Defining Electronic Authenticity: An Interdisciplinary Journey

Jean-François Blanchette
School of Library, Archival and Information Studies,
University of British Columbia
Jean-Francois.Blanchette@ubc.ca

Abstract

Since the mid-1990s, dozens of States, including those of the EU, have reformed their evidence laws so as to grant digital signature technologies the same proof value as handwritten signatures, as a mechanism for proving identity of authorship, consent to obligations, and integrity of electronic records after their transmission across time and space. Yet, several archival institutions (including the National Archives of Canada, Australia and France) have indicated they have no intention of preserving digitally signed records. This paper presents an overview of the development of the concept of digital signatures by the cryptographic research community, the process of its legal codification as evidence of contractual relations, and of its rejection by the archival community as a tool for long-term preservation of authentic records. The paper argues that acceptable definitions of "electronic authenticity" will emerge only when the relationships between technological solutions and social conventions are carefully articulated.

1. Introduction

Up until thirty years ago, cryptology essentially remained a military science, providing technologies to generals, diplomats, and spies wishing to communicate privately. In the 1960s, the security needs of the banking industry spurred the emergence of an academic cryptology research community, independent from the intelligence establishment. In 1976, this community made its presence widely known, with the publication of Diffie and Hellman's "New Directions in Cryptography" [1].

In this seminal paper, the authors simultaneously introduced a radically new method of key exchange, the concept of public-key cryptography, widely acknowledged as one of the most important development ever to occur in cryptography, and finally, suggested how public-key cryptography could be used to offer not only *confidentiality*, but also, *authentication* services: "In order to have a purely digital replacement for [written contracts], each user must be able to produce a message whose authenticity can be checked by anyone, but which could not have been produced by anyone else, even the recipient."

In a nutshell, public-key cryptography functions by assigning two keys to every user on a computer network: the *private key* can only be legitimately accessed by its owner, while the *public key* is made available to other users on the network through publicly accessible directories. The whole magic of public-key cryptography rests on the fact that while the private and public keys are mathematically related, *knowing the public key, it is computationally infeasible to deduce the private key*. To transmit a *confidential* electronic message over the network to user Bob, user Alice encrypts the message using Bob's public key, before sending it to him. Only Bob's private key will successfully decrypt the message. To "*sign*" a message, the role of each key is inversed: Alice encrypts the message using her private key before sending it to Bob. If Alice's public key successfully decrypts the message, Bob is then be convinced that only Alice could have signed that message.

The cryptological model for digital signatures is thus characterized by a signing algorithm, requiring the signer's private key, and a verification algorithm, requiring the signer's public key. Because the signer's public key is openly available on the network, users need not communicate prior to exchanging signed messages, thus providing an efficient system for securing commercial transactions. In practice, digital signatures are realized through public-key infrastructures (PKI), the enabling software, hardware and procedures providing the necessary key management, directory and revocation services.

2. Digital Signatures and Evidence Law

Clearly, widespread acceptance of the cryptological model of electronic signatures could only occur if the legal texts which specifically required that *written* signatures be used in transactions were modified. The

mid-nineties explosion of the Internet on the world scene, and the ensuing e-commerce "tidal wave" insured that, all over the world, governments lent a ready ear to calls for adapting their legislations in order to ensure the most favorable environment for the blossoming of e-commerce. Three texts played a particularly important role in the process of legal codification of the evidential value of digital signatures.

2.1. UNCITRAL Model Law on E-commerce

The United Nations Commission on Trade Law (UNCITRAL) is a UN organization with headquarters in Vienna. Created in 1966, the UNCITRAL is composed of thirty-six member States elected by the General Assembly, representative of the world's various geographic regions and its principal economic and legal systems. The UNCITRAL Model Law on electronic commerce was adopted in 1996, with the objectives of "falicitat[ing] the use of modern means of communications and storage of information, such as electronic data interchange (EDI), electronic mail and telecopy, with or without the use of such support as the Internet. It is based on the establishment of a functional equivalent for paper-based concepts such as 'writing', 'signature' and 'original'. By providing standards by which the legal value of electronic messages can be assessed, the Model Law should play a significative role in enhancing the use of paperless communication" [2].

