatest strength is also its greatest weakness when seen from another perspective, and it is helpful to look at each technology from several angles.

As mentioned, SIP is a rendezvous and session management protocol. Because SIP translates traditional call-setup and call-management functionality into the modern world of Internet communications, it has been widely adopted as the basis for voice and video services over the Internet. SIP is also being extended to handle some basic forms of messaging and presence via a protocol called SIP for Instant Messaging and Presence Leveraging Extensions (SIMPLE). However, because SIP was designed for session negotiation, the SIMPLE extensions under development are not optimized for the advanced exchange of structured data and to this point provide only a subset of messaging and presence functionality. For these reasons, SIMPLE will certainly interoperate with XMPP at a basic level but which is unlikely to ever offer the advanced data exchange capabilities which come standard with XMPP technologies.

RTP stands for Real-time Transport Protocol, however, it is optimized only for certain types of real-time transport, such as computationally intensive, loss-tolerant binary data streams. For example, transforming SOAP requests or RSS feeds into binary XML and then transferring the resulting data over RTP, might cause critical information to be lost. But for voice calls or video conferences, RTP is an excellent transport medium.

Similarly, XMPP is at its core an XML streaming technology, which is optimized for sending relatively small chunks of structured data between network endpoints. Structuring or encoding voice or video data in XML and sending it over XMPP would make little sense, since RTP provides a perfectly good transport for large amounts of binary data. However, XMPP excels at transporting XML data, from chat messages and RSS/Atom notifications to SOAP and RPC payloads, as well as specialized data formats for custom applications.

The key question for media-heavy applications is: how to get to RTP? SIP is one such technology. Another method is *Jingle*, a native XMPP extension for setting up peer-to-peer multimedia sessions between

XMPP entities for voice chat, video chat, and various types of media sharing. Jingle—a joint effort between Jabber, Inc. and the Google Talk team—is being standardized within the Jabber Software Foundation's (JSF) Jabber Enhancement Protocol (JEP) series, and has broad support from a wide range of XMPP vendors and developers. The intent of Jingle is not to replace SIP, but to provide a pure-XML, XMPP-friendly mechanism for negotiating and managing multimedia interactions between Jabber/XMPP users. Because Jingle negotiations often result in the use of RTP for media transport, XMPP entities that use Jingle can also communicate with SIP entities, typically through gateways that translate Jingle's XML syntax into SIP's plaintext syntax (since both Jingle and SIP are well-defined protocols, the syntax translation is relatively straightforward). Both entities then use RTP for the media exchange, resulting in interoperability across networks.

At the same time, Jingle provides benefits to those who have already implemented and deployed XMPP technologies. For one, Jingle re-uses existing XMPP identifiers rather than forcing the use of an XMPP ID and a SIP ID; since XMPP IDs are fully Unicode compliant, this is a strong selling point for service providers outside the English-speaking countries. Second, XMPP technologies in general have a stronger security profile than SIP technologies do; for instance, address spoofing is much more difficult in XMPP than in SIP and authentication is mandatory in XMPP but optional in SIP (the SIP community is working on resolving these issues). Third, Jingle provides pure XML formats, thus easing the burden of auditing and compliance through schema checking of the XMPP streams. Finally, Jingle's modular design makes it possible to interface with a number of popular technologies that are not based on SIP and RTP, such as Asterisk, a popular private branch exchange (PBX) project.

In general, interworking between XMPP and SIP is handled through gateways or, in more advanced implementations, through specialized connection managers for a given protocol. The latter approach enables, for example, a SIP/SIMPLE user agent to connect to a communications server that for all intents and purposes appears to be a SIP/SIMPLE proxy. Similarly, a Jabber/XMPP client could connect to the same server on a different port and use XMPP's streaming XML technology to communicate. This "smart server" would internally translate between SIMPLE and XMPP for messaging and presence, and between SIP and Jingle for multimedia session negotiation, but leave peer-to-peer media in its native RTP format. The result is seamless communication between XMPP entities and SIP entities for presence, messaging, and media interactions. (For instance, this is the approach taken in the Jabber Extensible Communications Platform (Jabber XCP) from Jabber, Inc.)

