White Paper

SIP and the new network communications model

NORTEL NETWORKS
BUSINESS WITHOUT BOUNDARIES

About this document

Introduction

The purpose of this document is to provide an overview of SIP—the Session Initiation Protocol—and explain how SIP is reshaping the communications landscape. SIP enables an Internet-based architecture used to manage communication "sessions" over IP networks, enabling converged voice and multimedia services while promoting natural communications between people, not devices. This document will also review the significant business implications of this new protocol.

Contents

What is SIP?

Basic introduction

The Session Initiation Protocol forms the basis of an Internet-centric multimedia communications architecture. SIP establishes sessions over IP networks for people wishing to communicate. In this context, a session is any interactive communication that takes place between two or more entities over an IP network, from a simple two-way telephone call or an instant message exchange, to a collaborative multimedia conference session. SIP is a Web paradigm protocol standardized by the Internet Engineering Task Force (IETF). It's similar to the two major Internet protocols—HTTP (World Wide Web) and SMTP (e-mail)—in that it uses symbolic addresses to represent people who wish to communicate.

SIP enables converged voice and multimedia services such as voice-enriched eCommerce, Web page click-to-dial, instant messaging with buddy lists, and much more. SIP session management is the key to enabling IP-based natural communications between people, not devices. By using SIP, users may locate and contact one another—regardless of media content or number of participants—using disparate computers, phones, televisions, and hand-held devices.

SIP was developed to serve as a mechanism to establish a wide variety of sessions. Therefore SIP does not dictate the details within a session but instead negotiates interaction based on the capabilities of participants. This simplicity means that SIP is scalable, extensible, and fits comfortably into different architectures and deployment scenarios.

History

SIP emerged in the mid-1990s from research conducted at Columbia University in an effort to standardize a method for inviting participants from other universities to large-scale multimedia conferences. As it was developed, it became apparent that the protocol was much more flexible than anyone expected. The IETF—the body responsible for administering and developing the mechanisms that comprise the Internet—adopted SIP as the standard protocol for establishing and terminating multimedia sessions in 1999.

Capabilities

SIP is a control protocol that initiates, modifies, and terminates communication sessions with one or more participants. The protocol enables participants to agree on a set of compatible media types and supports user mobility by proxying and redirecting requests to any user's current location. SIP enables the following functions:

- **Name translation and user location—**Allows people to find each other without knowing the details of each others' device addresses or physical locations.
- **Media negotiation—**Handles negotiations that enable all participants in a session to agree on common media and the technology details involved—including voice, video, audio, instant messaging, applications data exchange, or any combination thereof.
- **Session participant management—**Manages the adding, dropping, or transferring of participants.
- **Session feature changes—**Allows for changing the media used in a session while the session is in progress.