The most fundamental principle of the Model Law is that of "non-discrimination": Article 5 of the Model Law states that "information shall not be denied legal effect, validity or enforce- ability solely on the grounds that it is in the form of a data message." The Model Law offers a *functional* definition for signatures, that is, "the signing method must enable one to identity the signer, and indicate that the signer manifests his consent." The Model Law has been a very influential document, cited as a reference by most electronic signature legislations and the principles of "non-discrimination" and of a "functional" definition of signatures have enjoyed widespread dissemination, as effective legal devices to negotiate the transition between the requirements of the paper-and-ink world, and the promises of the new electronic worlds.

2.2. ABA's Digital Signature Guidelines

The American Bar Association (ABA), through its Information Security Committee, has offered a set of guidelines, aimed at helping and influencing (US) State legislatures in the elaboration of digital signatures bills [3]. The first US State legislation to cover digital signatures, the *Utah Digital Signature Act*, was conceived in the spirit of the ABA guidelines, and became itself a "model law" for other state legislatures. Perhaps the most striking characteristic of the guidelines is their exclusive definition of electronic signatures as those based on public-key cryptography: "Digital signature, as used in these guidelines, does not include the results of encryption and decryption by means other than an asymmetric cryptosystem, nor does it include a digitized version of a handwritten signature, a typewritten signature, such as 'John Doe,' the use of passwords or other practices for controlling access, or any other computer-based representation of identity or authentication." Thus, the guidelines literally suggest that legislators "hardwire" into their texts the usage of asymmetric cryptology as the basis for signature systems, to the exclusion of other technology.

Since the passage of the Utah Act, other state legislatures (Minnesota, Washington) have followed the ABA lead in equating digital signatures with public-key cryptography technologies, while others (e.g., California) have allowed for less restrictive definition of allowable technologies.

2.3. European Union Directive

The EU has adopted on December 13, 1999 "a European Parliament and Council directive on a common framework for electronic signatures." [4] Given the transnational potential of electronic commerce, the European Parliament sought to rapidly establish a harmonized legal framework and avoid any obstacles to the promised expansion of the European Internal Market. At the same time, European regulators hoped to repeat the economic miracle of the GSM cellular telephony standard and provide n regulatory framework which could kick-start the nascent market for electronic signature products and related services.

In order to achieve this dual objective, the Directive defines two distinct kinds of signatures:

- *Simple electronic signatures* are defined as "data in electronic form which are attached to or logically associated with other electronic data and which serve as method of authentication";
- *Advanced electronic signatures* "means an electronic signature which meets the following requirements: (a) it is uniquely linked to the signatory; (b) it is capable of identifying the signatory; (c) it is created using means that the signatory can maintain under his sole control; (d) it is linked to the data to which it relates in such a manner that any subsequent change of the data is detectable."

While the first definition allows for a wide range of technologies, the second one is clearly directed at cryptographic signatures, since it is the only one that fulfills mandate (d). To create an incentive for market adoption of cryptographic signatures, each type of

signature is granted a distinct evidential value: simple electronic signatures are admissible, but the Directive does not specify their proof value; advanced electronic signatures are not only admissible, but Member States must grant them a value equivalent to that previously accorded to handwritten signatures.

In the period between 1997 and 2001, dozens of countries around the world amended their evidence law in order to account for electronic signatures, with a significant number adopting regulatory schemes inspired by the European Directive.

3. The Electronic Signature Lifecycle

Documents with legal value are archived with the idea that they provide evidence that may be used in some potential future litigation. Governmental administrations, businesses, and individuals are expected to preserve the documents, letters, records of transactions, bills, and contracts which prove their rights, so that these may be used later as evidence when some dispute arises over a transaction.

Preservation involves protection against two different threats: decay and attempts to modify the information on records. In the case of paper, such protection involves well-know parameters: using adequate media and ink (protection against material decay), some form of cataloguing (protection against decay of institutional memory), access control (protection against malicious modifications), and the use of experts to ascertain the integrity of questioned documents.

In the case of electronic documents, the parameters are somewhat different, and our experience with such protection is much more limited. Signed electronic documents introduce yet another variable into this equation: the evidence created by the electronic signature must also be preserved along with the document itself. That is, the archiving process must now deal with the problem of simultaneously ensuring document and signature legibility.

This dual requirement is made more visible by looking at the *lifecycle* of a cryptographic signature, which can be broken into four distinct steps: (1) **creation:** the cryptographic signature is created by the signer; the signed document is then sent to the person meant to receive it; (2) **initial verification:** upon receiving the electronically signed document, the destinatory verifies the signature, and if a success, proceeds with the actions related to the document; (3) **archiving:** the signed document is archived with view of preserving it as evidence in potential future litigation; (4) **litigation:** litigation does occur, the document is presented as evidence in front of a judge, and the signature verified again, so that the identity of the signer and the integrity of the document ascertained.

