The future of communication using SIP

Peter Granström, Sean Olson and Mark Peck

In this article, the authors describe the session initiation protocol (SIP), which is the new standard for establishing multiparty, multimedia communication in IP-based networks, and some of the implications that its adoption carries for the future of communications. SIP-based solutions are shown to highlight Ericsson's competitive position in this new technology space.

SIP has been called the integrated services user part (ISUP) of next-generation networks. Certainly, the two protocols share many characteristics, but whereas ISUP embodies the evolution of common-channel signaling over the span of many years in the history of telephony, SIP has the potential to be a groundbreaking, disruptive force for change. We foresee that the adoption of SIP will usher in a new era of multimedia communications that leverage the strengths of Internet technologies to provide opportunities for innovation.

The authors discuss the basic concepts behind SIP, including some of the technologies on which it builds. This leads into a discussion of the features and capabilities in the protocol and its defined behavior, which enables operators to deliver rich functional content to their users.

The authors also consider the business drivers behind the adoption of SIP-based networks. These business drivers gave rise to the specification and the development of the IP Multimedia System (IMS), which is surveyed briefly in this article.

The authors conclude the article with an examination of a SIP server that can function as the basis for several SIP-based products in the near future. Although the product is initially targeted for the IMS, other products that use this server as a base can be positioned for a wide array of offerings and operating scenarios in the core networks of Ericsson's key operator customers. Ericsson is poised to capitalize on the revolution that SIP heralds, which offers an exciting promise for the communication networks of the future.

How SIP works

The session initiation protocol (SIP) is an application-level control protocol for establishing, modifying, and terminating multimedia sessions between one or more participants. It supports multimedia conferencing, Internet telephone calls, registration and redirection services, and is easily extended. It traces its roots to several multiparty conferencing initiatives in the history of the Internet Engineering Task Force (IETF) as well as to the World Wide Web (WWW) and Internet e-mail. As a result, SIP embodies a distinct protocol, with syntax and semantics for the messages to be exchanged, and a philosophy of end-to-end control over session establishment with support from servers in the network. Figure 1 shows the relationship of SIP to various other protocols in use in the Internet.

Basic session establishment

In the simplest case, two users want to communicate with one another using a variety of media types (such as audio, video, and text messages) over an IP network. The software application that enables the communication of each party is known as a user agent (UA). The UA could be running as a "soft client" on a PC, as the operating software in a mobile device, or in the firmware of a desktop SIP phone.

Party A, who wants to communicate with Party B, sends an INVITE message to Party B, who is listening on the official SIP port (5060). The INVITE body contains information encoded with the session description protocol (SDP), which indicates the types of media the receiving party is willing and able to use. Party B's response indicates his preferred media types. Once Party A returns an ACK message, each party is aware of the other's IP address and port numbers where the media streams will be received. Similarly, each knows which types and bandwidth of media the other is able to receive. When ACK has been sent and received, both ends begin transmitting data to the corresponding receiver ports, via a separate media connection using the real-time protocol (RTP) or some other appropriate transport protocol. Throughout the session, either party can make updates (indicating a new set of media types, addition of new parties to the session, or other changes), by sending additional SIP messages. At the end of the communication, either Party A or B can send a BYE message to indicate termination. When the other party responds the session is ended.

The user datagram protocol (UDP) is a required transport protocol primarily for performance reasons (TCP and other protocols are optional). But because of the unreliable nature of the datagram service in UDP, SIP contains its own retransmission mechanisms, including the three-way exchange between nodes for establishing sessions.

Syntax and addressing

Instead of being IP addresses, destinations in SIP can be represented with uniform resource indicators (URI), which have the same format as e-mail addresses. Accordingly, a valid SIP address might be *sip:Mark.Peck@ericsson.com*. This implies the use of the domain name service (DNS) to map host and domain names to IP addresses. Support for DNS is a key aspect of the integration of SIP with Web- and mailenabled technologies, which are already familiar with the concepts of URIs and their interpretations.

The close connection between SIP and DNS facilitates interoperability with telephone systems and addressing mechanisms. Support for E.164 numbering in DNS (ENUM) allows SIP servers and clients to send and receive telephone numbers in place of SIP URIs in messages, and to route them in a sensible fashion.