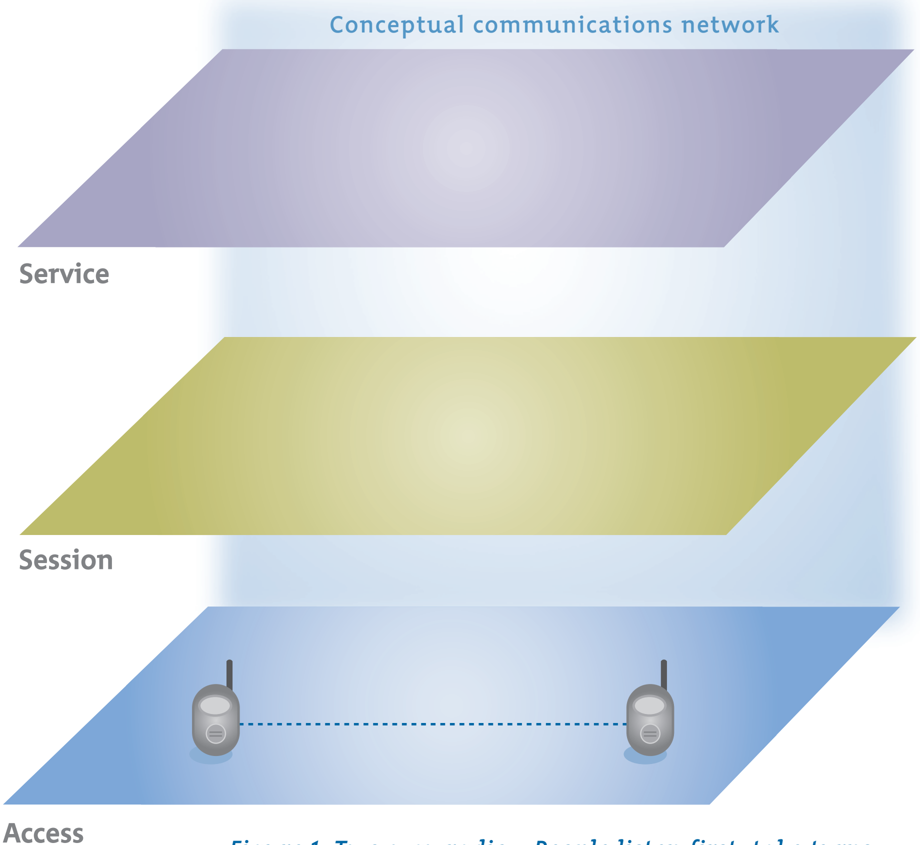

Figure 1. Two-way radio—People listen first, take turns talking, often signaling breakpoints verbally with "over"

The new network model

As communications have evolved, there has been a fundamental change in the way information is distributed. To truly appreciate the communication implications of the SIP protocol, it helps to consider its functionality within the context of network architecture. A useful conceptual model divides communications networks into a set of three functional "planes": the bottom access layer, the middle session layer, and the top service layer. Traditional communication methods fit within these layers as follows:

- **Two-way radio—**An early form of electronic communications, basic two-way radio exists solely on the bottom access plane. Radios broadcast on specific frequencies, and anyone with a receiver tuned to the proper frequency can listen. Users take turns transmitting to avoid "stepping" on each other's messages. There is no formal session control.

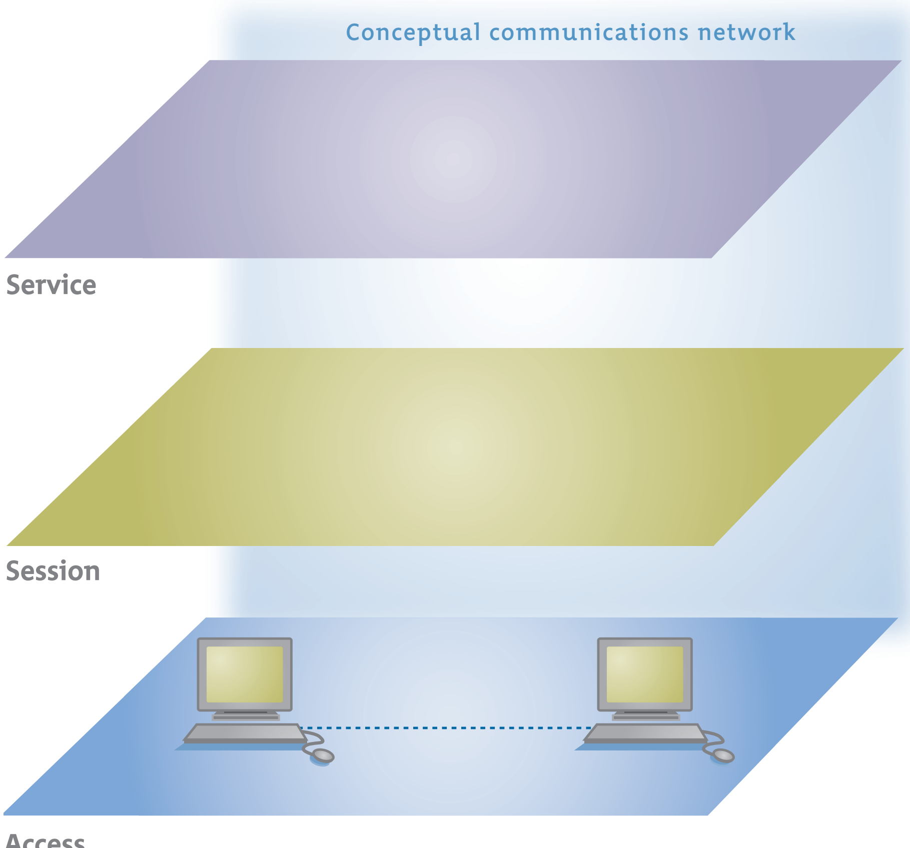

Figure 2. Simple Local Area Network (LAN)—Devices listen first, start talking, detect collisions and try again

- **Simple Local Area Network (LAN)—**
A simple Ethernet LAN between two computers has only rudimentary peer-to-peer session management. Each computer listens all the time and talks when it wants to, as long as it doesn't hear anyone else. It constantly watches for data "collisions" in case someone else starts talking at the same instant. If a collision is detected, the colliding parties both wait some random amount of time and try to send the data again, hoping the LAN will be available.