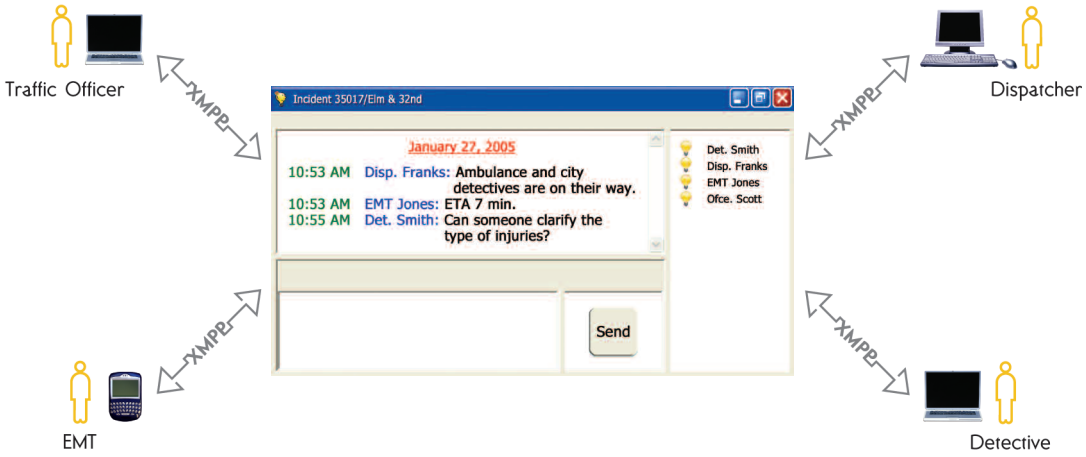
WHAT TO IMPLEMENT AND DEPLOY?

SCENARIO: RTP FOR MULTI-PARTY INTERNET VOICE CONFERENCING

A customer and supplier want to schedule a voice chat over the Internet to address a logistics problem. The supplier has a SIP-based infrastructure and the customer has an XMPP-based infrastructure. RTP is the best solution, negotiated and managed via SIP or Jingle.

