Integrating Authentication into Existing Protocols

Reiner Sailer
Phone: +49-711-121-2487 Telefax: +49-711-121-2477
E-Mail: sailer@ind.uni-stuttgart.de

University of Stuttgart
Institute of Communications Switching and Data Technics
Prof. Dr.-Ing. P. J. Kühn
Seidenstrasse 36, D-70174 Stuttgart

Abstract

Proving the identity of individuals is the basis for many communications services and belongs to the context of network security. Access Control regulates the rights of individuals to access services or data. It is necessary to be sure of the identity of the asking individual to realize effective Access Control. Furthermore, it is necessary to provide the facility to prove the identity of individuals accessing data or services in order to prosecute them if they misuse their rights and compromise systems. Proving one's identity is referred to as authentication. Due to the overwhelming importance of authentication there are several well known and investigated protocols realizing the prove of identity.

This paper shows a way to integrate authentication protocols into existing user-network signalling protocols and its application to integrate authentication into the ISDN user signalling. Authentication is implemented in a way that affects the complex signalling software as little as possible. The implementation shall introduce minimal additional delays[1].

[1]this work is funded by the Gottlieb Daimler und Carl Benz-Stiftung in Ladenburg, Germany

1 Introduction

The need for authentication in communications will increase in the future. It will be important to control access to remote services on the basis of users rather than impersonal phone numbers. One application of the authentication method presented within this paper is the control of access to distance learning services. People shall be able to access distance learning services no matter whether they are at home or within enterprise networks. However, there may be strong security requirements for sensitive training material of companies. Hence it will be necessary to be sure that only intended users will be granted access to these services. Moreover it might be necessary to encrypt the communication between users of the service and the server to secure the training material during its transmission over public networks as well as between the network termination and the user terminal. The principle communications services underlying these services are assumed to be ISDN (Integrated Services Digital Network [4]) or Broadband-ISDN communications services.

Another important application of authentication is the secure transmission of data - or a telefax - between computers over ISDN. The sender of a telefax wants to be sure that only the intended receiver will receive it. In order to satisfy this requirement, it is important to verify the identity of the person at the other end of the communications line. If the person has proven its identity a common encryption key for the secured transmission of the data over public networks needs to be installed.

Finally, it is important even for telephone or conventional telefax transmissions to ensure the identity of the peer if the contents of the call or telefax is confidential or if a certificate for a telefax transmission is needed. Simply misdirecting a telefax may cause severe financial and social consequences as it is vividly described in [6].

2 Authentication Protocols

Authentication protocols [12] describe the sequence and the composition of messages that must be exchanged in order to prove the identity of communicating entities. They also describe the way these messages are to be constructed including parameters and use of encryption. Both parties[1] shall be sure of the identity of the other if and only if the required messages are exchanged. The protocols usually base on a unique relationship between an object's knowledge of a secret and its identity. There are many other ways to prove one's identity[2] but we will focus on this one for the rest of the paper.

Attacks to be considered in the context of authentication are among others[3]:

- impersonation (faking an identity)
- replay of messages (using eavesdropped messages of former authentication procedures to gain unauthorized access)
- "chess grandmaster" attacks (using messages of other simultaneous authentication procedures to gain unauthorized access)
- key stealing (eavesdropping a session key while it is established or stealing it from where it is stored or used)

The next chapter will introduce the protocol used throughout the rest of the paper and explains how it counteracts the above threats.