SIP is a text-based protocol—it reuses the message structure found in the hypertext transfer protocol (HTTP) and simple mail transfer protocol (SMTP)—with numerous informational headers followed by a body (possibly multi-part). As a result, scripting languages such as Perl or Python are well suited for automating many session-processing tasks in a SIP server.

That SIP addressing lends itself to finding resources and people is an important part of the philosophy behind the protocol. When we consider real-time communications and the future of such services, we see that our focus is still mainly on communication between people. SIP is very useful for finding and "connecting to people," regardless of where they happen to be or what they are doing.

Indirection

Indirection is a key concept for understanding how SIP works. According to this concept, the current location of a user is hidden behind a permanent user identity or uniform resource locator (URL). A network-based SIP server binds the user's mobile identity to the permanent URL much in the same way a home agent functions in a mobile IP network. The two principal mechanisms in SIP that support this are redirection and proxying.

If Party A does not know the address of Party B, then Party A can send an INVITE to a redirect server, which returns a response indicating where Party B can be found (usually in the form of a SIP URI). Party A can then send a new INVITE to Party B.

The use of a proxy server allows users to have a node in the network that performs some intermediary function before the SIP messages are routed to their destination on behalf of the UA. If such a node exists, the SIP messages that it receives are forwarded to the appropriate destination and responses are forwarded in the reverse direction. Thus in terms of signaling, the proxy appears to each endpoint as if it were the other endpoint.

In cases that involve either a redirect or a proxy server, a location server might be consulted for information on the current SIP address of the indicated destination. The interface between the proxy or redirect server and the location server is not defined in the SIP RFC, but can be some appropriate querying interface, such as LDAP, HTTP, or DIAMETER. SIP supports real-time updates to the location server database via a

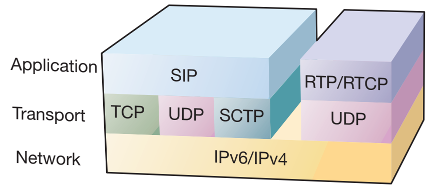

Figure 1
Relationship of various protocols used on the Internet.

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third-generation Partnership Project
3GPP2	Third-generation Partnership Project 2
AAA	Authentication, authorization and accounting
ALG	Application layer gateway
API	Application program interface
CDMA	Code-division multiple access
CORBA	Common object request broker architecture
CPL	Call processing language
CSCF	Call/session control function
DIAMETER	IETF-defined protocol for AAA functions, successor to RADIUS (Remote authentication of dial-in user services)
DNS	Domain name service
ENUM	E.164 numbering in DNS
GPRS	General packet radio service
HSS	Home subscriber service
HTTP	Hypertext transfer protocol
IETF	Internet Engineering Task Force
IMPP	Instant messaging and presence protocol
IMS	IP multimedia subsystem
IP	Internet protocol
ISP	Internet service provider
ISUP	Integrated services user part
LDAP	Lightweight directory access protocol
MGCF	Media gateway control function
MRF	Media resource function
NAT	Network address translator
OSA	Open service architecture
PBX	Private branch exchange
QoS	Quality of service
RFC	Request for comment
RTCP	Real-time control protocol
RTP	Real-time protocol
SCM	Service control manager
SCTP	Stream control transmission protocol
SDP	Session description protocol
SGW	Signaling gateway
SIMPLE	SIP for instant messaging and presence leveraging
SIP	Session initiation protocol
SMTP	Simple mail transfer protocol
SNMP	Simple network management protocol
SRV	Server location records extension to DNS
SS7	Signaling system no. 7
TCP	Transmission control protocol
TSP	Telecom server platform
UA	User agent
UDP	User datagram protocol
UMTS	Universal mobile telecommunications system
URI	Uniform resource indicator
URL	Uniform resource locator
VPN	Virtual private network
WAP	Wireless application platform

REGISTER message that indicates the user's current location via a SIP URI. The SIP server that receives the REGISTER messages and updates the location database is called a SIP registrar. Through information obtained from the registrar, the redirect or proxy server is able to reroute a SIP request to the destination where the user wants to be reached.