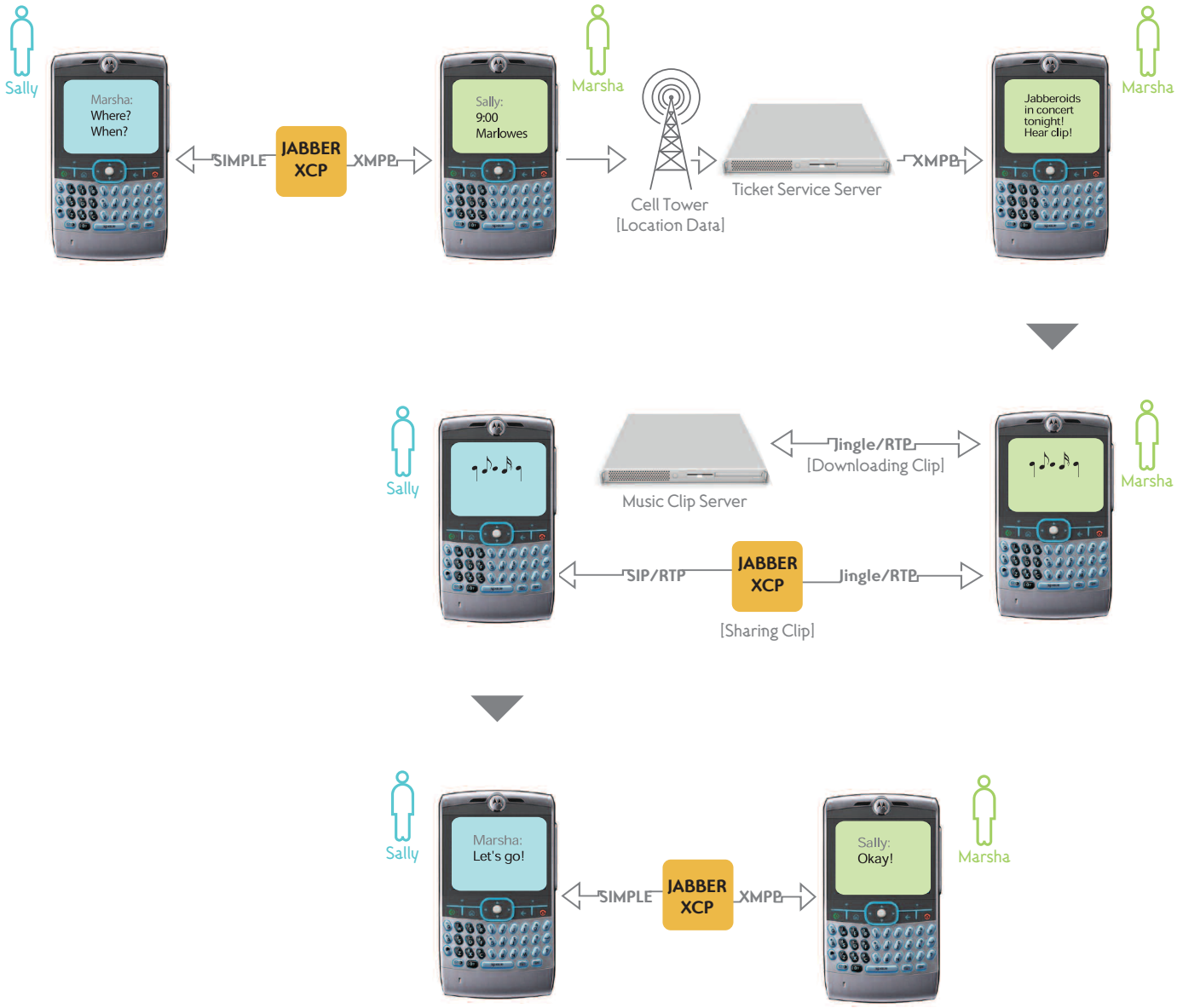
SCENARIO: XMPP FOR MULTI-PARTY TEXT CONFERENCING/GROUP CHAT

Emergency personnel from several government jurisdictions need to gather together in a virtual room in order to resolve an incident. Real-time incident data updates are piped directly into the room so that all participants share the same contextual information, all of the conversations are logged for post-incident analysis. XMPP-based text conferencing/group chat rooms are probably the best solution.



SCENARIO: XMPP, SIP, AND RTP FOR PERSON-TO-PERSON AND APPLICATION-TO-PERSON EXCHANGE

A young woman stops at a cafe in the city center before meeting a friend on a Friday. Using the XMPP client on her mobile phone, she chats with her friend Sally (who has a SIP client on her phone). Based on her personal profile and location, an XMPP-enabled content server sends her a time-sensitive offer of pre-sale tickets to a concert at a nearby music club. She quickly chats with her friend while leveraging Jingle/RTP to download some music clips of the band. After sharing one of the music clips, her friend agrees that the band is pretty good, so she reserves two tickets for the show.



As these brief scenarios illustrate, different protocols solve different problems. Which technology to implement and deploy in a particular application depends on many factors, including the user or application requirements, the deployed base of existing hardware and software, and the need for integration with back-end data stores and third-party applications. Because of this diverse landscape, no one technology or protocol will provide all of the answers within any given organization or service provider.

