Secure Voice-over-IP

Markus Dunte and Christoph Ruland,

Institute for Data Communications Systems, University of Siegen, Siegen, Germany

Summary
Voice-over-IP (VoIP) is one of the main applications or services, which dramatically increased the spread and usage of the Internet. VoIP describes the transmission and reception of voice and video data over networks that use the "internet protocol" (IP) as transport protocol. With the use of VoIP within company networks (intranet) investors hope to reduce the expenses for infrastructure and maintenance, because telephony and data networks will no longer be separated. The often-used term "internet telephony" describes the transport medium for VoIP, which will in this case be the Internet. Based on the widespread availability of broadband Internet access, VoIP is becoming more and more a cheap alternative to conventional telephone networks. The main drawback of VoIP is that there is no reliable authentication of the calling parties as well as no methods to prevent trace and capture of calls. Thus, with simply methods and low effort VoIP calls can be recorded. This fact leads to the idea to develop a VoIP communication software which offers secure and confidential communication. The details of design and implementation of this software as well as the chosen specifications are described in this paper.

Key words:
Voice-over-IP, confidentiality, encryption, integrity, SRTP, SIP, H.323

1. Introduction

Telephony services over networks based on the internet protocol are becoming increasingly important on the telecommunications sector for private end-users. Due to this reason it is necessary to provide services such as authentication, authorization and integrity. These services are partly specified in some of the communication protocols but not yet fully implemented.

In order to overcome the lack of powerful and reliant security mechanisms the above-mentioned services authentication, authorization and integrity were implemented in a small application for VoIP communication. In detail this means that existing protocols like "Real-time Protocol" (RTP), "Session Initiation Protocol" (SIP) [1] and H.323 [2] were modified and/or fully implemented to be used within the application.

2. Technical Description

The mentioned application was developed to be used within VoIP networks which are standard compliant. Within these networks the application offers services like confidentiality, data integrity and authentication to its users.

Confidentiality is used to prevent information being extracted from the communication between two or more VoIP entities. The use of confidential communication does not prevent data being extracted or captured from the communication but due encryption it prevents information being stolen. Moreover, it is essential to offer data integrity services in digital communication networks that enable users to detect active attacks to the system. Data integrity is used to restrict the communication to users and to exclude possible attackers. The actual threads in this case are manipulation of data as well as data replay. Authentication is then used for a reliable mapping of calling identity and end user.

It has been specified that the application should be able to deal with most important signaling protocols for VoIP, namely H.323 and SIP. Furthermore, the communication data between two or more entities should be transmitted using secure version of RTP (SRTP) [3]. Confidentiality and data integrity in SRTP are realized using encryption and "hash message authentication code" (HMAC) [4]. The necessary keys will be exchanged during the signaling phase of the communication procedure using cryptographic secure methods. Authentication will be guaranteed via a "public key infrastructure" (PKI).

3. Security Related Analysis

Since audio and video data in VoIP applications is transmitted via heterogeneous open networks thread analysis is totally different to the conventional telephone system. The main difference of both VoIP and "plain old telephone system" (POTS) is that there is no fixed route from sender to receiver as in POTS. VoIP data is transmitted in packets, which must not stick to a single route. This exhibits the possibility for man-in-the-middle attacks. The eavesdropper is therefore able to alter data

Manuscript received June 5, 2007
Manuscript revised June 20, 2007

and to transmit the changed packets to their destination. The heterogeneity of VoIP systems illustrates another source of threat just because every single component is a tender spot.

The following subsections will provide an overview on threats that have to be considered as well as security services that should be provided.

3.1 Goals

During the design phase of the above-mentioned software the following goals in terms of security have been defined:

- **Confidentiality:** to protect personal data and transmitted information against unauthorized access (no eavesdropping on calls)
- **Integrity:** to protect (and detect) personal data and transmitted information against being altered
- **Authentication:** to prevent fakes of communication partners identity and prove of data origin

The above-mentioned security goals cannot be considered separately because relations exist between all of the mentioned tasks. If for example there is no authentication of the identity of the communication partner, a possible attacker is able to receive confidential information under a faked identity (loss of confidentiality). Furthermore, the attacker is able to alter the information received (loss of integrity). This example shows that almost all security goals have to be considered in combination to provide a maximum in security.

3.2 Threats

This subsection is aimed at giving a short overview on possible threats and attacks within VoIP systems.

