FORUM NOKIA

SIP Frequently Asked Questions

Version 1.0; May 4, 2004

SIP

NOKIA

Contents

Change History

May 4, 2004	Version 1.0	Initial document release

1 SIP Background

The Session Initiation Protocol (SIP) has its origins in the early 1990's as a component of the "MBone" set of utilities and protocols. The MBone, or multicast backbone, was an experimental multicast overlaid network on top of the public Internet. It was used for distribution of multimedia content, including talks and seminars, broadcasts of space shuttle launches, and meetings of the Internet Engineering Task Force (IETF). One of its essential components was a mechanism for inviting users to listen in on an ongoing or future multimedia session on the Internet. Basically, SIP is a session initiation protocol. By that time SIP was only used to invite users to multicast sessions. Later, around 1995-96, the work to extend SIP to support unicast sessions (like VoIP) as well was started by Columbia University.

1.1 What Is SIP?

Session Initiation Protocol (SIP) is IETF's standard for multimedia conferencing over IP. SIP is a text-based, application-layer control protocol (defined in RFC 2543) that can be used to establish, maintain, and terminate sessions between two or more end points.

SIP is in some respects comparable to HTTP. However, service creation methods based on SIP have also been defined. Addressing and extensibility mean that SIP can be seen as a basis for the whole 3G architecture.

1.2 Why Is SIP Needed?

SIP is one of the emerging protocols that can be used for setting up an enriched communication session such as a real-time video/voice call (rich call). SIP can also be used in several other application areas, such as instant messaging (IM) and presence, but currently the primary focus area is enabling rich call.

1.3 Why Is SIP the Best of All Call Set-up Protocols? What Is the True Importance of SIP?

SIP has been chosen as the call set-up protocol for All-IP networks that will be standardized by the Third Generation Partnership Project (3GPP). This means that 3G All-IP terminals and networks will support SIP.

1.4 What Functionalities Does SIP Provide?

- Determining the location of the target end point—SIP supports address resolution, name mapping, and call redirection.
- Determining the media capabilities of the target end point—via Session Description Protocol (SDP), SIP determines the "lowest level" of common services between the end points. Conferences are established using only the media capabilities that can be supported by all end points.
- Determining the availability of the target end point—if a call cannot be completed because the target end point is unavailable, SIP determines whether the called party is already on the phone or did not answer in the defined number of rings. It then returns a message indicating why the target end point was unavailable.
- Establishing a session between the originating and target end point—if the call can be completed, SIP establishes a session between the end points. SIP also supports mid-call changes, such as adding another end point to the conference, or changing a media characteristic or codec.
- Handling the transfer and termination of calls—SIP supports the transfer of calls from one end point to another. During a call transfer, SIP simply establishes a session between the transferee and a new end point (specified by the transferring party) and terminates the session between the

transferee and the transferring party. At the end of a call, SIP terminates the sessions between all parties.

1.5 What Kind of Applications Can Be Implemented with SIP?

For example, the following applications can be implemented with SIP:

- The subject of the call is displayed immediately to the callee (for example, About Tom's birthday party).
- Ringing tone and caller image is delivered within signaling (SIP transports MIME payload).
- URLs can be passed within signaling, seamless e-mail/media-on-demand integration (for example, the call can be forwarded to rtsp URL: videomail answering service).
- Possibility to create richer profiles, for example, “If the caller is Bob, send him the soccerresults.html file (in SIP payload or by e-mail).”
- Simple scripts: “If the time is past 4 p.m. and the caller is the boss, forward the call to my voice mail.”

1.6 What Are the Key Characteristics of the SIP Protocol?

SIP supports the basic facets of multimedia communications:

- User location: determination of the end system to be used for communication, that is, where to send the requests for a user specified in the SIP URL.
- User capabilities: determination of the media and media parameters to be used, using, for example, Session Description Protocol (SDP).
- User availability: determination whether the called party is willing to be engaged in communications.
- Call setup: establishment of the call parameters at both the called and calling party.
- Call handling: transfer and termination of calls included.

1.7 Is SIP a Transport or an Architecture Protocol? Can It Be Compared with, for Example, SMS, WAP 1.x, or HTTP?

Whereas the WAP protocol is meant only for mobile networks, SIP is a protocol for both mobile and fixed networks.

SIP is not a transport protocol (although it can be used to carry content). Instead, SIP uses transport protocols such as UDP, TCP, SCTP, or even SMS. SIP is an application layer protocol.

1.8 Why the Fuss About All-IP, Considering That the Quality of VoIP in the Internet Has Not Been Good?

From the user perspective in the All-IP era, the primary benefit is the integration of services. Voice, video, instant messaging (IM), and presence are all simply different aspects of the general problem of interpersonal interactive communications. Many features and services used in one domain make a lot of sense in the other. Therefore, one common protocol and architecture will offer significant benefits.

SIP is totally different from early Internet voice demos, which had insufficient audio quality and proprietary telecommunication signaling protocols.

By using AMR voice codec and conversational Quality of Service (QoS) class, the target is to achieve at least the same voice quality as the current GSM circuit-switched technology can offer. By using header adaptation schemes to be defined in 3GPP Release 4 / Release 5 standardization, even basic voice calls over packet domain will remain spectrum-efficient.

1.9 What Kind of Protocols Are Needed for All-IP Services?

- A signaling protocol to establish presence, locate users, set up, modify, and tear down sessions between callers and callees (SIP/SDP (IETF), H.232 (ITU-T)).
- Media transport protocols for transmission of packetized audio/video (RTP...).
- Other transport protocols (UDP, TCP, SCTP).
- Supporting protocols (for example, gateway location, Quality of Service (QoS), Interdomain AAA, Domain Name System (DNS), and IP).

2 SIP and Operators

2.1 What Is SIP's Impact on 2G and 2G+ Networks?

Part of the "intelligence" normally residing in the network will shift towards terminals, that is, mobile terminals will become more intelligent in terms of call handling and services with SIP.

2.2 Why Are Operators Interested in SIP?

SIP has been accepted by 3GPP, so it is used in 3GPP Release 4/ Release 5 terminals.

In the all-IP architecture, the mobile operator sees itself offering services over various access networks, not only the mobile network. The business model is spread horizontally.

For example, as the mobile operator is able to map QoS levels, it can offer a company services over 3G wide area network and the intranet. In such a business, it is essential that the mobile operator manages the users' service profiles. SIP provides the tools for this.

