# BIOMETRICS AND DIGITAL SIGNATURES IN ELECTRONIC COMMERCE

**R. R. Jueneman**
**R. J. Robertson, Jr.**[*]

**CITATION:** R R. Jueneman and R. J. Robertson, Jr., Biometrics and Digital Signatures in Electronic Commerce, 38 Jurimetrics J. __-__.

## I. INTRODUCTION

In our society, traditional and accepted means for a person to identify and authenticate himself, either to another human being or to a computer system, are based on one or more of three general principles: what the person knows (some form of shared secret), what he possesses (some kind of unique token or key), or what he is (some aspect of his physical being). The written signature is regarded as the primary means of identifying the signer of a written document, based on the implicit assumption that a person's normal signature changes slowly and is very difficult to erase, alter, or forge without detection.

Today it is claimed that the technology known as digital signatures can authenticate both the originator and the content of an electronic document to levels approaching mathematical certainty, through the use of cryptography based on mathematical principles. In actuality, this confidence is subject to some unproven assumptions, in particular the difficulty of solving certain mathematical problems that have been studied for centuries. Moreover, it assumes that the user

---

* Robert Jueneman is Consulting Engineer and Security Architect, Novell, Inc.. An Associate Member of the ABA, he contributed extensively to the development of the ABA Digital Signature Guidelines and is a frequent contributor to various technical and legal discussion lists in the area of Public Key Infrastructure. R. J. Robertson, Jr., is Associate Professor of Law, Southern Illinois University School of Law. A.B., 1973; J.D., *cum laude*, 1976, University of Missouri–Columbia.

can and will maintain exclusive control over the cryptographic private key, perhaps through the use of a password to encrypt the key (what he knows) or a tamper-proof hardware token which is used to securely store the key and implement the digital signature algorithm (what he possesses). However, even assuming that the user's computer system could be made invulnerable to security weaknesses, a user's password could still be accidentally or deliberately disclosed to someone else, and even a secure hardware token could be lost, stolen, or given away.

Biometric identification and authentication techniques provide a more direct means of identifying an individual, normally by means of some form of physical measurements or indicia that are uniquely associated with that individual. Existing and proposed biometric techniques include the use of fingerprints, retinal scans, iris patterns, facial characteristics, voice prints, signature dynamics, and numerous others. Recently, it has been suggested that electronic documents authenticated using biometric techniques should be viewed as the legal equivalent of documents which are authenticated using cryptographic digital signatures. However, equating or conflating these two techniques risks serious confusion as to the respective merits of the two different technologies.

This paper argues that biometric techniques provide an excellent way of controlling or authenticating physical access to a local trusted agent, *e.g.*, an access control device which allows a user to enter a building or which confirms his identity to a clerk at a point of sale terminal, or which acts as a gatekeeper guarding access to the user's private key; but that biometric techniques alone cannot securely authenticate either the originator of an electronic document or the contents of that document at a point that is removed in space or time from its creation, when the document has been stored in or transmitted over an insecure medium.

However, if a secure implementation of a credible biometric technique were used in combination with cryptographic techniques to protect an electronic document from being tampered with while in transit or storage in an untrusted medium, it should be possible to establish beyond reasonable doubt that only the individual in question could have originated the document. This note discusses various forms of biometric identification and authentication and how they can be combined with digital signatures to overcome many of the technical and legal difficulties arising from the use of digital signatures alone.

## II. WRITTEN DOCUMENTS AND THE LAW OF SIGNATURES

In order to understand how authentication techniques will operate with respect to electronic documents, it will be helpful to review how American courts

and legislatures have treated an ink-on-paper signature as a form of authenticating a written document and the legal effect of such authentication.

## A. When Is a Signature on a Written Document Required by Law?

There is no general requirement that a party "sign" a "written" document in order for his promise to be enforceable—only certain types of promises must be evidenced by a "signed writing." The statutory requirement of a signed writing for promissory enforceability is known by the generic term "Statute of Frauds." The "traditional" Statute of Frauds is modeled after the English statute of 1677 and ordinarily requires a memorandum signed by the party against whom an action is brought to enforce specified categories of promises.[1] Presently, the most commercially significant writing requirement for contracts is section 2-201 of the Uniform Commercial Code ("the Code"), which requires a writing signed by the person against whom enforcement is sought in order to enforce a contract for the sale of goods for a price of $500 or more.[2] In addition to the requirement that some types of promises be evidenced by a signed writing, virtually all states also require a signed writing in order to validate other transactions, especially testamentary instruments such as wills.

## B. What Constitutes a "Signing" or a "Signature" on a Written Document?

The traditional Statute of Frauds does not define the terms "signed," or "signature," and courts have broadly defined these terms. Likewise, under the Code, the term "signed" is broadly defined as "any symbol executed or adopted by a party with present intention to authenticate a writing."[3] Hence, a "signing" or a "signature" can be generally defined as any mark or symbol affixed to a writing to manifest the signer's intent to adopt it as his own and to be bound by it.[4] The signature need not be the full name of the signer—it may be any mark[5] that is typewritten,[6] printed,[7] or made with a device such as a rubber stamp.[8]

---

1. *See, e.g.*,740 ILL. COMP. STAT. 80/1 (West 1997).

2. U.C.C. § 2-201 (1995).

3. U.C.C. § 1-201(39) (1995). "A 'writing' includes printing, typewriting or any other intentional reduction to tangible form." *Id.* at § 1-201(46).

4. *E.g.*, Just Pants v. Joeine Wagner, 617 N.E.2d 246, 251 (Ill. App. Ct. 1993); *cf.* U.C.C. §1-201(39) (1995) ("'Signed' includes any symbol executed or adopted by a party with present intention to authenticate a writing.").

5. *E.g.*, Zacharie v. Franklin, 37 U.S. 151, 162 (1838) (mark of "X" sufficient signature); Zenith Radio Corp. v. Matsushita Elec. Indus. Co., Ltd., 505 F. Supp. 1190 (D.C. Pa. 1980) (stamp from Japanese "chop" or signature seal may constitute a signature). *See also* U.C.C. §1-201(39), cmt. 39 (1995) ("signing" can be by initials or thumbprint).