- **Wired telephony—**Traditional telephony is highly dependant on hardware that spans the access and session planes, controlling all sessions. Telephone switches—the modern replacement for humans with patch cords—employ sophisticated digital logic to establish, bill for, and release circuits between devices. Additional architecture within the session layer enables the reliable, global exchange of session information. Sessions may connect to the service plane for telephony services such as voice mail.

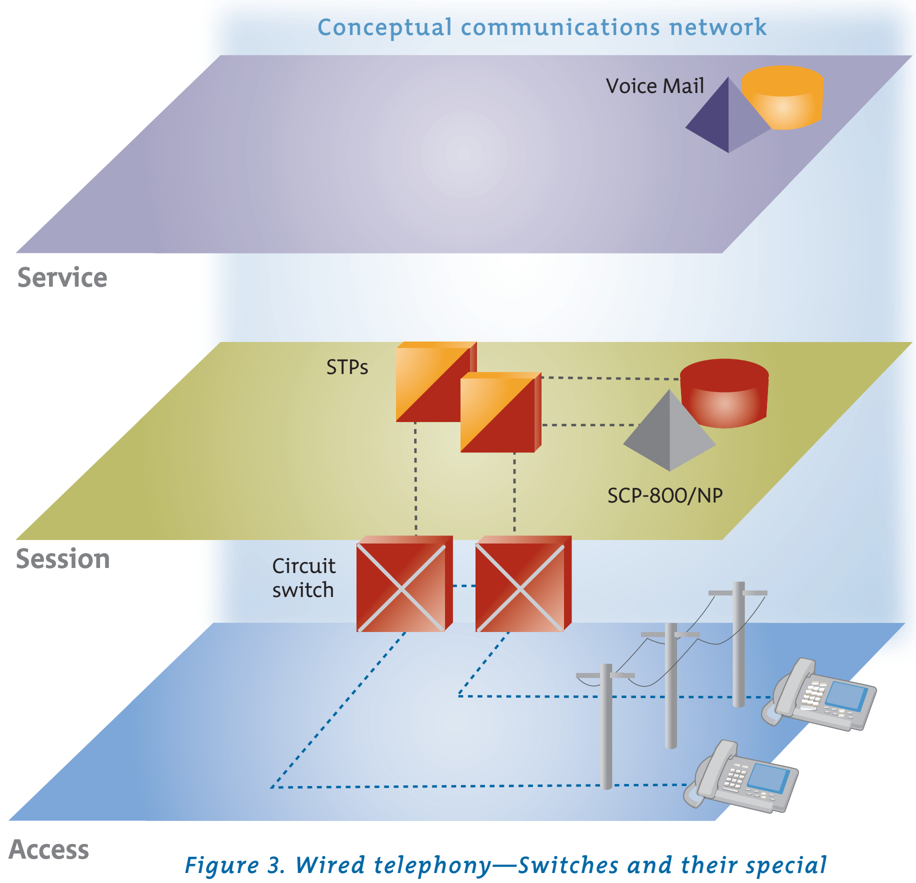

Figure 3. Wired telephony—Switches and their special signaling network control everything

- **Wireless telephony—**Modern wireless telephony is based on the concept of defined "cells" of radio coverage that are all interconnected by a sophisticated base station control network supervised by logic in a Mobile Switching Center (MSC). An MSC is a superset of a telephony switch and spans the access and session planes of the networking model in the same way. Its many specific session control functions are necessitated by the complications introduced by mobility. Various databases of real-time and customer information are maintained by session logic.

