ABA Defending Liberty Pursuing Justice

American Bar Association

Member Login

Home | JOIN | CLE | Lawyers | Education | Public | Store | Events | About | Contact

Search the ABA Web Site GO Advanced Search Print Page

HOME

- E-Commerce and IT Division
- Life & Physical Sciences Division
- Special Committees
- NCLS / AAAS
- *The SciTech Lawyer*
- *e-BLAST*
- *Jurimetrics*
- Publications
- Calendar of Events
- E-mail Discussion Lists
- Sites of Interest
- About the Section
- Contact Us
- Join the Section

Digital Signature Guidelines

Information Security Committee
Science and Technology Section
American Bar Association

American Bar Association
Section of Science and Technology
Information Security Committee

Digital Signature Guidelines Tutorial

Digital Signature Guidelines Press Release

Tutorial

In today's commercial environment, establishing a framework for the authentication <1> of computer-based information requires a familiarity with concepts and professional skills from both the legal and computer security fields. Combining these two disciplines is not an easy task. Concepts from the information security field often correspond only loosely to concepts from the legal field, even in situations where the terminology is similar. For example, from the information security point of view, "digital signature" means the result of applying to specific information certain specific technical processes described below. The historical legal concept of "signature" is broader. It recognizes any mark made with the intention of authenticating the marked document. <2> In a digital setting, today's broad legal concept of "signature" may well include markings as diverse as digitized images of paper signatures, typed notations such as "/s/ John Smith," or even addressing notations, such as electronic mail origination headers.

From an information security viewpoint, these simple "electronic signatures" are distinct from the "digital signatures" described in this tutorial and in the technical literature, although "digital signature" is sometimes used to mean any form of computer- based signature. These Guidelines use "digital signature" only as it is used in information security terminology, as meaning the result of applying the technical processes described in this tutorial.

To explain the value of digital signatures in legal applications, this tutorial begins with an overview of the legal significance of signatures. It then sets forth the basics of digital signature technology, and examines how, with some legal and institutional infrastructure, digital signature technology can be applied as a robust computer-based alternative to traditional signatures.

Signatures and the Law

A signature is not part of the substance of a transaction, but rather of its represen tation or form. Signing writings serve the following general purposes:<3>

- **Evidence:** A signature authenticates a writing by identifying the signer with the signed document. When the signer makes a mark in a distinctive manner, the writing becomes attributable to the signer.<4>
- **Ceremony:** The act of signing a document calls to the signer's attention the legal significance of the signer's act, and thereby helps prevent "inconsiderate engagements.<5>
- **Approval:** In certain contexts defined by law or custom, a signature expresses the signer's approval or authorization of the writing, or the signer's intention that it have legal effect.<6>
- **Efficiency and logistics:** A signature on a written document often imparts a sense of clarity and finality to the transaction and may lessen the subsequent need to inquire beyond the face of a document.<7> Negotiable instruments, for example, rely upon formal requirements, including a signature, for their ability to change hands with ease, rapidity, and minimal interruption.<8>

The formal requirements for legal transactions, including the need for signatures, vary in different legal systems, and also vary with the passage of time. There is also variance in the legal consequences of failure to cast the transaction in a required form. The statute of frauds of the common law tradition, for example, does not render a transaction invalid for lack of a "writing signed by the party to be charged," but rather makes it unenforceable in court,<9> a distinction which has caused the practical application of the statute to be greatly limited in case law.

During this century, most legal systems have reduced formal requirements,<10> or at least have minimized the consequences of failure to satisfy formal requirements. Nevertheless, sound practice still calls for transactions to be formalized in a manner which assures the parties of their validity and enforceability.<11> In current practice, formalization usually involves documenting the transaction on paper and signing or authenticating the paper. Traditional methods, however, are undergoing fundamental change. Documents continue to be written on paper, but sometimes merely to satisfy the need for a legally recognized form. In many instances, the information exchanged to effect a transaction never takes paper form. Computer-based information can also be utilized differently than its paper counterpart. For example, computers can "read" digital information and transform the information or take programmable actions based on the information. Information stored as bits rather than as atoms of ink and paper can travel near the speed of light, may be duplicated without limit and with insignificant cost.