Indirection is not limited to the user—by using DNS it can also be applied to the SIP servers themselves. A number of DNS mechanisms exist to map a symbolic name for a SIP server into an IP address where that server can be reached. Particularly interesting are SRV records (server location records extension to DNS), which allow the definition of one or more SIP servers that are the first point of contact for a given domain. For example, four separate SIP servers might share the load for the *ericsson.com* domain. The use of DNS makes it very easy to establish these network topologies. Therefore, the process by which a user is contacted using SIP involves

- identifying the server for that user; and
- determining the location of the server.

Indirection can be used in each of these steps to yield a very flexible and fluid communications network.

Forking

By means of a procedure known as forking, SIP proxies can simultaneously forward a SIP message to multiple destinations. A user can thus have multiple destinations registered (say, a mobile device as well as a desktop phone) and have each destination alerted simultaneously when a new session request arrives. The proxy server correlates the responses received from various branches and ensures that only a single upstream response is sent to the client.

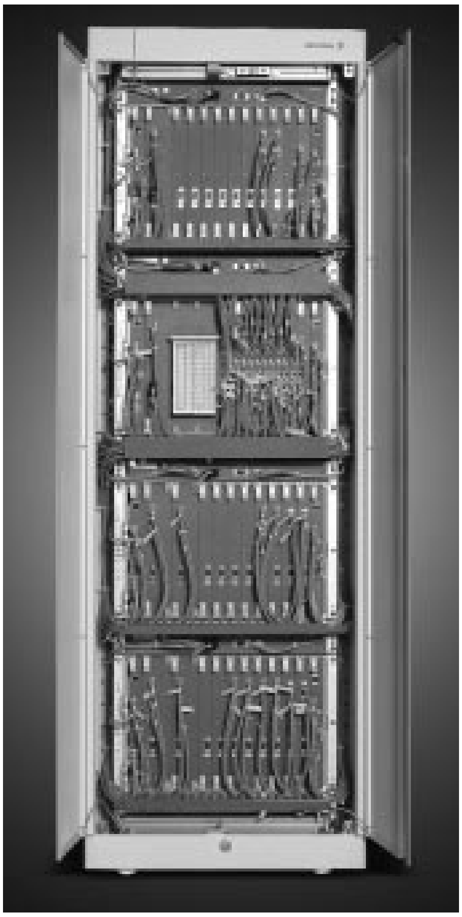

Figure 2
Telecom server platform (TSP).

Using SIP in various environments

To bring useful SIP implementations to market, some hurdles must be overcome in particular environments. Knowing these hurdles and their solutions helps explain Ericsson's architectural choices for the solutions currently under development.

SIP on the public Internet

SIP on the public Internet is the default case, given the high bandwidth and low latency between network elements (making QoS mechanisms less important) and the lack of differentiated domains that would introduce elements such as firewalls, network address translators (NAT) or other gateways. In this environment, emphasis can be put on terminal-based or end-to-end solutions that embody the "pure" philosophy of the Internet. Most of the general descriptions of SIP in this article are based on this scenario, and it is from here that we will add domain-specific extensions that accommodate the unique needs of the other scenarios.

SIP in protected domains

For many users, the default scenario described above is incomplete. Most notably, the network is often not "transparent" end-to-end at or above the IP layer. Provisions must thus be made to ensure unchanged behavior from an end-user perspective.

For security purposes, many IP networks are protected from external traffic by the installation of a firewall, which solely allows packets to flow through designated secure points. For security purposes and address conservation, many ISPs and enterprises allocate private IP addresses to their users. These private addresses are then mapped to public IP addresses by a network address translator. For "pure" client-server applications, such as the Web, file transfer, or e-mail, in which the precise real-time network address of the end-user is irrelevant, application-level proxies can provide unimpeded access to the services needed, and firewalls and NATs need not be show-stoppers. However, although SIP differentiates the signaling from the media in establishing a real-time communication session, the two are highly interdependent, and firewalls and NATs can pose a problem. A SIP proxy that is employed to pass messages through a firewall, for example, must provide indirection support for locating users on either side of the firewall, and have some means of instructing the firewall to allow media to pass through it (since the media might use an entirely different combination of IP addresses and port numbers). Similarly, if a NAT transforms the private IP address used by a SIP UA into a public address, then the private media address described by the session description protocol (SDP) in the SIP message will be unusable by the remote UA. Therefore, an application layer gateway (ALG) must be coupled to the NAT to ensure that the SDP is rewritten to reflect the public address that will be used to transit the media across the domain boundary.

