

rt-solutions.de
networks you can trust.



Voice over IP Security

An Overview

N. Simantirakis
Cologne, March 2006

Veröffentlichung mit freundlicher Genehmigung der
rt-solutions.de GmbH*
Oberländer Ufer 190a
D-50968 Köln

*rt-solutions.de GmbH ist Technologiepartner der MCA GmbH mit wissenschaftlicher Expertise der Universität Magdeburg, der Fachhochschule Wiesbaden und der Fachhochschule der Wirtschaft Bergisch-Gladbach. Kompetenzfelder: Security Consulting, WAN & Application Performance, Advanced Wireless und Networks in Automation.

Table of Contents

Introduction	3
A. Protocols	3
B. End-user devices	7
C. Network	7
Threats	8
A. Social Threats	8
B. Eavesdropping	9
C. Interception and Modification	9
D. Intentional Interruption of Service	9
E. Unintentional Interruption of Service	10
Related Work & Protocols	10
A. VPNs	10
B. VLANs	11
C. SIP Security	11
A. <i>Secure RTP (SRTP)</i>	11
B. <i>DTLS over RTP</i>	11
C. <i>MIKEY</i>	12
D. <i>sdescriptions, SIPS, S/MIME</i>	12
E. <i>ZRTP</i>	13
F. <i>DTLS Handshake</i>	13
D. Skype Security, Google Talk Security	14
Open Issues.....	15

Introduction

Voice over Internet Protocol (Voice over IP or VoIP) is the fast-growing technology which consists of routing voice conversations over the Internet or any other IP-based network. The voice data is transmitted over a general-purpose packet-switched network, instead of traditional dedicated, circuit-switched voice transmission lines. The convergence of data and voice networks can lead primarily to cost reduction and added services, such as parallel intelligent access to the data in the network. VoIP has already acquired a significant market share, though security concerns have unfortunately not yet properly been addressed.

A. Protocols

The most common protocols used for VoIP communication are H.323 and SIP. SIP has clearly begun putting the older H.323 aside, largely due to its comprehending architecture. SIP is a signaling protocol for virtually every kind of media stream using a request-reply (client-server) model for call setup, while its messages are in a human-readable format, pretty similar to HTTP ones. Two logical entities, a *User Agent Client (UAC)* and a *User Agent Server (UAS)*, are used to initiate a SIP request or to reply to one, respectively. A *User Agent (UA)* is a logical entity able to act both as a UAC as well as a UAS and is typically a SIP device. A UA sends its requests to the local *Proxy*, which then forwards them to the proxy of the UAS. A SIP Proxy can also act as both a UAC and UAS and can rewrite message headers. A Proxy basically routes messages to another entity (Proxy) closer to the target UAS, and can be used to enforce policy. It additionally allows for server-side authentication to the UA. A UA also receives requests (as a UAS) through its Proxy. In case a local Proxy does not exist for one of the communication partners, his/her UAC can connect to the UAS Proxy directly. SIP devices can also directly communicate with each other, if the other's URL is known. Authentication can be a problem in both cases. With the use of a *Registrar Server*, UAs can register their current location, which is then published to a *Location Server*. A *Redirect Server* can then supply this information to other UAs and Proxies. Different server entities can be combined into an SIP server.