Basically there are two main groups of attacks, which can be differentiated [5]:

- Passive attacks: read, trace and evaluate data
- Active attacks: manipulate data, creation of new faked information

The main “holes” for possible attacks in H.323 are faking identities and manipulation of transmitted data. During the signaling phase attackers are able to change the destination and source address for the multimedia stream. This leads to the fact that the data stream during the call can be routed anywhere but to the intended destination.

The above-mentioned attacks can equally be defined for SIP. Moreover the attacks can be performed much easier due to the fact that SIP messages are transmitted in plain text with standard ASCII characters.

The previously mentioned application, on which this paper is based on, was designed to resist a specific set of threats and possible attack, which will be described in the following list. Generally, the intention was to protect the identity of calling parties, the transmitted and saved information as well as the detection of manipulation.

- **Spoofing:** Detection of faked messages or packets is essential for VoIP applications. If the message origin and its integrity cannot be proved it will significantly lower the trust and is equal to compromise the whole system.
- **Replay:** Retransmitting captured messages which violate integrity must be prevented.
- **Man-in-the-middle:** Prevention against man-in-the-middle attacks where the attacker has unlimited access to the data transmitted. Intruders will not be able to capture, trace and manipulate data being masqueraded.
- **Attacks to VoIP middleware:** VoIP middleware was not extended by security services due to the fact that the overall system should be standard compliant and users with no security extension will also be able to use the VoIP middleware. Specific services such as “lightweight directory access protocol” (LDAP) [6] were extended by “server side secure sockets layer” (SSL) or “transport layer security” (TLS) [7] for server authentication. For the complete client-server authentication an additional client certificate is used.

4. Security for H.323 and SIP

This chapter will give an explanation on the security features and mechanisms that were used throughout the implementation. The focus is set on the signaling phase of the underlying protocols.

4.1 SIP Security

According to RFC 3428 [8] S/MIME is used to encrypt and sign the “session description protocol” (SDP) [9] portion of SIP packets. The header is still transmitted as plain text. S/MIME guarantees end-to-end security, which means, that only the calling parties are able to get the transmitted information. Originally S/MIME was developed for authentication and encryption of email but can be used in other application as well. S/MIME provides mechanisms for the secure end-to-end delivery of message bodies within IP based networks. Asymmetric key pairs realize encryption and authentication in S/MIME, where the private key of the sender is used to sign the message and the public key of the receiver for encryption. With these facts S/MIME offers the following security services for SIP messages:

- Authentication of sender.
- Integrity of information within the SDP portion of the SIP packet.
- Confidentiality for data in SDP. This is essential because within the SDP part keys will be exchanged for media data security.

During the signaling phase of SIP master key and salt key for media data encryption will be exchanged; this will be described later in detail.

4.2 H.323 Security

The "International Telecommunication Union" (ITU-T) has specified H.235 [10] for security services within the H.32X series. These security services aim at providing secure channels for signaling and further parameter exchange in the pre-call phase. H.235 is basically an extension of previous well-known protocols and describes the implementation of security services for existing protocols.

There are several possibilities to protect signaling channels of H.323. Basically they can be divided into two main groups. The first group deals with securing the underlying network layers and the second one uses special mechanisms to protect the content of the channels.

Due to the fact that H.225.0 [11] is the first channel that is established, there is no possibility for on-the-fly security parameter exchange; this must be done in prior. To avoid information exchange before the actual call takes place, the H.225.0 channel is connected via a secure connection. TLS provides the secure connection for TCP data. Therefore, integrity and confidentiality can be guaranteed. If certificates will be used, there is the possibility to additionally authenticate the communication partner.

Protection of communication channels for H.323 can be done in various ways. One of these methods is specified in H.235 Annex D. Annex D is based on a pre-shared secret, which is in most cases a password that is know two all calling parties. Based on that, simple integrity and authentication can be realized. Integrity is guaranteed by HMAC where the hash value is computed over the password concatenated with the actual message. The fact that the participating clients should only know the pre-shared password, authentication can be provided because the correct hash value can only be computed by knowing the password. Optionally specified by Annex D there is an extension called "Voice Encryption Security Profile" which supports session keys for multimedia data via Diffie-Hellman key exchange.

Additionally, H.235 Annex E can be used as an extension to Annex D but is not mandatory. From the technical side of view this standard is a great step ahead of Annex D because security is no longer provided by simply HMAC calculations but by X.509 [12] certificates. These certificates may also be pre-shared or can be exchanged on connection set-up. In detail, Annex E uses RSA signatures for the signaling messages which guarantee integrity. Authentication is provided by the authenticity of the certificate. Annex E has just as Annex D the optional extension for session key exchange.