2.3 What Are the Management and Operational Costs of Running a SIP-Enabled Network Like?

The management and operational costs of running a communications service are vastly reduced, since the costs of adding presence and instant messaging (IM) to an existing SIP network is nowhere close to the costs of building and running a completely separate, parallel network. The equipment costs themselves are reduced, since only one set of proxy and location servers are needed, rather than two. System capacity is improved, since resources can be shared across many services. Infrastructure investments in, for example, firewalls and mobility services that have been put in place for supporting SIP can also be reused. Service providers have also invested a serious amount of intellectual capital in SIP; thus reusing SIP means that these providers have a much shorter learning curve.

2.4 In Short, What Is the Main Benefit of SIP to the Operator?

By implementing SIP the operator gets the whole array of IP based services in the same network that is easy to implement, manage, and control. New service development and introduction is made rather simple when compared to other protocols.

Naturally, HTTP and SIP complement each other.

2.5 Doesn't the Use of SIP for Call Set-up Cause Additional Overhead?

There are two issues: the size of messages and the number of RTTs needed to set up a session.

There is a lot of mandatory information that is needed anyway (for example, addresses of end points and authentication information). If needed, additional (extra) information can be left out thus reducing the size of the message in the expense of a richer call set-up.

The encoding of SIP messages is text, while some telecommunication protocols use binary encoding. However, the increased complexity and debugging problems do not justify the use of the slightly more efficient binary encoding (H.323 being a terrible example).

3 SIP and End Users

3.1 Does the Consumer Benefit from SIP?

Consumers benefit from SIP by having the same address while choosing between the most suitable interfaces, for example, PC, mobile terminal, and PDA. While SIP is a generic protocol over all access networks, consumers can use the most capable access network available, for example, in terms of price, bandwidth, and usability.

3.2 Does SIP Enable Service Personalization?

Services are also personalized, which makes the user experience appealing. Furthermore, consumers can easily select the desired modality for communication based on the presence that fits the occasion.

3.3 How Does SIP Enable New Consumer Experience?

The primary advantage that can be seen with the SIP proposal is the way in which it can integrate voice, video, and other interactive communications services with instant messaging and presence.

SIP will greatly enhance person-to-person communications by combining multiple media types and communication methods with presence and community information. With SIP, adding “richer media” (video streams, live video, or application sharing), presence (availability and preferred media type), or community information is more straightforward because all the services use the same communication protocol.

4 SIP Architecture

4.1 What Are Clients and Servers in the SIP Network?

From an architecture standpoint, the physical components of a SIP network can be grouped into two categories: clients and servers. Figure 1 illustrates the architecture of a SIP network.

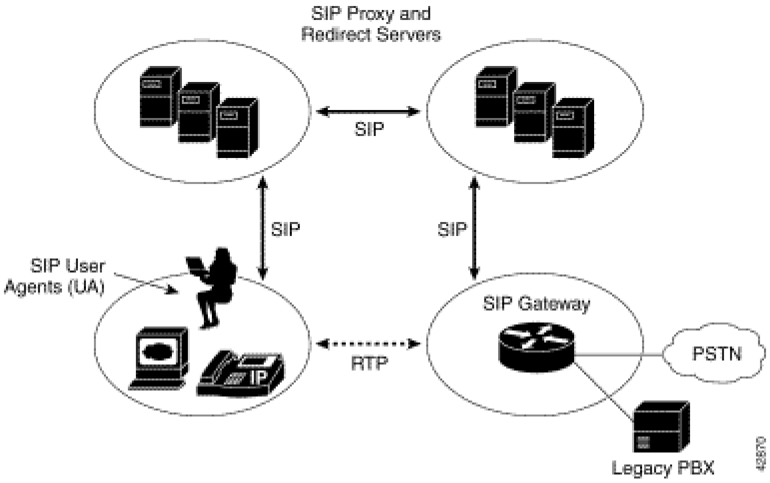


Figure 1: SIP architecture

SIP **clients** include:

- **Phones,** which can act as either a user agent server (UAS) or user agent client (UAC). Softphones (PCs that have phone capabilities installed) and Cisco SIP IP phones can initiate SIP requests and respond to requests.
- **Gateways,** which provide call control. Gateways provide many services, the most common being a translation function between SIP conferencing endpoints and other terminal types. This function includes translation between transmission formats and communications procedures. In addition, the gateway also translates between audio and video codecs and performs call setup and clearing on both the LAN side and the circuit switched network side.

SIP **servers** include:

- **A proxy server,** which is an intermediate device that receives SIP requests from a client and then forwards the requests on the client's behalf. Basically, proxy servers receive SIP messages and forward them to the closest SIP server in the network. Proxy servers can provide functions such as authentication, authorization, network access control, routing, reliable request retransmission, and security.
- **A redirect server**, which provides the client with information about the next hop or hops that a message should take, and then the client contacts the next hop server or UAS directly.
- **Registrar server,** which processes requests from UACs for registration of their current location. Registrar servers are often co-located with a redirect or proxy server.

4.2 What Is the Role of the SIP Application Server?

The application server provides access to service logic through a CGI, Servlet, or proprietary interface. The service logic is free to access business logic and databases using, for example, Corba, SOAP, JDBC, or Socket. The application server might be in a third party domain.

This is the key element for Nokia, especially for Nokia mobile phones since this enables new exciting and cool services and a tempting service creation environment. One problem is that if the terminal market becomes too heterogeneous, interoperating is difficult. Terminal profiles might be useful (for example, defining a set of codecs, SIP extensions, and display options).

4.3 What Is the Role of the Proxy Server?

A SIP proxy operates in a similar way to a proxy in HTTP and other Internet protocols. A SIP proxy does not set up or terminate sessions, but it sits in the middle of a SIP message exchange, receiving messages and forwarding them. There may be multiple proxies in a signaling path. The proxy server is not really "in the call." It facilitates the two end-points locating and contacting each other, but it can drop out of the signaling path as soon as it no longer adds any value to the exchange.

Note that the media session is always end-to-end and does not go through a proxy. In SIP, the path of signaling messages is totally independent of the path of the media. In telephony, this is described as the separation of control channel and bearer channel.

4.4 What Is the SIP Gateway and What Is Its Role?

A SIP gateway is an application that interfaces a SIP network to a network utilizing another signaling protocol. In terms of the SIP protocol, a gateway is just a special type of user agent, where the user agent acts on behalf of another protocol rather than human. A gateway terminates the SIP signaling path and can also terminate the media path, although this is not always the case. For example, a SIP to H.232 gateway terminates the SIP signaling path and converts the signaling to H.232, but the SIP user agent and H.232 terminal can exchange RTP media information directly with each other without going through the gateway.