6. Peoples Bank v. Northwest Georgia Bank, 228 S.E.2d 181 (Ga. App. 1976) (Article 9

## C. The Functions of a Signature on a Written Document: Evidence and Ceremony

The requirement of a signature serves a number of functions. The primary function can be described as "evidentiary"—a person's signature on a written document provides presumably reliable evidence of that person's assent to the terms contained in the writing.[9] This evidentiary function relates to two matters. First, a distinctive form of signature is considered to be reliable evidence of the *identity* of the person signing the writing. In other words, it serves the purpose of "signer authentication,"[10] even if the signature is nothing more than an illegible scrawl. Second, because of the assumed semipermanent nature of ink on paper, a person's signature is also considered to be very reliable evidence that the signer assented to the *terms* contained in the written document. In other words, it serves the purpose of "document authentication."[11] This function is premised on the difficulty of altering a written document without leaving detectable (by physical or chemical means) marks of alteration and is the basis for the ordinary rule that the signer cannot avoid the legal consequences of the terms contained in the written document by claiming that he did not read or understand it.[12]

Additionally, the act of signing a written document serves what has variously been described as a "ceremonial," "psychological" or "cautionary"[13] function. This function recognizes that "signing on the dotted line" is an act that calls to the signer's attention the fact that he is entering into a transaction that has legal consequences and may deter the signer from entering into hasty or ill-considered transactions.[14]

---

financing statement); First Security Bank v. Fastwich, Inc., 612 S.W.2d 799 (Mo. Ct. App. 1981) (Article 3 promissory note).

7. *E.g.*, Welch v. Mitchell, 351 So. 2d 911, 915 (Ala. Civ. App. 1977); Merrill, Lynch, Pierce, Fenner & Smith, Inc. v. Cole, 457 A.2d 656, 662-63 (Conn. 1983); Kohlmeyer & Co. v. Bowen, 192 S.E.2d 400, 404 (Ga. Ct. App. 1972).

8. *E.g.*, Parshalle v. Roy, 567 A.2d 1927 (Del. Ch. 1989).

9. Joseph M. Perillo, *The Statute of Frauds in Light of the Functions and Dysfunctions of Form*, 43 FORDHAM L. REV. 39, 64-69 (1974).

10. INFORMATION SECURITY COMMITTEE, SECTION OF SCIENCE AND TECHNOLOGY, AMERICAN BAR ASSOCIATION, DIGITAL SIGNATURE GUIDELINES: LEGAL INFRASTRUCTURE FOR CERTIFICATION AUTHORITIES AND SECURE ELECTRONIC COMMERCE 6 (1996) [hereafter DIGITAL SIGNATURE GUIDELINES].

11. *Id.*

12. Merit Music Serv. v. Sonneborn, 225 A.2d 470, 474 (Md. 1967) ("the law presumes that a person knows the contents of a document that he executes and understands at least the literal meaning of its terms.").

13. DIGITAL SIGNATURE GUIDELINES, note 10, at 4; Perillo, note 9, at 45, 53.

14. *See generally* RESTATEMENT (SECOND) OF CONTRACTS § 72 comment c, ch. 5 statutory note at 286 (1982).

## D. Proving the Fact of a Signature on a Written Document

Before a written document can be introduced into evidence at trial, it must be "authenticated," a term of art in the law of evidence. A written document is "authenticated" by introducing "evidence sufficient to support a finding that the matter in question is what its proponent claims."[15] Authentication can be accomplished by a variety of methods. For example, it may be accomplished by eyewitnesses who saw a defendant sign the document,[16] by comparison of the challenged signature to other examples of handwriting known to have been executed by the defendant,[17] by expert witness testimony about the challenged signature based on similar comparisons,[18] or by circumstantial evidence from the contents of the document which would indicate that the author knows facts that could only be known by the defendant.[19] The basis for the requirement of authentication has been nicely summarized by a leading treatise:

> In the everyday affairs of business and social life, it is the custom to look merely at the writing itself for evidence as to its source. Thus, if the writing bears a signature purporting to be that of X, or recites that it was made by X, we assume, nothing to the contrary appearing, that it is exactly what it purports to be, the work of X. At this point, however, the law of evidence has long differed from the commonsense assumption upon which each of us conducts her own affairs, adopting instead the position that the purported signature or recital of authorship on the face of a writing will *not* be accepted as sufficient preliminary proof of authenticity to secure the admission of the writing in evidence.[20]

Once a written document has been "authenticated" the trier of fact may consider the document and its contents. However, "authentication" does not mean the document is, in fact, "authentic." Rather, the finder of fact must determine whether the document is, in fact, what it purports to be and who, in fact, signed it. Ordinarily, a plaintiff alleging that a defendant signed a document has the burden of persuasion on that fact. The evidence that might lead the trier of fact to conclude that the defendant signed the document is the same kind of evidence that could be used initially to authenticate the document.

---

15. FED. R. EVID. 901(a).

16. MCCORMICK ON EVIDENCE § 219(a) at 38 (4th ed. John William Strong ed. 1992).

17. The comparison can be done by a lay witness if it is based on familiarity not acquired for purposes of litigation. FED. R. EVID. 901(b)(2); United States v. Gallagher, 576 F.2d 1028, 1048 (3d Cir. 1978) (bank's vice president who had supervised defendant for ten years qualified to give opinion as to genuineness of defendant's signature).

18. FED. R. EVID. 901(b)(3).

19. MCCORMICK ON EVIDENCE, *supra* note16, § 225 at 49.

20. *Id.* § 218 at 37.

## E. Special Evidentiary Rules About Signatures on Written Documents: Notaries and Negotiable Instruments

Because each written document and the signature affixed to it represent a unique set of physical attributes, the law of evidence does not usually differentiate between types of documents or signatures or assign different evidentiary rules to those types. In two special cases, however, the ordinary evidentiary rules about proving a signature are altered.

First, some signatures are treated differently with respect to the requirement of authentication. Under the Federal Rules of Evidence, for example, a notarized signature is treated as "self-authenticating."[21] This means only that the written document bearing a notarized signature is admissible without the requirement of independent evidence of "authentication." This rule does not create a presumption that the document or the signature is, in fact, authentic.[22]

Second, a few statutes provide that certain signatures are presumed to be genuine. For example, under the Code, a holder who produces a negotiable instrument is entitled to a presumption that all signatures on the instrument are authentic and authorized.[23] It is important to note that the effect of a presumption of authenticity is substantially different from the evidentiary "authentication" of a written document. If a document is "authenticated," it means only that the proponent of the document has introduced evidence sufficient to support a finding that the document is what the proponent claims it to be. Even if no contrary evidence is introduced, the trier of fact may disbelieve the proponent's evidence and find that the document is not what the proponent claims it to be. However, if a signature is "presumed" to be authentic, the presumption *requires* the trier of fact to find that the signature is genuine unless and until the person against whom the presumption operates introduces evidence sufficient to "rebut" the presumption.

A presumption is "rebutted" by introducing evidence contradicting the presumed fact. There is some disagreement about how much evidence is necessary to rebut a presumption. The majority rule is the so-called "bursting bubble" doctrine, which provides that the mandatory effect of the presumption disappears when any evidence to the contrary is introduced.[24] Under such a rule, the presumption that a defendant's signature on a written document is genuine would be rebutted if the defendant testifies he did not sign it. A minority rule, the so-

---

21. FED. R. EVID. 902(8).

22. 5 WEINSTEIN'S FEDERAL EVIDENCE § 902 .02[3] at 902-10 (2d ed. Joseph M. McLaughlin ed. 1997).