4.3 Multimedia Security

Protection of multimedia data (voice and video) during phone calls is performed independent of the chosen signaling protocol. If the user has decided to communicate using security features from the application user interface, the multimedia stream is protected using SRTP irrespective to the signaling protocol.

In the case of SIP as signaling protocol the master key and salt keys are exchanged during call set-up as previously mentioned.

In the case of H.323 as chosen protocol, the exchange of SRTP parameter is performed using H.235.8. Within this standard the overall procedure for SRTP set-up is described. It is assumed that the channel used for call set-up and security context was secured in prior and provides authentic communication.

In order to provide a secure multimedia data communication the following steps are absolutely necessary:

- Diffie-Hellman key exchange
- Exchange and negotiation of security features
- Negotiation of security algorithms
- Exchange of the security context

The fact that SRTP is a well-known and often discussed topic it will not be explained in any detail in this paper.

5. Authentication

This chapter is dealing with mechanisms to provide authentic communication for VoIP.

Considering classic communication networks for telephony, the caller can almost be sure that the number to which the connection is established belongs to the intended communication partner. This is true because traditional networks are closed and managed thus unauthorized attacker cannot easily get access to the infrastructure. Manipulation needs tremendous effort, technical equipment and access to restricted infrastructure components.

Voice-over-IP allows access and manipulation with much less effort due to the fact that signaling information as well as multimedia data of telephone calls are being transmitted over open networks such as the Internet. Appropriate software tools can be used to capture, trace, reroute, initiate and cancel calls. Even calling party identification by telephone number transmission is not a reliable feature. Therefore, robust authentication needs more sophisticated methods.

As previously mentioned the developed communication is able to set up calls using H.323 und SIP. Due to the different function these protocols are providing, authentication has to be considered separately. In the following a few mechanisms will be presented, discussed and evaluated.

5.1 Authentication for H.323 and SIP

The documentation of H.323 lists three possible implementation techniques to guarantee authentication.

Both techniques "IP security" (IPSec) and TLS are used on the network or data link layer to protect the whole communication channel. Furthermore, with H.235 the ITU-T specifies its own mechanisms especially for authentication of signaling information.

SIP specifies also the use of techniques like TLS; alternatively S/MIME can be used. It should be noticed that in contrast to H.323 SIP signaling links are not solely established between the two endpoints. The signaling messages might pass along the way several parts of the VoIP infrastructure such as SIP proxies, which are necessary to retrieve the communication endpoints. Thus, TLS will not serve end-to-end but rather hop-to-hop security. Using S/MIME both hop-to-hop and end-to-end security can be implemented.

Detailed explanations to IPSec will be omitted due to interoperability problems with the current standards of H.323 and SIP.

TLS was specified by the "Internet Engineering Task Force" (IETF) in 1999 as enhancement of the SSL protocol. The main tasks performed by TLS are confidentiality (encryption) of data and authentication of client and server. In contrast to IPSec TLS does not need to change the actual packets sent, but introduces an additional layer between "transmission control protocol" (TCP) and the application layer above. Since TLS is a connection-oriented protocol, it must be used in combination with TCP and is not suitable for "user datagram protocol" (UDP). During the specification phase of software development TLS was chosen due to the fact that both SIP and H.323 are able to use it in order to perform authentication. Furthermore, the Federal Office for Information Security (BSI) in Germany and the ITU-T (within the H.235 standard) recommended the usage of TLS. Note that TLS guarantees in the context of SIP only hop-to-hop authentication. The following picture shows the TLS handshake procedure.