4.5 What Does the Application Server Do?

For example, a Game server enables new games. Users can request what games are available, download them, or they can request the availability for other players to join an ongoing game. SIP can also be the actual game transfer protocol (for example, chess moves). Moreover, other SIP features (for example, presence) can be easily integrated, so during the game the user can use SIP calls, SIP messaging, SIP presence, and SIP gaming features.

Furthermore, a chat server may provide one or more chat services (for example, different chat groups for different customer groups).

4.6 What Does the Rich Call Server Do?

The rich call setup server provides rich call setup features. For example, a user may want that for every outgoing call, the Star Wars theme is attached to the message as in-line content for ringing tone purposes. This saves the air interface resources since there is no need to send it each time over the air interface. Other complex call-setup service scripts can also be executed here. This saves Call State Control Function (CSCF) resources.

4.7 What Does the Location Server Do?

The server for location-based data can provide location-based information (for example, maps or commercials) to consumers.

4.8 What Does the Messaging Server Do?

The messaging server has a similar functionality as the Short Message Service Center (SMSC). All messages coming from the outside world are forward to this server. It analyzes the message, checks service scripts (such as “If size of inline image>20k, send it to me”) and access rights from the subscription profile (for example, the user is not allowed to receive any messages or the user is out-of-reach), and consults the charging server. It is also possible that the message content is stored in the server and re-sent later, or that the content is transcoded (for example, gif-to-jpeg). However, in session-based messaging the messaging server is bypassed as messages go directly from user A to user B, with, for example, an instant messaging (IM) application.

4.9 What Does the Presence Server Do?

A presence server is a SIP server that maintains the presence status. The presence information can be accessed or updated with other protocols as well. The user can inform his or her presence status to the server, and query the presence status of other users. The presence status can be simple (“online” or “offline”) or more detailed, such as available, available for chat, or in office. It can also include additional information, such as terminal capabilities (for example, gif, jpeg, amr, or h.263 qcif) or general features (for example, mobile/fixed and bandwidth). The presence server can utilize information from the home subscriber server (HSS). This is, however, possible only for operator-hosted servers.

5 SIP Applications and Services

5.1 Why Is SIP Suitable for Services?

Since the underlying elements of SIP have much in common with HTTP, creating network-based services, such as time-dependent call forwarding, is quick and straightforward. Designing and implementing new SIP-based voice services is as easy as creating Web pages.

5.2 What Benefits Does SIP Provide for Application Developers?

By not requiring major hardware upgrades to application servers, but rather enabling new software-based services using SIP, service providers can reduce the time associated with deploying new features from months to days. For subscribers, this means ever-improving communications service, plus lower initial and recurring telephone service costs.

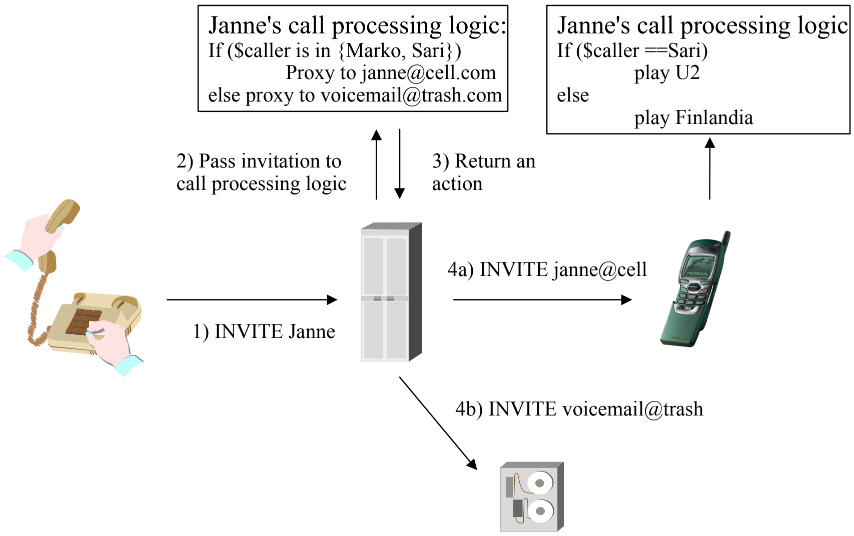
5.3 How Are Applications for SIP Developed?

SIP applications are developed with basic Web methods, such as Servlet, CGI, and call processing language (CPL).

5.4 What Are Examples of Call Processing Services?

- “Discard all calls from Marko during my business hours.”
- “Redirect authenticated friends to my mobile phone, anyone else to my secretary. ”
- “If I’m busy, return my homepage and redirect the caller to my voice mail. ”

5.5 What Would Be an Example of Call Processing Logic?



5.6 Why Is SIP Better for Service Creation When Compared to Other Protocols?

With a SIP enabled network, it is possible for the operator to seamlessly integrate speech, chat, location, messaging, URL-based addressing, and resources. SIP also allows efficient and easy 3rd party service creation.

5.7 What Are Instant Messaging (IM) and Presence?

IM and presence are extremely popular applications in the fixed Internet. Enhancing today's Short Messaging Service (SMS), IM enables messaging that is enhanced by presence information. In this model, users are aware of each other's status, which provides better service and supports new types of use scenarios. Messages can be sent real-time to the B-party whose on-line status can be checked with the presence feature.

The B-party can select various presence modes (for example, busy, available for gaming, or meeting) and share this information with predefined friends or colleagues.

IM can be categorized into fixed and wireless instant messaging. In the fixed arena, AOL is today a clear market leader in USA. Other IM providers in the Internet are Mirabilis (ICQ, owned by AOL), Yahoo, Cybiko, and Bantu. On the wireless messaging side, Ericsson is rather active with its wireless version of iPulse.

5.8 Can SIP Be Used for Instant Messaging (IM) and Presence? If Yes, What Are the Basic Capabilities in SIP That Make It Suitable?

Yes, SIP can be used for IM and presence. A dedicated MESSAGE method, which can carry any MIME content (such as text/plain or image/gif), has been defined by IETF for messaging purposes.

Presence can be described as dynamic status information about the user (addresses, call/connection status, terminal capabilities, and location) that can be used for controlling services and viewed by other parties.

Note that SIP enables the use of presence features. However, the management of presence characteristics, such as buddy list editing, is typically carried out with tools based on HTTP or WAP.

With SIP, instant or multimedia messaging can be seamlessly integrated with conversational communications thus creating the concept of *ubiquitous communications*; the seamless usage of various real-time communications media and messaging services depending on the presence information of the parties.