23. U.C.C. § 3-308(a) (1995).

24. MCCORMICK ON EVIDENCE, *supra* note 16, § 344(A) at 462-63. This is also the rule under the Code. *See* U.C.C. § 1-201(31) (1995).

called "burden-shifting" doctrine, provides that the presumption continues until the defendant persuades the trier of fact that it is more probable than not that the presumed fact does not exist.[25] In effect, this approach shifts the burden of persuasion on the disputed issue to the defendant. Under such a rule, the presumption that a defendant's signature on a written document is genuine would not be rebutted until the defendant proves it is more likely than not that he did not sign the document.

## III. ELECTRONIC DOCUMENTS AND THE LAW OF SIGNATURES

### A. Reforming the Law of Signatures to Accommodate Electronic Commerce Transactions

The prospect of electronically concluding contracts and other legally-significant transactions raises a number of technical and legal questions about how to establish the genuineness of electronic documents. Because electronic documents[26] consist solely of streams of binary digits or "bits"—a seemingly endless series of ones and zeroes—they lack the distinctive, semipermanent physical attributes of a written document. Hence, proof of the originator and content of an electronic document poses problems not heretofore encountered.

In the last two years, there have been many legislative initiatives to reform existing legal requirements regarding the enforceability and proof of written documents to give legal recognition of electronic documents. This section of the

---

25. MCCORMICK ON EVIDENCE, *supra* note 16, § 344(B) at 470-74.

26. The terms "electronic document," "electronic message," or "electronic record" are often used interchangeably. In general, "electronic" should not be taken to mean exclusively "electrical," but may also include other forms of document preparation, transmission, and storage, including fiber optic transmission lines. As used in this paper, "electronic document" refers to a *digital* representation of information , where the human-readable characters and images have been reduced to a unique set of binary digits or bits—ones and zeros—which represent those characters. The term "electronic document" does not refer to a stored or transmitted *image* of a document, such as a photographic or microfilm copy of a document, a traditional analog tape recording or videotape, or even an analog electrical transmission of a scanned image such as takes place over a facsimile machine. The difference between a digital, electronic document and an analog image of the same document is that the digital document has effectively captured the raw keystrokes used to create it, whereas in the case of a written document it is the *image* of the document that has been captured. In a sense, it is though the digital electronic document contained the hexadecimal character value and description of every character it contains, *e.g.*, "Wingdings Bold 'A' 8 point," whereas the analog record would contain the actual image of the character, *e.g.*, the "smiley face" **J** character. Digital electronic documents are normally stored and transmitted in computer-readable form only. The term "written document" will be used when it is necessary to differentiate between a traditional written document, whether recorded on paper or carved in stone, versus an electronic document in digital form, even if such a document were recorded on some semi-permanent medium such as a writeable CD-ROM.

paper addresses two basic issues raised by this reform legislation: (1) is an electronic document bearing some electronic mark the legal equivalent of a "signed writing" that would satisfy the Statute of Frauds?; and (2) if so, does the recipient's use of a security procedure alter the normal rules of evidence about proving the identity of the sender of the electronic document?

### B. Equating Electronic Documents with "Signed Writings" under the Statute of Frauds

A fundamental question is whether an electronic document qualifies as a "writing" that is "signed" as required by the Statute of Frauds. To ensure that an electronic document is not denied legal effect merely because it is in that form, most reform legislation specifically equates electronic documents with "signed writings" under the Statute of Frauds by equating virtually any electronic document with a "writing" and some equate virtually any electronic mark or symbol with a "signature."[27]

### C. Special Rules for Proving the Originator and Content of Electronic Documents During Employment of a Security Procedure

Even if an electronic document can constitute a "writing" and even if it contains an "electronic mark," the Statute of Frauds requires the plaintiff to prove that the "electronic mark" was executed or adopted by the defendant. Even if a promise is not covered by the Statute of Frauds and does not require a "signed writing" to be enforceable, if a plaintiff relies on an electronic document as the source of the promise, he must still show that the promise was made by the defendant and that the electronic document introduced into evidence has not been altered after it left the defendant's control.

Security procedures such as those discussed in subsequent sections of this paper greatly reduce the chances that a person can successfully impersonate another as the source of an electronic document or alter the content of an electronic document. Thus, some, but not all, reform legislation creates special evidentiary rules about proving the originator and content of an electronic document when a security procedure is used.

---

27. *E.g.*, U.C.C. §§ 2B-2-102(a)(37), 2B-102(a)(3), 2B-113 (March 1998 Draft); UNIFORM ELECTRONIC TRANSACTIONS ACT §§ 201, 301 (March 23, 1998 Draft); FLA. STAT. ch. 282.73 (Supp. 1998); ILLINOIS ELECTRONIC COMMERCE SECURITY ACT §§ 202(a), 203(a) (Jan. 16, 1998 Final Version) <http://www.mbc.com>; R.I. GEN. LAWS. §§ 42-127-3(a), 4(a) (Supp. 1998). However, the Utah statute provides that only electronic documents bearing digital signatures satisfy the legal requirements of a "signed writing," UTAH CODE ANN. §§ 46-3-401, 46-3-403 (Supp. 1997), although it also provides: "[n]othing in this chapter precludes any message, document, or record from being considered written or in writing under other applicable state law." *Id.* § 46-3-403(2).

Some examples of reform legislation take a "hands off" approach and contain no provisions about whether the use of a security procedure should affect the ordinary evidentiary rules about these matters. This "minimalist" legislation either says nothing about how to prove the originator and content of an electronic document[28] or says that the trier of fact may consider any relevant evidence and circumstances in determining whether an electronic mark was executed or adopted by a particular person.[29]

Other examples of reform legislation do provide special evidentiary rules resulting from the use of a security procedure. In general, this type of reform legislation provides that, if the recipient of an electronic document has used a security procedure and that procedure indicates that the electronic document was originated by an individual and has not been altered, there is a rebuttable presumption that the electronic document was sent by that individual and has not been altered.[30] Examples of this type of reform legislation contain different provisions about two issues: (1) what forms of security procedures give rise to the presumptions?; and (2) how can the presumptions be rebutted?

The pioneering Utah Digital Signature Act ("Utah Act") addresses the first issue by providing these presumptions only if the electronic document bears a digital signature.[31] Most other reform legislation, however, does not specify any particular technology, but affords the presumptions to any commercially reasonable security procedure that the parties have agreed to use.[32]

The proposed Illinois Electronic Commerce Security Act (the "Illinois Act") also provides for a security procedure that is certified by the Illinois Secretary of State as being capable of creating an electronic signature that has a number of attributes of reliability.[33] The purpose of this provision is to define, in a techno

---

28. *E.g.*, FLA. STAT. ch. 282.70 *et seq.* (Supp. 1998); GA. CODE ANN. § 106-3401 *et seq.* (Supp. 1997).

