ON THE NEED TO DIVIDE THE SIGNATURE CREATION ENVIRONMENT

Jorge L. Hernandez-Ardieta, Ana I. Gonzalez-Tablas, Benjamin Ramos and Arturo Ribagorda
SeTI Research Group, Department of Computer Science, University Carlos III of Madrid
Av. Universidad 30, 28911 Leganes (Madrid), Spain

Keywords: Non-repudiation, Digital evidence, Electronic signature, Vulnerabilities.

Abstract: Electronic signatures have been legally recognized as the key element for boosting e-commerce under secure conditions. Several legislations throughout the world establish electronic signatures as legally equivalent to hand-written signatures, assigning them the property of evidence in legal proceedings. In addition, international standards define electronic signatures as non-repudiation evidence respecting the signed information. Bearing this in mind, it is obvious that the reliability of electronic signatures is paramount. However, the results show that several attacks on signature creation environments are feasible and easy to perform. As a result, the reliability of evidence is drastically undermined. We claim that the division of the environment becomes the most effective solution to counteract current threats. The formal proofs that support this statement are given along with an overview of the legal background and a summary of main potential threats on signature creation environments.

1 INTRODUCTION

Many efforts have been made to boost e-commerce, especially those related to the enhancement of e-commerce security. Maybe the most remarkable one has been the support given to the electronic signature (e-signature) by Governments and the IT Industry. An e-signature is generally considered as data in electronic form which are attached to or logically associated with other electronic data and which serve as a method of authentication (European Directive, 1999). Furthermore, several legislations throughout the world set the e-signature as legally equivalent to hand-written signatures, as well as assign them the property of evidence in legal proceedings (European Directive, 1999; Federal Trade Commission, 2001; Government of Canada, 2000).

International standards also consider e-signatures as non-repudiation evidence in electronic transactions, either when applied to e-commerce or other contexts (ISO/IEC DIS 13888-1, 1996). This scenario clearly places the e-signature as a key technology, and therefore its security is a critical issue which should be carefully analyzed.

However, several results show that the security of the vast majority of e-signature applications is not as trustworthy as it should be. The feasibility of some published attacks should make the reader consider if the property of non-repudiation and legal equivalence of e-signatures should be maintained under current conditions. If no secure means can be provided to the end user, a signatory should not be held liable for the commitment made in a signed message (i.e. a purchase order, a contract, an email). The signatory could allege that the corresponding private key was compromised or the document she intended to sign had been modified before generating the e-signature. If the case went to court, the signatory could easily prove on the balance of probabilities (in a civil action) or beyond reasonable doubt (in a criminal action) (McCullagh, 2000) that any potential attack could have been carried out to successfully forge the signature. However, the mere possibility of having to deal with the consequences arisen from forged signatures obviously undermines the user's confidence in technology, especially once legislations have legally reversed the onus of proof for e-signatures. The proof is now set on the signatory instead of on the verifier, contrary to traditional commerce.

We claim that an untrustworthy environment cannot generate reliable signatures, and thus cannot assure the non-repudiation of evidence. As every new proposal is always welcomed with a new attack, a completely different approach must be taken in or-

L. Hernandez-Ardieta J., I. Gonzalez-Tablas A., Ramos B. and Ribagorda A. (2009).
ON THE NEED TO DIVIDE THE SIGNATURE CREATION ENVIRONMENT.
In *Proceedings of the International Conference on Security and Cryptography*, pages 375-380
DOI: 10.5220/0002241037503380

der to enforce the reliability of evidence. We consider that dividing the signature creation environment can be the most effective solution to counteract current threats. By imposing the usage of different environments to create the legally binding evidence, the probability of a successful attack is highly decreased, even when those environments are untrustworthy on their own.

The article is organized as follows. The next section provides an overview of the legal background on e-signatures. Section 3 introduces the security problem for PKI-based e-signatures. The proofs that demonstrate the benefits of the division of the signature creation environment are given in section 4. We conclude the article in section 5.