5.9 What Elements Are Needed for Using Presence?

The following requirements must be fulfilled for using presence:

- The network must be able to identify users independent of their location.
- Subscription requests can be forwarded to the server handling that user.
- The user must be able to indicate the server his or her location and other presence data.
- The network can forward notifications to subscribers.
- The network must be scalable.
- The network is able to deliver messages real-time.

5.10 What Does SIP Provide for Presence?

- Identifies users independent of their location
- Forwards requests (INVITE or otherwise) to the server handling the user
- The REGISTER method allows the network to the tell server its location and other information
- Can forward messages back to originators in reverse direction
- Scales
- Delivers messages in real-time

5.11 What Would Be an Example of Presence?

User A registers to follow user B's presence information. When user A logs on to the network, user B gets a notification of it by an instant message.

User B wants to call user A. Same as before, but A's service logic checks A's availability. It finds out that A is in a meeting and accepts only instant messages and e-mails. B gets a message back saying that A accepts only instant messages and e-mails. B sends an instant message to A (Message). A answers with an instant message and a chatting session is going on. The local CSCF is charging 15 cents per message.

5.12 What Would Be an Example of the Jukebox Service?

User B browses a web page and finds a video clip that he wants to play. He clicks on the link (sip:mikael@jukebox.nokia.com), and an INVITE message is sent to the jukebox which selects the appropriate codec from the invite Session Description Protocol (SDP) and sends back a 200OK with no media (one way stream). The video starts playing. When B wants to stop the video, a BYE message is sent to the server (the same SIP rules as for a call apply).

5.13 What Would Be an Example of the Automatic Reply Service?

User B is a keen soccer fan and he offers an automatic soccer result service. When other users send him a SIP message which includes the keyword "soccer," the reply is an automatically generated HTML or ASCII page including the latest soccer results. The keyword can be, for example, in the Subject header, or inside another header. These kind of services should be easily offered by, for example, teenagers using their standard Nokia mobile phones.

5.14 What Would Be an Example of Instant Messaging (IM)?

The user checks the availability of the respondent with his or her presence status, decides to send a message with his homepage URL, and sends it. The receiver receives the message instantly, and has the possibility to proceed to the given URL by clicking the hyperlink in the message.

A simpler (and more typical) example is just send an instant message to somebody without first checking the presence.

5.15 What Would Be an Example of Rich Call?

The basic definition of "rich call" is that when establishing a connection between two users, that is, making a call, elements other than voice can be added to the communication. Now, when you call someone on a GSM mobile phone, the caller's phone number is usually displayed on the display of the recipient's phone. Instead of just having the number there, the caller could send his or her business card, a photo, or any other extra element. Furthermore, during the phone call, the callers could exchange information such as documents, data or images, or other media. In short, rich call is the term for broadening the definition of a phone call.

5.16 Why Is SIP Better in 3G Networks than the Well-Proven IN Technologies?

IN was developed for 2G networks and protocols, and especially for services related to call handling. In 3G networks the mobile phone is nothing but just one terminal among others (for example, TV, laptop, or PDA), that can be used to access information and services. IN cannot easily serve these terminals, too. In addition to that, IN systems are rather closed, thus making it difficult to build services easily on top of them.

5.17 Real-Time Voice Will Surely Remain Circuit-Switched for Many Years. Therefore, IN Will Have a Dominating Role, Won't It?

The transition to IP voice in All-IP networks will take time. Most likely new services are introduced first. Surely, there will be operators who offer VoIP straight from the beginning in their 3G networks.

However, it is possible to signal a circuit-switched call with SIP.

5.18 Having a Mobile Network Full of Customers Surely Means That the Service Architecture Cannot Handle Them All?

Remember that part of the service intelligence will shift towards terminals, thus reducing the network load.

Scalability is always an issue – and an opportunity for those players who get a reputation for reliable systems.

In SIP, basically the same methods are used as in HTTP (for example, load balancing for creating scalable solutions).

5.19 How and Where Are 3rd Party Services Provided?

3rd party services can be provided using SIP in the same way as Internet-based services are being provided at the moment. There are many choices for that. For example, operator-hosted (and trusted) services can be installed locally inside the 3G core network, usually on separate workstations (Unix, PC) which include basic SIP proxy software and some additional software on the top of the basic stack. The SIP request can be forwarded to the 3rd party service provider using the record/route mechanism to be aware of the future actions.

3rd party services can also be located elsewhere in the Internet (for example, Yahoo). Users can communicate with these service providers directly, but operator-specific information is usually not available.

5.20 Who Is Responsible for Providing 3rd Party Services?

Anybody, in the same way as services are provided in the Internet at the moment.

5.21 What Would Be a Typical Call Scenario with SIP?

Consider the scenario in Figure 2. In our example, the caller (jdrosen@dynamicsoft.com) wishes to place a call to joe@columbia.edu. Jdrosen sends his SIP INVITE message to the proxy for dynamicsoft.com (Step 1). This proxy then forwards the request out to Columbia, where it reaches the Columbia.edu server (Step 2). This server is actually not a proxy, but a similar device called a redirect server. Instead of forwarding calls, a redirect server asks the requestor to contact the next server directly. The Columbia.edu server looks up Joe in its database, and determines that today, Joe is on sabbatical to foo.com. It therefore sends a special response, called a redirect, to the dynamicsoft.com proxy, instructing it to try joe@foo.com instead (Step 3).

The dynamicsoft proxy then acts on this response, which means it directly tries to contact joe@foo.com. So, its sends the INVITE to the foo.com server (Step 4). This server consults its database (Step 5), and learns (Step 6) that Joe is actually in sales. So, it constructs a new URL, joe@sales.foo.com, and sends the INVITE to the sales.foo.com proxy (Step 7).

The proxy for the sales department then needs to forward the INVITE to the PC where Joe is currently sitting. How does it know which PC Joe is at? SIP defines another request, called REGISTER, which is used to inform a proxy of an address binding. In this case, when Joe turned on his SIP client on his PC, the client would register the binding sip:joe@sales.engineering.com to sip:joe@mypc.sales.foo.com. This would allow the proxy to know that Joe is actually at mypc, a specific host on the network. The bindings registered through SIP are periodically refreshed, so that if the PC crashes, the binding is eventually removed.