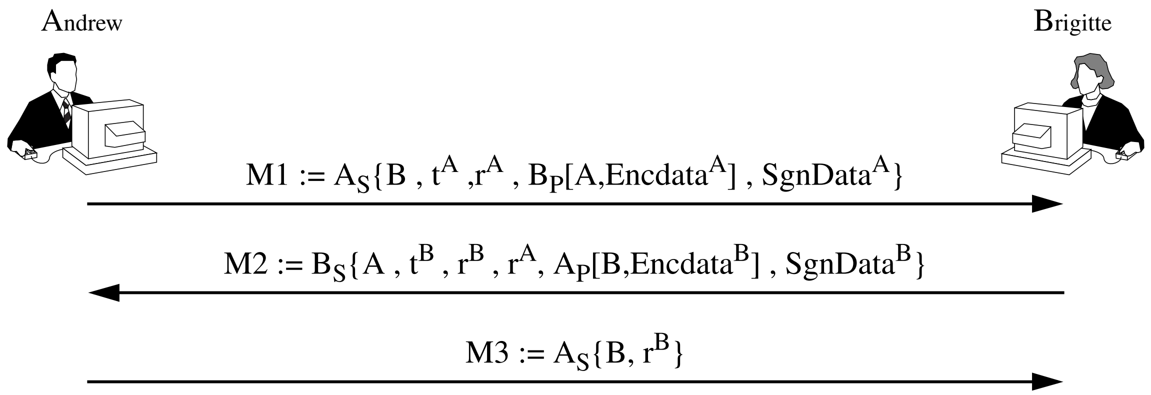
[1] we speak of two parties but most holds with little adaptation for more than two parties, too
[2] see also [7],[11]
[3] see also [1], [10]

2.1 The X.509 Three-Way-Handshake - Protocol for Strong Authentication

The X.509 Authentication Framework [1] is based on public key cryptosystems, namely the RSA cryptosystem [8]. This cryptographic scheme assumes that each individual possesses a unique secret key (also known as private key) and a corresponding public key. The private key acts as the secret to prove the identity of the owner. When signing messages, the secret key is used for signing whereas the public key is used to check the signature. When used for encryption, the public key of the intended receiver is used to encrypt the message and the receiver uses its secret key to decrypt the messages.

To distinguish between signed and encrypted messages we will use the notation $X_S\{Y\}$ to describe that data Y is signed by X using the secret key X_S of X. Similarly we will use notation $X_P[Y]$ for data Y that is encrypted by the sender for receiver X using the public key X_P of X.

Fig. 1 shows the protocol that is used to authenticate the calling party Andrew and the called party Brigitte. The exchanged messages are named M1, M2, M3 respectively to facilitate referencing the messages in following descriptions and signalling diagrams.



A ... Calling Party Identity Encdata ... encrypted user data t^x ... period of validity
B ... Called Party Identity Sgndata ... signed user data r^x ... random number

Figure 1: *X.509 Based Authentication Protocol*

The first message M1 is constructed by the party initiating the authentication protocol. This message includes the identity of the intended receiver, a timestamp t^A and a random number r^A that serves as a challenge for message M2. M1 is signed with A's secret key. Therefore B can be sure that the message must have been constructed by A - but it may be a replay of a former message sent by someone who wants to impersonate A. In a second step, B responds with message M2. M2 is constructed in the same way as M1 including r^A as a prove of the freshness of M2. By including r^A in M2, the sender of M2 shows that he actually knows B_S hence being B. As B has not yet a freshness prove of M1, M3 is necessary to verify that M1 has not been replayed. M3 comprises of the so called challenge r^B and the identity of the receiver. As this message is signed by A, B can be sure that the sender of M3 knows the secret key A_S hence being A.

As far as described, the protocol proves A the identity of B and vice versa. The parameters not yet described (Encdata) are used for initial key exchange because there is no benefit in knowing the intended receiver's identity unless you know that this is the only receiver. To make sure, that only the intended parties A and B will participate in the communication, it will be necessary to use cryptographic functions to protect all data exchanged.

For this purpose, EncData is included in M1 and M2. A tells B one part of the future common key $Encdata^A$ by encrypting it under B's public key and B tells A the second part of the common key $Encdata^B$ by encrypting it under A's public key within M2. The common secret key for this communication may be calculated by adding these parts modulo 2.

Parameter Sgndata may serve as a basis for exchanging additional information that can be used to decide whether the call will be established or not. This offers means for decisions that will be made not only on the basis of the identity of the peers.