29. *E.g.*, R.I. GEN. LAWS. § 42-127-4(c) (Supp. 1997).

30. U.C.C. §§ 2B-116(b), 2B-117(a) (March 1998 Draft); UNIFORM ELECTRONIC TRANSACTIONS ACT §§ 202(a)(2), 203(a), 302(b)(1) (March 23, 1998 Draft); ILLINOIS ELECTRONIC COMMERCE SECURITY ACT §§ 301, 302, 304 (Jan. 16, 1998 Final Version) <http://www.mbc.com>; UTAH CODE ANN. § 46-3-406(3)(a), (b) (Supp. 1997)

31. In Utah, the presumption arises if the digital signature has been verified by the public key listed in a valid certificate issued by a licensed certification authority. UTAH CODE ANN. § 46-3-406(3)(a) (Supp. 1997). Similar provisions apply in other states that have used the Utah legislation as a model. *E.g.*, WASH. REV. CODE § 19.34.350(3) (1998).

32. *E.g.*, U.C.C. § 2B-102(2) (March 1998 Draft); UNIFORM ELECTRONIC TRANSACTIONS ACT §§ 102(18), 202(a)(2), 203(a), 302(b)(1) (March 23, 1998 Draft); ILLINOIS ELECTRONIC COMMERCE SECURITY ACT §§ 301(b)(1), 302(b)(1) (Jan. 16, 1998 Final Version) <http://www.mbc.com>.

33. The resulting electronic signature must be one that:

> (i) is unique to the signer within the contest in which it is used;
> (ii) can be used to objectively identify the person signing the electronic record;
> (iii) was reliably created by such identified person,(e.g., because some aspect of the procedure involves the use of a signature device or other means or method that is under the sole control of such person), and cannot be readily duplicated or compromised, and

logically neutral way, the characteristics that a security procedure must have to be entitled to the presumptions in cases where the parties have not previously agreed to use a security procedure. In order to be so certified, the technology underlying the security procedure must be completely open and disclosed to the public in order to allow comprehensive review in the scientific or information security community, and must also be generally accepted in the applicable scientific or information security community as meeting the standards of reliability.[34] If all of these factors are present, the Illinois Act refers to the resulting signature as a "secure electronic signature." A secure electronic signature is entitled to the presumption that the originator is the person indicated by the security procedure if the proponent of the electronic signature establishes that the security procedure was also: (1) commercially reasonable under the circumstances; (2) applied by the relying party in a trustworthy manner; and (3) reasonably and in good faith relied upon by the proponent.[35]

The requirements to rebut the presumptions are not specified by some states.[36] Other statutes provide that a presumption is rebutted by the introduction of evidence sufficient to support a finding of its non-existence, *i.e.*, the "bursting-bubble" theory.[37] The Illinois Act, on the other hand, specifically adopts the "burden-shifting" theory.[38]

## IV. DIGITAL SIGNATURES AS A SECURITY PROCEDURE

---

> (iv) is created, and is linked to the electronic record to which it relates, in a manner such that if the record or the signature is intentionally or unintentionally changed after signing the electronic signature is invalidated.

ILLINOIS ELECTRONIC COMMERCE SECURITY ACT § 302(b)(2) (Jan. 16, 1998 Final Version) <http://www.mbc.com>. Similar language appears in other states. *E.g.*, CAL. GOVT. CODE §16.5(a) (West Supp. 1998); GA. CODE ANN. § 106-3403 (Supp. 1997).

34. ILLINOIS ELECTRONIC COMMERCE SECURITY ACT § 307(a)(1)-(2) (Jan. 16, 1998 Final Version) <http://www.mbc.com>.

35. *Id*. at § 302(a)(1)-(3).

36. For example, the Utah statute merely provides that "a court of this state shall presume" the facts. UTAH CODE ANN. § 46-3-406 (Supp. 1997). However, the commentary to the statute says that "[t]he effect of the presumptions provided in this section is merely to allocate the burden of going forward with allegations and evidence to the party challenging the digital signature . . . ." DIVISION OF CORPORATIONS AND COMMERCIAL CODE, UTAH DEPARTMENT OF COMMERCE, UTAH DIGITAL SIGNATURE LAW: TECHNICALLY AND LEGALLY SECURE ELECTRONIC COMMERCE 69 (Nov. 1995). This language is consistent with the "bubble-bursting" theory of presumptions. See *supra* text accompanying notes 24-25.

37. Proposed Article 2B of the Code does not contain a specific provision on this issue, but under section 1-201(31) of the Code, "presumption" is defined as having this effect. U.C.C. § 1-201(31)(1995). *See also* UNIFORM ELECTRONIC TRANSACTIONS ACT §302, Reporter's Note 4 (March 23, 1998 Draft).

38. ILLINOIS ELECTRONIC COMMERCE SECURITY ACT § 304(c) (Jan. 16, 1998 Final Version) <http://www.mbc.com>.

Because much reform legislation gives special evidentiary weight to electronic documents verified by a "security procedure," persons engaged in electronic commerce should understand that not all security procedures provide identical levels of reliability about the originator and content of an electronic document. The security procedure that has received the most attention is known as a "digital signature."

The legal analysis that forms the basis for much of the current and proposed use of digital signatures in electronic commerce is largely the result of work of the Information Security Committee of the American Bar Association's Section of Science and Technology (the "ISC") that drafted the *Digital Signature Guidelines* ("the *Guidelines*") over the period 1992 to 1995, with final publication in August 1996.[39]

## A. Overview of the Technology of Digital Signatures

A "digital signature" is a specific term of art within the technical community that has been used consistently since the landmark publication describing public key cryptography by Whitfield Diffie and Martin Hellman in 1976[40] and its implementation in its most popular form, the RSA algorithm.[41] Although the term "digital signature" is, unfortunately, sometimes misused by lawyers, regulators, and others outside the technical community to mean more generic methods of identifying the originator of an electronic message,[42] the authors will use the term "digital signature" as having its original, specific meaning.

A digital signature is not a "digitized" version of a handwritten signature; instead, it is typically a transformation or reduction of the text of an electronic document which is logically appended to the document itself.[43] The digital signature algorithm is based on the use of "public key cryptography" and involves the use of two codes (known as "keys") that are used by the signer to authenticate the source and content of his electronic documents, and by the recipient to validate

---

39. DIGITAL SIGNATURE GUIDELINES, *supra* note 10.

40. W. Diffie & M.E. Hellman, *New Directions in Cryptography*, IEEE TRANSACTIONS ON INFORMATION THEORY, vol. IT-22, No. 6, Nov. 1976, at 644-54.