The sales.foo.com proxy consults this registration database, and forwards the INVITE to joe@mypc.sales.foo.com (Step 8). This INVITE then reaches Joe at his PC. Joe can then respond to it (thus the request-response model). SIP provides many responses, and these include acceptance, rejection, redirection, busy, and so on. The response is forwarded back through the proxies to the original caller (Steps 9, 10, 11, and 12). An acknowledgement is sent (another type of request, called ACK) in Step 13, and the session is established. Media can then flow (Step 14).

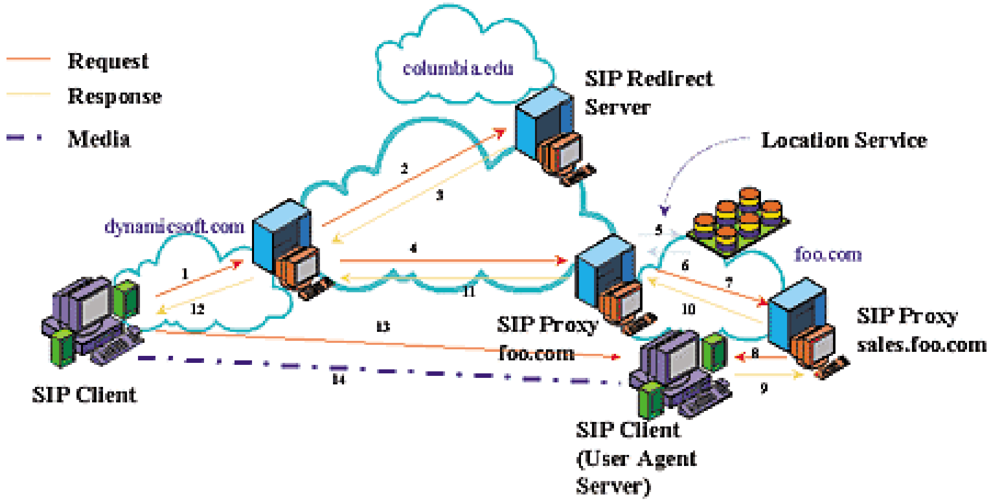


Figure 2: Call phases with SIP

5.22 How Can Applications Utilize SIP? Is SIP Alone Sufficient for a Mobile Terminal – HTTP Connection?

SIP applications in the terminal are built on top of the SIP API. The SIP API provides the access to the terminal's SIP user agent functionality, that is, applications can send and receive SIP messaging. The SIP user agent needs a mechanism that routes incoming SIP messages to the correct application, for example, MESSAGE would be routed to the instant messaging (IM) application. Other features (such as codecs and content parsers) are needed on application basis.

SIP user agent implementation requires:

- specified standard support
- flexibility to function with all kinds of SIP servers (even with faulty ones)
- a method how applications can broaden their SIP support, for example, in form of new methods, headers, and content formats.

5.23 What Is SIP's Position in the 3G Networks?

SIP software is mandatory in a 3G terminal (which supports the ALL IP network architecture, 3GPP Release 4/Release 5).

5.24 When Does Using SIP Start?

SIP will impact 3GPP releases as described in the following table. Note that the table has been compiled from a real-time IP voice call point of view.

SIP could naturally be used today, for example, for messaging over High-Speed Circuit Switched Data (HSCSD) with a suitable implementation. In fact, it can be assumed that competition will be out in SIP-over-GPRS solutions.

3GPP Release	Key new functionality for real-time IP call	Specification/ infrastructure ready	Comments
Release 3	QoS defined for 3G PS services, including real-time.	Infrastructure 2001-2	Release 99 specification does not adequately cover packet-side real-time handover.
Release 4	RTP header compression, enhanced packet side handover.	Estimate: specification 3/2001 and infrastructure 9/2002[1]	QoS for IP voice works in the intra-operator domain, which is actually fine for operators. De facto implementation for SIP would be needed for terminals.
Release 5	Usage of SIP standardized, potentially end-to-end QoS.	Specification 12/01, infrastructure 06/03	Network implementation may support both de facto Release 4 and 3GPP standardized Release 5 SIP terminals.

[1] Assuming 18 months from specification completion to infrastructure implementation. 3GPP schedule is subject to change.

6 SIP in Relation to Other Protocols

6.1 Will 3GPP Create Its Own Version of SIP? If Yes, Why?

3GPP will use IETF SIP as the basis for 3GPP signaling. However, IETF SIP includes a lot of optional functionality and it does not define the exact set of signaling needed to set up a QoS-enabled and authenticated session. Different environments (such as cable modem networks or 3G networks) will use the basic SIP and define the necessary extensions and limitations before using it. For example, quality of service (and authentication) is provided in a different way in different environments and defining how it relates to SIP is a network-specific issue.

3GPP has defined that SIP over UDP is used for call control (TCP is not supported for call control, which has been defined in the IETF SIP specification). Similar decisions are expected from 3GPP for other SIP details.

3GPP will create a 3G-profiled version of the SIP toolkit.

6.2 What Is SIP's Relation to WAP?

On the surface, SIP seems to compete with WAP. However, SIP can be positioned as an evolution of WAP today; WML pages can be sent in SIP payload and WML scripts can be forwarded to the Internet (SIP, e-mail). Client-to-client communication and multicast are possible in SIP, and an application can use WML, HTML, or whatever content depending on the situation.

In addition to that, WAP security mechanisms can be utilized for SIP services and WAP user agent (UA) profiles can be converged with SIP.

6.3 Are WAP and SIP Competing?

WML could be carried inside SIP messages, thus combining the best features of WAP and SIP. WML is a nice and effective syntax, for example, for menus, and SIP is the one and only transport method which can also include intelligent features. Depending on the situation, SIP can carry WML or another payload (for example, HTML or gif).

SIP and other IP services will probably be included in 3G as well and they will replace WAP services if WAP does not evolve.

However, it is possible to combine WAP with these IP services and thus protect existing WAP investments and improve WAP further. SIP offers additional advantages over WAP (for example, direct client-to-client communication, multicast, and better and more seamless Internet interoperability, such as forwarding SIP/WML messages in the Internet and loop detection). On the other hand, WAP offers strong a security framework.

6.4 Are There Any Similarities Between SIP and HTTP?

SIP and HTTP stacks have a lot in common. An advanced SIP implementation usually provides a simple HTTP stack. One example of HTTP and SIP co-operation is that after SIP signaling, a large file is fetched over HTTP, using the HTTP GET method. Moreover, a SIP client (for example, a mobile phone) can host a simple HTTP server and offer, for example, files so that other parties can download information from the client (embedded Web server).