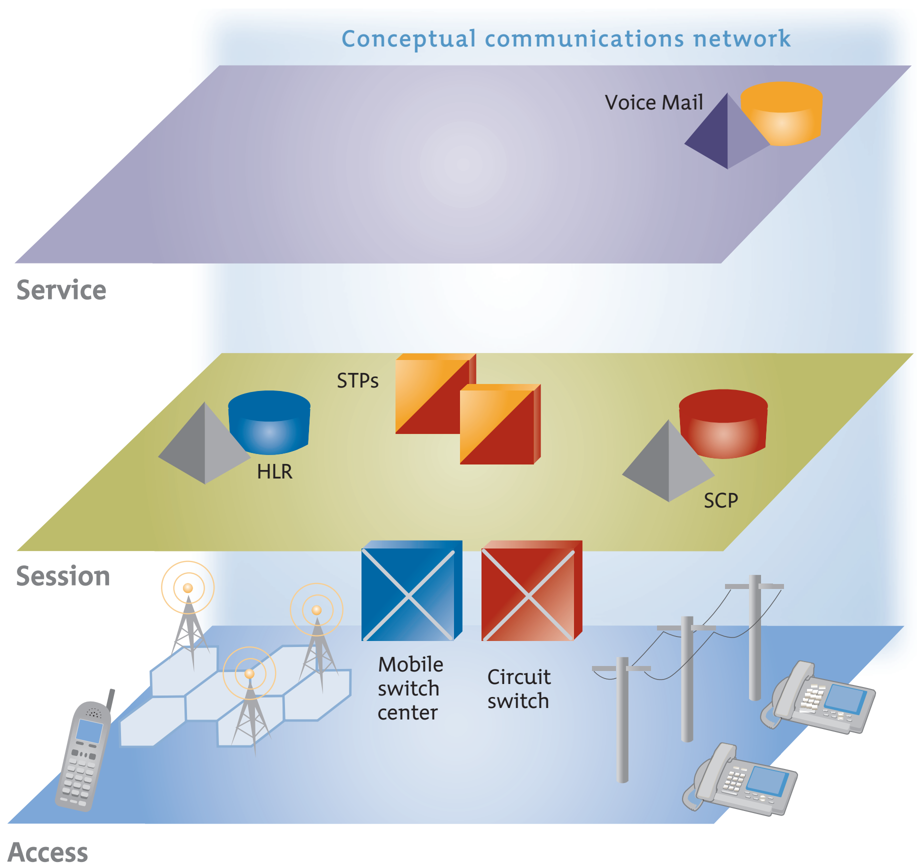

Figure 4. Wireless telephony—Additional network elements deal with the complications of mobility

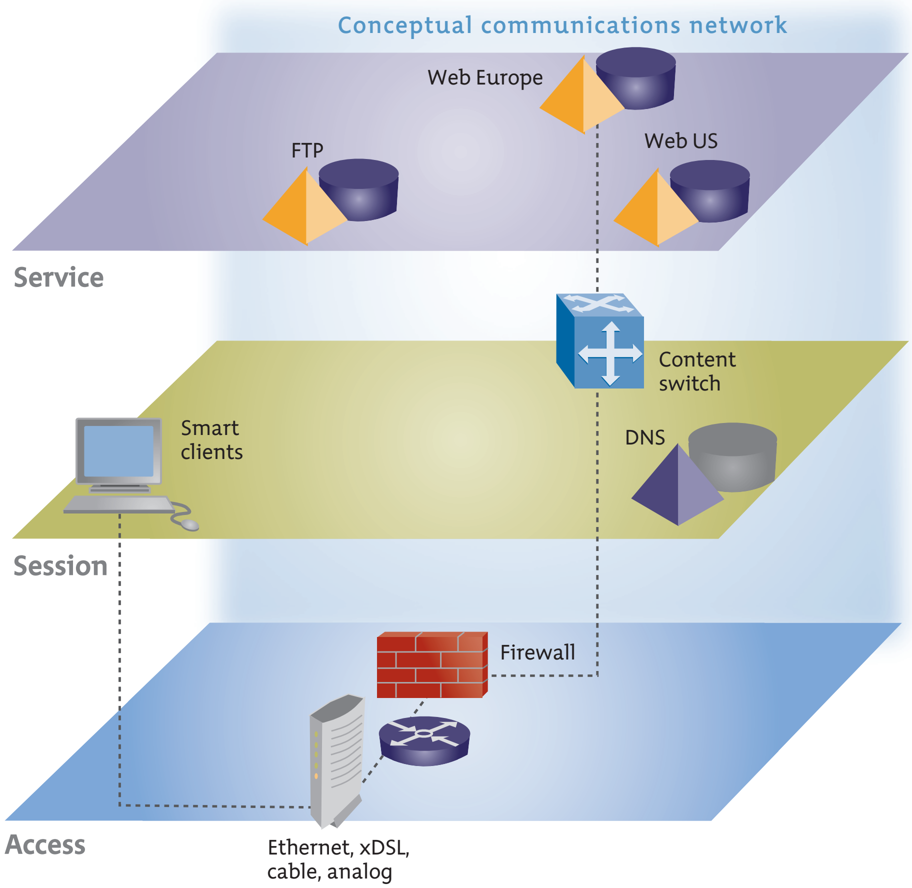

Figure 5. The Internet—Session control is distributed; smart clients play a significant role

- **The Internet—**In the early days of the Internet, each computer did all the session management and no centralized session control was involved. A computer essentially negotiated a session with a specific server by address as needed. As more people started using the Internet and the address space became significantly complicated, an address resolution service was implemented. This Domain Name Service (DNS) resolves names that are meaningful to humans into actual IP addresses on demand and resides in the session layer with the client computer. The most recent session-related element to enter the Internet picture is the Content Sensitive Switch, which resides in the session or service layer and directs traffic to various similar servers based primarily on the real-time availability of servers, their proximity to clients, and support for things such as the appropriate language for the person requesting information.

