Network Architecture: Session Internet Protocol
Will SIP Make Ubiquitous Computing a Reality?

by **Jay Gitomer**

In this article...

In the never-ending search for the elusive Killer App, the latest direction is towards convergence. Convergence in IT-speak is the merging of multiple technologies onto a single device, or through a single interface. To make this happen, certain protocols, or formats, have to be agreed upon to ensure that all these technologies can communicate with each other. Without protocols, data such as web pages and email could only be read by computers that used the exact same software (and sometimes hardware) as that used by the sender.

Session Internet Protocol, or SIP, is a signaling protocol for use on the Internet. It enables conferencing, telephony, presence, events notification and instant messaging. It can be used to move application data across networks, including wireless networks, and it can interact with next-generation wireless networks as a component of payment and security systems. There are some interesting ways that SIP might be used in the near future that have the potential to change the way we use the Internet for business, education and leisure. But before going into the blue skies, a little grounding in the basics of SIP will be necessary.

When data is transmitted over the Internet, it has to pass through gateways. A gateway is a network point that acts as an entrance to another network. To initiate a connection, two gateways have to negotiate an exchange. In the case of multimedia files, a multimedia packet (including voice) has to be negotiated.

SIP is a lightweight packet-voice call set-up protocol, which means that it can make a connection between network endpoints without passing a lot of information back and forth. Before SIP, network administrators relied on the standard that is supported by the International Telecommunication Union, the H.323 standard. The H. 323 standard was developed primarily to support desktop video conferencing and whiteboarding applications, which were the lead contenders for killer app status at the time. As a result, H.323 requires many messages to travel between two gateways before a communications link can be established, so there's usually a several-second delay in set-up time and sometimes even ongoing latency throughout the session. H.323-based products haven't proved able to interoperate as anticipated, which means users have to find ISPs and regional carriers that use the same IP telephony gateways in order to transmit data.

SIP, on the other hand, was designed primarily to serve voice-over packets. It only requires the exchange of, at most, six messages to establish a communications link, so set-up time is reduced dramatically. In plain English, SIP makes it possible for computer users to easily and instantly communicate over the Internet, better than they can now. SIP will make it easier to do things like:

- View a concert online
- Get instant messages on your handheld
- Attend a business conference online with attendees from all over the globe
- Run a call center based entirely on voice-over IP
- Use a unified messaging system, getting email, voice and fax messages in a single box accessible from any web browser, including mobile browsers

But all of these things can be done today, to some extent, so what's the real point of SIP? Is it of interest to anyone other than network administrators, and will it still be around next year?

SIP Today: Conferencing, Telephony, IM -- So What?

Some of the current uses of SIP are bound to continue to be of interest. Conferencing, for instance, has the potential to make it easier for businesses to communicate with field offices or global customers. People from anywhere can see, hear and talk with people in different locations, and can share documents through software that creates a virtual whiteboard. Conferences can be held using combinations of audio, video, multimedia and screen-sharing.

Virtual conferences are inexpensive to run, and they can be held in real-time from desktop computers. In addition to being inexpensive in themselves, virtual conferences also provide a business benefit by reducing travel expenses and, in the case of audio conferences, eliminating long distance fees.

Telephony is and will continue to be an important aspect of SIP usage. Telephony is anything shared between computers that would have traditionally been delivered through a telephone, such as voice communications and faxing. Currently, the audio quality of calls placed over the Internet is of low quality, and while there are no long distance charges for Internet calls, the rapidly diminishing cost of traditional long distance telephone tolls makes the value proposition of SIP-based long distance pretty light. When you can call your friend across the country for four cents a minute through a landline, or as part of a monthly, all-the-long-distance-you-can-use mobile phone subscription, why hassle with an Internet call and the attendant fuzzy sound and headset/mike hassles?

Of course, there's always chat. SIP is great for chat, and as long as there are teenagers and computers, there will be a desire for chat technology. Chat hasn't found a strong business niche yet, although some businesses use it to provide instant support to customers and or as a B2B marketing tool.

The three services that SIP is most likely to enable first will be the ability to make a normal voice call that will ring at the location of the person called, wherever they may be, at the cost of a local call, the ability to send faxes at local call prices through the Internet, although only in major cities, and the ability to send voice messages along with text email.

These services are just the beginning, though. The real value of SIP is in the development of the "presence" business. Presence describes how we'll be using the Internet in the upcoming years. This is a relatively new concept, so the ideas being thrown around now are vaporware. But the potential of this concept is so powerful, and the technology is close enough, that it seems likely that

the next stage of computing will be the realization of true anywhere, anytime computing.

SIP Tomorrow: Intelligent Connections

Anyone who's used an instant messaging system like AOL Instant Messenger knows that the interface shows which of your associates are logged onto the IM system at the moment. Presence will be able to do that, and also will be able to find users at their home, office, or other location.

Presence knows when a user is ready to accept communications or on another call. This means no more busy signals or rounds of phone tag. In addition to knowing if someone else is online, it can tell which medium that user prefers - voice, IM, email or video - and transmit the message in that medium.

Subscribers on a network would have the ability to see if another subscriber was on a call, and request permission to join in that call. If permission were granted, the two-person call would automatically become a conference call between three. A subscriber could even see if a number of other people were online, and if they were, a conference call would be immediately initiated among them.

Presence may include the identity of the calling party, and can respond to different users in different ways. Your boss's call might be put through immediately, but that chatty cousin might get an 'in a meeting' response. Or vice versa, of course.