SIP is working well in conjunction with the Web (MIME, Uniform Resource Identifier (URI), and Domain Name System (DNS)). De facto, a lot of SIP services (such as click to call or Web phone book) could be originated by Web pages. SIP services can also use the resident client Web/WAP browser to present information (for example, call presentation service, messages, or pictures). SIP also enables “multicast Web,” for example, sending Web pages on top of SIP.

6.5 Can Java™ Applets Be Transported as Payload?

SIP and Java are closely linked together. Java applets can be transported as SIP payload, and Java applets running in a mobile phone can use SIP based services in communications.

A simple use case would be distributing a mobile game, checking availability for gaming, setting up the game, and playing against another person over SIP. In a game such as chess, the moves can be transported using SIP. In games requiring more “bandwidth,” SIP can be used for signaling TCP/UDP connection.

6.6 Does SIP Support Servlets?

A SIP Servlet Application Server (AS) is, in principal, comparable to CGI. The advantage of the SIP Servlet AS is in the use of the Java tools and security. The Java service logic as access to enterprise Java beans, database connection (JDBC), Jini. Standard development tools (for example, JBuilder or Visual Café) can be used for service creation. Servlet development is well known by many developers. Finally, Servlet is more effective than CGI.

6.7 Can SIP Replace E-mail Protocols in the Mobile Phone?

In theory, e-mail can be delivered to mobile phones in SIP payload just like in i-mode mail or, for example, WAP 1.X MMS can be used for e-mail transport.

This is an issue that markets will decide.

6.8 What Is the Relationship Between SIP and Existing/Developing Standards Such As UPnP, Bluetooth, 802.11, HAVi, HomePNA, HomePlug, and HomeRF?

SIP is used for locating devices within a given logical environment, which may well consist of disparate physical networks, with widely different characteristics and underlying protocol systems (for example, any of the above-mentioned protocols). SIP for Networked Appliances is intended to give a uniform means of access to appliances in a realm without being tied to any particular one of them.

6.9 Can H.323 and SIP Be Used Together?

Yes. SIP can locate the called party and determine its capabilities, including H.323. H.323 is then used to connect the two parties. Unfortunately, there is currently no specification on translating between the two. Conversion is made more difficult by the multiple versions of H.323 (v1, v2, v3). However, there is at least one product (Lucent PacketStar IP) that allows SIP and H.323 terminals to call each other.

There will be SIP/H.323 gateways in All-IP mobile networks. However, it is unlikely that there will be cellular H.323 terminals in the market, and it is likely that eventually SIP will overcome H.323 in fixed and WLAN implementations.

7 More Detailed Questions About the Technology

7.1 Extensions – What Do They Mean?

There is a lot of work in extending SIP in IETF and in the PacketCable forum. 3GPP is also starting its work to extend SIP (at least some minor extensions are needed).

SIP is easy to extend since it does not break interoperability. SIP can be extended in several ways. First, new methods (such as MESSAGE, PRACK, and DO) can be defined. This is the recommended way to provide new features in many cases because the cost of the new method is minimal. Second, new SIP headers (such as Dcs-Billing or Alert-Info) can be defined. Third, new content formats can be used in addition to the Session Description Protocol (SDP) (for call signaling) but basically this is not a SIP specific issue since any MIME content format can be carried in SIP payload. The content format is defined in the SIP header Content-Type:. Examples of new MIME formats relevant to SIP are presence format, animation formats, and next generation Session Description Protocol (SDP).

These extensions do not need to be understood by SIP network elements since they only forward SIP messages. However, some network elements may parse SDP information from SIP payload so they must understand also other formats than SDP in payload if they are used.

Another form of SIP extension work is to define the usage of SIP in a new context. For example, there is on-going work in IETF to use SIP to control networked appliances (for example, “turn the lamp on”). A new method (DO) has been proposed for this idea. Similar ideas (using SIP for new kind of services) will probably follow. Whether SIP is the correct solution for these new areas is still unclear.

These extensions can also be company-proprietary, for example, Microsoft clients might include Microsoft specific headers in SIP messages thus enabling better service, if both users are using Microsoft products. Note that this concerns e-mail and HTTP as well.

7.2 Can't It Be Said That SIP Has Been Extended, and Will Be Extended, from Its Original Scope?

That is correct. We must study how SIP performs, for example, if there is payload in call setup messages. On the other hand, having payload (such as images) we can create a service with consumer benefit.

7.3 Is SIP Messaging Reliable?

SIP has a defined reliability mechanism, which allows the use of unreliable transport layer protocols such as user datagram protocol (UDP). When SIP uses TCP, these mechanisms are not used, since it is assumed that TCP will retransmit the message if it is lost and inform the client if the server is unreachable. Reliability mechanisms in SIP include:

- retransmission timers
- increasing command sequence Cseq numbers and
- positive acknowledgements.

7.4 Can Geographical Location Information Be Transported in SIP?

Yes, geographical information could be transported as payload in a SIP message.

7.5 What Does SIP Being a Peer-to-Peer Protocol Mean?

The peers in a session are called User Agents (UAs). A UA can function in one of the following roles:

- User agent client (UAC): A client application that initiates the SIP request.
- User agent server (UAS): A server application that contacts the user when a SIP request is received and that returns a response on behalf of the user.

Typically, a SIP end point is capable of functioning as both a UAC and a UAS, but functions only as one or the other per transaction. Whether the endpoint functions as a UAC or a UAS depends on the UA that initiated the request.

7.6 What Does SIP Being a "Client-to-Client" (vs. HTTP Being Client-to-Server) Protocol Mean?

In HTTP, the browser always acts as a HTTP client, and the responding server always as a HTTP server. However, in SIP the entity may have both roles. When the sender originates the INVITE request, he or she is acting as a SIP client. When the receiver responds to the request, he or she is acting as a SIP server. The roles are opposite, when a media session has been established and ended, and the receiver establishes a BYE request. This is why a SIP enabled device must contain both SIP server and client software. In SIP, an endpoint will switch back and forth during a session between being a client and a server.

SIP is not trying to replace HTTP as a protocol—they complement each other. Also in SIP, a WWW style service creation is enabled.

7.7 Is Streaming Possible with SIP? What Is SIP's Role?

Yes, streaming is possible with SIP. Using the INVITE message you can open up media streams (for example, to collect video from a remote camera). However, real-time streaming protocol (RTSP) is the actual streaming protocol.