Of course, while phase four may only occur rarely, the entire point of the archiving process (apart from questions of institutional memory) is to provide for just such an event. A number of important problems arise because of the significant time which may elapse between step 2 and step 4. That is, while the initial verification may occur within seconds, minutes, or days of the signature creation, the later verification will occur potentially years after signature creation, and in the context of an archived document. What does this imply in terms of the evidence provided by a cryptographic signature?

Three distinct implications may be distinguished: (1) the decay of security as a consequence of scientific advances in cryptanalysis; (2) the availability, over long periods of time, of signature verification software; (3) the interaction between document legibility and integrity. These considerations have received uneven consideration from the technical community.

4. Technical Responses

The EESSI consortium (a standardization effort which seeks to translate the requirements of the European Directive on electronic signatures into European standards) has sought to address the need for ensuring the long-term preservation of cryptographically signed documents through its standard on "Electronic Signature Formats" [5]. The format distinguishes between two signature validation moments, "initial validation" and "late validation" (corresponding respectively to steps 2 and 4 of the signature lifecycle defined above). The format for late validation encapsulates all of the information that can be eventually used in the validation process, such as revocation information, timestamps, signature policies, etc. This information is gathered at the stage of initial validation.

The designers of these electronic signature formats were concerned with one primary security threat to the validity of the signature, one induced by decay in cryptographic strength (implication 1 above): "before the algorithms, keys and other cryptographic data used at the time the [electronic signature] was built become weak and the cryptographic functions become vulnerable, […] the signed data […] should be timestamped. If possible this should use stronger algorithms (or longer key lengths) than in the original timestamp. The timestamping process may be repeated every time the protection used to timestamp a previous [electronic signature] become weak."

That is, the primary security concern here is modeled as one where advances in cryptanalysis could make it possible, some years after the moment of signature creation, to deduce the original private signing key. Cryptographic signatures would then no longer provide credible evidence suitable for litigation

purposes, since such a scenario reproduces the conditions of a symmetric key cryptosystem—where signer and verifier both have access to the same key. To guard against this threat of decay, EESSI signatures are regularly timestamped afresh, with signing algorithms and key sizes appropriate to state-of-the-art cryptanalytic methods.

The problem of software longevity has been addressed in a 2000 report by EESSI, which introduced "Trusted Archival Services", a new type of commercial service that would be offered by yet to be specified competent bodies and professions, in order to guarantee the long-term preservation of cryptographically signed documents [6]. The report lists a number of technical requirements such archival services should provide, among them, "backward compatibility" with computer hardware and soft ware, through either preservation of equipment and/or emulation: "Trusted Archival Services (TAS) should maintain a set of applications (viewers as well as signature validation applications) together with the corresponding platforms (hardware, operating systems, etc) or at least an emulator of such applications and/or platforms in order to guarantee that the content of the documents can still be viewed and that the signature on these documents can still be validated years later (even if the technology is not available anymore at that time)."

While the cryptographic community has sought to deal with implications 1 and 2 (as defined above), it has not faced squarely implication 3, that of the interaction between document legibility and integrity. That issue has however received considerable attention in a community very much concerned with the long-term preservation of documentary evidence— that of archivists.

5. Archival Responses

While digital signatures have enjoyed great success in the legislative arena, this success has not translated into market share, and the PKI industry has repeatedly failed to realize the predictions of forecasters. However, several government—US, Canada and Australia, among others—have established significant PKI initiatives in order to make governmental programs and services available on-line in a secure manner.

Accordingly, several national archival institutions (among them, NARA, the National Archives of Canada, and Australia) have issued guidelines which seek to guide governmental agencies in the steps necessary to preserve digitally signed records, as required by the various rules governing such agencies. As well, archivists have initiated research projects, such as InterPARES (see www.interpares.org) designed to develop their understanding of the problem of preserving authentic electronic records, and the role which digital signatures might play in solving it.

Both professional and academic archivists soon discovered that digital signatures pose a fundamental dilemma: for signature verification to succeed, the integrity of the document must be preserved. It cannot be modified in any ways, whether through malicious intervention or through procedures aimed at countering the effects of hardware and software obsolescence, such as logical encoding migration. Such procedures necessarily tamper with the bitwise integrity of the document. Thus, cryptographic signatures freeze the signed document in its original state, forever forbidding any modification that would entail the inevitable failure of the signature verification process.