The benefits of the authentication figured out in Fig. 1 are

- identities A and B are proved
- a common session key is installed [1]
- additional signed data is exchanged between A and B

At last we have to mention that the X.509 protocol that underlies our authentication protocol has some drawbacks. Anson and Mitchell have described some security defects of the X.509 protocol [9]. They are avoided in our protocol by adding the identity of A and B respectively within the encrypted part of M1 and M2. This is the only difference between the illustrated protocol and the one described in X.509 authentication framework.

2.2 Embedding Authentication into ISDN

This chapter describes how to apply the above authentication protocol to ISDN services. Fig. 2 shows a communications scenario using ISDN services. There are two terminals, each connected to the network termination (NT) of its ISDN line.

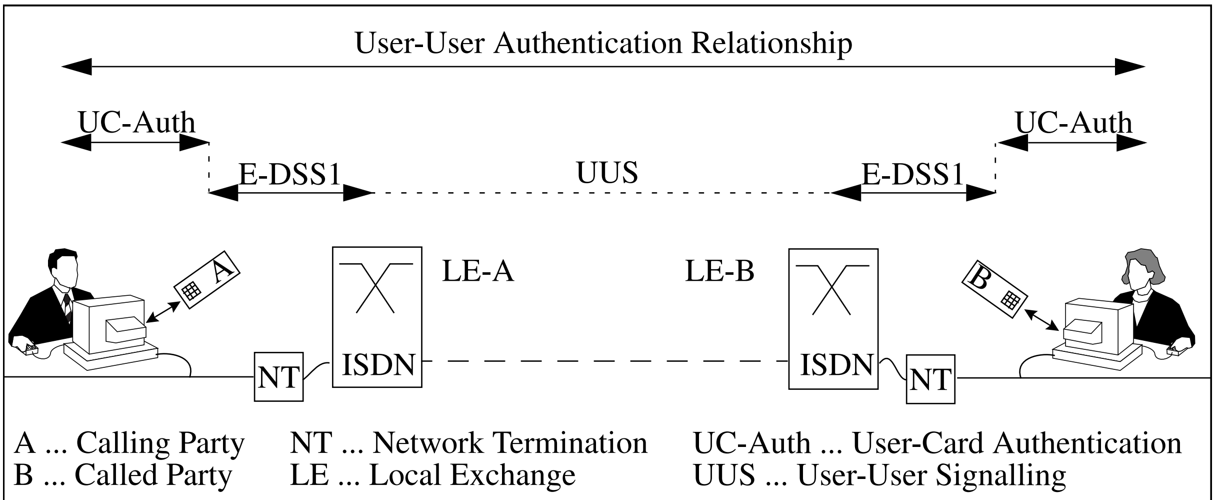


Figure 2: *Realization of User-User Authentication*

We realize user-user authentication in different steps. At first we authenticate the users against their cryptocards by passwords. These passwords are fed into the cards through mini keyboards included in the cryptocards. This authentication will activate the cryptocard for a specified time or until it is removed from the cryptocard reader. Afterwards, the cryptocards will authenticate each other on the behalf of the users. They implement the cryptographic functions that use the private key, stored within the cryptocard because the private key must never leave the secure cryptocard. The services it offers are at least signing and decrypting data.

[1] this key is used for cryptographic functions associated with this session

The terminals use the protocol of Fig. 1 for peer authentication. The exchange of messages between the terminals is embedded into the E-ISDN subscriber signalling E-DSS1 (European Digital Subscriber Signalling System No. 1 [2],[3],[5]). E-DSS1 is defined for the user-network interface but offers some parameters, called user-user information, that are transparently transferred through the network during call establishment.

By implementing authentication using these parameters only, authentication will be transparent to the underlying network services. Consequently we do not need any changes within the ISDN network.

We include it into layer 3 of the E-DSS1 to keep it transparent to higher layer applications using E-ISDN. This makes it possible to include authentication and associated security services in plug-in cards for the terminals without changing the applications.