7.8 Is Messaging Possible with SIP?

Yes, There are three different ways to implement messaging services with SIP:

- Content inside SIP messages (in-line content)
- Content in separate data (TCP/UDP) session signaled using SIP
- Content in a place specified by a URL (for example, in real-time streaming protocol (RTSP), SIP, or Web server)

Messages can be sent as an in-line attachment in SIP signaling in the MESSAGE method. This is a useful scheme for small messages, but problematic for bigger ones. This can, however, be enhanced so that small messages are carried in-line, and bigger messages are only referenced (URL in MESSAGE method) for retrieval. The user may have set in his profile whether to fetch the content instantly on screen or to send it to the mailbox for later retrieval.

Another way is to use SIP only to establish a new uni-directional data session. This way, when the messaging server receives a message targeted for the mobile client, it sends the SIP INVITE message to the client specifying that a uni-directional data session using codec X (for example, jpeg) is needed.

7.9 Can Users Be Prioritized with SIP? What Does It Mean?

The proxy (or CPS) can offer better service for prioritized users (for example, by providing better QoS or by giving prioritized handling inside CPS/UMS). Also, in case of congestion, users with low priority can be dropped first. These decisions are made in CPS and are out of SIP's scope.

7.10 What Is CPL?

CPL stands for Call Processing Language (IETF RFC 3050). It is specified by the IP Telephony working group in the IETF standardization organization. CPL enables screening, forwarding, filtering, and notification types of services. For example, a CPL specifying that calls from Joe are forwarded to a unified messaging server after 5 p.m., and all other calls are routed to a PDA can be written. The specification of the service is method independent; this means that this CPL and all the related CPL infrastructure in the network would allow this service to be extended to instant messaging with no additional work. This would allow users to customize their own instant messaging (IM), presence, and multimedia services with the same tools.

7.11 What Information Is Visible in Each SIP Message?

An intercepted SIP message reveals both parties' SIP URLs and IP addresses, the fact that two parties have established a call or another type of session, and the IP addresses and port numbers associated with the media, thus allowing eavesdropping.

However, SIP messages can be encrypted or they can include network-modified information (for example, the terminals' IP addresses are replaced by proxies' addresses).

7.12 Does SIP Support Encrypting?

Yes, SIP supports the encryption of both message bodies and message headers. The encryption of message bodies makes it more difficult for an eavesdropper to listen in. Also, an uninvited third party, knowing all the Session Description Protocol (SDP) information could guess the real-time transport protocol synchronization source (RTP SSRC) number and send unwanted media to either party, practicing so called "media spamming." Encryption of headers can hide the parties involved in the session.

In addition to that, actual session stream (such as voice) can be encrypted as negotiated by SIP.

7.13 Is SIP Addressing Secure?

By placing this information in the username portion of a SIP URL, encryption (for example, using PGP) of the "sensitive" information is enabled while still retaining the routing capability of the SIP proxies. Of course the base 64 encoding would not be required if you encrypt the field and all the characters are RFC compliant (for example, [slp:/d=lamp,r=bedroom,u=stanm]@stanm.home.net ==> 5n3PwPFV793Bd6EmgO@stanm.home.net).

7.14 How Is the Caller ID Implemented?

The caller ID is provided by the *FROM* SIP header containing the caller's name and "number." The number would most likely be placed in the user field of a SIP URL or appear in a *tel: URL*.

Since the callee generally does not know or trust the caller's server, only cryptographic signatures can be used to ensure that the information is valid. For example, the outgoing proxy might be operated by an Internet service provider (ISP), enterprise, or phone company and sign for the identity of the caller, using the *signedby* parameter, with the identity of the company verified by a public key certificate similar to those used by Web sites.

7.15 How Is Authentication Done in SIP?

Authentication in SIP can take two forms. One is authentication of a user agent by a proxy, redirect, or registration server. The other is the authentication of a user agent by another user agent. A proxy or redirect server can require authentication to allow a user agent to access a service or feature.

7.16 How Does SIP Get Through a Firewall?

There are several possible approaches to SIP-capable firewalls. One of the difficulties is that, unlike for, say, HTTP, connections are originated both by hosts inside and outside the firewall. A likely arrangement is that a SIP proxy sits “on” the firewall and relays SIP requests between the Internet and the intranet. This proxy would also open up the necessary ports in the firewall to let audio and video flow through, for example, using Socks V5.

As an alternative, if a firewall or network address translator (NAT) allows outgoing TCP connections, the inside client can open up a TCP connection to an outside proxy. All outgoing and incoming calls would then be handled by that TCP connection. (The client would still have to use SOCKS or a similar mechanism to convince the firewall to let RTP packets through.)

7.17 Is SIP Used As a General Purpose RPC Mechanism?

SIP is not used as a general purpose Remote Procedure Call (RPC) mechanism. However, SIP is used as a messaging system. There is a distinction: in a general purpose RPC mechanism, the message body defines a function and parameters and the response returns a function, possibly structured. However, the body of the SIP message does not define a function or parameters, and the response does not return a function. In the case of SIP, the originating User Agent (UA) does not 'call' a function at the terminating UA, but it sends control information to it, carried in the SIP message body. Similarly, the response is not a function, but another message. The terminating UA is not in any way under the control of the UA sending the SIP message.

7.18 What Are SIP's Request Methods?

There are several methods in SIP and new methods are created all the time. IETF RFC 2543 (basic SIP) defines:

- INVITE - Indicates that a user or service is being invited to participate in a call session.
- ACK - (Acknowledge) Confirms that the client has received a final response to an INVITE request.
- BYE - Terminates a call and can be sent by either the caller or the callee.
- CANCEL - Cancels any pending searches but does not terminate a call that has already been accepted.
- OPTIONS - Queries the capabilities of servers.
- REGISTER - Registers the address listed in the To header field with a SIP server. Gateways do not support the REGISTER method.

All SIP methods, including the INVITE method, can be carried via TCP or UDP to its destination. The following table gives a detailed description of the SIP INVITE method.

Line	Description
INVITE sip:12125551111@192.168.1.2 SIP/2.0	Method type, request URI, and SIP version.
Call-ID:917300000239@192.168.1.2	Globally unique ID for this call.
Content-Length:171	Length of the body of the SIP method.
Content-Type:application/sdp	The body type (an SDP message).
From:sip:17325551234@192.168.1.2; tag=c0-a8-1-2-3ea24151-35f3	User originating the request.

To:sip:12125551111@192.168.1.2	User being invited into the call.
Via:SIP/2.0/UDP 192.168.1.2:7300	IP address and port of previous hop.
	Blank line separates header from body.
v=0	SDP version.
s=Incoming phone call from smayer	The name of the session.
p=+1 732-555-1234	Phone number of caller.
c=IN IP4 192.168.1.2	Connection information.
t=3126288799 3126289399	Time the session is active.
m=audio 49170 RTP/AVP 0	Media name and transport address.