2 LEGAL BACKGROUND

The European Directive on a Community framework for e-signatures (European Directive, 1999), the US e-Sign Act (Federal Trade Commission, 2001), the Canadian Personal Information Protection and Electronic Documents Act (Government of Canada, 2000) and the UNCITRAL Model Law on Electronic Signatures (United Nations, 2001) recognize electronic signatures as the means for promoting e-commerce under secure conditions.

After the approval of these laws, e-signatures are regarded as legally equivalent to hand-written signatures only if the requirements for hand-written signatures are fulfilled. In a nutshell, these requirements imply the signatory's intention of signing and the ability to uniquely identify the signatory based on her signature mark, from which a reasonable difficulty to forge the signature is derived. These requirements are transposed in a technology-neutral viewpoint by the laws. However, if we take a look at these requirements, it will be easily noticed that, taking into account the current state-of-technology, only digital signatures based on Public Key Cryptography (PKC) and a Public Key Infrastructure (PKI) satisfy them. In fact, current legislations implicitly consider PKC and PKI as the underlying technologies. For instance, the UNCITRAL Model Law establishes PKI and PKC as an example of implementing technologies for compliant e-signatures.

On the other hand, it is obvious that sensitive information is normally exchanged in an Internet transaction between the participant parties. In order to protect the parties against the other's misbehavior, digital evidence is generated during the transaction. Evidence is information that, either by itself or when used in conjunction with other information, is used to establish proof about an event or action. Zhou et al. have established certain conditions to be fulfilled by digital evidence (Zhou, 1997). The origin and the integrity of the evidence must be verifiable by a third party and the validity of the evidence must be undeniable. When evidence is acting as non-repudiation evidence, the party can not later successfully repudiate having participated in the transaction. ISO defines non-repudiation service as a service that protects the parties involved in a transaction against the other party denying that a particular event or action took place. In particular, ISO specifies non-repudiation services by means of asymmetric cryptography (ISO/IEC 13888-3, 1997). Therefore, PKC is considered again as the key technology for creating non-repudiation evidence in the form of digital signatures. When we say digital signatures, it can be automatically extended to the legally-supported e-signatures.

Other important issues are the differences between physical and digital realms. Although obvious, sometimes it seems that these differences have not been taken into account when writing previous laws. Governments throughout the world have directly transposed the legal validity of physical evidence like a hand-written signature to digital evidence like an e-signature. In fact, the legal consequences are exactly the same. However, the environment where the evidence is produced is completely different, as well as the threats that may arise. Furthermore, the environment where non-repudiation evidence is generally produced, the user's Personal Computer, is almost always untrusted.

3 SECURITY PROBLEM DEFINITION

Electronic signature technology used by end users is prone to suffer from a wide variety of attacks. Some of the most relevant threats that may subvert the security of the signature creation process are briefly referenced in this section. The aim is to provide an overview of practical and feasible attacks that can diminish the reliability of legally binding signatures.

3.1 Scenario

We consider a scenario where the user is in possession of a PKC key pair as well as a PKI digital certificate which bounds her identity with the public key. The user can store her private key either in a software keystore, like those managed by Web Browsers, or a

hardware device, such as a smart card (named Signature Creation Device, SCDev). The application that uses that private key for signing purposes is called the Signature Creation Application (SCA). The Signature Creation Environment (SCE) is the user's Personal Computer (PC), the most common environment used nowadays to communicate and purchase through the Internet.

The e-signature can be calculated over a local document (i.e. Microsoft Word document, PDF document, XML document, etc.), a Web content (i.e. data of a web form) or any other kind of information (i.e. local database information, raw data, etc.).

The context where the signature is created can be an e-commerce transaction, a contract signing or any other context which implies a legal commitment made by the signatory on the signed information. Therefore, the e-signature has some sort of legal effectiveness which cannot be repudiated by the signatory.

In this scenario, two assumptions are made:

- The user's PC is untrusted. It is not possible to obtain an certain level of assurance on the security of the PC since the user normally has no technical knowledge, the applied protection measures on the PC tend to be null (e.g. installation and periodical update of anti-virus and anti-malware programs, checking the SSL certificates when accessing to e-commerce Web sites, installation of the operating system security patches, etc.) and, although applied, these measures are not completely protective taking into account the number of Trojan horses, viruses and any kind of malware that can potentially infect the computer.
- The attacker has knowledge, resources and motivation enough for successfully carrying out the potential threats identified in section 3.2, provided that they are technically feasible according to the current state-of-technology.