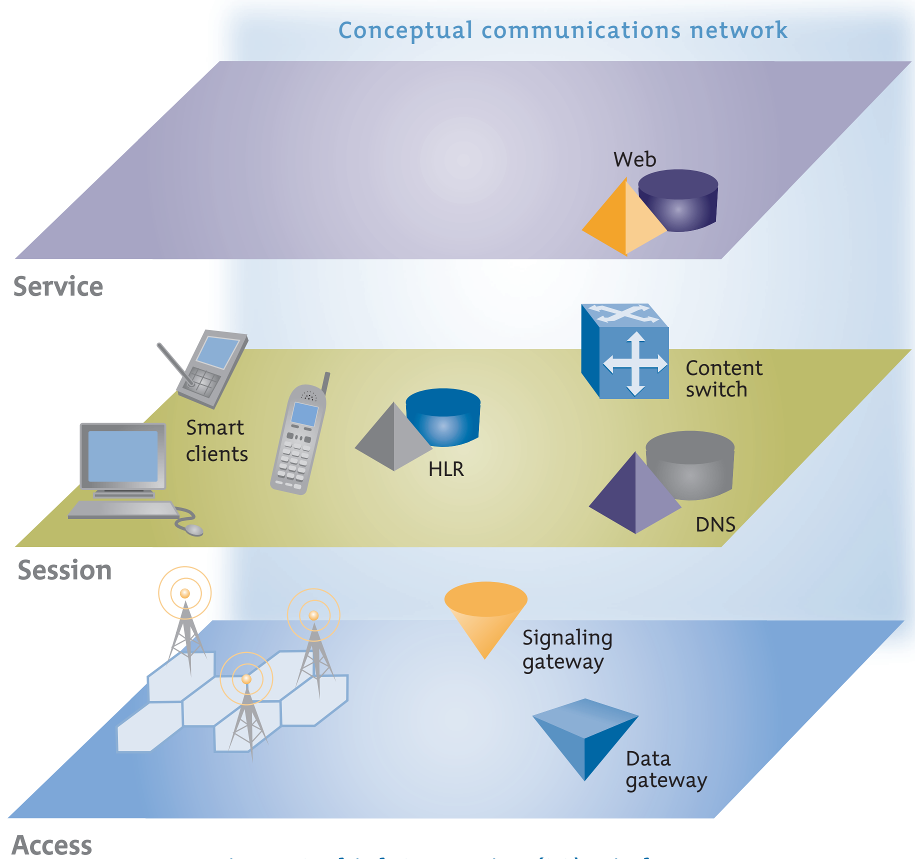

Figure 6. Third-Generation (3G) Wireless Internet—Extending mobility infrastructure to support IP networks

- **Third-Generation (3G) Wireless Internet—**In order to provide wireless high-speed data access to the Internet, signaling gateways and data gateways were introduced into the wireless infrastructure, reducing the reliance on mobile switching functions in establishing data connectivity. The 3GPP wireless standards body has adopted SIP, and once it is completely implemented, 3G mobile networks will be fully capable of participating transparently in multimedia services.

- **Distributed multimedia network—** All the pieces come together as optical and access bandwidth continues to expand and become more economical, enabling a dramatic increase in the media content available. Both physically attached and mobile smart clients are actively involved in session management, while proxy servers and content switches help establish sessions dynamically and redirect sessions to the most available appropriate content sources. Signaling gateways, softswitches, and media gateways of many types work together to provide interaction with other network technologies.

It is apparent that the communications dynamic is evolving as the interaction between network layers increases and session management becomes highly distributed. Just as the computer has always played a significant role in session initiation over the Internet, edge devices will play a significant role in session initiation and control in next-generation networks. Session management can reside completely with smart clients or may be distributed among a few or many cooperating network elements, including smart clients.

By storing information mapping device addresses on a SIP registrar, communications can be addressed to a person's name instead of a complex number scheme. A person will simply register one or more devices with the network and become reachable, wherever he or she may be, independent of the details of the networks and devices involved.