SIP in a mobile environment

Limited bandwidth over the air interface means that the amount of signaling must be kept to a minimum. The size of SIP messages might become quite large due to the number of steps taken in routing between endpoints and because SIP is a text protocol rather than binary. This necessitates the establishment of a node that can customize signaling for efficiency without affecting end-to-end transparency. Such a node (SIP proxy) can also serve as a useful default inbound and outbound proxy for address-translation services, the invocation of locally significant services (such as taxi locators, weather, or news), and guaranteeing quality of service (QoS) for network resources.

Consequences

Because of the particular requirements imposed on SIP-based networks in these environments, there is great benefit to be derived from an architecture that supports some or all of the requirements in a flexible manner. Several specialized nodes have been identified by (among others) the Third-generation Partnership Project (3GPP) to accommodate these requirements and provide operators with a powerful framework for adding value to the end-user experience.

Capabilities supported by SIP

In the paragraphs that follow we list some examples of services and service building blocks that are made possible by the basic mechanisms of SIP. We will also describe scenarios in which they add value to the end-user experience.

Add/drop media

SIP supports the ability to add new media types (or remove unwanted media) in the middle of a session. Additionally, disparate media types can be supported at the various endpoints without degrading the user experience (that is, lowest-common-denominator communications are not necessary). Consider a surgeon who, while in the middle of a procedure, wants to consult with a colleague in another part of the country. They begin discussing the case over a traditional voice connection, but midway through the session, the specialist determines that she needs to see the patient's condition. A one-way video feed is established from the operating room, and the specialist begins to type in a series of instructions that show up in a separate chat window on the surgeon's viewing screen. The basic SIP mechanisms that support this exchange require no extensions or specialized hardware apart from an image-capture device in the operating room.

Find me/follow me

The specialist from our previous example can be simultaneously registered (with a SIP registrar) in multiple locations according to her daily patterns. Perhaps she has a fixed SIP-enabled terminal on her desk that rings when an INVITE message is sent to her public SIP address. She can also register

- with her mobile terminal, which can receive audio calls or instant messages; and
- with her secretary, who might be instructed to take any audio calls that have not been answered after a given number of signals.

Finally, she can register with a SIP video client on her desktop PC to receive video feeds from surgeons while she is talking to them on the phone. Audio calls might ring simultaneously at all four locations. When she answers one of the devices the other three devices stop ringing. Note that none of what we have described thus far requires specialized PBX equipment!

Presence and instant messaging

In the example above we mentioned the possibility of sending and receiving instant messages via SIP, and its use for instant messaging and presence has also been specified by the SIMPLE working group of the IETF. SIP is currently the leading candidate to fulfill the requirements of the instant messaging and presence protocol (IMPP) working group, which has specified an overall instant messaging and presence framework. Several major operators and vendors, including AOL and Microsoft, have announced plans to support SIP for interworking between presence and instant-messaging domains. Because it is well suited for this purpose, SIP will be one of the leading choices for transport of presence and instant messaging information between users.

Conferencing and distance working

The operating-room scenario above could be extended to include a class of medical students who observe the surgery in progress on a remote video feed and ask questions of the specialist over an audio connection. A lecturer located on a college campus could provide distance-learning support to students spread around the city. The necessary equipment includes little more than video capture equipment, a conferencing server to mix the audio, and a chat-room environment in which students' questions can be queued and answered.

The abolition of Class-5 services

Traditional voice services that previously provided substantial revenue to incumbent local operators are made trivial by SIP. Services such as caller ID, call waiting, and call hold are handled by basic SIP mechanisms within the UA and require no input or control by the network. While this poses a threat to the existing business model, in the new business model of SIP-enabled networks, it becomes a competitive advantage since users are given greater control over the behavior of their communications services at minimal cost to the operator. For this reason, the simplification and trivialization of formerly significant services gives operators with new networks a competitive advantage over their legacy competition.