SIP can be used to communicate the state of mobile phones within a subscriber network. For example, if one subscriber sends a message to another who happens to have his phone off, the first sender could receive a notification when the recipient powered on. Or the sender might have his phone automatically re-ring the call when the recipient came on the network. In theory, the status of the mobile phone could be used to see the location of subscribers on the network, possibly by identifying the nearest mobile phone base station, or in the future, possibly by use of GPS coordinates. There are some obvious benefits to a usage like this, but it's also easy to imagine how it could be abused. Interesting privacy issues are certain to surround the early years of location tracking.

Even today, with mobile phone quality and reliability far less than perfect, a growing number of people are using their mobile phones as their primary number, and only keeping landlines in their homes to connect computers to the Internet. There is much convenience to having a single number, a single bill, free long distance and electronic voice-mail, along with other cost-added features like call forwarding or conference calling. If the trend continues along these lines, it would seem likely that the next step will be to have a single number for work and home. This would be of particular interest to self-employed individuals, who may prefer to pay a single bill or maintain a single mobile phone contract.

While SIP can make the shedding of extra phones a feasible proposition because it offers location-based screening. Location-based screening lets subscribers to a network select which callers can reach them during certain hours, or when they are at certain locations. A side benefit to being able to have a single number serve both business and personal needs is that it will ease the strain of having to come up with more area codes and phone numbers every few years. On the other hand, a scheme like this would require telephone service providers to completely rehaul their billing systems, since users would certainly demand some way to sort business from personal calls in order to submit expenses to an employer or itemize the phone bill at tax time.

Many people who've spent time around field technicians or anyone else with a Nextel phone have experienced intercom telephony. A call placed through intercom telephony first buzzes the recipient to see if they're willing to accept communications, and then puts the call through. It has the effect of

seeming to provide an open connection between parties because it allows communication to be instantaneous. It's like having a conversation with someone in the next room; you can't see them, but you can talk to them with perfect clarity and no lag. In the future, when video displays are common on mobile phones, it will be like having a conversation with someone in the same room.

Who's Using SIP?

Despite the market downturn, commercial availability of SIP-based VoIP (Voice-over Internet Protocol) became a reality in the first half of 2001, and the two major US carriers who have rolled it out, Level 3 Communications and WorldCom, have announced plans to aggressively expand their reach in upcoming months. Other providers are also beginning to deploy SIP infrastructure and SIP-based applications. Microsoft's new operating system, XP, will ship with a SIP User Agent, and the company's next generation of instant messaging clients will also be based on SIP. This will put SIP on the vast majority of Internet users' computers within the next few years.

A big target for SIP products is the young IP Centrex market. IP Centrex is a new way to handle traditionally costly and unwieldy Centrex services. Centrex is a service that lets businesses lease phone facilities from their provider's central office instead of having to maintain their own. IP Centrex is the same thing, only for VoIP services. It is gaining attention as more potential customers have access to converged broadband IP connections that can carry both data and voice traffic, like DSL. It enables all voice and data services to work over a single connection (as opposed to multiple loops) from a provider's central office into an enterprise, and allows telephony and data to be combined to provide new services. The potential growth of converged broadband access has spurred IP Centrex vendors to bring carrier-grade products to market quickly. Convergence eliminates complexity for end-users, and carriers hope it will change the way their customers do business by simplifying communications.

But most recent changes in the telecom industry have shifted away from innovative upstarts and back toward traditional Regional Bell Operating Companies (RBOCs). With the disappearance of so many small next-generation carriers, there are fewer early adopters likely to commit to SIP infrastructure build-outs. The build-out will occur eventually; there are too many other factors in its favor for temporary market conditions to prevent the adoption of SIP.

Using SIP to implement presence reduces overall technical complexity for service providers by offering a single architecture, a single protocol dealing with a single stack, and a single persistent data store containing subscription and subscriber information.

It reduces costs because it reduces the need for hardware because one system can run call set-up. Overhead on the presence system is minimized by having a single integrated system and possibly a single integrated management console. Staffing costs are also reduced because just one skill set is needed in each technical role, rather than having to hire multiple people or expensive people with multiple skills.

Will SIP Trip?

If SIP can do half of the things it is intended to do and do them affordably, there is no doubt that consumer interest in SIP-based products and services will be strong. The caveat is that users will have to be on the same subscriber network in order to take advantage of most of the features that SIP will make possible. If the market for ubiquitous computing follows the model of the mobile phone market, which seems the likeliest model at this time, there will be two or three large networks, which will make things frustrating for the first years. An obvious next step would be for the presence providers to make their systems interoperable.One can imagine how many years that would take, but without true interoperability, SIP won't reach its full potential. In addition to tech hurdles, these providers will have business challenges to face. For instance, a practicable billing scheme will have

to be developed, and each provider's own billing systems will have to be rehauled to handle the new services.

There are many other obstacles on the path to SIP acceptance. Rival standards might emerge, and end users might resist the innovation. Neither of those things seems likely to develop into a major stumbling block, though. There is the danger, however, that SIP won't be used to its potential, but instead will be co-opted to mimic existing IP services - applying the new technology to perform old-fashioned tasks, while losing sight of the fundamental benefits that SIP has to offer. SIP must be accepted not simply as a connection protocol, but as an entirely new way of thinking about and using the Internet, if it is to be drive the Internet to the next level - ubiquitous computing.

Web Links for More Information:

The SIP Center: http://www.sipcenter.com/
SIP Forum: http://www.sipforum.org/

About the Author

Jay Gitomer is an Internet strategy consultant. Formerly, she was the manager of corporate web development of USinternetworking. Prior to that she was on the marketing staff of Silicon Graphics, where she was responsible for creating an online presence for the Silicon Graphics Federal Marketing Division. Ms. Gitomer has a B.A. from Sarah Lawrence College and an M.A. from New York University, and is a regular contributor to Faulkner Information Services.

[back to convergence quarterly main page]
[