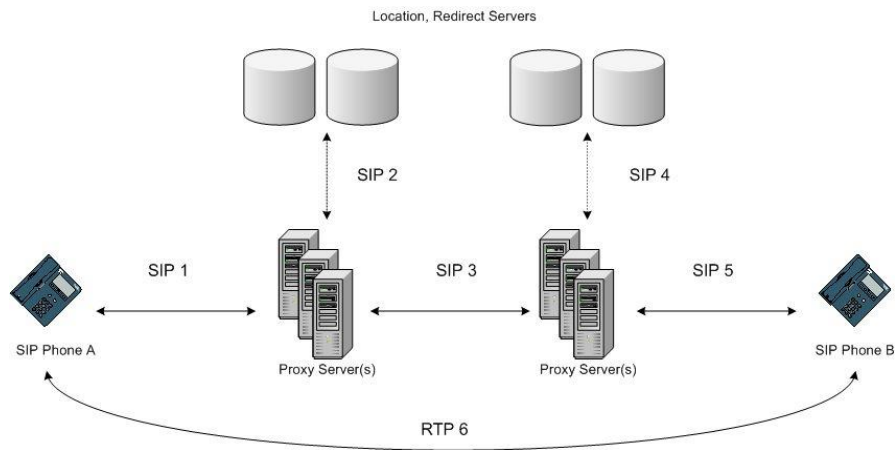


Figure 1, SIP architecture

A typical SIP call setup message flow is shown in figure 2: a SIP device initiates a call by sending a SIP “INVITE” request for another SIP device to the Proxy. The Proxy forwards the message to the Proxy of the called SIP device (if the device is not connected to the same proxy), which then forwards it to the device. The device responds with a “200 OK” message and an RTP channel (media stream) is set up. With a “100 Trying” message a Proxy can inform another Proxy or a UA that the end SIP device is being looked for but not yet located, while “180 Ringing” messages inform the caller that his/her request has reached the called user’s phone and the user must pick up to initiate the media stream. As shown in Figure 1, the RTP stream must not follow the same path as SIP signaling and will almost certainly use a much faster path between the SIP devices. The call is (normally) terminated with a “BYE” message from the UA wishing to end the call and is followed by a “200 OK” message from the other end.

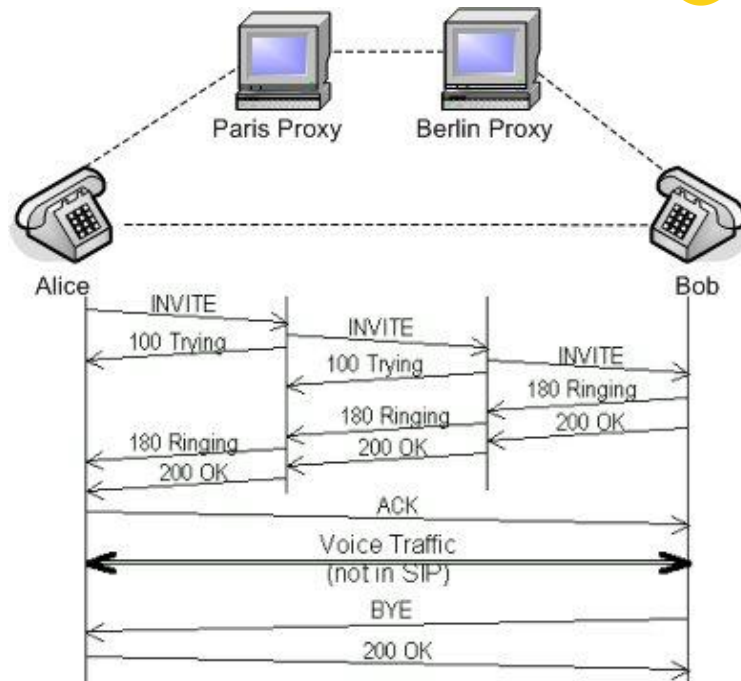


Figure 2, a SIP call

With the use of *Gateways*, SIP devices can communicate with PSTN- or ISDN end-devices (Figure 3). Gateways are responsible for routing calls between networks with different signaling schemes (SIP to PSTN (SS7) or ISDN and vice-versa) and adapting the signaling parameters.

