Authentic Data Collection in an Untrustworthy Computer Environment

Thomas J. Wilke
tjw@prz.tu-berlin.de

24.05.2002

Abstract

A central matter in e-commerce and e-government is digital contracting and signing. The processes to be supported are to ensure a trustable and provable sequence of acts between the parties concerned following legal rules in a fast and flexible manner. Complex infrastructures like PKIs have been established to cope this task. But still there is a big lack: creating signatures in an untrustworthy computer environment. In this paper the problem of using desktop computer systems with or without card readers for signing documents containing binding declarations of intends for legal transactions will be discussed.

Starting with analysing the process of digital signing and its key issues, the representation will point out that today's common way of having documents digitally signed can be compromised despite the usage of so called class 3 card readers with smart cards and certificates of trustworthy CAs. The weak points of the employed signing process will be rolled out and a method will be introduced that guarantees the authenticity of signed data. As practical solution of the method explained a mobile device implementation pattern is presented. Finally an outlook will be given how to arrange processes and systems in order to provide e-contracting services in a flexible and trustable way.

Contents

1 Introduction

One commanding trend in the development processes of information technologies today is the activity to shift the partial information processing to a continuous process in order to provide faster, associative and more flexible services which reduce costs and processing time. Especially companies and public agencies are interested in these kinds of technologies since they intend on one hand to get closer to their customers by providing comprehensive and easy to use services which can be used anywhere and anytime fitting to the particular needs of each customer. On the other hand they want to tighten their internal operational processes to be more efficient. Usually the data processed in such a context is associated with real monetary or legal values. In order to protect these values it should be a need to ensure the trustworthiness of the computation in respect of availability and liability of the processed data. A method to prove the authenticity of digital data representations is digital signing. Commonly this method is used to sign software codes or documents. Digital document signing has a specially significance for e-commerce and e-government today; it provides the technical base to make binding declarations using exclusively digital data representations. The German legislator has staid abreast with these technical options by passing a law called "Gesetz über die Rahmenbedingungen für elektronische Signaturen (SigG)". This law has been amended in May 2002 by putting handwritten signatures par digital signatures. The regulations defining the signing process base on a certificate and a technical unit which is called "sichere Signaturerstellungseinheit" (secure signing unit). The common method of digital signing as practiced today does not meet the requirements the legislative demands for. This was shown by a research group at the University of Bonn which has found a way to successfully compromise digital signatures [HEI01].

The following sections will analyse the common way of creating digital signatures and will show how the signing process can be compromised. A method will be presented that guarantees the authenticity of signed data. As a practical solution of the method explained a mobile device implementation pattern is presented. Finally an outlook will be given how to arrange processes and systems in order to provide e-contracting services in a flexible and trustable way.

2 The process of digital signing

The basic mechanism used for digital signing today was first published by W. Diffie and M. Hellman 1976 in their article "New Directions in Cryptography" [DH76] describing a method to provide a key exchange for symmetric cryptography over insecure paths. The idea was to have a 2 key cryptography mechanism, called "Public-Key Cryptosystem", that could ensure that any person could send a cryptography key to a definite identity in a way, that only a person associated with that identity could get access to the key in plain text.

For digital signing this mechanism can be used in the other direction: an identity (person with a secret key) encrypts a message with it's secret key. The encrypted message can only be decrypted with the other matching key which has to be publicly available. Since the private key is only available to the specific identity and since the keys have to match for successful en- and decryption, the person using the public key can be sure that the successfully decrypted message comes from the specific identity. In order to reduce computation effort and to ensure a uniform length of each digital signature, not the whole message is encrypted but a unique value - called hash or hash value - which describes the whole message one to one.

With the fulfilment of the preconditions discussed in the following subsection the digital signing can be used to generate trustable control information that enables any person in conjunction with the signed data to prove the signing identity and to prove whether the data representation has been modified after signing process or not.

2.1 Preconditions for a trustworthy usage of digital signatures

Looking closer to the signing method described above the following preconditions are to ensure the trustworthiness of a digital signature:

1. the secret key should only be accessible to the person it has been dedicated to, since it represents its digital identity,

2. the public key distribution has to be trustworthy,

3. the used algorithms and key length should be adequate to the common security standards due to usage time in an insecure environment,

4. the process of generating a digital signature has to be done in a trustable environment,

5. the process of signature verification has to take place in a trustworthy decryption environment.