Although the basic nature of transactions has not changed, the law has only begun to adapt to advances in technology. The legal and business communities must develop rules and practices which use new technology to achieve and surpass the effects historically expected from paper forms.

To achieve the basic purposes of signatures outlined above, a signature must have the following attributes:<12>

- **Signer authentication:** A signature should indicate who signed a document, message or record,<13> and should be difficult for another person to produce without authorization.
- **Document authentication:** <14> A signature should identify what is signed, <15> making it impracticable to falsify or alter either the signed matter or the signature without detection.

Signer authentication and document authentication are tools used to exclude impersonators and forgers and are essential ingredients of what is often called a "nonrepudiation service" in the terminology of the information security profession. A nonrepudiation service provides assurance of the origin or delivery of data in order to protect the sender against false denial by the recipient that the data has been received, or to protect the recipient against false denial by the sender that the data has been sent. <16> Thus, a nonrepudiation service provides evidence to prevent a person from unilaterally modifying or terminating legal obligations arising out of a transaction effected by computer-based means. <17>

- **Affirmative act:** The affixing of the signature should be an affirmative act which serves the ceremonial and approval functions of a signature and establishes the sense of having legally consummated a transaction.
- **Efficiency:** Optimally, a signature and its creation and verification processes should provide the greatest possible assurance of both signer authenticity and document authenticity, with the least possible expenditure of resources.

Digital signature technology generally surpasses paper technology in all these attributes. <18> To understand why, one must first understand how digital signature technology works.

How Digital Signature Technology Works

Digital signatures are created and verified by cryptography, the branch of applied mathematics that concerns itself with transforming messages into seemingly unintelligible forms and back again. Digital signatures use what is known as "public key cryptography," which employs an algorithm using two different but mathematically related "keys;" one for creating a digital signature or transforming data into a seemingly unintelligible form, and another key for verifying a digital signature or returning the message to its original form. <19> Computer equipment and software utilizing two such keys are often collectively termed an "asymmetric cryptosystem."

The complementary keys of an asymmetric cryptosystem for digital signatures are arbitrarily termed the private key, which is known only to the signer <20> and used to create the digital signature, and the public key, which is ordinarily more widely known and is used by a relying party to verify the digital signature. If many people need to verify the signer's digital signatures, the public key must be available or distributed to all of them, perhaps by publication in an on-line repository or directory where it is easily accessible. Although the keys <21> of the pair are mathematically related, if the asymmetric cryptosystem has been designed and implemented securely <22> it is "computationally infeasible <23> to derive the private key from knowledge of the public key. Thus, although many people may know the public key of a given signer and use it to verify that signer's signatures, they cannot discover that signer's private key and use it to forge digital signatures. This is sometimes referred to as the principle of "irreversibility."

Another fundamental process, termed a "hash function," is used in both creating and verifying a digital signature. A hash function is an algorithm which creates a digital representation or "fingerprint" in the form of a "hash value" or "hash result" of a standard length which is usually much smaller than the message but nevertheless substantially unique to it. <24> Any change to the message invariably produces a different hash result when the same hash function is used. In the case of a secure hash function, sometimes termed a "one-way hash function," it is computationally infeasible <25> to derive the original message from knowledge of its hash value. Hash functions therefore enable the software for creating digital signatures to operate on smaller and predictable amounts of data, while still providing robust evidentiary correlation to the original message content, thereby efficiently providing assurance that there has been no modification of the message since it was digitally signed.

Thus, use of digital signatures usually involves two processes, one performed by the signer and the other by the receiver of the digital signature:

- **Digital signature creation** uses a hash result derived from and unique to both the signed message and a given private key. For the hash result to be secure, there must be only a negligible possibility that the same digital signature could be created by the combination of any other message or private key.
- **Digital signature verification** is the process of checking the digital signature by reference to the original message and a given public key, thereby determining whether the digital signa ture was created for that same message using the private key that corresponds to the referenced public key.

To sign a document or any other item of information, the signer first delimits precisely the borders of what is to be signed. The delimited information to be signed is termed the "message" in these Guidelines. Then a hash function in the signer's software computes a hash result unique (for all practical purposes) to the message. The signer's software then transforms the hash result into a digital signature using the signer's private key. <26> The resulting digital signature is thus unique to both the message and the private key used to create it.