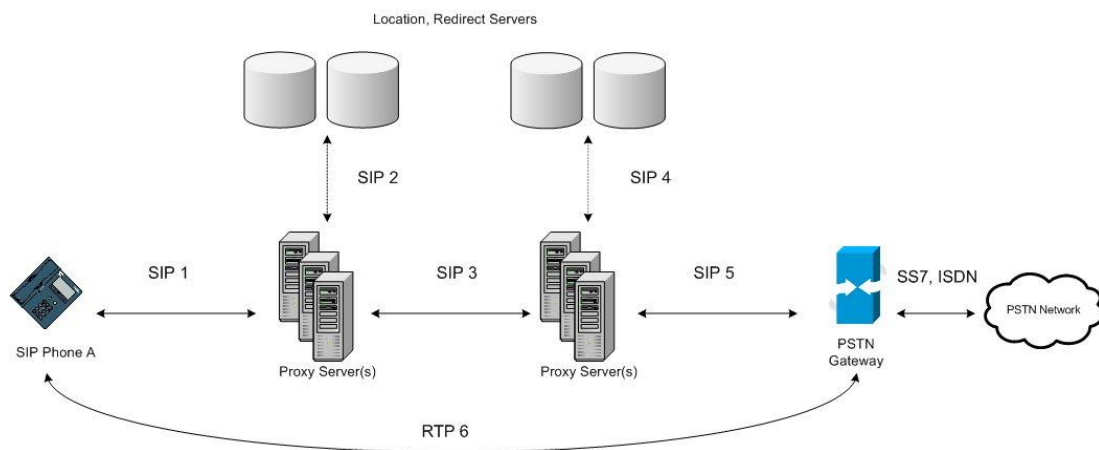


Figure 3, SIP architecture with PSTN gateway

Skype and Google have also claimed an important market share using their own proprietary protocols. The Computer Science Department of Columbia University was the first to study the network architecture of Skype. Based on their analysis one can get a very good idea of what the Skype network looks like. Skype is a peer-to-peer VoIP client, claiming it can work almost seamlessly over NAT and firewalls. In the Skype net-

work, there are ordinary hosts (Skype clients) and supernodes. Every node receives candidate status for becoming a supernode if it fulfills some requirements: has a good bandwidth connection, no firewall and adequate processing power. The Skype network is an overlay network, so each client Skype client should build and refresh a table of reachable nodes. Supernodes are used to route traffic in a fashionable manner. Each user logs in against the Skype login server (not a Skype node), who then passes the login information on to the supernodes. Each supernode knows almost all other supernodes, which are estimated today at an approximate 20.000. SkypeIn and SkypeOut are the premium features allowing a user to accept and place calls to the telephone network, respectively.