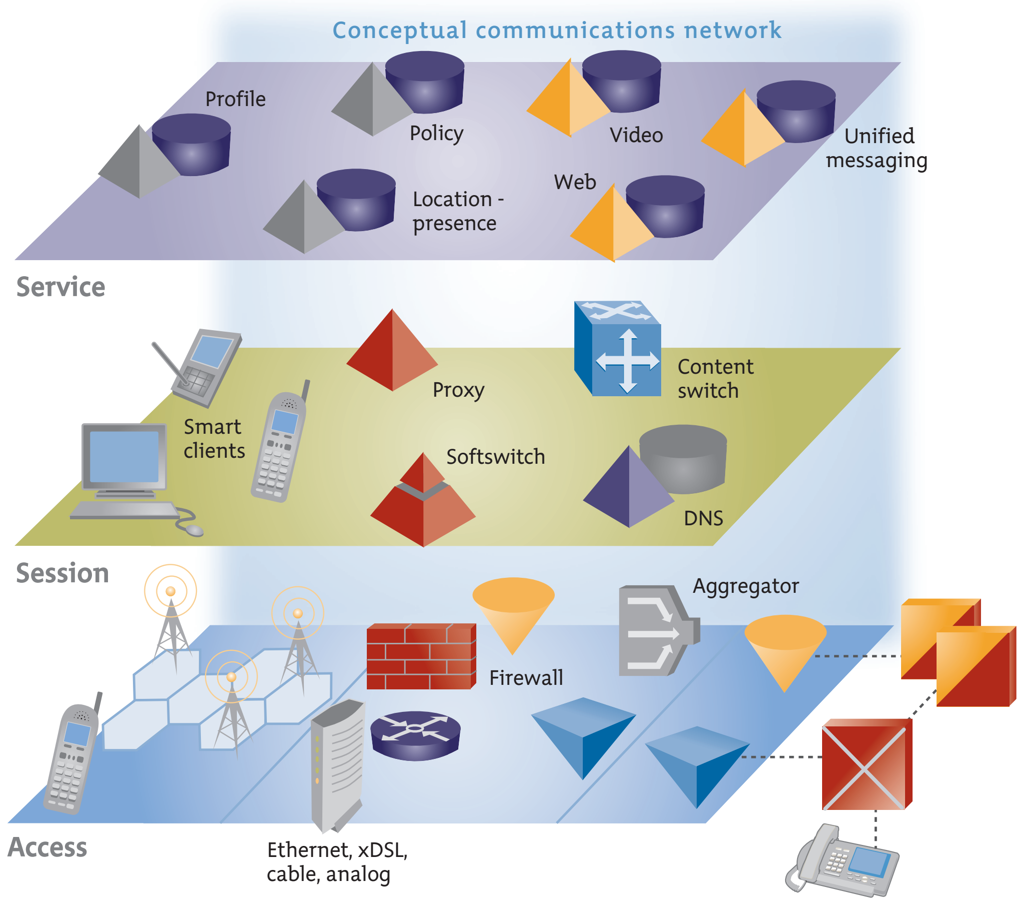

Figure 7. Distributed multimedia network—SIP enables distributed session management

The SIP value proposition

Protocol attributes and business implications

There are several attributes of SIP that have a direct impact on both personal and enterprise communications.

Collaboration

Because SIP is inherently multimedia, it supports a wide variety of the media people use to collaborate interactively, from audio and video to instant messaging. SIP also supports sessions involving interactive applications such as games.

Using SIP, it is possible to establish multimedia collaborative sessions between people dispersed all over the globe, relatively independent of differing client devices. Within a single session, some participants may be on a cellular phone, while others use PC interfaces, and some may even be using a PDA or set-top-enabled televisions. Different PC functionality is supported as well, as some may be set up for video conferencing, some may use IP Telephony, and some may only have instant messaging capabilities. With the proper applications, participants in this session could co-browse the Web, share interactive white boards, and transfer files instantly. SIP enables the creation of an environment that truly removes distance as a barrier to collaboration.