Multiparty gaming

SIP can also be used to transport a wide range of real-time information, including gaming events for multi-player games. Additionally, SIP (whether supported by the game client itself or via a separate communicator application) can be used during the game for audio, video, and chat between players.

VPNs made simple

The use of DNS, ENUM, and the indirection capabilities of SIP make it easy to develop and manage virtual private networks (VPN). A single SIP proxy can provide address mapping and forwarding services for a remote location, giving users the perception that they are located within the same corporate domain as their colleagues (even when they are not).

Business motivation for SIP-based systems

We are currently witnessing a dramatic change in the way we communicate. The monolithic incumbent telecommunications operators, who until only a few years ago provided all or most of our communications needs, are now under constant pressure from

- technology turning one of their core assets—network bandwidth—into a commodity;
- regulators and governments encouraging competition in all areas of operator businesses;
- dramatic price drops being driven by regulatory changes, competition, and technology;
- the Internet, and IP in general, which is moving communications into a totally new era of packet-switched technologies;
- the need to make large investments to meet users' demands for broadband access;
- the threat that mobile telephony will overtake and replace fixed telephony, which challenges operators' core service and revenue sources; and
- the relative ease with which a user can obtain services from third parties in an IP-based network, relegating the role of the operator to that of "bit-pipe" provider.

Within the next decade operators will need a new generation of networks, primarily based on IP technology. These will pave the way for new opportunities to earn revenue and make customer-driven services the key to profitability. All the same, these new networks do not have the same value chain as traditional telephony, and the business models have not yet been established. Operators will thus need to learn how they can take advantage of these new networks and move their businesses into a totally new space. Simply providing the same set of services using new technology will not enhance the end-user value. In fact, as we have seen, a re-implementation of traditional services will only drive margins down and eventually lead to operator losses. The deployment of SIP-based networks can be a remedy. Compared to existing systems, elements of the SIP technology and business value proposition include but are not limited to

- presence;
- combinational services;
- access independence;
- new charging models;
- quality of service; and
- security.

Presence

By presence we mean that a group of individuals can share information (status) on their current availability. SIP provides new and creative ways of developing services based on presence information. The value of presence information is not found in providing it as an individual service, but rather in combining presence with other multimedia capabilities, such as combinational service offerings.

Combinational services

With SIP, it will be easy to combine conversational multimedia services with other categories of services, such as directory information, Web browsing, positioning, and presence. For example, a location-based service can be developed to combine a conversational communication session with positioning information and maps to provide information that is pertinent to the geographical location of the parties involved.

Access independence

Being an application-layer protocol in the IP-based suite, SIP is access-independent and offers seamless service capabilities between fixed and mobile networks. This is a key element in making the promise of fixed-to-mobile convergence a reality.

New charging models

Operators will be able to define new charging models based on actual media usage. For example, if the communication between two parties begins with a real-time voice session and video is later added, it is possible to charge for the sessions individually and for actual media usage during the sessions.

Figure 3
Prototype SIP telephone.

Quality of service

Given the importance of conversational multimedia services (where sessions which are real-time sensitive, and which involve voice and video streams, are shared with streams that are not as time-sensitive) then sufficient QoS mechanisms must be in place to guarantee a rich end-user experience. New product offerings will provide a robust and flexible architecture that supports the required QoS and security requirements in mission-critical applications. SIP-based products sit on top of the IP network and take advantage of the capabilities of the underlying network to provide QoS.

Security

Security mechanisms are major concerns for operators who deploy IP-based networks. SIP can encrypt and authenticate signaling messages; RTP supports the encryption of media. Together these two protocols provide cryptographically secure communications. Important benefits will be found in providing the necessary security functionality for the operator, including authentication, access control, confidentiality and integrity. These capabilities enable operators to deploy secure next-generation networks built on IP technology.