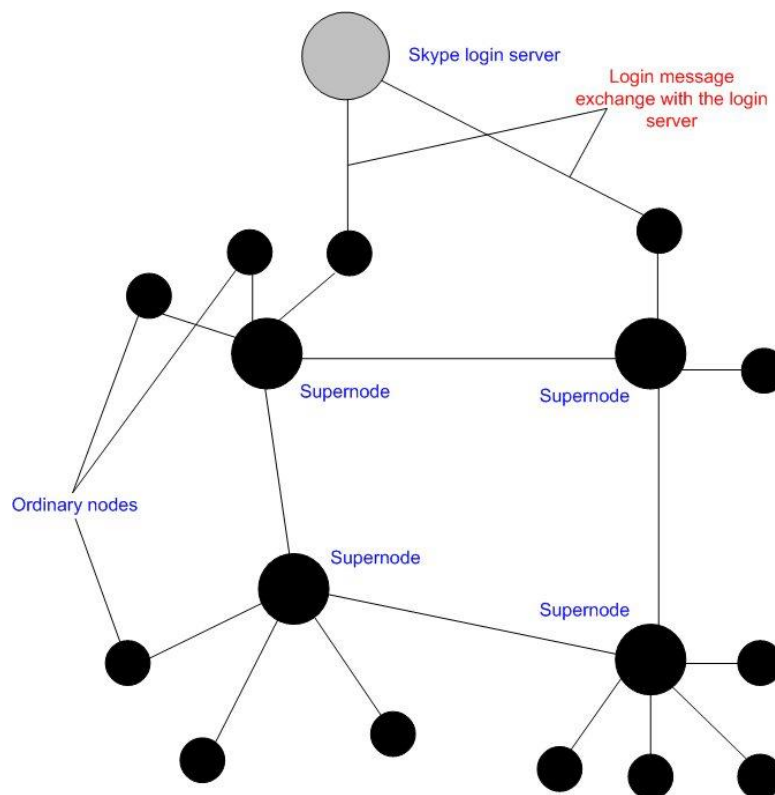


Figure 4, Skype architecture

Google Talk (a.k.a. GTalk) is the IM (including voice) product of Google. It is directly competitive to Skype but still in beta phase. Little is known on the network architecture of GTalk. It is also a peer-to-peer network like Skype and is based on the XMPP protocol (XML-based). As such, it can also cooperate with other XMPP-based clients like Jabber. Google has announced that it will include SIP in the forthcoming versions.

The diversity of available protocols is a matter of concern when it comes to security, whether the latter has to do with each individual protocol or with the interoperability between them.

B. End-user devices

The end-user devices can also be different from case to case. Soft and hard phones are the most advantageous end devices when using VoIP: they integrate the VoIP functionality together with the added value services directly available to the user, and offer a great degree of customization. Nevertheless, soft phones are easier to upgrade than hard phones and new functionalities can blend in without being confined by the embedded hardware platform capacity. This is an advantage when integrating security features. In return, a soft phone also needs a host computer to function, while a VoIP hard phone is an autonomous integrated platform. Users can also keep their PSTN- or ISDN-telephones as end devices and use a VoIP adapter to connect them to a network and reap the benefits of VoIP functionality. Alternatively, a company's PBX can be connected to a VoIP gateway with users retaining their legacy end devices here, too. These solutions limit the deployment of added VoIP services but retain the advantage of communication cost reduction with a possible additional lower cost for hardware and skill adaptation. The upgrade overhead is, nevertheless, significant, especially as far as VoIP adapters are concerned.

C. Network

Another important factor for security is the underlying IP network carrying the voice data. This can be a wholly-owned network (e.g. the case of a company network interconnecting corporate sites) or a public IP network (e.g. the Internet, when using the infrastructure of a VoIP service provider). It is obvious that the first case limits the risks considerably. Last but not least is QoS: the benchmark is set by the accustomed quality of landline communication. In-call quality is defined by a latency up to 150ms and a jitter of maximum 25 ms. Packet Loss must not exceed 5% with a BER of 0,25%, an important factor considering of UDP communication (ITU). There is also a benchmark for call setup quality: SS7 (PSTN) call setup time has settled to a maximum time of 1,0 second. It could be required from VoIP calls to stay close to it.

Threats

Every new technology brings new chances as well as new risks with it. In the case of VoIP, the risks threatening traditional communication coexist with some new, technology-related ones. Landline telephone networks are closed networks, with access to the network infrastructure only for authorized personnel and with each terminal device accorded its own circuit when the caller wishes to place a call. It is therefore difficult to get access to the physical layer and/or network components (though the Main Distribution Frames, concentrators are often in public places) or to tap a call (or data) destined for another terminal device (though it can happen by chance). The telephone companies do have a great authority on calls flowing through their networks and are able to tap and store calls and to keep track of the calls made and received by each end device, when necessary. Mobile phone networks allow, on the other hand, easy access to the wireless physical layer, but, with the exception of old analog cellular telephony in the U.S., interception is not trivial in the wireless part. For the wired part of the communication, the same rules as for landlines apply. Telecom carriers have the capability to intercept calls and store call data in cellular networks, too.

