The Digital Signature Dilemma: To Preserve or Not to Preserve

Jean-François Blanchette
School of Library, Archival and Information Studies
University of British Columbia
Vancouver, Canada

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, consentement 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 digital signatures by the cryptographic research community, and the process of its legal codification as evidence of contractual relations. It argues that the process overlooked the problems induced by the need to preserve digital signatures over the long-term. It presents currently offered solutions to digital signature preservation, suggesting that they are in fact profoundly at odds with the principle of trusted custodianship at the heart of the archival profession.

I. 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.

II. Digital Signatures and Evidence Law

Clearly, widespread acceptance of the cryptological model of electronic signatures could only have occurred based on a number of factors: (1) legal texts which specifically required that *written* signatures be used in transactions had to be modified; (2) the strict controls regulating the use of cryptological technologies had to be softened, or altogether abandoned. Given the nature of the institutions in play (law, intelligence agencies), such changes should have taken decades to achieve, but 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 much readier ear to calls for adapting their legislations

and softening up cryptology control laws, 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.

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 "[i]nformation 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.

ABA's Digital Signature Guidelines

The American Bar Association (ABA), through its Information Security Committee, has offered a set of guidelines,[3] aimed at helping and influencing (US) State legislatures in the elaboration of digital signatures bills. 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.

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.

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

IV. Technical Responses

The technical responses to this problem have (so far) fallen under three distinct headings: trusted archival services, so-called "resignature", and canonicalization.

Trusted Archival Services

The concept of "Trusted Archival Services" was introduced in the context of the EESSI standardization effort, which seeks to translate the requirements of the European Directive on electronic signatures into European standards. It refers to 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 integrity of cryptographically signed documents.

An EESSI report[5] 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)." Such requirements may seem surprising at first. Why should TAS act as information technology museums or invest in emulation strategies? What is the rationale for such stringent requirements, which do not even account for the simplest and most widely accepted archival strategy, migration?

The answer simply lies in the fundamental dilemma facing archivists seeking to preserve both document legibility and cryptographic signature legibility: 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 archival procedures, such as logical encoding migration, which necessarily tamper with the bitwise integrity of the document. 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.

This essential characteristic of cryptographic signatures has failed to surface in the technical literature, which has preferred to settle on the more familiar issues of cryptographic key strength, whatever their plausibility or actual relevance.

Resignature

The EESSI consortium has also sought to address the need for ensuring the long-term integrity of cryptographically signed documents through its standard on "Electronic Signature Formats".[6] 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, while initial validation is used to gather this information to construct late validation format.

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: "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.

Canonicalization

The Internet Engineering Task Force (IETF) has developed specifications to another approach to dealing with the issue of long-term preservation of cryptographic signatures, that of canonicalization. In computer science, canonical refers to the process of conforming to an authoritative or authorized definition. In this case, canonicalization refers to the process of translating an encoded text into a version conformant with some canonical definition of that encoding.

The perceived usefulness of canonicalization for digital signatures is made clear in the case of the S/MIME secure messaging format,[7] which defines the various data structures making it possible to cryptographically sign plain text email messages. Unfortunately, there are no universally adopted standards for representing plain text on computing platforms: even though ASCII is available on most modern computing platforms, it is a standard for *character*, not *text* encoding. Thus, Windows, Mac, and Unix platforms all use different characters for indicating end of lines. This poses very real problem for cryptographic signatures, which cannot tolerate any modification of the original message — even one involving a change of invisible characters.

Thus, the S/MIME standard specifies that: "each MIME entity MUST be converted to a canonical form that is uniquely and unambiguously representable in the environment where the signature is created and the environment where the signature will be verified. […] The most common and important canonicalization is for text, which is often represented differently in different environments. MIME entities of major type "text" must have both their line endings and character set canonicalized." Thus, the S/MIME compliant sending agent processes the email message so that it conforms to the canonical encoding of plain text required by the standard. This will enable the receiving agent to adequately process the message and to verify the signature.

In practice, the effect of using canonical formats is to perform a format migration before the signature occurs, thus minimizing the effect of logical format decay. In this way, documents that have undergone canonicalization are less susceptible to simple transformations of the logical format (such as whitespace normalization), which immediately invalidate digital signatures.

V. Archival Responses

Faced with either legislation granting special evidential value to digitally signed documents or with government-wide PKI development projects (as is the case with the US, Canada and Australia), archival institutions have had to determine how they would deal with cryptographically signed records.

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 records which may be digitally signed, as required by the various rules governing such agencies. As well, archivists have initiated research projects, such as InterPARES, designed to develop their understanding of the problem of preserving authentic electronic records, and the role which digital signatures might play in solving it.

United States

The United States have not, as suggested by the American Bar Association, enacted federal rules of evidence explicitly granting cryptographic signatures special status as evidence. However, The National Institute of Standards and Technology (NIST) is leading the development of a Federal Public Key Infrastructure, in coordination with industry and technical groups. The National Archives and Records Administration thus issued in 2000 guidelines intended to help agencies expecting to produce, retain, and eventually transfer to NARA, digitally signed documents.[8]

The guidelines distinguish between the content, context and structure of electronic records, noting, "for a record to remain reliable, authentic, […] it is necessary to preserve its content, context, and sometimes structure." Arguing that digital signatures are simultaneously part of the content, of the context, and of the structure of a digitally signed document, the guidelines go on to suggest that "an agency may determine that it is necessary to maintain the structure of the electronic signature. In that is case, it is necessary to retain the hardware and software that created the signature (e.g., chips or encryption algorithms) so that the complete

record could be validated at a later time." The guidelines do not offer advice as to when an agency may come to such a determination.