With the establishment of trust centers, certification authorities, the usage of smart cards, the employment of capable encryption as well as the procedures to generate, hand over and access keys the first three conditions are meet with available technology today. Modern smart cards are very important to hold up the chain of trustworthy since they are used to generate, store and ensure the legitimate usage of the private keys. The private keys will never be outside the cards because they are generated on the cards and normally no function is implemented to give away these keys. They are just used for internal encryption operations. Moreover the legitimate usage of a card has to be proved by a user. If this cannot be done the card would not operate.

Condition 4 and 5 is commonly performed on personal computers or workstations and in ideal case with smart cards. Inherently the architectures of these computer systems are designed to be very open, flexible and complex in function. The base organization scheme is a layered one which implies that a modification of base functional modules will take effect on above layered modules. Beyond this, the functionality can be changed while these systems are in operation. Using this kind of systems we are faced with the problem of secure booting. This means that it is not for sure that a user is always able to realize a change of functionality of such systems. Moreover it can happen that the system is took over by an attacker who can act under the identity of the user. Especially systems which can get data from untrustworthy sources - for example being connected to the internet, or being equipped with removable data storages like CD-, floppy disk drive and so on - are endangered. Even if a smart card is used to calculate the hash value and to encrypt the hash value, a user cannot be sure whether the smart card receives the authentic data he wants to sign. A similar situation is given when data is to be proved to be authentic.

3 The problem of authentic data collection

The secure booting problem and the problem to maintain the functional integrity during operation makes a common computer system untrustworthy in context with digital signing. Scientists at the University of Bonn have proved that the untrustworthiness is not a theoretical problem. They developed a Trojan horse which successfully compromised software used for digital signing [HEI01]. Even before the problem was known. The industry developed so called class 3 card readers to ensure the authenticity of bank transactions. These card readers have a little LCD-Display and an extended numeric keyboard with control keys build in. A user can check the parameters of the transaction displayed on the LCD and will have to commit them by pressing a control key. Since this solution is dedicated for verifying very little amount of data it is not capable to be used for digital signing.

A way to get a trustworthy system could be a restrictive configuration of the PC or workstation as far as functionality and flexibility are concerned. This could be done by reading executable codes and configuration data from read-only sources like a CD-ROM. Moreover functional complexity would have to be reduced drastically to provide only some very simple and elementary tasks in order to ensure functional integrity. Such a system would not be capable to meet the needs we are used to have on a common desktop computer. In consequence the strength of these systems in daily use would have gone. Another approach would be to dedicate a special machine only for the purpose of signing and verifying digital signatures. But this would also not be very comfortable for daily use and would handle digital signing and signature verification only.

As a result of considerations on a trustworthy digital signing process on a insecure desktop computer system the following operation sequence will have to be protected:

1. data collection from a trustable device (for example a keyboard) and transfer to a temporary storage

2. temporary storage of the collected data

3. hash calculation

4. public key encryption of the hash value

Thereby not only each operation is to be protected but also the whole operation sequence. If an attacker has the chance to compromise during operation steps or between these actions than trustworthiness would not exist.

Digital signing is a very special case where trustfully data collection is needed. As explained in the example to ensure the correctness of the bank transactions there are many other applications where a trustable data collection method is needed.

A more general way to guarantee a trustable data collection process is:

1. get the data to be collected from a source

2. check the correctness of the collected data

3. transform the collected data into a representation that can be proved to be authentic

In case the source and data delivery are trustable step 2 is not necessarily needed. Like the previous case of trustworthy data collection for digital signing, it is important to protect each step as well as the whole operation sequence. To do so the operation sequence has to be processed in a trustable area being defined as "Trusted Point" which protects the data collection process against unnoticed compromising attacks.

Definition: A "Trusted Point" is a computing environment, that only performs a dedicated functionality with ability to verify whether it has been manipulated or not.

To show an analogy the trusted point can be looked at as a secure chamber. Let a regent send commands to its general at war. To get the commands delivered the regent instructs a trustable knight. The command briefing takes places in a secure chamber to ensure that no one else but the trustable knight gets the commands. In order to assure the trustworthiness of the knight regent and general have agreed on a special suit of armour. Then the regent must be present in the secure chamber observing the messenger dressing the right amour. If this would not have taken place in the "secure chamber", the regent could not be sure that he has not been betrayed and that not anybody else than the right knight has been dressed up with the corresponding suit of armour.