41. R. L. Rivest et al., *A Method of Obtaining Digital Signatures and Public-key Cryptosystems*, COMMUNICATIONS OF THE ACM, vol. 21, No. 2, Feb. 1978, at 120-26.

42. *E.g.*, 1998 Colo. H.B. 1043 § 11(1) (Introduced Jan.16, 1998) (a "'digital signature' means an electronic identifier, created by computer, intended by the party using it to have the same force and effect as the use of a manual signature"); CAL. GOV'T CODE § 16.5 (West Supp. 1998) (same).

43. More precisely, the electronic document is first condensed into a shorter form of digital representation, known as a "hash result," or more precisely a "message digest" that is then transformed by the originator of the electronic message using the digital signature algorithm, and it is the transformed or signed version of the message digest that is appended to the message and known as the "digital signature." WARWICK FORD & MICHAEL S. BAUM, SECURE ELECTRONIC COMMERCE § 4.1 at 113-114, 115-16 (1997).

their correctness. One of a pair of keys (which are generated at the same time), the "private" key, is kept solely in the possession of the signer of an electronic document and is used to encode the text of the document into the digital signature.[44] Another key, the "public" key, is made publicly available via some trustworthy publication procedure to any person ("relying party") who may deal with the originator of the document. The public and private keys are mathematically related, but the relationship is so complicated that is "computationally infeasible" to deduce the private key solely from knowledge of the public key[45] or to create a signed message which can be verified by application of the public key without the knowledge of the private key. Hence, the relying party, having trustworthy access to the public key, can validate the documents as having been signed by someone who had knowledge of the corresponding private key, but cannot deduce the private key from the public key, nor create such a signature without the private key.[46] Because the keys only allow the digital signature created by one of the keys to be decrypted or validated by the other key, a person receiving a digitally signed document that is verified by use of the public key knows that the document was signed by a person possessing the private key.[47] The digital signature thus provides very reliable algorithmic evidence of the source of an electronic document, assuming that the relying party has a reliable way of verifying the identity of the person with whom the private key is associated and assuming that the secrecy and control over the private key has been maintained.

The method of verifying the person with whom the private key is associated is through the use of a trusted third party known as a certification authority ("CA"). The CA's role is to verify the identity of a person who possesses a key pair and then publish a "certificate"—an electronic record which lists the public key as the subject of the certificate and which confirms that the prospective signer identified in the certificate holds the corresponding private key. This certificate is then made publicly available in a "repository"[48] maintained by the CA or someone else. A relying party will access the certificate and determine that

---

44. DIGITAL SIGNATURE GUIDELINES, *supra* note 10, at 10.

45. *Id.* at 9 n.23.

46. *Id.* at 10-11.

47. The recipient verifies the digital signature by taking the text of the electronic document and converting it into a "hash result" using the same "hash function" as the originator and then applying the public key to the hash result. The process will result in verification if and only if the hash result of the original electronic document was encrypted by use of the private key to which the public key is related, to an extremely high level of confidence. *Id.* at 11-12.

48. A "repository," also referred to as a "directory," is an on-line database of certificates and other information available for retrieval and use in verifying digital signatures. In order to minimize the amount of trust that must be placed in the repository itself, the certificates issued by a CA are themselves validated by a chain of certificates issued to that CA and signed by a higher level CA, and so forth, until a self-signed certificate created by top level or root CA is encountered. The top level certificate must be installed in the relying party's software by a trusted, out-of-band process, reflecting the trust that is accorded that CA. *Id*. at 16.

a public key is associated with a private key possessed by a particular person, obtain a copy of that public key, and then use that public key to decrypt the digitally signed document he had received. If the public key decrypts the digital signature, that is extraordinarily reliable evidence that the document was in fact sent by a person in possession of the private key that the CA has verified as being associated with that public key.

A digital signature is also very good evidence that the document has not been tampered with since it was sent. Because the digital signature is a reduced encoding of the document itself, altering a document that is associated with a digital signature in any way will cause the public key to be unable to verify the digital signature, thus providing irrefutable evidence to the recipient that the document has been altered since it was digitally signed. Since the typical "message digest" or "hash results" that are compared are 160 bits in length, it would require an attacker to generate and/or search though approximately $2^{80}$ pairs of messages in order to have an approximately even chance of finding even a single pair of messages that would produce the same message digest but yet not be precisely identical, down to the bit level.[49] That is $1.2 \times 10^{24}$, or approximately a trillion trillion messages that would have to be examined—a patent impossibility.

Hence, assuming that the CA has done an appropriate job of verifying the identity of the entity associated with a public key (the "subject")[50] and assuming

---

49. The probability of randomly selecting two objects from a pool of N objects such that the two objects are identical is called the 'Birthday Problem' in statistics texts, after the cocktail party game where everyone in the room calls out their birth date to see if anyone else was born on the same day. As the number of objects in the pool increases, the probability of finding any two objects that match approaches 50% after approximately the square root of the number of objects in the pool have been sampled. Hence, if there are $2^{160}$ possible values for a message digest algorithm, as is the case for the Secure Hash Algorithm (SHA-1), then approximately $2^{80}$ random messages would have to be generated and their message digests calculated and compared in order to have a 50% probability of finding a duplicate message digest. Although calculating that number of message digests might (barely) be feasible, the problem of sorting or otherwise comparing that many objects is quite intractable, because of the huge amount of memory that would be required and the speed of available sorting algorithms. For an historical perspective and background on message authentication, see R. R. Jueneman, *Electronic Document Authentication*, IEEE NETWORK MAGAZINE, v.1, n.2 17-23 (Apr 1978); G. Yuval, *How to Swindle Rabin*, CRYPTOLOGIA, v. 3, n. 3, 187-190 (Jul 1979); R. R. Jueneman, *Analysis of Certain Aspects of Output-Feedback Mode*, ADVANCES IN CRYPTOLOGY: PROCEEDINGS OF CRYPTO 82 17-23 (1983); R. R. Jueneman, S. M. Matyas, and C.H. Meyer, *Message Authentication with Manipulation Detection Codes*, PROCEEDINGS OF THE 1983 IEEE COMPUTER SOCIETY SYMPOSIUM ON RESEARCH IN SECURITY AND PRIVACY 733-754 (1983); R. R. Jueneman, S. M. Matyas, and C. H. Myer, *Message Authentication*, IEEE COMMUNICATIONS MAGAZINE, v.23, n.9 29-40 (Sep 1985); R. R. Jueneman, *A High Speed Manipulation Detection Code*, ADVANCES IN CRYPTOLOGY—CRYPTO '86 PROCEEDINGS, 327-346 (1987). For a current assessment of message digest algorithms and public key cryptography technology, see BRUCE SCHNEIER, APPLIED CRYPTOGRAPHY, SECOND EDITION: PROTOCOLS, ALGORITHMS AND SOURCE CODE IN C 429-502 (1996).