Ericsson's IMS delivers on the promise of SIP

Recognizing the opportunities and challenges of SIP-based networks, and seeing the

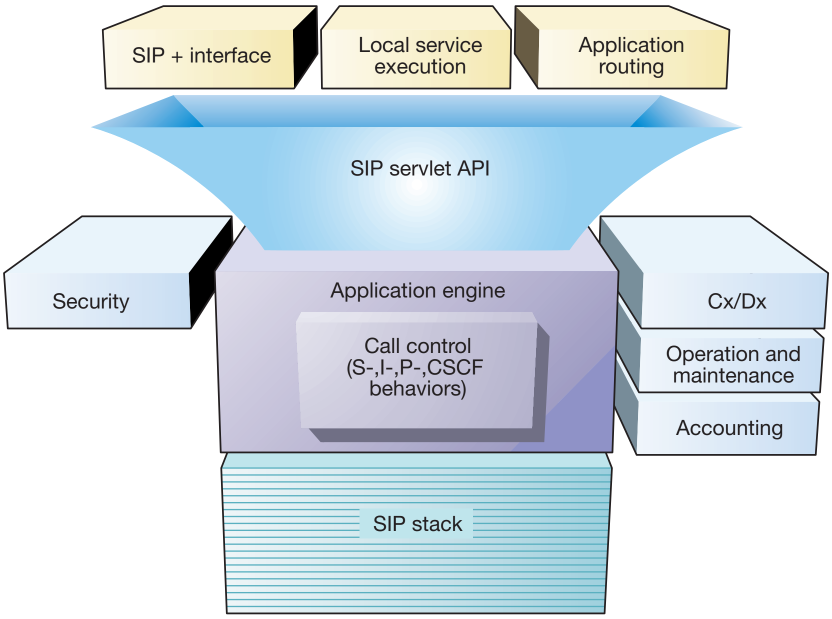

Figure 4
The CSCF architecture.

potential for dramatic change in mobile communications, the 3GPP has defined a new domain that takes advantage of the packet-switched capabilities provided by GPRS and EDGE. This domain is called IP Multimedia. Although the leading mobile standards body drives the standardization work, IP Multimedia is independent of any access technology and is also intended to be suitable for fixed-access networks.

Components of the IMS

Ericsson is building a 3GPP-compliant IP Multimedia System (IMS) that will bring the benefits of SIP-based communications to operators and end-users in the near future. The commercial IMS is not a standalone offering, but is intended for delivery as part of solutions from several of Ericsson's major business units.

The remarkable success of the ENGINE concept has given Ericsson a powerful market message for fixed operators around the world. The ENGINE concept continues to evolve from an ATM product focused on telephony into an IP-based solution: ENGINE Integral + IP. The next major step, called ENGINE Multimedia, offers a multimedia-enabled solution. At the heart of this solution sits the IMS.

The standards work for CDMA2000-based mobile systems is carried out in 3GPP2, which has defined a network architecture for multimedia that is very similar to that of 3GPP. The products and applications deployed in the IMS will also be part of the offerings from Ericsson to CDMA-based operators in the US and elsewhere.

Delivering on the promise of interoperable multimedia

During 2002, Ericsson will release IM 0.9, which will allow operators and end-users to see

- close alignment with the 3GPP architecture, including the CSCF, MRF, MGCF, SG, HSS, subnetwork manager, and other nodes;
- integrated operations and management;
- improved system characteristics, including capacity, reliability and throughput; and
- new services.

Beyond release IM 0.9 lies the reality of commercial systems delivered as part of total solutions that offer robust security, flexible accounting and charging, differentiated QoS, and support for the wide array of media types and devices that will be connected to the system.

SIP Server example: CSCF

The call/session control function (CSCF) is to be deployed as part of Ericsson's IMS, and as such it can be used in next-generation wireline and wireless networks using SIP as the signaling protocol. However, it is not intended to be a standalone commercial offering. Instead, it has been designed to be suitable for a wide range of applications, including the call/session control function defined by 3GPP, and the session control manager (SCM) defined by 3GPP2, as well as a number of call control and application servers described in various Internet drafts.