Collaborative features SIP can enable include:

- **Ad hoc conferencing:** A user already in a session who receives or initiates an additional session invite (a "call" in telephony terms) can easily combine multiple sessions into a multi-party conference.
- **Meet-me audio conferencing:** Users have their own private conferencing resources available for meetings at any time. Participants dial in and are put on hold until the conference is activated by the arrival of the chairperson. Instant messages that announce when anyone joins or leaves the conference greatly enhance conference management.
- **Sharing functions:** Users with PC clients can utilize the following features as part of an active call or as a standalone session:

  Clipboard: Users send and receive the contents of a PC clipboard

  File exchange: Session participants send and receive files directly with each other

  Web push: Users send a URL that opens a Web browser on the recipient's PC

  White board: Participants enter and exchange text and graphics in a shared workspace

Mobility and personalized communications

As instant messaging has become more and more popular, the concept of "presence" has entered into our personal communications. People are accustomed to checking their buddy lists to see if someone is available to chat. SIP pushes the concept into other media such as voice and video, allowing users to view and act upon real-time information about other users' status in the network before attempting to contact them. People can even be available for one media such as instant messaging while engaged in a session involving another media such as voice, a common occurrence when someone is on a conference call.

SIP inherently enables mobility and location independence. SIP routes session initiation requests around the network based on dynamically updated information about the availability of a user's registered devices. So session requests aren't placed to a device, in the hopes of reaching a person. Requests are placed to people and the network locates them.

It's a truly personalized communications model. Using SIP, it is possible to create a network-based agent to act on the behalf of a person 24 hours a day, 365 days a year. Users simply set personal configurations to control how, when, and by whom they are contacted, using a combination of automatic "find-me, follow-me" and call screening to control their time while assuring availability to important callers. Internet clients, wireless, and landline phones can all be easily included in a custom communications mix.

With an application of this nature, a person can specify whether he wants to be contacted via voice, instant messages, e-mail, or some other means on a caller-by-caller basis. Preferences can also be adjusted based on time of day. For example, all calls during the lunch hour go straight to a voice mail or an e-mail inbox, unless it is a call from home, in which case it rings the user's cellular phone. Both the availability information displayed to others and the routing of inbound session attempts can be dynamically adjusted based on various inputs related to actions a person naturally engages in. For example, if a person places a cellular phone in "meeting mode" to prevent disturbing an ongoing meeting, this information could be conveyed to the network and used to adjust parameters in the SIP domain. Users can set their own status for times when they are away, busy, inactive, or temporarily unavailable. Presence status can even be updated automatically, such as when a user's PC is inactive for a period of time or when the user is on the phone.

The implications of this capability to the enterprise are considerable. A business with a mobile workforce can allow every employee to customize his or her communication preferences, easily updating them via the Web as often as necessary. Location ceases to matter, as telecommunications follow a mobile workforce everywhere and are filtered to minimize interruptions while keeping people in touch with those who matter most. In a personal communications context, similar benefits exist for the soccer mom or the active teenager.

Productivity and information interactions

A SIP-centric system enables numerous time-saving network capabilities, with the potential to affect both efficiency and productivity. Some potential functions include:

- **Address books and contact lists:** A personal address book can provide a network-based directory of addresses that a user can edit and access from any client. A change, addition, or deletion of an address book entry or buddy list made with one client will automatically update information in all other clients. So if a user updates a contact's information on a PDA, that change would be reflected on the user's PC as well as all other clients.
- **Software integration:** Most SIP-centric services can be integrated with existing office productivity applications, allowing contacts and other common information to be shared.

- **Data stream delivery:** Data such as stock quotes can be sent at regular intervals, directly to the display of various clients. Web services are easily integrated with SIP to deliver real-time information updates on virtually any subject.
- **Telephony service extensions:** While traditional telephony services such as hold, call forward, call waiting, and caller ID are readily enabled by the SIP protocol and inherent in any well implemented SIP system, many multimedia enhancements are possible as well:
  - Photo caller ID—"Caller-I-See"
  - Calls with e-mail-style subject lines
  - Call rejection with context-sensitive reasons for declining the call
  - One-click calling from an address book
  - Voice-activated dialing
  - Text-to-speech conversion (including IM-to-voice conversions)
  - Audio/video voicemail
  - Music streaming
  - Entertainment, including music or video on hold (local or remote)
  - Voice, video, or Web-based announcements
  - Corporate dialer—VoiceXML
  - Interactive voice/video response (next-generation IVR)
- **Educational innovation:**
  - Multimedia distance learning easily delivered anywhere
  - Professors in remote countries educating abroad, without travel
  - Students using presence to know which teacher's assistant is available and interacting via instant messaging, voice, or white board sessions
  - Mobility around campus while staying connected with peers and educational resources—dorm room, library, classroom, or courtyard via wireless LAN
- **Retail and travel planning enhancement:** Touch-screen kiosks to review information such as reservations, information about products, or how-tos related to using products. Need more help? Select the "talk to an agent" key on the kiosk and immediately talk with a service agent or product expert.