The guidelines do not provide definite answers to the problem of digital signature preservation. Two distinct approaches are suggested: on the one hand, retaining contextual information to adequately document the processes in place at the time the record was electronically signed. That is, "the agency's preserves the signature's validity and meets the adequacy of documentation requirements by retaining the contextual information that documented the validity of the electronic signature at the time the record was signed." Such an approach is deemed more appropriate for records with long-term retention requirements, as it is less subject to the effects of technological obsolescence.

On the other hand, agencies may preserve the ability to validate signatures, that is, preserving both the contextual and structural information of the record: "this approach is potentially more burdensome, particularly for digitally-signed records with long retention needs, due to issues of hardware and software obsolescence." The guidelines do not offer guidance as to what may constitute "long" retention needs in the context of digitally signed records.

Canada

The 1999 Throne Speech announced an ambitious plan to make all federal programs and services available on-line by 2005. A key element of such a plan has been the establishment of the Government of Canada Public Key Infrastructure project to meet the security requirements of federal electronic services delivery. Equally importantly, the project would provide a key market for the nascent Canadian PKI industry, in particular, Entrust, an offshoot of the now defunct Bell-Northern Research.

The Canadian National Archives have thus issued guidelines relatives to the preservation of digitally signed documents.[9] The guidelines 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."

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."

InterPARES

The International Research on Permanent Authentic Records in Electronic Systems, known as the InterPARES 1 project, took place from 1999 to 2001. Its goal was to develop the theoretical and methodological knowledge essential to the permanent preservation of authentic records generated and/or maintained electronically, and, on the basis of this knowledge, to formulate model policies, strategies and standards capable of ensuring that preservation. It was composed of archivists, both from academia and from major archival institutions, among others, NARA, The National Archives of Canada, Australia, France, the Netherlands, Sweden, the UK, and China.

The report of the Authenticity Task Force,[10] entrusted with the mission of identifying "conceptual requirements for assessing and maintaining the authenticity of electronic records," adopts an firm position with regard of the role of digital signature technologies and PKI as a means of ensuring the authenticity of records: "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 *across 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*."

InterPARES has indicated that further research is necessary to establish the impact of digital signature technologies on electronic record management: "What are the implications of their use [digital signature technologies] for the management of authentic electronic records over the long term? Will their implementation impede the long-term management of authentic electronic records? Can the use of digital signatures be adapted and extended to support the long-term preservation of authentic electronic records. What specific adaptations and extensions would be necessary?" Such questions are currently being investigated in the context of the InterPARES 2 project, which is expected to conclude in 2006.

VI. Discussion

The security afforded by digital signatures thus poses archival institutions an impossible dilemma: either preserve the ability to validate the signatures (and the proof value of documents, if so legislated), and risk that the documents themselves become unreadable as logical encoding formats evolve; or migrate documents in order to maintain their legibility and in the process, immediately invalidate their signatures.

Once this fundamental dilemma is understood, the equivocal nature of the various guidelines is better understood. On the one hand, preserving the electronic signatures and the means for their validation over time is beyond the technical means available to archival institutions (or to anyone else for that matter). In addition, preserving the signatures means that archivists have to forego any preservation strategy that involves migrating the logical formats of the signed documents, an impossible choice. On the other hand, resorting to the time-tested archival method

of recording contextual information—by capturing metadata about the events of signature creation and validation—negates much of the perceived usefulness of digital signatures as a means of evidence.

Thus, the various guidelines examined in the paper all reach the somewhat disappointing insight that digital signatures offer little in the way of preserving a record's integrity, reliability, and authenticity over time—or rather, do so at the cost of sacrificing its legibility. While, as the InterPARES project remarks, the ability of digital signatures to preserve the authenticity of documents transmitted *over time* is in question, their value as means to do so *across space* is not. Obviously, digital signatures remain an unsurpassed technology for verifying that a document indeed originates from a given person, and that it has not been modified in transit. Their limitations must, however, be squarely addressed, so that they can be adequately integrated within records management policies in a such a way that they contribute to, rather than further complexify, the problem of preserving authentic electronic documents.

References

1. D. Wittfield and M. E. Hellman, "New Directions in Cryptography", *IEEE Trans. on Inf. Th.* **22** pg. 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. O. Libon, A. Mitrakas *et al.*, "European Electronic Signature Standardization Initiative—Trusted Archival Services" (2000).
6. "Electronic Signature Formats ES 201 733", ETSI (2000).
7. "RFC 2311: S/MIME Version 2 Message Specification," IETF (1998).
8. "Records Management Guidance for Agencies Implementing Electronic Signature Technology", NARA (2000).
9. "Guidelines For Records Created Under a Public Key Infrastructure Using Encryption And Digital Signatures," National Archives of Canada (2001).
10. InterPARES, *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (2002).

Biography

Jean-François Blanchette received a B.Sc. (1995) and a M.Sc. (1997) in computer science with a specialization in cryptography, both from the Université de Montréal, and a Ph.D. (2002) in social studies of science and technology from Rensselaer Polytechnic Institute. In 2001, he served on a task force mandated by the French Ministry of Justice to make proposals regarding the adaptation of French evidence law to electronic signatures. He is currently a Post-doctoral Fellow at the School of Library, Archival and Information Studies, where he is working on the articulation of information technology policies drawing on both computer and archival science principles.

Jean-François Blanchette
School of Library, Archival and Information Studies
University of British Columbia
301-6190 Agronomy Rd, Vancouver, Canada V6T 1Z3
Tel.: +1 604 822 2694; Fax: +1 604 822 6006
Email: Jean-Francois.Blanchette@ubc.ca
Web: http://www.slais.ubc.ca/jeanf/