7.19 What Is MIME and Does SIP Support It?

MIME stands for Multipurpose Internet Mail Extensions. MIME extends the format of Internet mail to allow non-US-ASCII textual messages, non-textual messages, multipart message bodies, and non-US-ASCII information in message headers.

SIP supports MIME headers.

7.20 What Does Multicast Mean, and Is SIP Multicast Capable?

With a multicast transport service, a single node can send data to many destinations by making just a single call to the transport service, that is, multicast is point-to-multipoint messaging.

SIP is multicast capable.

7.21 When I Send a Message in the SIP Network, Do I Know if the Message Reached the Recipient?

Yes, you get a 200 OK reply. Otherwise, the error is reported by the user agent (for example, “user could not be reached”).

7.22 What Does Abstract Addressing/Naming Mean, and Does SIP Support It?

Devices behind the residential gateway/firewall/network address translator (NAT) may not be directly addressable because their addresses may not be known, they may have private IP addresses, or even non-IP addresses. For these cases, SIP provides the abstract addressing method.

7.23 What Communications Methods Does SIP Provide?

SIP supports the following different communications mechanisms for devices:

- Queries (for example, device state) – “What is the temperature in the house?”
- Asynchronous events ("notification") – “Notify me when the security alarm goes off.”

7.24 What Does All-IP Mean?

All-IP means building the mobile system fully using packet oriented Internet Protocol (IP) technology, and in particular that the mobile devices will be IP terminals. This boils down to that all mobile services and applications will be developed using IP, the very underlying technique of the Internet. This will enable the mobile operators to provide a much richer variety of end user services. All-IP will make Nokia's vision of the Mobile Information Society a reality, where end users are able to access mobile Internet services seamlessly, independent of time and place.

7.25 What Is Meant by IP Multimedia?

IP multimedia means the exchange of any digitalized information between the communication parties with a variety of different communication modes, such as interactive communication, streaming, and sharing. Examples include video call, sharing a whiteboard or a Web page during a phone call, or sharing a still image during a phone call.

7.26 Is IPv6 the Environment in Which SIP Will Be Needed? How About the Ipv4 Environment?

Technically, SIP will function across both IPv4 and IPv6 networks. However, Nokia believes that only IPv6 will be able to deliver the optimum environment and user experience to proliferate rich call services. Hence we would like to take this opportunity to further endorse our views on IPv6. For further information, please visit http://www.nokia.com/ipv6/

7.27 What Is SIP's Relationship to Multimedia Messaging?

The Multimedia Messaging Service (MMS) specification covers both the presentation and the transmission of a multimedia message. SIP is one possible future transport protocol for MMS messages, but this technology choice has not yet been made. Today, WAP was the chosen transport solution for MMS, because it was the best solution available. In 2.5G with GPRS and in the beginning of 3G, the WAP transport will be optimal for MMS. This is because mobile networks are evolving and also because of the maturity of the WAP specification. Furthermore, the WAP specification is also evolving and the convergence of technologies can be expected. Nokia's intention is to work in open standardization bodies, such as the WAP Forum and IETF, to create the best possible solutions for future services.

8 SIP Forum

8.1 What Is the SIP Forum?

The SIP Forum is the promoter association of SIP. The SIP Forum does not participate in the standardization work. The Internet Engineering Task Force (IETF) is the body developing the protocol.

8.2 How Many Members Does the SIP Forum Currently Have?

About 20-30 companies, including most other key players in the telecommunications industry, have already joined the SIP Forum. For more information on the members, please visit http://www.sipforum.org/

8.3 What Are the Business Areas They Represent?

Members of the SIP Forum are, for example, telecom vendors, software companies, and IP equipment vendors.

8.4 All the Other Leading Mobile Telecom Companies Seem to Be There Already. Why Did Nokia Join So Late?

Nokia has been active in developing basic technologies for SIP for a long time already. We have also participated in the interoperability testing events, the so-called bake-offs. Having carefully evaluated SIP Forum's mission and objectives, Nokia concluded that the company wants to actively participate and drive the work of the SIP Forum.

8.5 Why Do We Need Yet Another Acronym, SIP? Couldn't These Issues Be Addressed in Other Existing Forums?

At times, it may seem confusing that new acronyms keep popping up all the time. The reason for this is, however, fairly straightforward. As digital technologies converge, different companies from different fields become involved. In the SIP Forum, for instance, various IP infrastructure manufacturers are heavily represented, and hence this is the right team to address SIP related questions to. Likewise, the WAP Forum, for instance, has a different objective and involves a different set of companies.

9 Where Can I Find More Information on SIP?

- Columbia FAQ http://www.cs.columbia.edu/sip
- SIP Forum homepage http://www.sipforum.org

10 Terms and Abbreviations

Term or abbreviation	Meaning
3GPP	Third Generation Partnership Project.
AOL	American Online.
API	Application Program Interface.
ASCII	American Standard Code for Information Interchange.
CGI	Common Gateway Interface.
CPL	Call Processing Language.
CSCF	Call State Control Function
DNS	Domain Name System.
ETSI	European Telecommunications Standards Institute.
HLR	Home Location Register.
HTTP	Hypertext Transfer Protocol.
IETF	Internet Engineering Task Force.
ITU	International Telecommunication Union.
IM	Instant Messaging.
IP	Internet Protocol.
LDAP	Lightweight Directory Access Protocol.
MIME	Multipurpose Internet Mail Extension.
PDA	Personal Digital Assistant.
PSTN	Public Switched Telephone Network.
QoS	Quality of Service.
RFC	Request for Comments.
RTP	Internet-standard protocol for the transport of real-time data, including audio and video. It can be used for media-on-demand as well as interactive services such as Internet telephony. RTP consists of a data and a control part. The latter is called RTCP.
RTSP	Real Time Streaming Protocol - The Real-Time Streaming Protocol allows controling multimedia streams delivered, for example, via RTP. Control includes absolute positioning within the media stream, recording and possibly device control.
SDP	Session Description Protocol - SDP is an ASCII-based protocol that describes multimedia sessions and their related scheduling information.
SMS	Short Message Service.
SMSC	Short Message Service Center.
TCP	Transmission Control Protocol.
UAC	User Agent Client.

UAS	User Agent Server.
UDP	User Datagram Protocol.
UPnP	Universal Plug'n'Play.
URL	Uniform Resource Locator.
VoIP	Voice over IP.
VPN	Virtual Private Network.
WAP	Wireless Application Protocol.