The VoIP Security Alliance (VOIPSA) is a forum bringing together people from the industry and research in order to fill today's gap of VoIP security resources. It has published a draft serving as a comprehending taxonomy of threats for VoIP technologies. According to the draft, VoIP security threats can be divided into 5 major categories: Social Threats, Eavesdropping, Interception and Modification, Intentional Interruption of Service as well as Unintentional Interruption of Service. The categories in overview:

A. Social Threats

Security and Confidentiality are treated as important social needs and their balance against other vital needs is the first topic of discussion. Important factor is the demand for security and confidentiality both against other individuals as well as governmental agencies to the extent of the law. One topic here is the *Misrepresentation of Identity, Authority, Rights or Contents* with the intent to mislead persons or computers. False Caller Identification is the most prominent form of Identity Misrepresentation. Current VoIP phones have the ability to setup a call with an arbitrary Caller ID, unlike PSTN and ISDN phones, where Caller ID is a matter of the switching centre. Many cases of using fake Caller ID to support other illegal or unauthorized actions have been reported.

The other two topics concerning Social Threats are *Theft of Services*, which consists of attacks intended for financial benefit either by not paying for a service or by charging someone else for its use, and *Unwanted Contact*. Under the latter falls, among others, VoIP SPAM. VoIP SPAM, widely known as SPIT (SPAM over Internet Telephony), is the equal of Email SPAM for VoIP. VoIP brings the opportunity to use another medium for unsolicited advertisement and its broadcast capabilities can be a liability. Only minor cases have yet been reported. Therefore, some form of authentication, both of the user and the network, and integrity of the data are useful to counter this kind of threats.

B. Eavesdropping

Eavesdropping threats describe the possibility for an attacker to monitor the entire signaling and/or data stream without altering it. Cases include *Call Pattern Tracking* (which can also be practiced on VoIP signaling only), *Traffic Capture*, *Number Harvesting* and the *Reconstruction of Conversations*, *Voicemail*, *Fax*, *Video and Text*. Often it is enough for an attacker to know who the call was destined for or how long it lasted. By capturing and storing encrypted traffic it is also possible to decrypt it when an algorithm vulnerability becomes known or when computers simply become fast enough for the key length used (and some people can wait that long). Tools freely available in the Internet like “Cain & Abel” can easily capture VoIP traffic and rebuild a whole conversation. Eavesdropping is mostly used to support other means of unauthorized practices such as theft of information. Confidentiality protection, both of the signaling and the media streams, is here a means to tackle the threats.

C. Interception and Modification

This category describes attacks by which an attacker can see the entire signaling and/or data stream between two endpoints and can also modify this traffic. Types of attacks are *Call Black Holing* (also “call black-holing”), which consists of dropping essential protocol information to prevent or terminate communication and (unauthorized) *Call Rerouting* in order to exclude some authorized nodes and/or include some unauthorized ones. *Alteration*, *Degrading*, *Impersonation and Hijacking of Conversations* are also forms of Interception and Modification. A Conversation Degrading attack aims to limit or frustrate communication, while the other forms aim to modify content, identity, presence and/or status of any of the parties. Authentication of network devices and users as well as integrity are important premises on which attackers will possibly look the other way.

D. Intentional Interruption of Service

The two major subcategories of Intentional Interruption Service are *Denial of Service (DoS)* and *Physical Intrusion*. Physical Intrusion includes the threats posed by unauthorized physical access to the VoIP components' premises as well as to the Physical Layer (OSI Reference Model). The threat to VoIP functionality coming from Denial of Service can itself be further divided into two distinct parts. The first one is *VoIP specific DoS*, where the attacker aims at protocol- or technology-specific assets. The other one is practically *Network specific DoS*. This is a hard to neglect vulnerability and a major difference from traditional telephone networks. The fact that the voice flows over IP data networks expose voice communication to the same threats data is exposed to. This includes the communication as well as the network components with their Operating Systems and/or firmware. The operating system is often a general-purpose one, like Windows, for which many viruses and worms are in circulation. VoIP is thus more vulnerable to attacks because it builds on networks that are already susceptible to a wide variety of threats. Enforcement of General Operational & Security Policies in a network is therefore a must for anyone willing to implement and operate VoIP. When it comes to single users communicating over a private-owned broadband connection like xDSL, part of the responsibility remains with the provider.