3.2 Potential Threats

An attacker will always try to compromise the security of a system by focusing on the weakest element. If the attacker wants to subvert the security of an e-signature process, like those developed to support e-commerce transactions, the weakest point is the end user. The reason is twofold: on one hand, the security measures implemented in the end user's PC are generally low compared to those implemented by server systems. On the other hand, the user usually lacks of any sort of security knowledge, what remarkably aggravates the situation. Therefore, threats herein mentioned are only focused on undermining the security of the end user's PC where the signature creation process is to be carried out.

In this sense, most attacks try to compromise the cryptographic private key in order to generate forged signatures without the user's consent and knowledge (Dasgupta, 2007; Girard, 2003; Marchesini, 2005). Other attacks are focused on deceiving the user to sign a message posing to be the original one (Spalka, 2002). In this way, it is not possible to assure a reliable signature if complex document formats are being used, because the WYSIWYS (What You See Is What You Sign) property is completely undermined. WYSIWYS (Scheibelhofer, 2001) is a security measure that provides the signatory with a last step verification by means of a graphical representation of what is going to be signed. Once the signatory confirms the displayed information, it is the one supposed to be sent to the SCDev, and therefore the information on which the digital signature is computed. Nevertheless, if the document format allows the inclusion of complex data structures, active code or hidden text, then the semantic can vary depending on specific conditions, conditions that can be manipulated by the attacker. Therefore, the signatory could sign a document with the desired semantic meaning while the verifier could visualize a substantially different document. Because the syntactic of the signed information is maintained, the signature is correctly verified. This security problem has been widely studied in the literature (Alsaid, 2005; Jøsang, 2002; Kain, 2003), and has become one of the most dangerous issues in e-signatures.

Other attacks, called side-channel attacks, exploit the information leakage from physical characteristics of the hardware during the execution of the cryptographic algorithm. The aim of these attacks is to extract the private key from the SCDev. Depending on the hardware characteristic analyzed, these attacks are classified in Timing Analysis attack (Kocher, 1996; Schindler, 2000; Brumley, 2003), Power Analysis attack (Kocher, 1999; Fahn, 1999; Le, 2008), Electromagnetic Emanation attack (Quisquater, 2001; Gandolfi, 2001; Tanaka, 2008) or Microarchitectural attack (AciiÇmez, 2007a; AciiÇmez, 2007b; AciiÇmez, 2007c).

Finally, it is worth noting that design flaws and code bugs in the SCA or the underlying software are an endless source of vulnerabilities that an attacker can exploit to break the security of the signature process. Standard security evaluations like Common Criteria or Federal Information Processing Standards (FIPS) 140-2 can be applied to obtain a certain level of assurance on the security of the system. However, most manufactures are not keen on them due to their

high costs, and, though applied, just a level of assurance, and not a level of certainty, is achieved.

4 DIVIDING THE SIGNATURE CREATION ENVIRONMENT

It is obvious to recognize that perfect security does not exist. There will always exist a risk. A single signature creation environment will have a higher or lower probability of suffering an attack, but the probability is never null, specially under the scenario described in section 3. Our proposal consists in drastically reducing the probability of an attack by using several environments, instead of one. Obviously, there must be a trade-off between the added complexity and the security improvement. Formal proofs for our proposal are given further.

We consider that the legally binding evidence does not consist of a single e-signature, but several. Some protocols take into account a multi-signature based evidence, like fair non-repudiation (Kremer, 2002; Hernandez-Ardieta, 2008) or contract signing (Backes, 2006) protocols. Thus, these protocols are a perfect candidate to use more than just one signature creation environment.

4.1 Provable Benefits of Using Several Environments

Next, the proofs of the benefits of using several environments for the generation of the digital evidence are given. From here on, *environment* corresponds to a signature creation environment (i.e. a PC, a mobile device, etc.), and *signatory* to the end user that needs to generate an evidence based on PKI-based e-signatures.