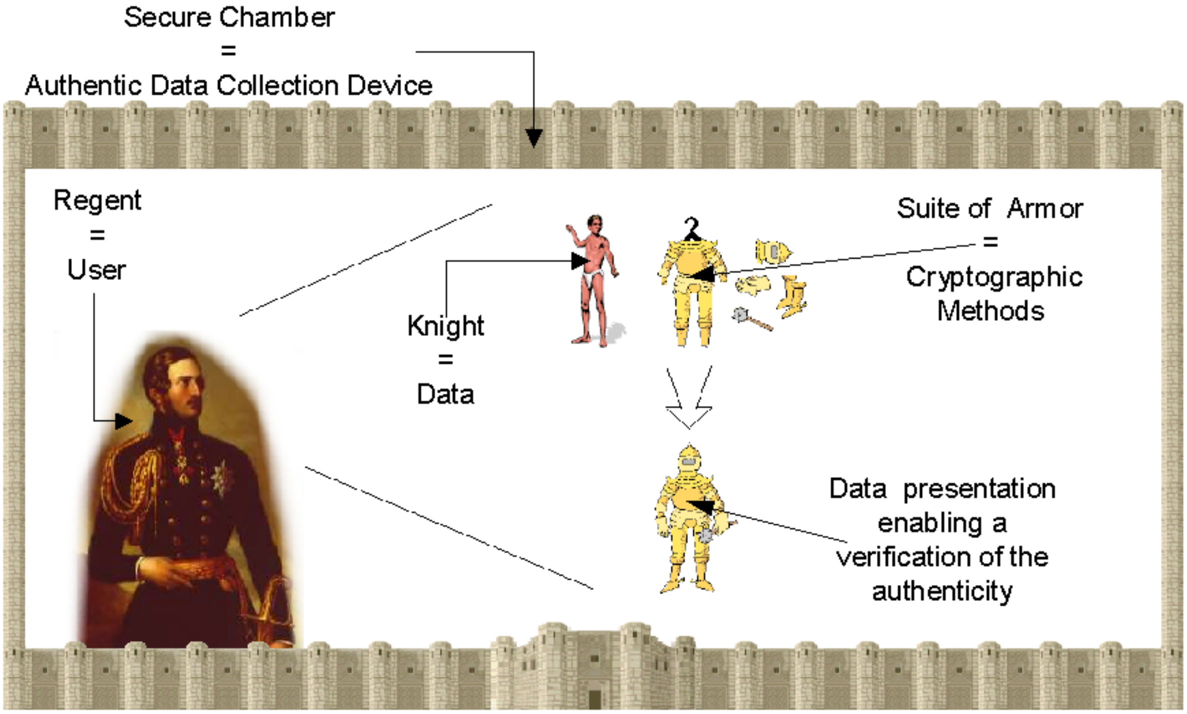
Secure Chamber
=
Authentic Data Collection Device
Regent
=
User
Knight
=
Data
Suite of Armor
=
Cryptographic Methods
Data presentation enabling a verification of the authenticity

Figure 1: Trusted Point Analogy

4 A practical solution for data collection in an untrustworthy computer environment

Till now it is not formally proved that a “Trusted Point” can be implemented by software means only because we face the problem of corrupt software computation on a malicious host. Therefore a dedicated hardware is proposed to implement the authentic data collection device.

4.1 Authentic Data Collection Device (ADD)

The first approach was to implement a keyboard with an enhanced functionality. A common keyboard has to be extended with the following additional units:

1. card reader
2. crypto unit
3. radio controlled clock and GPS receiver
4. video unit to fade in a data display over the regular video signal of the computer
5. RS232 interface

The modified keyboard can operate as regular keyboard or as authentic data collection device (ADD). If a chip card is inserted and the legitimate usage of the chip card is proved successfully by the user, the keyboard will change to the authentic data collection device operation mode. Otherwise the keyboard operation mode is activated. Since the keyboard provides different data representations due to the different operation modes a special keyboard driver has to be installed on the computer system. Depending on the operation mode the driver has either to act like a regular keyboard driver or in case of authentic data collection mode it has to convert the data received from the ADD to regular representation. Next to the transformation task the driver has to keep track between the regular and authentic provable data representation version.