E. Unintentional Interruption of Service

Interruption in VoIP Service can also occur accidentally or as a side-effect of other malicious activity. IP networks are predominantly general-purpose networks, where protocols and network components are designed with interoperability in mind. Redundancy is hardly a major topic and network monitoring is a complicated issue, mainly because the network is owned by many participants at a time. This makes it prone to errors and unpredictable fall-outs. A *Loss of Power* happening when no UPS provisioning has been made results to a loss of VoIP functionality, even when just individual network components are affected. *Resource exhaustions*, whether memory-, CPU- or communication link-related, can have devastating effects.

Particularly important to Voice over IP is *Performance Latency*, affecting both signaling and media in different ways. Perceived Quality degrades when Latency, Jitter and Error Rate exceed some strict threshold values.

Authentication, Integrity and Confidentiality are the most important security factors for VoIP communication, too. Though these were not generally provided for at landline and cellular networks for normal, it is vital to know that they are more at risk with running over the corporate network or over the Internet. It is equally important to enforce Security Policies for the network, because mostly general-purpose operating systems and hardware are used in the case of VoIP. Having said that, we should now take a look at the main approaches used to secure VoIP communication.

Related Work & Protocols

The first thing to notice about security-related on VoIP is that there are no widely accepted protocols and solutions. The fact is due on one hand to the variety of VoIP protocols. Then again, it has to do also with the indifference of the VoIP providers to integrate security in the first place. They expect a lack of interest from the end users and opt for eye-catching “functions” instead. Thereby should be mentioned that, although some form of advanced security such as encryption is not offered by landlines either and only partially offered by mobile telephony, VoIP communication is, as seen previously, considerably more at risk than these two communication technologies, for a number of reasons.

A. VPNs

VPNs were among the first protocols proposed for secure voice communication. Tests have shown A VPN only secures communication to and from network. Once inside it, communication is carried out in clear. Furthermore, VPNs are not suitable for real-time communication such as VoIP, because they introduce unacceptable additional latency.

B. VLANs

Virtual LANs are able to separate voice and data flows inside a physical network. Though this technique can be used to enhance QoS and prioritize voice packets

C. SIP Security

The next step in securing VoIP is to develop new protocols, specifically suited for it. These protocols are mostly signaling technology-dependent. SIP, as an open standard, has led to (many) open protocols for security, too. It is important to note, that, the fact that the RTP stream flows over an alternative path, other than the signaling one, offers some security, too, because an attacker must first locate the route of the media channel (and be able to have access to it). Of course, such protection should never be judged as adequate. Encrypting the media is therefore a precaution one should take, because data flowing over an IP network, even over a switched network, can be intercepted, modified and eavesdropped.

Media Channel Encryption

A. *Secure RTP (SRTP)*

SRTP, an IETF ratified protocol (RFC 3711), is responsible for symmetrically encrypting the media stream in each direction. It usually uses AES for encryption (a default key length of 128 bits in counter mode) and is fully integrated with RTP (it is actually a “profile” of RTP). SRTP has thus a very small overall overhead, with zero header overhead (it does not encrypt the header), while the packets appear to be RTP ones. The protocol is quite easy to implement and the relevant technology well understood. It is directly applicable to the media stream and suitable for UDP real-time communication. It has therefore been established as the standard protocol for media encryption, with small competition (see DTLS). SRTP offers Integrity and Confidentiality for the communication, as well as protection against replay attacks. Nevertheless, there is a basic problem lying in the association management, including key establishment, peer authentication and encryption algorithm selection for the media stream. Thus, a form of handshake must take place. A design choice is the use of the signaling channel or the use of the media channel for the handshake. Another one is the use or not of a Public Key Infrastructure.

B. *DTLS over RTP*

DTLS (Datagram Transport Layer Security) over RTP is recent alternative to SRTP (IETF drafts have recently been made public). TLS normally only functions over connection-oriented transport protocols like TCP. DTLS aims to change that, by adapting TLS to a connection-less protocol like UDP. Its main advantage is the use of the well-known TLS protocol. It RTP and RTCP. It can also function in an SRTP-compatible mode to prevent incompatibility issues and can cooperate easily with a DTLS authentication (so no additional authentication scheme is needed). Nevertheless, it has a much greater overhead than SRTP.