Definition 1. An attack on an environment is an attack carried out by a malicious agent (active intruder or resident malware) which purpose is to obtain some benefit from the signature generation capabilities of that environment. Methods used by the malicious agent include obtaining the signing private key and deceiving the signatory to sign data different than the purported one.

Definition 2. The probability of a successful attack (PSA) on an environment depends on both the probability of a malicious agent (attacker) to gain access to that environment (undermine the environment's security measures) and the probability of that attacker to subvert the specific security measures implemented by the environment to protect the signature capabilities from unauthorized usages.

Claim 1. Increasing the number of environments needed in conjunction to generate the evidence enhances the reliability of the resultant evidence.

Proof 1. Suppose a set of environments $Set(E)$ of size $n \geq 2$, being n the number of possible environments available to the signatory, each of which with a specific PSA. The PSA on $Set(E)$ is given by next equation:

$$PSA(Set(E)) = \prod_{i=1}^{n} PSA(E_i) \quad (1)$$

We are considering the resultant PSA as the probability of occurrence of n independent events. However, subverting the security of a process in which several environments are needed implies a kind of collaborative attack from the attacker's side. As a consequence, the actual PSA would even be lower. Notwithstanding, we will maintain this value of PSA for the analysis.

Let $PSA(E)$ be the probability of a successful attack on a single environment E.

The PSA of $Set(E)$ is always lower than the PSA of a single environment E if an environment E' member of the set $Set(E)$ has a PSA lower than or equal to the PSA of the environment E, and at least one of the rest of the environments members of the set has a PSA lower than 1.

$$PSA(Set(E)) < PSA(E), if \exists E' \in Set(E) / PSA(E') \leq PSA(E) \wedge \prod_{i=1}^{n-1} PSA(E_i) \neq 1, E_i \in Set(E) \quad (2)$$

The direct consequence of *Proof 1* is that adding new environments - either equal to the former environment, and thus with equal PSA, or different to it, and thus with equal or different PSA - will always improve the security of the system by decreasing the final PSA. The assumption of adding environments to the set with a PSA lower than 1 is reasonable, as the signatory would never use an environment which is known a priori to be compromised.

Another obvious conclusion that can be derived from 1 is that if n tends to infinite, the PSA tends to 0, providing that new environments added have a PSA lower than 1.

$$\lim_{n \to \infty} \prod_{i=1}^{n} PSA(E_i) = 0, PSA(E_i) \neq 1 \quad (3)$$

4.2 Provable Benefits of Using Heterogeneous Environments

Section 4.1 has proved that using several environments increase the level of security of the system. This section analyses the impact of configuring a set $Set(E)$ of environments to be used for the evidence generation in case the set consists of either homogeneous or heterogeneous environments.

Definition 3. We define homogeneous environments as those environments that can be attacked by the same type of attacker. That is, their implemented security measures and the type of potential attacker are the same. As a result, the PSA for those environments remains the same. On the contrary, we define heterogeneous environments as those that, either due to their nature or the implemented security measures, different types of attacker must be considered. In this case, heterogeneous environments can have the same or different PSA.

Claim 2. Replicating the same environment in the set of environments $Set(E)$ (homogeneous environments) always provides a higher level of security than a configuration based on heterogeneous environments providing that the chosen environment is the most secure one among all possible environments.

Proof 2. Let $PSA_{hom}(E)$ be the resultant PSA of n homogeneous environments:

$$PSA_{hom}(E) = PSA(E).PSA(E)...PSA(E)$$

$$PSA_{hom}(E) = \prod_{i=1}^{n} PSA(E) = (PSA(E))^n \quad (4)$$

Let $PSA_{het}(E)$ be the resultant PSA of n heterogeneous environments:

$$PSA_{het}(E) = PSA(E_1).PSA(E_2)...PSA(E_n)$$

$$PSA_{het}(E) = \prod_{i=1}^{n} PSA(E_i) \quad (5)$$

From 4 and 5 we can deduce that:

$$PSA_{hom}(E) < PSA_{het}(E), if\ \exists\ PSA(E_j) < PSA(E_i), \quad \forall i = 1...n\ and\ j \in \{1...n\} \quad (6)$$

Claim 3. In a more general manner, replicating the same environment in the set of environments $Set(E)$ (homogeneous environments) provides a higher level of security if the resultant PSA in (4) is lower than that obtained from a configuration based on heterogeneous environments (5).