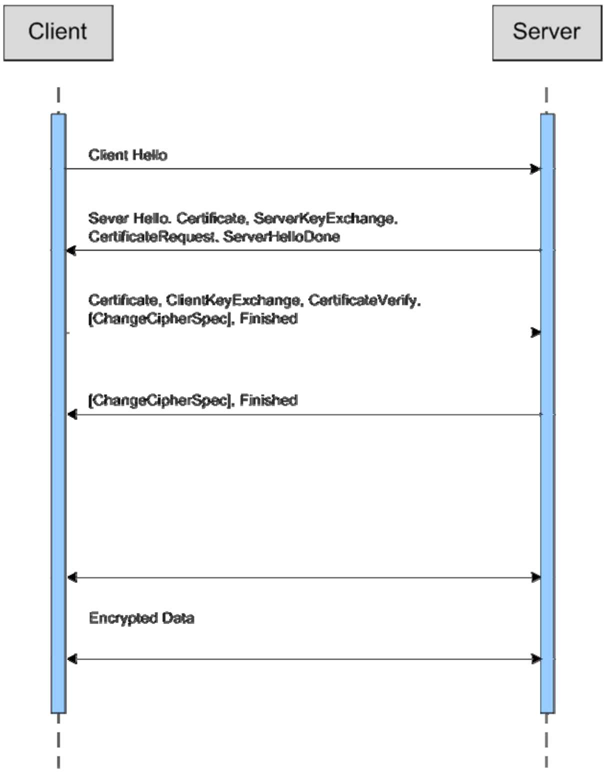

Client
Server
Client Hello
Sever Hello. Certificate, ServerKeyExchange.
CertificateRequest. ServerHelloDone
Certificate, ClientKeyExchange, CertificateVerify.
[ChangeCipherSpec], Finished
[ChangeCipherSpec], Finished
Encrypted Data

Fig. 1 – TLS handshake

As an enhancement of the MIME concept, S/MIME was developed, which is today the de facto standard for secure email communication. S/MIME is used for encryption and digital signatures based on certificates for messages. Due to the fact that SIP messages do not differ too much from email, the S/MIME concept can be used for SIP as well. The general structure of both email and SIP messages consists of header and body separated by a blank line. The version 2.0 of the SIP standard specified in RFC 3261 lists several mechanisms, which are related to email security concepts. These mechanisms are the following:

- Encryption of the SDP body
- Encryption and signature of the SDP body
- SIP tunneling with signed body
- SIP tunneling with encrypted and signed body

S/MIME is at the moment the only mechanism that offers end-to-end integrity, authentication and confidentiality for SIP. Therefore, compromised network infrastructure will lose its threat to the system and the key exchange for multimedia encryption can be done more reliable.

After the detailed study of security mechanisms for H.323 and SIP it was chosen to implement S/MIME for authentication and confidentiality in SIP and the use of TLS in case of H.323. With this decision, the developed software is able to perform secure call set-up for H.323 and SIP. The following figure shows the call set-up procedure.

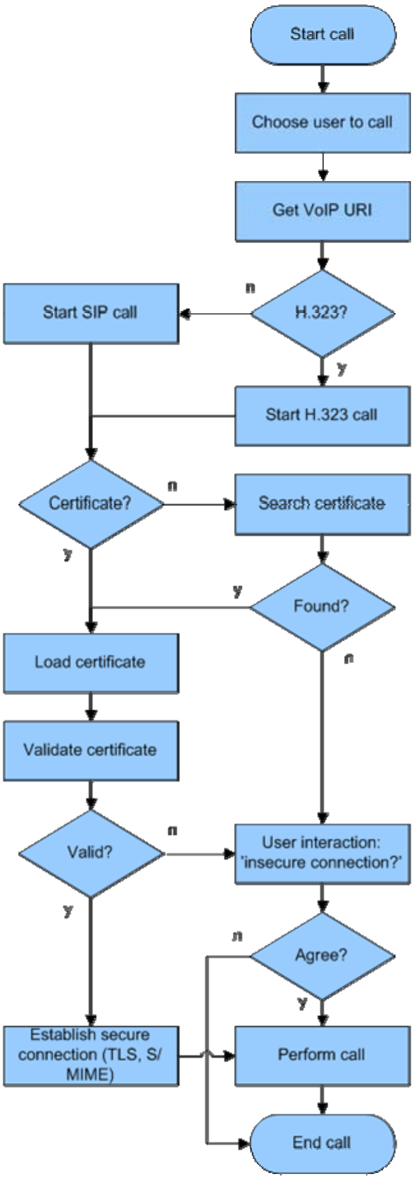


Fig. 2 – Call set-up

5.2 Public-Key Infrastructure

The implemented security features, which are used in the application developed, are directly or indirectly based on the usage of certificates. Certificates are used to encrypt that data transmitted and to authenticate the calling parties. The certificate management demanded for a public-key infrastructure, which has to be set up. The infrastructure components are “certificate authority” (CA) and a LDAP server that hosts all user certificates. The CA is able to administer an arbitrary number of users. It is aimed at creating, issuing and revoking certificates for the users.

5.3 Other Approaches

Based on the well known fact that until now there is no reliable authentication and/or encryption for VoIP, a lot of attempts to implement security have been made, some of them not worth mentioning.