If the authentic data collection operation mode has been activated only data are sent from ADD to the computer system which have such a data representation that can be proved to be authentic. The operation mode can only be switched to regular keyboard operation board by removing the chip card. Next to the data collection via the keyboard in authentic data collection mode, data can be sent from the computer system via the RS232 interface to the ADD. The data are faded in the regular monitor signal to enable the user to verify the data in a trustworthy environment. If the user recognizes the displayed data to be correct then he initiates the ADD to transform the data and transmit them to the computer.

The following state diagram will illustrate the operation modes.

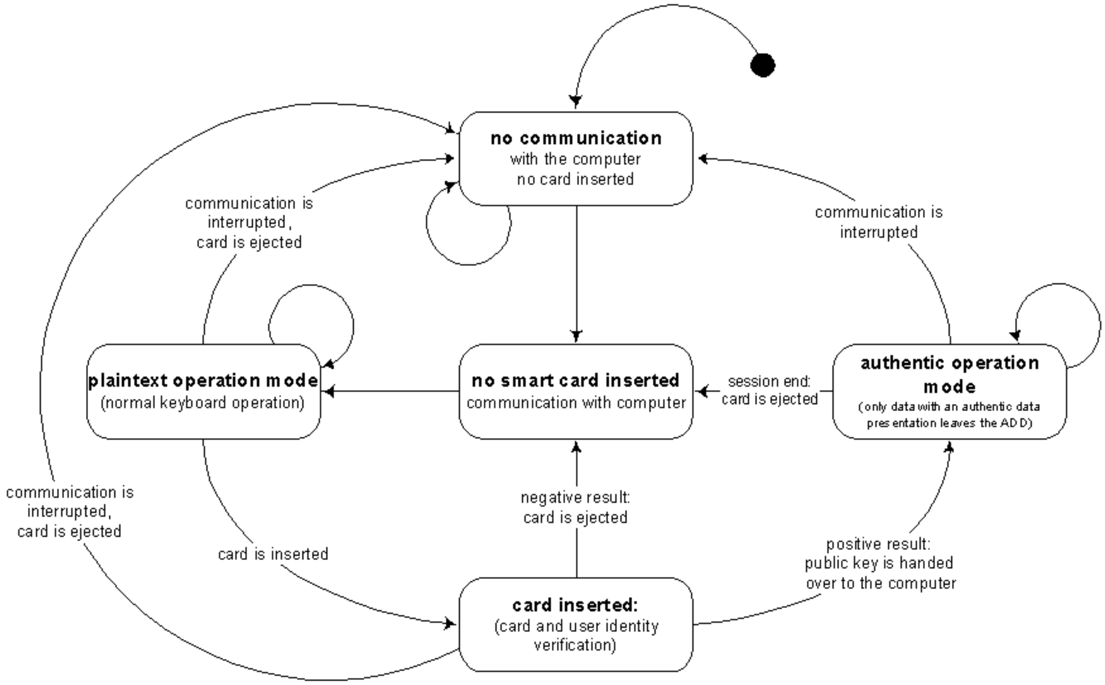

Figure 2: State Diagram of the Authentic Data Collection Device

The first design of the authentic data collection device in 1996 was originally dedicated for applications like e-mailing, bookkeeping systems, security trading systems, inter-bank transactions and so on which are done with desktop computer systems at an office workplace.

Due to the

- very complex modifications at the keyboard and the keyboard driver,
- the inability to have a computer platform neutral implementation,
- and the immobility of the device as well

it turned out that this implementation version would not fit to the common needs of today's information technology usage. In future we may have to handle much more trustworthiness tasks instead of only authentic data collection like e-contracting and e-payment. Therefore a more universal version in respect of system interoperability and possible fields of applications demanded re-design of the ADD which will now be called "Mobile Authentic Data collection Device (MADD)".