Proof 3. Basing on $PSA_{hom}(E)$ and $PSA_{het}(E)$ given in 4 and 5 respectively, there can be a configuration of homogeneous environments where:

$$(PSA(E_j))^n < \prod_{i=1}^{n} PSA(E_i),\ j \in \{1...n\} \quad (7)$$

However, in practice the signatory generally uses her PC as the signature creation environment. Once the PC is considered a highly risky environment (high PSA), Claims 2 and 3 are clearly difficult to be achieved. The conclusion is that, if the signatory uses her PC as one of the environments, using additional heterogeneous environments (i.e. a mobile device) will surely provide a higher level of security.

5 CONCLUSIONS

PKI-based e-signatures are legally binding according to International laws. Moreover, these signatures act as non-repudiation evidence in electronic transactions, preventing the parties denying that a particular event or action took place. However, several attacks can be easily performed on the environments where these signatures are produced. As a consequence, the reliability of the evidence is drastically undermined.

In this paper we have demonstrated that the division of the signature creation environment is an effective approach that can substantially decrease the probability of a successful attack, even though the environments used by the signatory are untrusted. As a result, the reliability of a multi-signature based evidence is clearly enhanced.

Our proposal can be applied to a wide variety of Internet protocols, like e-commerce or contract signing protocols, where several signatures must be performed by each participant. In (Hernandez-Ardieta, 2009) we propose a fair exchange protocol where the environment division principle has been applied to enhance to reliability of the protocol evidence.

ACKNOWLEDGEMENTS

This research has been partially supported by the Ministry of Industry, Tourism and Trade of Spain, in the framework of the project CENIT-Segur@, reference CENIT-2007 2004. (https://www.cenitsegura.es)

REFERENCES

AciiÇmez, O. (2007). Yet another MicroArchitectural Attack: Exploiting I-cache. In Proc. of the 2007 ACM workshop on Computer security architecture.

AciiÇmez, O., Ko, Ç. K., and Seifert, J.-P. (2007). On The Power of Simple Branch Prediction Analysis. 2007 ACM Symposium on Information, Computer and Communications Security (ASIACCS'07).

AciiÇmez, O., Schindler, W., and Ko, Ç. K. (2007). Cache Based Remote Timing Attack on the AES. Topics in Cryptology - CT-RSA 2007 (pp. 271-286.). Springer-Verlag, LNCS, series 4377.

Alsaid, A., and Mitchel, C. J. (2005). Dynamic content attacks on digital signatures. Information Management & Computer Security, 4 (13), 328-336.

Backes, M., Datta, A., Derek, A., Mitchell, J. C., Turuani, M. (2006). Compositional analysis of contract-signing protocols. Theoretical Computer Science 367, 33-56.

Brumley, D., and Boneh, D. (2003). Remote Timing Attacks are Practical. In Proc. of the 12th Usenix Security Symposium.

Dasgupta, P., Chatha, K., and Gupta, S. K. S. (2007). Vulnerabilities of PKI based Smartcards. In Proc. of the IEEE Military Communications Conference 2007 (MILCOM 2007).

European Directive 1999/93/CE of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures.

Fahn, P., and Pearson, P. (1999). IPA: A New Class of Power Attacks. In Proc. of CHES 1999 (pp. 173-186). Springer-Verlag, LNCS, series 1717.

Federal Trade Commission, Department of Commerce, United States of America. (2000). Electronic Signatures in Global and National Commerce Act.

Gandolfi, K., Mourtel, C., and Olivier, F. (2001). Electromagnetic Analysis: Concrete Results. In Proc. of the Cryptographic Hardware and Embedded Systems (pp. 251-261). Springer-Verlag, LNCS, 2162.