The CSCF offers an ideal platform for deploying next-generation communication services in multimedia-enabled networks. It provides support for popular Internet protocols and open, industry-standard APIs on a scalable, high-performance platform that draws from Ericsson's experience of telephony systems (Figure 4).

High-level functionality

The CSCF, which supports the establishment, modification and release of IP multimedia sessions using the SIP/SDP protocol suite, provides the following capabilities:

TRADEMARKS

CDMA2000 is a trademark of the Telecommunications Industry Association (TIA).

- subscriber registration;
- invocation of multimedia sessions (originating and terminating)—the CSCF supports SIP mechanisms for invoking one or more IP multimedia sessions;
- capability negotiation at session invocation—the CSCF supports SIP mechanisms for establishing the capabilities of a session;
- modification and clearing of multimedia sessions;
- forwarding, redirection, and rejection of multimedia sessions;
- notification of multimedia session events to the service network; and
- interfaces to an IP policy control function.

Carrier-class technology

The CSCF has been designed to take advantage of the capabilities of TSP. This is considered a critical part of the value proposition of this product to Ericsson's customers, because of their need for high availability, cost-effective scalability, and best-in-class capacity for running large-scale networks. Notwithstanding, a focus on carrier-grade characteristics does not limit the functional capability of the server.

Routing and addressing

The CSCF is capable of routing according to standard SIP mechanisms for session establishment, clearing, and modification. It can query DNS servers to map E.164 numbers (using ENUM) or SIP URIs to network addresses. The CSCF also supports specialized routing mechanisms (as defined in 3GPP, 3GPP2, or other standards bodies) that involve other nodes in the network (such as the HSS or AAA server). It is also able to route multimedia session attempts to and from non-3GPP and non-IMS systems. This gives operators the ability to provide their customers with open and interoperable services.

Customized services

The CSCF supports the invocation of services either remotely (remote invocation) or locally (local invocation). For remote invocation the CSCF supports an external API (such as Parlay/OSA) carried over CORBA for access to remote application servers. Applications that want to make use of the capabilities of the CSCF can also use this interface. For local invocation the CSCF supports the execution of services on the node by means of at least two mechanisms:
- through the execution of scripts written in the call processing language (CPL); and
- through the execution of Java SIP servlets. The SIP servlet engine allows programmers to write applications in Java that control the behavior of the CSCF in response to SIP messages. This gives operators extensive opportunities to customize and deploy basic services that enhance the establishment of multimedia sessions.

Charging

The CSCF supports the collection of accounting information for time- and event-based charging. Using DIAMETER it forwards this information to the accounting server, where it can be aggregated with data collected from other nodes.

Availability/reliability

The CSCF has been designed for uninterrupted operation—it takes advantage of the facilities offered by TSP to enable high availability of hardware and software. All hardware and software upgrades can be performed on the node while in operation. Availability is expected to exceed 99.995%. A single hardware failure will not stop the operation of the CSCF. Moreover, the network architecture of the IMS provides additional redundancy in the event of catastrophic failure of a single node—alternate CSCFs can be selected without loss of service.

Conclusion

SIP is a central part of the value proposition of future multimedia networks. This value stems from several key aspects of the protocol, including
- the flexibility of addressing, routing and modifying messages using the protocol;
- support for a wide range of media types, simultaneously invoked or selectively added as the need arises;
- the wealth of information that can be communicated to all nodes in the network, fostering an end-to-end view of services and applications; and
- ready integration with Web-enabling technologies that springs from the origins of SIP in the IETF.

Ericsson is bringing several new products and solutions to bear in this new technology space. These offerings will enable operators to position themselves in new markets and to realize new revenue streams while protecting their investment in legacy systems.

REFERENCES

1 Handley, Mark et al. The Session Initiation Protocol, draft-ietf-sip-rfc2543bis-05.txt, http://www.ietf.org, October 2001
2 Camarillo, Gonzalo. SIP Demystified, McGraw-Hill, 2002
3 Rosenberg, J., et al. SIP extensions for presence, Draft-ietf-simple-presence-04.txt, http://www.ietf.org, November 2001
4 Rosenberg, J., et al. SIP extensions for instant messaging, Draft-ietf-simple-im-01.txt, http://www.ietf.org, July, 2001