50. The *Digital Signature Guidelines*, and some state laws and regulations, use the term "subscriber" to identity the entity that is named in a certificate. Technically, however, the entity that

that the subscriber has exercised reasonable care to prevent the loss or compromise of the private key, the use of digitally signed documents provides an extraordinarily reliable method of validating both the originator and the content of an electronic document. As a result, the digital signature would apparently serve the "evidentiary" purpose[51] immeasurably better than a manual signature on a written document.

However, as in any human endeavor, nothing is ever perfect or without risk. That is true for cryptography, and for digital signatures as well. There are technical and other risks involved in the use of digital signatures which ought to be understood, even though on balance those risks appear to be small.

## B. Technical Risks of Digital Signatures

### 1. *Risk of an Unexpected Cryptanalytic Breakthrough*

The history of cryptography is replete with examples of cryptosystems that were once widely believed to be unbreakable and are now viewed as shattered beyond repair. There is, therefore, the possibility that some sudden, completely unforeseen insight will occur to someone and one or more of the leading digital signature algorithms will be shown to be fatally flawed.

The only way to guard against such a possibility is to be vigilant in examining the initial evidence of a cryptographic system's strength and to continue to monitor both academic and industrial research in the area. In particular, any user of a cryptographic system should insist on a full and complete disclosure of all of the relevant details of both the algorithm and the implementation, in order to provide some level of assurance against the occurrence of a hidden flaw. In this regard, proprietary algorithms are regarded as being *much* more likely to contain some fundamental flaw, because they have not undergone the usual intensive peer review and scrutiny.

However, given the extensive study that has been devoted to the RSA algorithm and the Digital Signature Standard, and that is being devoted to the emerging Elliptic Curve algorithm, it is unlikely that a cryptanalytic breakthrough would reveal a fatal flaw in one of them, and even if such a flaw were found with one algorithm, it would be relatively simple to switch to another one. Moreover, even if such a flaw were found it would not *ipso facto* invalidate previous digital signatures, it would merely make them somewhat more suspect.

---

is so named is called the "subject," since a certificate could identify either a person (legal or natural), or a process or machine. The term "subscriber" is therefore appropriate terminology for the person or organization who contracts with the CA to have a certificate issued, and is directly or indirectly responsible for the care and use of the private key associated with the subject of that certificate. Of course, the subscriber and the subject will be the same entity in many cases.

51. See *supra* text accompanying notes 9-10.

### 2. *Risk of Cryptanalytic Attack by Increased Computing Power*

Just as there is some risk of a new or immature algorithm falling before an unexpected insight type of cryptanalytic attack, there is also the problem of algorithms becoming obsolescent because of improved computational capability. There is an rule of thumb in the computer industry known as "Moore's law," to the effect that the number of logic elements in an integrated circuit tends to double every 18 months, while the cost stays roughly constant. This has two effects. First, the density of circuits per square inch tends to increase at the same rate, and that tends to shorten the distance between the active elements and thereby decrease the electrical propagation time and allow the clock speed to increase. Second, the increased number of circuits permits an increased use of parallelism, where multiple events can take place simultaneously, and/or a more complex design, where more complex instructions can be accommodated within a given clock cycle. In addition to advances in raw hardware speeds, there are also advances in the speed of algorithms and new approaches to solving old problems. The net result is that approximately every ten years, the overall effective computer speed available at a given cost increases by as much as a factor of a thousand or more.

In general, the effect of these developments on cryptography is that longer and longer key lengths are needed to provide a comfortable margin of strength against cryptanalytic attacks. Less than 10 years ago, a 384 bit RSA digital signature key was considered adequate for commercial use. Five years ago, a 512 bit key was considered advisable. Today, 1024 bit RSA keys are considered the norm, and 2048 bit keys are recommended for long-term, highly important uses. Fortunately, the difficulty for the cryptanalyst who is trying to break a system increases even more rapidly with increasing key length than does the time necessary to carry out the basic cryptography itself, so increasing the key length almost always has a beneficial effect on the strength of the system.[52]

### 3. *Compromise of a Digital Signature Through Inadequate Computer or Cryptographic Security*

The *Guidelines* require the certification authority ("CA") to use a "trustworthy system," in performing its services.[53] However, there is no requirement that

---

52. Even with respect to a document digitally signed with a key of a length that is subsequently shown to be vulnerable to attack, it would be possible to have such a document archived and periodically re-signed by a trusted third party using a suitably stronger key.

53. DIGITAL SIGNATURE GUIDELINES, *supra* note 10, at § 3.1. A "trustworthy system" is defined as "[c]omputer hardware, software, and procedures that: (1) are reasonably secure from intrusion and misuse; (2) provide a reasonably reliable level of availability, reliability and correct operation; (3) are reasonably suited to performing their intended functions; and (4) adhere to generally accepted security principles." *Id.* at § 1.35.

the subscriber use a "trustworthy system" to protect or use his private key when the digital signature is being produced.[54] In retrospect, this was a regrettable oversight. Without such a requirement, the security afforded by digital signatures can be undermined if the subscriber does not maintain computer and cryptographic security.[55] In addition to the possibility of a private key being inadvertently disclosed to an unauthorized individual, perhaps through some form of computer virus or Trojan horse program, there is also the possibility that the conversion of original human input into the electronic document or message form, or the conversion from electronic form back into human-readable form, whether on a computer screen or on paper, may not be performed in an accurate and reliable manner.[56]

These technical risks are primarily concerned with the generation and proper storage of the private key. Although much of the discussion within the technical community has focused on this issue, it is perhaps important to realize what is *not* at issue, and that is the security of the digital signature once it has been created. Unlike any other system of writing developed throughout history, the use of a digital signature with a sufficiently long key length can virtually guarantee that *no* alteration or modification of the text of the document would go undetected by

---

54. The *Digital Signature Guidelines* do require a subscriber who generates the key pair to use a "trustworthy system" for that purpose. *Id*. at § 4.1.