Girard, P., and Giraud, J-L. (2003). Software attacks on smart cards. Information Security Technical Report, 8 (1), 55-66.

Government of Canada, Department of Justice. (2000). Personal Information Protection and Electronic Documents Act.

Hernandez-Ardieta, J. L., Gonzalez-Tablas, A. I., Alvarez, B. R. (2008). An Optimistic Fair Exchange Protocol based on Signature Policies. Computers & Security, 27 (7-8), 309 - 322. Elsevier.

Hernandez-Ardieta, J. L., Gonzalez-Tablas, A. I., Ramos, B. (2009). Formal Validation of OFEPSP+ with AVISPA. Joint Workshop on Automated Reasoning for Security Protocol Analysis and Issues in the Theory of Security. Springer-Verlag, LNCS. (to appear)

International Organization for Standardization. (1996). ISO/IEC DIS 13888-1. Information technology - Security techniques - Non repudiation - Part 1: General model. ISO/IEC JTC1/SC27 N1503.

International Organization for Standardization. (1997). ISO/IEC 13888-3 Information technology - Security techniques - Non repudiation - Part 3: Mechanisms Using Asymmetric Techniques.

Jøsang, A., Povey, D., and Ho, A. (2002). What You See is Not Always What You Sign. In Proc. of the Australian UNIX User Group. Melbourne.

Kain, K. (2003). Electronic Documents and Digital Signatures. Master Thesis.

Kocher, P. C. (1996). Timing attacks on Implementations of Diffie-Hellman, RSA, DSS and Other Systems. In Advances in Cryptology - CRYPTO '96 (pp. 104-113). Springer-Verlag, LNCS series 1109.

Kocher, P., Jaffe, J., and Jun, B. (1999). Differential Power Analysis. In Proc. of CRYPTO 1999 (pp. 388-397). Springer-Verlag, LNCS series 1666.

Kremer, S., Markowitch, O., Zhou, J. (2002). An intensive survey of fair non-repudiation protocols. Computer Communications, 25, 1601-1621.

Le, T-H., Canovas, C., and Clediere, J. (2008). An overview of side channel analysis attacks. In Proc. of the 2008 ACM Symposium on Information, Computer and Communications Security (ASIACCS 2008).

Marchesini, J., Smith, S.W., and Zhao, M. (2005). Keyjacking: the surprising insecurity of client-side SSL. Computers & Security, 24 (2), 109-123.

McCullagh, A., and Caelli, W. (2000). Non-repudiation in the digital Environment. First Monday, 5 (8).

Quisquater, J.-J., and Samyde, D. (2001). ElectroMagnetic Analysis (EMA): Measures and Counter-measures for Smart Cards. In Proceeding of the International Conference on Research in Smart Cards (pp. 200-210). Springer-Verlag, LNCS, 2140.

Scheibelhofer, K. (2001). What You See Is What You Sign - Trustworthy Display of XML Documents for Signing and Verification. In Proc. of the IFIP TC6/TC11 International Conference on Communications and Multimedia Security Issues of the New Century.

Schindler, W. (2000). A Timing Attack against RSA with the Chinese Remainder Theorem. In Proc. of the Cryptographic Hardware and Embedded Systems (pp. 110-125). Springer-Verlag, LNCS, series 1965.

Spalka, A., Cremers, A. B., and Langweg, H. (2002). Trojan Horse Attacks on Software for Electronic Signatures. Informatica, 26, 191-203.

Tanaka, H. (2008). Evaluation of Information Leakage via Electromagnetic Emanation and Effectiveness of Tempest. IEICE Transactions on Information and Systems, 91 (5), 1439-1446.

Tiri, K. (2007). Side-channel Attack Pitfalls. In Proc. of the 44th ACM IEEE Design Automation Conference.

United Nations. (2001). UNCITRAL Model Law on Electronic Signatures with Guide to Enactment.

Zhou, J., and Gollmann, D. (1997). Evidence and Non-repudiation. Journal of Network and Computer Applications, 20 (3), 267-281.