Distributed model-enabled innovation

There are inherent advantages in the way the key elements of SIP are distributed. An end user with the right smart client can implement services from that end point, without the need for a centralized server. Therefore, the new service development process is fast, safe, flexible, and scalable. It's an Internet service delivery model where it's easy to introduce a new service, and it's easy to grow it. In addition to the flexibility of end-point service nodes, network-based SIP elements enable services to be readily available at all times, from any location. Services can start out highly focused on a specific problem for a limited set of individuals and very easily expand to serve a much larger community with similar needs.

For example, a little league baseball coach could code a simple SIP application that responds to an instant message requesting schedule information with details about the next practice or game. Once the application is placed on a SIP-enabled home office PC, the coach's players could access it from any SIP client via the Internet. Then, if other coaches in the league became interested in the application, it could be migrated to a hosted service and expanded for more wide-scale usage, enabling the coaches to create information specific to their teams, without having to know how to create the actual application. The service could be further enhanced to allow requests to come in via phone calls and responses played as audio announcements, all driven by the simple file each coach creates with the updated info.

Beyond IP Telephony

Because SIP can be used to enable IP Telephony, there are some who mistakenly consider it to be nothing but a software-based telephony switch. It is true that SIP can function as a voice protocol quite well. However, SIP is not simply an IP Telephony protocol. It is truly much, much bigger than that. SIP is a whole new communications model. A SIP proxy server utilizes a protocol architecture in which signaling intelligence is distributed across the network. Features and applications are integrated at the session and service layers, independent from access constraints and the processes of message transport. In SIP networks, voice is just another media, albeit a very powerful one.

Open architecture

SIP is being embraced today by all major communications equipment manufacturers and many software companies. Because SIP is an Internet Engineering Task Force standard, it is inherently an open architecture, which serves to quicken its acceptance. The protocol readily enables voice and data convergence and is quickly becoming the backbone protocol for interpersonal interactive communications.

Summary

The SIP control paradigm provides the key functions that will enable an entirely new network communications model, changing the way people communicate forever. SIP represents exciting possibilities for personal and enterprise communications. As more users adopt IP-compatible smart clients (phones, PCs, PDAs, mobile handsets), SIP-enabled sessions utilizing IP Telephony, rich-media conferencing, push-to-talk, and location-based services will become more and more prevalent. In this brave new world, users will be able to locate and contact one another with little regard to physical location, media content, or the number of participants in a session. These users will enjoy a wealth of new services as well, as service providers and enterprise IT departments will be able to dramatically lower the cost of designing and deploying innovative new IP-centric services for their customers.

Appendix

About the author

John Yoakum is a Director of Business Development at Nortel Networks, where he builds collaborative relationships with customers and partners that capitalize on emerging opportunities and disruptive technologies. Since joining Nortel Networks in 1987, John has been instrumental in identifying and architecting leading-edge multimedia communications products, building external technology partnerships and modernizing software development environments. John has also been a key instigator of efforts to help the industry define signaling standards, helping Nortel Networks enable true multi-vendor multimedia communications networks. With numerous communications technology patents, publications, and presentations to his name, John has proven to be a true innovator and visionary in the field.

Related documents

For more information on SIP, try the following sources:

www.cs.columbia.edu/sip
www.sipcenter.com
www.sipforum.com

In the United States:
Nortel Networks
35 Davis Drive
Research Triangle Park, NC 27709
USA

In Canada:
Nortel Networks
8200 Dixie Road, Suite 100
Brampton, Ontario L6T 5P6
Canada

In Caribbean and Latin America:
Nortel Networks
1500 Concorde Terrace
Sunrise, FL 33323
USA

In Europe:
Nortel Networks
Maidenhead Office Park
Westacott Way
Maidenhead Berkshire SL6 3QH
UK

In Asia:
Nortel Networks
6/F Cityplaza 4,
Taikooshing,
12 Taikoo Wan Road,
Hong Kong

NORTEL NETWORKS

BUSINESS WITHOUT BOUNDARIES

www.nortelnetworks.com

Nortel Networks is an industry leader and innovator focused on transforming how the world communicates and exchanges information. The company is supplying its service provider and enterprise customers with communications technology and infrastructure to enable value-added IP data, voice and multimedia services spanning Wireless Networks, Wireline Networks, Enterprise Networks, and Optical Networks. As a global company, Nortel Networks does business in more than 150 countries. More information about Nortel Networks can be found on the Web at:

For more information, contact your Nortel Networks representative, or call 1-800-4 NORTEL or 1-800-466-7835 from anywhere in North America.

NN106700-121903