55. The National Computer Security Center, an organization within the National Security Agency, rates commercially produced software submitted to it in accordance with the TRUSTED COMPUTER SYSTEM EVALUATION CRITERIA (TCSEC), *infra* notes 67 & 68. Popular commercial operating systems including Novell's NetWare, Microsoft's NT server, and Sun's Solaris have received a C2 rating, which is considered adequate for most commercial applications. In addition, the National Institute of Standards and Technology within the Department of Commerce has developed the FIPS 140-1 standard for cryptographic implementations, *infra* note 69, and rates cryptographic implementations according to that standard. The Entrust line of products has achieved a FIPS 140-1 level 1 rating, and Netscape recently announced that its browsers and servers have received a FIPS 140-1 level 2 rating, the highest possible without using hardware cryptography. The combination of a C2 or higher computer security valuation and a FIPS 140-1 level 2 rating is considered by the authors to be a reasonable and practical definition for a minimally trustworthy system intended for commercial use. By way of contrast, Novell makes use of a system with a TCSEC rating of B3 (three rating levels higher than C2, and considered "substantially resistant to penetration") for generating keys and certificates to be used for code signing in compliance with U.S. export laws and for other important functions. And the recent German digital signature law, ordinance, and technical catalog suggest that a CA use a computer system with an ITSEC rating of E4, approximately comparable to a TCSEC B3. *Cf. infra*, notes 71-72. Unfortunately, the most popular consumer-level operating systems, *e.g.*, Microsoft's Windows 3.1 and Windows 95, have not been evaluated and would generally not be considered sufficiently trustworthy to be used for electronic commerce, except perhaps by consumers operating under applicable consumer protection legislation.

56. For example, even if an electronic document is accurately prepared using a word processor, and is not modified, it is at least possible that the printing fonts that are used to print the document might be modified within the computer in such a way as to change a dollar sign to a yen sign, or otherwise tamper with the meaning of the output.

a relying party, no matter how carelessly the document might be transmitted, handled, or stored, unless the person attempting to alter the document had knowledge of the private key. Maybe that degree of assurance is not quite as strong as our belief that the sun will come up tomorrow morning, but it is nearly so.

The issue of the computer and cryptographic security afforded the private key is a legitimate one, but one that is potentially exaggerated. Computer security experts perform a detailed risk analysis by asking and answering three questions: (1) what is the technical *vulnerability* of the system to attack?; (2) what is the *threat, i.e.*, the probability that one or more adversaries will actually carry out a successful attack against a particular system or key?; and (3) what is the *risk, i.e.*, what would be the possible consequences of a successful attack?

With respect to most commercially available and affordable computer systems, there is no question that they are potentially vulnerable, especially if the attack were carried out by an insider who had physical access to the machine. However, the vulnerabilities can be ameliorated through such protective measures as virus detection programs, firewalls and other isolation devices, and good security practices, such as not executing programs that are received from untrustworthy sources. In that regard, programs are being developed to apply digital signatures to commercially produced software and check the digital signatures of all programs after downloading them and before executing them. These techniques are expected to substantially diminish the vulnerability of systems to external attack.

In addition, the use of tamper-resistant smart cards or other hardware tokens will very sharply reduce the threat of a key compromise. The technology is available today, and is being used increasingly widely. At present the cost of the smart card and the ancillary card reader is too high to expect that every consumer will have one right away, but that is presumably only a matter of time, and increasing demand. In the mean time, CAs can and should include protective caveats and notices in the certificates they issue, restricting the issuance of certificates which are intended to validate high value transactions to those systems which use tamper-resisting hardware for the private keys. Such notices can be included in the certification practice statement, in the certificate policy, and/or as text in the certificate itself, so that keys that are only provided with a moderate amount of protection are flagged in the certificate as not being appropriate for high-value commercial transactions.

## 4. *Compromise of Digital Signatures on Uncontrolled Platforms*

The preceding analysis assumed that the hardware and software platform used to create the digital signature was under the administrative control of the subscriber and that the subscriber is motivated to use a trustworthy system in his or her own best interests. Although we have discussed the possibility of an out-

sider covertly infiltrating the user's system, the basic assumption is that the applications running on the system are benign. However, this assumption would *not* necessarily be valid if the user were to use someone else's computer, *e.g.*, one belonging to a dishonest merchant.

Consumers routinely swipe their debit cards through the merchant's card reader and then enter their secret PIN when charging their groceries. Given this habit, it is conceivable that a naive subscriber would put his private key on a diskette protected by a password and use the diskette to digitally sign documents by inserting it into a merchant's computer and entering his password. The danger, of course, is that the software on the merchant's computer might copy the key and the password required to unlock it and the merchant would then be able to forge the user's digital signature at will.

The use of tamper-proof smart cards would prevent this, because the private key is stored securely within the smart card and cannot be read out. The merchant's computer might make a copy of the PIN, but without the smart card it would do no good. However, there is another threat. Although the private key cannot be extracted from the smart card, the smart card essentially does what it is told to do by the computer. A protocol is executed between the computer and the smart card that essentially says, "sign this," and the smart card does it. Many smart cards do not even execute the message digest function over the text to be signed, because that would require sending the entire text to the smart card over a rather slow interface. But even if the smart card did calculate the message digest, *the smart card does not know what text the computer is displaying that the user is supposedly signing.* As a result, it would be possible for someone to use a smart card at a merchant's computer intending to pay for a $20 book, but for the text that is actually sent to the smart card to be signed to be for a $20,000 oriental rug.

Proposed uses of smart cards and digital signatures therefore have to be examined carefully to prevent such an attack. One possible solution might be for the smart card to require some confirmation of the legitimacy of the merchant's computer, perhaps confirmed by some kind of a branding procedure, through an authentication protocol that is executed before the smart card would sign anything. Just as the MasterCard, Visa, or American Express decal on the merchant's door confers a certain amount of legitimacy through the credit card company's branding and oversight function, a similar branding function might serve to authenticate the merchant's computer. Another possibility would be for the smart card to contain some built-in auditing capability, so that it would record on the card itself the identity of the merchant's computer, together with a description of the goods being purchased and the amount. Although the merchant's computer could not be trusted to display this information correctly, the user could at least check it against her receipts when she got home, using her own relatively trusted computer.

## C. Other Risks of Digital Signatures

Although there are technical risks associated with the use of digital signatures that might undermine their reliability, the principal risks are not technical issues, but issues of human and institutional behavior. The reliability of digital signatures is dependent on the appropriate conduct of both the CA and the subscriber in performing their roles.

### 1. *Errors and Omissions by CAs in Binding the Subject's Identity to the Key*

Because the relying party must depend on the CA's accurate binding of the identity of the subject to the public key, if the CA fails to accurately verify the subject's identity, the digital signature will not yield an accurate identification of the originator of an electronic document. Hence, states like Utah have enacted statutes that impose substantial regulation on a licensed CA's activities, including the use of trustworthy personnel as well as trustworthy equipment, the posting of appropriate security bonds, and a requirement for annual performance audits.[57] Other reform legislation is silent about the regulation of CAs or delegates the matter to some administrative agency.[58]

It is equally important to note, however, that not every transaction involving a digital signature involves a sum of money that would justify or compensate the CA for exercising the ultimate level of due diligence in identifying the subject. Because our society has essentially rejected any form of national identifier, the CA must make an individualized determination of identity in every case[59] and the level of effort involved may vary based on the purpose for which the subject plans to use the certificate. For example, the purchase of a $20 book over the Internet would make little sense if it could only be effectuated by the subject's purchase of a certificate that costs $5,000 because of the CA's need to verify the

---

57. UTAH CODE ANN. §§ 46-3-201, 202 (Supp. 1997).

58. *E.g.*, ILLINOIS ELECTRONIC COMMERCE SECURITY ACT § 802 (Jan. 16, 1998 Final Version) <http://www.mbc.com> (limited to CAs dealing with state agencies).