In 2006 Phil Zimmermann released the first version of “Zfone” [13], a new secure VoIP phone software product which allows encrypted phone calls over the Internet. The main idea behind Zfone is to provide security independent of the signaling protocol. Zimmermann used a newly developed protocol for this task ZRTP, which is now a draft for the IETF [14]. This protocol is not only used for media data transport, but also for the key management. ZRTP does not rely on a PKI, in fact, it does not use persistent public keys at all. It rather uses ephemeral Diffie-Hellman with hash commitment as stated in [13]. The overall key and security negotiation is purely peer-to-peer realized by ZRTP. According to the founders information Zfone is working with any SIP/RTP phone. The main difference to the solution presented in this paper is, that the standard signaling which has to be done between clients and infrastructure is not touched at all.

Services that are currently used by many VoIP providers are often called “Secure SIP” (SSIP). Although this term does not guarantee standardized services, there are core elements that most of the providers share when offering SSIP. In most cases SSIP systems offer signaling links secured by TLS and media data transport realized by SRTP as specified in [1].

6. Conclusion

In contrast to the most VoIP application available at present, the described software allows confidential and secure communication over open networks. Authentication of calling parties can be guaranteed for both protocols H.323 and SIP. Using H.323 the call set-up is secured with the use of TLS to encrypt the underlying communication layer. SIP signaling is secured by encryption of the SDP portion of the messages giving the advantage that the network infrastructure can be used without any changes. The integrity of multimedia data (audio and video) is provided by the usage of SRTP as transport protocol from one client to another.

In conclusion it can be said that the system described in this paper offers the possibility of confidential and secured communication. Furthermore, the communication links established are standard compliant for H.323 and SIP. This enables unsecured communication with any other software that is compliant to the standard of H.323 and/or SIP. Finally, there is no need to adapt the network infrastructure to secure communication due to the fact that all security mechanisms are designed for end-to-end communication. Ideally, network infrastructure will even not recognize that the communication link is secured.

References

[1] IETF RFC 3261, "SIP: Session Initiation Protocol", June 2002

[2] ITU-T Recommendation H.323, "Packet-based multimedia communications systems", June 2006

[3] IETF RFC 3711, "The Secure Real-time Transport Protocol (SRTP)", March 2004

[4] IETF RFC 2104, "HMAC: Keyed-Hashing for Message Authentication", February 1997

[5] Federal Office for Information Security (BSI), "Studie zur Sicherheit von Voice over Internet Protocol", viewed July 2006, http://www.bsi.de/literat/studien/VoIP/index.htm

[6] IETF RFC 2251, "Lightweight Directory Access Protocol (v3)", December 1997

[7] IETF RFC 4346, "The Transport Layer Security (TLS) Protocol", April 2006

[8] IETF RFC 3428, "Session Initiation Protocol (SIP) Extension for Instant Messaging", December 2002

[9] IETF RFC 4566, "SDP: Session Description Protocol", July 2006

[10] ITU-T Recommendation H.235, "Security and encryption for H-series (H.323 and other H.245-based) multimedia terminals", August 2003

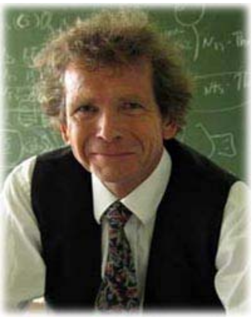
[11] ITU-T Recommendation H.225.0, "Call signalling protocols and media stream packetization for packet-based multimedia communication systems", May 2006

[12] IETF RFC 3280, "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile", April 2002

[13] "The Zfone Project Homepage", viewed 25 June 2007, http://zfoneproject.com/index.html

[14] Zimmermann, P. et al., "ZRTP: Media Path Key Agreement for Secure RTP", viewed 25 June 2007, http://zfoneproject.com/docs/ietf/draft-zimmermann-avt-zrtp-03.txt, March 2007

Markus Dunte born in 1979 in Kassel, Germany, received the B.Eng. in Electronic Engineering from the University of Hull in 2003 and the German engineering diploma in 2005 from the University of Duisburg. In 2005 he started work with the Institute for Data Communications Systems at the University of Siegen as research associate. His research areas focus on security aspects for multimedia communication. Special attention is given to Voice-over-IP communication and scalable multimedia content delivery.



Christoph Ruland, Professor, Dr., born 1949 in Hamburg, Germany, studied mathematics, physics and computer science at the University of Bonn. He received a diploma in mathematics as well as doctor degree. He was Professor at the University of Applied Sciences in Aachen in 1982 and a full Professor at the University of Siegen since 1992. He is the director of the Institute for Data Communications Systems at the University of Siegen. His main research area is the integration of security into communication systems on all layers. He has written books and many publications about information security in networks and is an active member in the ISO "Security Techniques" committee for 15 years. Professor Ruland founded the "Company for Cryptographic Communication Security and Communication Technology" (KryptoKom) in 1988.