But that does not mean interoperability will suffer. It is becoming clearer how some of the leading technologies—such as RTP, SIP/SIMPLE, and XMPP/Jingle—will co-exist and complement each other. Already, industry leaders such as Google, Apple, and Jabber, Inc. are providing products and services that take an agnostic approach by using the best technology for the job at hand. Exactly how all of today's protocols and technologies will work together in the future, no doubt supplemented by new entrants into the mix, is something that will be determined over time as more and more organizations and service providers deploy real-time services to their stakeholders and customers. But it is clear that those technologies will interoperate and complement each other in productive and profitable ways.

JABBER XCP: THE WAY FORWARD

What's the best way to move ahead in deploying sophisticated real-time communication offerings that combine both multimedia and structured data capabilities? More and more organizations are turning to Jabber XCP, which uses the power of presence to provide a rich feature set "out of the box" as well as flexibility that makes it easy to develop unique functionality that meets each organization's distinct requirements. Jabber XCP speaks the most important protocols for real-time communications, through both native support and Web services Application Program Interfaces (APIs), including support for Instant Messaging and Presence Service (IMPS) and SIP/SIMPLE. Furthermore, the technology experts at Jabber, Inc. continue to build out that support, with an ongoing SIP project that will enable truly seamless communications across today's heterogeneous environments, including a full SIP presence server and a Jingle-to-SIP protocol translator for voice and video calls between users of SIP and XMPP networks.

Future directions for Jabber XCP include further integration with existing real-time communications technologies, such as Web conferencing, application sharing, voice, and video; more advanced structured data handling for workflow applications and electronic commerce; intelligent and personalized content distribution services; and sophisticated "extended presence" capabilities, including location-based information offerings. Jabber, Inc. is a market leader that has proven its commitment to building the interoperability necessary for a wide range of presence-powered applications, and will continue to innovate as requirements evolve in the fast-changing world of real-time communications.

CONCLUSION

The powerful promise of the real-time Internet is already being realized by those who are implementing and deploying advanced, standards-based communication and presence solutions today. Those who wait for a single unifying technology to magically appear may find that their market opportunities have evaporated. Those who implement now will enjoy competitive and organizational advantage.

Thankfully, open standards are providing enterprises, service providers, original equipment manufacturers (OEMs), and government agencies with the freedom to innovate. SIP, RTP, and XMPP can work together, with RTP providing a transport for multimedia information and XMPP providing an advanced transport for a wide variety of structured data exchanges. And increasingly, leading products such as Jabber XCP are providing a comprehensive real-time communications platform by supporting all of these technologies.

GLOSSARY

Extensible Messaging and Presence Protocol (XMPP)—An XML streaming protocol used for IM and other real-time exchange of structured data. Developed originally by the Jabber open-source community, XMPP has been formalized by the IETF in Request for Comments (RFC) 3920 and RFC 3921. In addition, XMPP extensions continue to be defined in the JSF.

Jingle—A set of XMPP extensions for negotiating and managing multimedia sessions between Jabber/XMPP entities, resulting in a "handoff" to dedicated streaming protocols such as Real-time Transport Protocol (RTP) and Inter-Asterisk eXchange (IAX). Jingle is a JSF technology built on top of XMPP.

Real-time Transport Protocol (RTP)—A binary data streaming protocol mainly used for providing voice and video services over the Internet. RTP is an IETF technology defined in RFC 1889.

Session Description Protocol (SDP)—A data format, defined by the IETF in RFC 2327, that is used to specify information about a session for purposes of session negotiation and management. An XMPP transport for SDP is defined in JEP 0111: A Transport for Initiating and Negotiating Sessions (TINS).

Session Initiation Protocol (SIP)—A rendezvous protocol used to negotiate and manage data streams, resulting in a "handoff" to a dedicated streaming protocol such as RTP. SIP is an IETF technology mainly defined in RFC 3261.

SIP for Instant Messaging and Presence Leveraging Extensions (SIMPLE)—A set of SIP extensions currently under development to send IM and presence information over the SIP rendezvous channel.