Signaling Channel Association

C. MIKEY

MIKEY is a proposed association management method for SRTP (and in general for secure multimedia sessions) and is described in RFC3830. It uses the signaling channel to authenticate users and exchange keys. It can be implemented in four different modes: Public Key Mode, Diffie-Hellman Mode, Diffie-Hellman HMAC Mode, and RSA-R Mode. The Public Key and Diffie-Hellman HMAC Modes require the use of a pre-shared key for authentication. The other two modes rely on a PKI for the same cause. For this reason, it is either considered costly when a PKI is required or inapplicable when a pre-shared key must be used for the authentication between two strangers.

D. *sdescriptions, SIPS, S/MIME*

The Session Description Protocol used by SIP to describe multimedia sessions (media streams) can also be the carrier for SRTP initialization. A new attribute named "crypto=" has been proposed for the signaling and exchanging of cryptographic parameters (and not only for SRTP). This method is more lightweight than MIKEY as it doesn't need a PKI among others, but it has a serious disadvantage: The SDP message must additionally be secured by a data security protocol lest it will be transmitted in the clear. TLS is the most favourable solution to guarantee authentication and confidentiality for the SDP message cryptographic parameter exchange. None the less, it must be supported by the SIP proxies as well as by both the communication partners.

SIPS URIs (RFC 3261 for SIP) attempt to establish a TLS-secured signaling channel between the communication partners, by seeking to secure each communication hop, based on transitive trust. As many architectures only secure each hop from the UAC to the proxy of the UAS and then use other techniques to secure the last hop, the requester can never be sure that TLS security on every hop (virtually end-to-end) is in place and the protocol leaves a loophole open to attackers. Overall, *SIPS* is more trivial to use than MIKEY and has already been implemented by many VoIP vendors.

S/MIME is a solution providing this end-to-end secure channel directly between the UAC and the UAS and builds on PKI. It is described in the original RFC 3261 for SIP and works pretty much the same way *S/MIME* for mail security does. The UAC sends a signed request including his/her certificate to the UAS. The UAS should verify the signature and respond in the same way. Once this reply is verified by the UAC, the two partners are authenticated and can exchange session keys. The need for a PKI and well-known attacks against the *S/MIME* scheme have rendered this authentication scheme for SIP redundant.

Media Channel Association

E. ZRTP

ZRTP is a project from Philip Zimmerman, creator of PGP. It is based on the principle of legacy cryptophones and is still a draft. The partners start a communication in the clear and exchange the cryptographic parameters inside the media stream. This solution does not need a PKI, instead a Diffie-Hellman authentication is used and forward secrecy is provided by destroying the keys at the end of the communication. Diffie-Hellman is by default susceptible to MitM attacks. Therefore, the communication partners have the possibility to read their corresponding Diffie-Hellman keys aloud. Even if they refrain from it, they have the possibility to do it later: the process uses key continuity to ensure that, by means of using parts of previous authentication keys for the next ones. This means that, if the keys of any n-th session of the same communication partners do not match, nor have the keys of the n-1 previous sessions between them. Moreover, the media channel association has the advantage of making security establishment just as hard to trace in the first place as the media channel itself, due to the alternative path chosen for the RTP stream. Nevertheless, a call might be originally diverted to another UA rather than the expected one, so a mechanism ensuring the opposite must be in place prior to communication.

F. DTLS Handshake

Besides the TLS authentication mechanism used by SIPS, there is also the possibility of conducting the authentication over an RTP stream, protected by DTLS. This solution is based on self-signed certificates, and must be integrity-protected in other ways, suggested are the *Enhancements for Authenticated Identity Management*. Through the use of the latter, it is possible to cross-check the self-signed certificates against the otherwise authenticated SIP headers. It allows running a normal SIP call when no encryption is involved. Furthermore, there is the advantage of security protection establishment over the media path.