Typically, a digital signature (a digitally signed hash result of the message) is attached to its message and stored or transmitted with its message. However, it may also be sent or stored as a separate data element, so long as it maintains a reliable association with its message. Since a digital signature is unique to its message, it is useless if wholly disassociated from its message.

Verification of a digital signature is accomplished by computing a new hash result of the original message by means of the same hash function used to create the digital signature. Then, using the public key and the new hash result, the verifier checks: (1) whether the digital signature was created using the corresponding private key; and (2) whether the newly computed hash result matches the original hash result which was transformed into the digital signature during the signing process. The verification software will confirm the digital signature as "verified" if: (1) the signer's private key was used to digitally sign the message, which is known to be the case if the signer's public key was used to verify the signature because the signer's public key will verify only a digital signature created with the signer's private key; <27> and (2) the message was unaltered, which is known to be the case if the hash result computed by the verifier is identical to the hash result extracted from the digital signature during the verification process.

Various asymmetric cryptosystems create and verify digital signatures using different algorithms and procedures, but share this overall operational pattern.

The processes of creating a digital signature and verifying it accomplish the essential effects desired of a signature for many legal purposes:

- **Signer authentication:** If a public and private key pair is associated with an identified signer, the digital signature attributes the message to the signer. The digital signature cannot be forged, unless the signer loses control of the private key (a "compromise" of the private key), such as by divulging it or losing the media or device in which it is contained.
- Message authentication: The digital signature also identifies the signed message, typically with far greater certainty and precision than paper signatures. Verification reveals any tampering, since the comparison of the hash results (one made at signing and the other made at verifying) shows whether the message is the same as when signed.
- **Affirmative act:** Creating a digital signature requires the signer to use the signer's private key. This act can perform the "ceremonial" function of alerting the signer to the fact that the signer is consummating a transaction with legal consequences. <28>
- **Efficiency:** The processes of creating and verifying a digital signature provide a high level of assurance that the digital signature is genuinely the signer's. As with the case of modern electronic data interchange ("EDI") the creation and verification processes are capable of complete automation (sometimes referred to as "machinable"), with human interaction required on an exception basis only. Compared to paper methods such as checking specimen signature cards -- methods so tedious and labor-intensive that they are rarely actually used in practice -- digital signatures yield a high degree of assurance without adding greatly to the resources required for processing.

The processes used for digital signatures have undergone thorough technological peer review for over a decade. Digital signatures have been accepted in several national and international standards developed in cooperation with and accepted by many corporations, banks, and government agencies. <29> The likelihood of malfunction or a security problem in a digital signature cryptosystem designed and implemented as prescribed in the industry standards is extremely remote, <30> and is far less than the risk of undetected forgery or alteration on paper or of using other less secure electronic signature techniques.

Public Key Certificates

To verify a digital signature, the verifier must have access to the signer's public key and have assurance that it corresponds to the signer's private key. However, a public and private key pair has no intrinsic association with any person; it is simply a pair of numbers. Some convincing strategy is necessary to reliably associate a particular person or entity to the key pair.

In a transaction involving only two parties, each party can simply communicate (by a relatively secure "out-of-band" channel such as a courier or a secure voice telephone) the public key of the key pair each party will use. Such an identification strategy is no small task, especially when the parties are geographically distant from each other, normally conduct communication over a convenient but insecure channel such as the Internet, are not natural persons but rather corporations or similar artificial entities, and act through agents whose authority must be ascertained. As electronic commerce increasingly moves from a bilateral setting to the many-on-many architecture of the World Wide Web on the Internet, where significant transactions will occur among strangers who have no prior contractual relationship and will never deal with each other again, the problem of authentication/nonrepudiation becomes not merely one of efficiency, but also of reliability. An open system of communication such as the Internet needs a system of identity authentication to handle this scenario.