3 Integrating Authentication Protocol into ISDN Signalling

The implementation of the authentication protocol of Fig. 1 is shown in Fig. 3. Additional parameters and signals are emphasized. It is important, that there is no need for new signals or parameters. We use the High Layer Compatibility parameter (HLC) within the setup signal to indicate the called terminal an incoming call that is to be authenticated. This parameter is conveniently used by the terminals to decide whether they accept an incoming call or not. Phones will not accept an incoming call with HLC parameter including a telefax code. In a similar way, only terminals with the facility to support authenticated calls will accept an incoming call including an authentication code (Auth) within its HLC parameter.

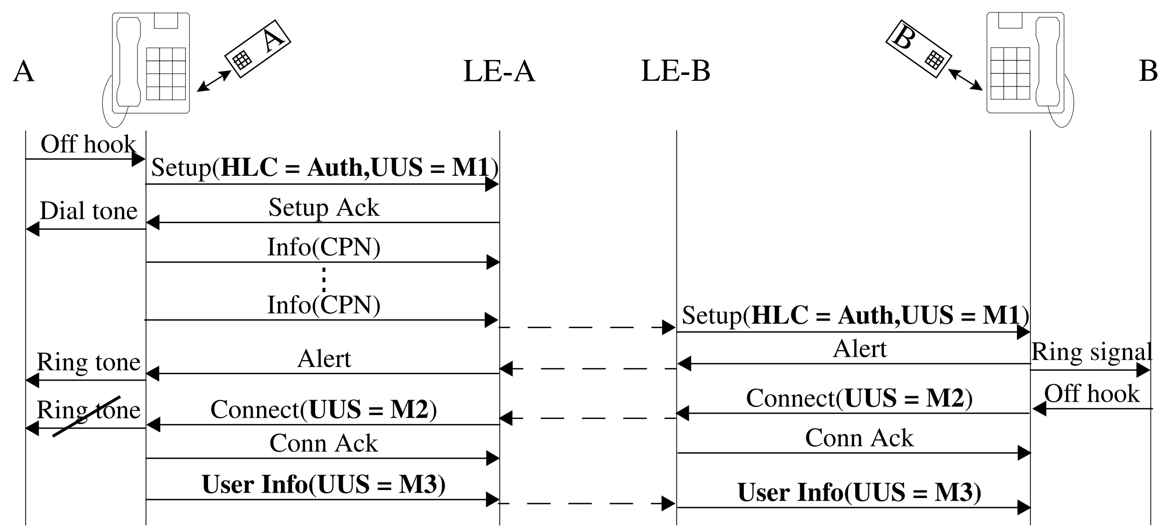


Figure 3: *Signalling Diagram of User-User Authentication within E-DSS1*

In the scenario described below phones are used as an example for terminals. Although this scenario can be used for general terminals.

At first the calling party puts its cryptocard into the phone and authenticates itself against the cryptocard. After going off hook, we use the UUS parameter of the setup signal to transmit the first message and the HLC parameter to indicate an authenticated call. When recognizing this incoming setup signal with HLC parameter including Auth, the receiving terminal will indicate the need for authentication to the called party. The user will notice the incoming call by the ring signal and the wish to authenticate by display information or special ring tones. Now the called party has to insert its cryptocard into the terminal (if not yet done) and authenticates itself against the cryptocard (to protect the card when it is lost).

Afterwards the called party hooks off. Now the second message is constructed by the terminal using the cryptocard. This message is transferred within the UUS parameter of the connect signal. This message is examined by the calling party and as a last step message three is composed and sent to the called party to complete the authentication procedure.

Unfortunately this last message must be transferred by an extra signal that has to be sent after connection establishment. This implies a small time delay in which the connection is established but the called party cannot be sure of the identity of the calling party whereas the calling party is sure of the identity of the called party after verifying message M2.

In the context of a phone call this is not a serious problem because the signal will take only little time to be transferred to B. When looking at data communications with terminals being computers the called party terminal has to wait for the third message on the signalling channel before it can be sure of the identity of the calling party resulting in a waste of bandwidth for a few milliseconds in one direction of the data channel.