4.2 Mobile Authentic Data Collection Device (MADD)

The experiences with the first ADD design and the very wide range of new digital based devices as well as the new trends in mobile application computing, e-government and e-commerce defined the design environment of the “Mobile Authentic Data collection Device (MADD)”. Main design issues have been:

1. same basic features as the ADD
2. high flexibility in interoperation with different kinds of devices for authentic data collection and authentic data provision (mobile phones, scanner, audio recording devices, digital cameras, etc.)
3. an intuitive and very easy to use man machine interface
4. providing a trustworthiness environment for other elementary applications (digital signing, e-contracting, e-payment, advanced access control)
5. functional networked services (e-payment and contract supervision)
6. an approved and standardized hardware base which can be used as mobile device
7. card interface for removable non volatile memory cards
8. optionally: advanced person identification mechanisms

Architectural hardware base of MADD is a modified PDA with a smart card reader and a bluetooth-, infrared-, keyboard-, serial- and USB-interface integrated. Like other PDAs it has a headphone plug, a build in speaker and a LCD display. It supports data collection via

- writing with an stylus on the display
- the interfaces
- the build-in microphone or an external microphone

Other than common PDAs the case is protected to detect any manipulations. Functionality of MADD is fixed and can not be changed or modified by deleting or installing of executable software. Moreover special verification mechanisms are implemented to check whether authentic function of the hard- and software components is given or not. In case the authentic function can not be guaranteed the MADD will stop its operation and will indicate the detected compromised state. Like the ADD only data can leave the MADD via the memory card or any other interfaces that have an authentic provable data representation.

In order to get the operation of the MADD started a smart card has to be inserted into the integrated card reader. Then the user has to prove the smart card and/or the MADD to be legitimated to use the identity stored on the smart card via the trustworthy interfaces of the MADD. Optionally advanced man identification methods can be used like finger prints, voice and/or face recognition. Via the interfaces the MADD can receive different kinds of data. Since the MADD is trustworthy the user can check the visual and/or audio data and instruct the MADD to transform it into a authentic provable representation. Moreover data packages can be composed and certified to belong together. The modified representation can then be transmitted to another device through any interface of the MADD.

A new function of the MADD is to archive trustworthy data on a trustable memory card. The archive data can only be accessed from the identity which previously has archived the data. The memory can only be copied or removed from the MADD if the user instructs the device to do so. Before the memory is used the MADD checks the trustworthiness of the memory. Moreover the MADD can also be used to prove the authenticity of data.

5 E-contracting services, an outlook

What is called e-commerce today is an electronic supported commerce, which is only used to exchange not-binding information. This is caused by the untrustworthiness of the employed systems. Then, in the legal sense, the real circumstances cannot be proved by electronic documents.

With the MADD and the other described facilities a technical method is given to sign trustworthy contracts via electronic representation. Although we have a trustworthy base to do so, there are still some organisational issues to be discussed. For example if the parties have signed a contract it's not clear how each party can archive the electronic contract. How can the authenticity of the signatures be ensured after the certificates have been out-dated?

At the time being we are preparing the MADD to overcome these problems. The archiving issue will have to be solved in a two step mechanism. The whole signing process will be logged. In the first step the log, the signed contract and the certificates to prove the signatures will be stored on the local removable memory. In step two the contracting data are externally stored while ensuring privacy. This is either done with services in the internet or by duplicating the archived data to another memory card. To keep track of the certification outdating either the MADD and/or the Internet service has to remind in time for re-signing the contract. An alternative could be to automatically resign such documents. This could not be done offline, as the issuing trustcenter will have to manifest these kinds of mechanisms in its operation policy.

References

[HEI01] Heise News, *Digitales Signieren unsicher*

[Wil98] Thomas J. Wilke, *Verfahren und Vorrichtung zur Erfassung von Daten und deren Übermittlung in authentischer Form*, Offenlegungsschrift DE 197 03 970 A1, Jun. 1998, Deutsches Patentamt

[Wil97] Thomas J. Wilke, *Konzeption zur authentischen vertraulichen Kommunikation über offene Rechnernetze*, Sep. 1997

[Wil96] Thomas J. Wilke, *Konzeption eines modularen und vernetzten Zugangskontrollsystem auf Basis der Produktfamilie SIPORT*, Diplomarbeit, Jun. 1996

[DH76] W. Diffie, M.Hellman, *New Directions in Cryptography*, IEEE Trans. Inform. Theory, 1976, pp. 644-654