To that end, a prospective signer might issue a public statement, such as: "Signatures verifiable by the following public key are mine." However, others doing business with the signer may for good reason be unwilling to accept the statement, especially where there is no prior contract establishing the legal effect of that published statement with certainty. A party relying upon such an unsupported published statement in an open system would run a great risk of trusting a phantom or an imposter, or of attempting to disprove a false denial of a digital signature ("nonrepudiation") if a transaction should turn out to prove disadvantageous for the purported signer.

The solution to these problems is the use of one or more trusted third parties to associate an identified signer with a specific public key. <31> That trusted third party is referred to as a "certification authority" in most technical standards and in these Guidelines.

To associate a key pair with a prospective signer, a certification authority issues a certificate, an electronic record which lists a public key as the "subject" of the certificate, and confirms that the prospective signer identified in the certificate holds the corresponding private key. The prospective signer is termed the "subscriber. <32> A certificate's principal function is to bind a key pair with a particular subscriber. A "recipient" of the certificate desiring to rely upon a digital signature created by the subscriber named in the certificate (whereupon the recipient becomes a "relying party") can use the public key listed in the certificate to verify that the digital signature was created with the corresponding corresponding private key. <33> If such verification is successful, this chain of reasoning provides assurance that the corresponding private key is held by the subscriber named in the certificate, and that the digital signature was created by that particular subscriber.

To assure both message and identity authenticity of the certificate, the certification authority digitally signs it. The issuing certification authority's digital signature on the certificate can be verified by using the public key of the certification authority listed in another certificate by another certificate authority (which may but need not be on a higher level in a hierarchy) <34>, and that other certificate can in turn be authenticated by the public key listed in yet another certificate, and so on, until the person relying on the digital signature is adequately assured of its genuineness. In each case, the issuing certification authority must digitally sign its own certificate during the operational period of the other certificate used to verify the certification authority's digital signature.

A digital signature, whether created by a subscriber to authenticate a message or by a certification authority to authenticate its certificate (in effect a specialized message) should be reliably time-stamped to allow the verifier to determine reliably whether the digital signature was created during the "operational period" stated in the certificate, which is a condition upon verifiability of a digital signature under these Guidelines. <35>

To make a public key and its identification with a specific subscriber readily available for use in verification, the certificate may be published in a repository or made available by other means. Repositories are on-line databases of certificates and other information available for retrieval and use in verifying digital signatures. Retrieval can be accomplished automatically by having the verification program directly inquire of the repository to obtain certificates as needed.

Once issued, a certificate may prove to be unreliable, such as in situations where the subscriber misrepresents his identity to the certification authority. In other situations, a certificate may be reliable enough when issued but come to be unreliable sometime thereafter. If the subscriber loses control of the private key ("compromise" of the private key), the certificate has become unreliable, and the certification authority (either with or without the subscriber's request depending on the circumstances) may suspend (temporarily invalidate) or revoke (permanently invalidate) the certificate. Immediately upon suspending or revoking a certificate, the certification authority must publish notice of the revocation or suspension or notify persons who inquire or who are known to have received a digital signature verifiable by reference to the unreliable certificate.

Challenges and Opportunities

The prospect of fully implementing digital signatures in general commerce presents both benefits and costs. The costs consist mainly of:

- **Institutional overhead:** The cost of establishing and utilizing certification authorities, repositories, and other important services, as well as assuring quality in the performance of their functions.
- **Subscriber and Relying Party Costs:** A digital signer will require software, and will probably have to pay a certification authority some price to issue a certificate. Hardware to secure the subscriber's private key may also be advisable. Persons relying on digital signatures will incur expenses for verification software and perhaps for access to certificates and certificate revocation lists (CRL) in a repository.

On the plus side, the principal advantage to be gained is more reliable authentication of messages. Digital signatures, if properly implemented and utilized offer promising solutions to the problems of:

- **Imposters**, by minimizing the risk of dealing with imposters or persons who attempt to escape responsibility by claiming to have been impersonated;
- **Message integrity**, by minimizing the risk of undetected message tampering and forgery, and of false claims that a message was altered after it was sent;
- **Formal legal requirements**, by strengthening the view that legal requirements of form, such as writing, signature, and an original document, are satisfied, since digital signatures are functionally on a par with, or superior to paper forms; and
- **Open systems**, by retaining a high degree of information security, even for information sent over open, insecure, but inexpensive and widely used

channels.