4 Conclusion

Implementation
Our proposed implementation integrates the authentication protocol into the protocol of layer 3 subscriber signalling. There are no extra states and signals introduced. It is transparent to the network and meets the international standard for E-ISDN. Communication without authentication is possible with every terminal. To establish an authenticated call the terminals of all parties must support the authentication feature.

Terminals should check the HLC parameter of the setup signal to decide whether an incoming call is an authenticated call or not. Further it seems necessary for terminals including authentication functionality to indicate to the user whether the call is authenticated or not.

Security
The protocol that is the basis of our protocol has been subject to investigation for many years [9],[13]. Hence we suspect that also dedicated weaknesses are public and can be avoided e.g. by including the sender's identity within the encrypted data field as described in Fig. 1. We expect compatibility with different implementations because X.509 is an international well known standard. From the security view it seems to be most important to focus on the correct implementation including error handling.

If authentication fails, this can be interpreted as a rejected connection. The terminals may send a disconnect to clear the call if a message is not the expected one. It will be useful to add the cause of the rejection into the cause parameter of the disconnect signal to enable the peer terminal to differ between failed security functions and other failures.

Outlook
The next step consists in integrating further security mechanisms into or onto ISDN-Services that are transparent to the application and use the common key installed throughout the authentication procedure. Our protocol relies heavily on the availability of reliable certification hierarchies that periodically supply the terminals with the public keys of the involved parties. If this hierarchy cannot be installed or the signatures of the messages need too much time to be computed, the public key cryptosystem can be substituted by a symmetric cryptosystem.

It would be interesting to include authentication as a service even within established connections. Our protocol can easily be changed in order to realize closed user groups. This may be implemented by cryptocards that share a common private / public key or a symmetric key to authenticate each other. In this case there is no need for a certification hierarchy.

References

[1] ITU-T X.509,"The Directory: Authentication Framework", *ITU-T Recommendation X.509*, 1993

[2] ITU-T Q.931, "Digital Subscriber Signalling System No. 1, Network Layer", *ITU-T Recommendation* Q.931, 3/1993

[3] 1TR6,"Kennzeichenaustausch zwischen DIVO(ISDN)-Vermittlungsstellen und ISDN-Teilnehmereinrichtungen - ISDN-D-Kanal-Protokoll (Schicht 2 und 3)", Ausgabe 1.90, *Fernmeldetechnisches Zentralamt*

[4] Gerd Siegmund,"Grundlagen der Vermittlungstechnik", 2nd Ed., *R.v.Decker*,1992

[5] 1TR67, "Zeichengabe im Euro-ISDN an der Benutzer-Netz-Schnittstelle. Auswahl der Netzoptionen sowie ergänzende Festlegungen.", Ausgabe April 1994, Deutsche Telekom AG

[6] M. Warwick, "Feeling Insecure?", *Communications International*, January 1996, pp. 37

[7] B. Miller, "Vital signs of identity", *IEEE Spectrum*, February 1994, pp. 22-30

[8] R. L. Rivest, A. Shamir, L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems", *Communications of the ACM*, Volume 21, No. 2, February 1978, pp. 120-126

[9] C. I. Anson, C. Mitchell, "Security Defects in CCITT Recommendation X.509 - The Directory Authentication Framework", *ACM Computer Communication Review*, Vol. 20, No. 2, April, 1990, pp. 30-34

[10] A. Beutelspacher, J. Schwenk, K. - D. Wolfenstetter, "Moderne Verfahren der Kryptographie", Vieweg Mathematik, 1995, pp. 96 ff

[11] D. W. Davies, W. L. Price, "Security for Computer Networks", 2nd ed., *Wiley series in communication and distributed systems*, 1989

[12] A. Liebl, "Authentication in Distributed Systems: A Bibliography", *ACM Operating Systems Review*, Vol. 27, No. 4, October, 1993, pp. 31-41

[13] A. Liebl, "Formale Methoden zur Analyse von Authentifizierungsprotokollen", *Proc. Sicherheit in Informationssystemen*, Zürich-Irchel, März, 1994, pp. 74-88