Therefore, archivists can either seek to preserve the document's legibility, performing the necessary format migration as made necessary by hardware and software evolution, or focus on preserving the bitwise integrity of the signed document, thus ensuring that the signature will be verify correctly, even if the document has, over time, become an unreadable and meaningless bit string.

The Canadian National Archives guidelines relatives to the preservation of digitally signed documents offer perhaps the bluntest assessment of the archival position with respect to the role of digital signatures in ensuring the evidential value of records: "for National Archives' purposes, the integrity and authenticity of records will continue to be inferred from their placement within an organization's record-keeping system during the normal course of business, and from proof of that organization's reliance on records kept within their record-keeping system" [7].

Such an assessment implies that, from the archivist's point of view, whatever security role digital signatures may have played prior to their transfer to the archives, they will have by then outlived their usefulness. Thus, "the National Archives will not attempt to maintain the capacity to re-verify a digital signature after transfer to its control, nor to preserve the traces of a digital signature generated under the current federal PKI system."

In its final report, the InterPARES project concluded that "digital signatures and public key infrastructures (PKI) are examples of technologies that have been developed and implemented as a means of authentication for electronic records that are transmitted *aross space*. Although record-keepers and information technology personnel place their trust in authentication technologies to ensure the authenticity of records, these technologies were never intended to be, and are not currently viable as a means of ensuring the authenticity of electronic records *over time*" [8].

6. Discussion

The journey of digital signatures, from mathematical discovery, to legally codified means of evidence, to failed technology as means of preserving the long-term authenticity of electronic records is interesting, if only for the enthusiasm with which the legal community embraced this new means of evidence.

An similar infatuation can be observed in the case of DNA profiling technology. While it was initially described as an irrefutable proof of identification, "a signature more credible than any other declaration", the technology encountered its first major challenge (and public humiliation) in the O.J. Simpson trial. As three sociologists of science remark, this was largely due to the fact that " […] the genetic fingerprint serves as a competent witness *if and only if* the chain of transactions during the extraction, transport, preservation, computation and analysis of the sample is attested to by witnesses, certified and duly registered by accountable personnel. To be considered as such, the truth contained in the automatic signature (the genetic bar-code) must be accompanied, surrounded, by a series of bureaucratic traces: handwritten signatures on standard forms, bar-codes attached to sample bags, etc" [9].

The situation is similar for electronic records: they can serve as "competent witnesses" of a legal act only if they are accompanied by other records, the "bureaucratic traces" which document all of the operations which the legal records might be subjected to: creation, modifications, annotations, signature, format migration, backup, copy, transfer, disposition, etc. To be credible, these operations must be performed by trusted information processing systems, that is, conforming to the criteria established by the archival community for the creation, management and preservation of electronic records. By surrounding an electronic record with a set of corroborating evidence, these traces guarantee its authenticity and integrity, even if these notions are no longer reducible to that of *physical integrity* of a bit string.

With regard to the issue of defining electronic authenticity, electronic records present the easiest case, as they are fairly locally standardized in form and content, and the legal world has already been dealing for some time with the technologies which have stretched the functionalities of paper — faxes, photocopiers, microfilms. In the case of more complex digital objects — e.g., databases, GIS, electronic music, engineering drawings — the challenges will lay in the careful articulation of the relationship between technological solutions and the social conventions which will ultimately determine what counts as authentic digital artifacts.

7. References

[1] W. Diffie and M. E. Hellman, "New Directions in Cryptography", *IEEE Trans. on Inf. Th.* **22** pp. 644–654 (1976).

[2] "UNCITRAL Model Law on Electronic Commerce and Guide to Enactment", UNCITRAL (1997).

[3] "Digital Signature Guidelines," American Bar Association (1996).

[4] European Parliament and Council, "Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community Framework for electronic signatures", *OJEC* **L13**, pg. 12–20 (2000).

[5] "Electronic Signature Formats ES 201 733", ETSI (2000).

[6] O. Libon, A. Mitrakas *et al.*, "European Electronic Signature Standardization Initiative—Trusted Archival Services" (2000).

[7] "Guidelines For Records Created Under a Public Key Infrastructure Using Encryption And Digital Signatures," National Archives of Canada (2001).

[8] InterPARES, *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (2002).

[9] M. Lynch, R. McNally and P. Daly, « Le tribunal, fragile espace de la preuve », *La Recherche* **300,** July-August 1997, pp. 112-115.