Overall is SIP security still an open topic, though things seem to have found their way. Topics of discussion are still, as mentioned, the use of a PKI or not, and the use of the signaling or the media channel for association management. PKI offers a complete authentication but it brings a great overhead with it when it comes to every day use. As for the channel selection, it doesn't seem to make a difference, though the use of the media channel makes security establishment hard to trace in the first place.

D. Skype Security, Google Talk Security

Philippe Biondi and Fabrice Desclaux have recently tested Skype security. Skype has only revealed sparse information on the security aspects of their product, purporting to use AES encryption with 256-bit keys. Biondi and Desclaux claim that Skype uses a centralized PKI for the authentication and session key exchange between the users. Each user logs in with username and password using a hash method and hard-coded RSA public keys. Then the users authenticate to each other using the PKI. Once authenticated, the users have the opportunity to have their public keys certified (private/public keys are used for one login session only) and exchange session keys. Replay attack, confidentiality and authentication are therefore provided for. Though the whole enterprise seems well implemented according to the testers, there are a few catches: a security policy can not be enforced when using Skype, while the Skype servers can intercept the session key of a communication (Man-in-the-Middle attack), decrypt and disclose it. This is particularly convenient for governmental agencies requiring the cooperation of Telecommunication Providers. And the biggest problem remains with the “security by obscurity” method of keeping the architecture details of the proprietary network protocol undisclosed. No known security assessment for Google Talk exists. The fact that it is based on open standards (XMPP) makes it easier to study and secure.

Open Issues

There is still no security solution for intra-protocol security and also no solution for end-to-end encryption when placing a VoIP call to a PSTN- or ISDN-phone or vice-versa. Intra-protocol security might be harder to cope with, because Skype and Google are unwilling to disclose their proprietary protocols (and so to cooperate in this direction). Even if each protocol will eventually be very well individually secured, attackers can still take advantage of the protocol transition to initiate attacks, so that the gateway becomes in fact an Achilles' heel. This is also the case when bridging to PSTN or ISDN. The degree of security they offer might be acceptable, but the combination of a secure VoIP with PSTN or ISDN means that somebody other than the end user gets to decrypt the media stream.

But still, more important is the fact that, up to now, the incursion of VoIP has not been accompanied by a responsible strategy for security. It is true that the telecommunication security needs of each user are different. Private and even business users have long lived with the minimal security provided by landlines and the somewhat more advanced one of cellular networks. End-to-end encryption has never been a matter for the masses. Today, however, confidentiality in communications (as in other areas) is gaining more and more attention. Besides the usual groups interested in intercepting or disrupting communication by illegal means, human rights experts warn of the infringement of the citizens' confidentiality by governmental agencies with the use of modern technologies, not only in traditionally unfree & oppressive regimes, but also in modern democracies. The use of VoIP makes it easier for someone to capture, store and process voice data. VoIP providers are obliged by law in many countries to store communications data for a period of time. There is thus the urgent need to catch up on the security demands, but this, of course, cannot take place before people agree on how much security a user needs. The latter should not hold up things for too long, though. Vendors have already begun integrating security protocols into their products and it would be a pity to have incompatible devices in the market just because we can't agree on how much security measures up.

The choice of the manufacturers seems to lie with SIPS/TLS for the authentication and SRTP for the encryption. On the other hand, some VoIP providers have also begun implementing and offering SIP security, with the SIPS/TLS and SRTP remaining the main protocols of choice. SIPS/TLS is a trade-off between security and ease-of-use acceptable for the private end-user and SRTP has, up until now, been virtually the only choice for the media stream encryption. Business users might want to go a step further and rely on S/MIME, the latter itself relying on the corporate Public Key Infrastructure. It is difficult to say what the future holds in store for inter-protocol communication. One can never emphasize enough that every incompatibility offers much less security than "less security" or "no security at all", whether this has to do with different security implementations within the same protocol or with securing communication between end-devices using different protocols.