59. This effort requires more than merely verifying a subscriber's identity. A CA should require the subscriber to provide convincing evidence of his possession of the private key which corresponds to the public key which is to be bound in the certificate. Otherwise, a subscriber could extract the public key from someone else's certificate and create a certificate request as though that public key was his own. If the CA were to create a certificate containing the subscriber's name and another person's key, the subscriber could potentially pass off documents that were digitally signed by the other person as though they were his own, even though he could not forge the other person's signature to newly created documents. For this reason, all CAs should be required to have the subscriber demonstrate proof of possession of the private key by signing some piece of innocuous text ("Mary had a little lamb.") which can then be validated by the public key contained in the certificate.

subject's identity to a high level of assurance. Accordingly, CAs will issue different classes of certificates that reflect the level of effort undertaken by the CA in identifying the subject or place conditions or limitations in the certificate or a certification practice statement. Therefore, the relying party will have to educate herself about these conditions and limitations before determining whether to accept a certificate in a particular transaction.

Alternatively, this risk could be ameliorated by an insurance mechanism. If the relying party were to effectively state in the certificate status request, "I value this transaction as having value $X, and I am willing to pay the standard rate of Y% to have you insure this transaction against Errors and Omissions by the CA as well as undetected key compromises and other potential faults as stated in Policy Z," then the CA could insure that transaction and collect an appropriate fee for doing so. There is some indication that some companies are planning to offer such a service, which is similar to the business model used by credit card companies.

### 2. *Misuse of a Private Key by an Authorized User*

Obviously, if the subscriber allows an unauthorized person access to his private key, that unauthorized person can impersonate the subscriber to the detriment of relying parties. Hence, reform legislation has frequently dealt with the subscriber's obligation not to disclose his private key. The *Guidelines* are deliberately silent on the precise standard of care applicable to a subscriber's duty not to divulge the private key,[60] yet most reform legislation provides that the subscriber must use reasonable care to protect his private key or for consequences if the subscriber does not.[61]

The issue of what level of due diligence should be required of the subscriber to protect against such abuses, and who should be liable if an error occurs is a matter of public policy and requires a political solution to balance the competing interests. It seems likely that consumer protection legislation will provide some reasonable limitation on the liability that a relatively uninformed user might face for routine credit card transactions, whereas something like strict liability might be required in cases of a major bank dealing with another major bank.

There is also the problem of misuse by an authorized user. For example, assume that a husband and wife share the use of the husband's private key to

---

60. The drafters of the *Guidelines* believed that this was a public policy decision that ought to be resolved by appropriate legislation, but the drafters did observe that "[p]ersons who intentionally discloses [sic] their private keys, with or without fraudulent intent, should be held to a higher standard than an involuntary discloser." DIGITAL SIGNATURE GUIDELINES, *supra* note 10, at cmt. 4.3.4.

61. *E.g.*, UTAH CODE ANN. § 46-3-305(1) (Supp. 1997); U.C.C. § 2B-116(c) (March 1998 Draft); ILLINOIS ELECTRONIC COMMERCE SECURITY ACT §§ 305(2), 306 (Jan. 16, 1998 Final Version) <http://www.mbc.com>.

control their joint banking account. After the husband dies, the wife learns that the husband has left her only a minimal amount of property in his will and also discovers that the husband owned, as separate property, substantial assets in an investment account. After revising her computer's clock backward a couple of weeks, she digitally signs (using her husband's private key) a document directing the mutual fund to transfer ownership of the account to husband and wife as joint tenants.

### D. Summarizing the Risks

In short, most of the technical risks of digital signatures are either relatively unlikely or can be ameliorated with prudent use of available security features. However, the human and institutional risks cannot be so easily remedied or dismissed. These risks, *i.e.*, the risk of a CA misidentifying a subscriber or the subscriber allowing another person access to his private key, each raise the issue of an impersonator obtaining a private key and using it to impersonate the subscriber. In order to limit the likelihood of these risks, it is well worth considering other measures that could be used to prevent, or at least mitigate, the possibility that someone's private key might be misused and this brings us to the heart of this paper and a discussion of biometrics.

## V. BIOMETRICS: PANACEA OR PANDORA'S BOX?

Biometrics involves the use of techniques that measure something about the physical nature of a person. It is hoped that such techniques could be used to provide the third approach to individual authentication, *i.e.*, "something you are," to complement measures of "something you know" and "something you possess."

### A. Identification vs. Authentication

Because most biometric techniques involve a pattern matching or pattern recognition algorithm and because the human body changes over time, it is difficult to obtain precision from such devices. One of the most reliable devices for actually identifying people, as opposed to validating a claimed identity, is a fingerprint reader that "classifies" an image of a fingerprint using the same criteria that law enforcement agencies use.[62] The device determines the relative location of the minutia points (the points where the various fingerprint lines come together or change direction), together with a basic classification of the lines as

62. Fingerprinting is the only technique for which there is empirical, as opposed to theoretical, evidence of accuracy, by virtue of the millions of fingerprints that have been collected without matching any previous fingerprint.

ridges, loops, whorls, etc., and automatically classifies the fingerprint. That classification can then be used to search a database of fingerprints, such as the FBI's. Although such an approach is very useful for law enforcement, it requires access to a huge database of fingerprints, and that raises substantial privacy issues for more general uses.[63]

## B. Pre-enrollment vs. After-the-fact Confirmation

Most biometric techniques use a voluntary enrollment process, whereby the individual who wishes to be identified goes through a process where multiple measurements of the particular biometric indicia are made, in order to establish a baseline for future comparison. However, not all biometric techniques would necessarily require enrollment, in particular if there is a rather small likelihood that the biometric identifier would be challenged. For example, in "signature dynamics" the user signs his name with a stylus on a tablet that electronically tracks the position of the pen. In the process, not only is the path of the signature traced so that the signature could be reproduced, but the rate and acceleration of the stylus is also measured, and potentially the minute variations in pressure as well.

Although a forger might be able to duplicate the graphical pattern of a signature by tracing an existing signature, he could not duplicate the acceleration and speed of the stylus during the signature, as even the user herself does not consciously control those aspects. Therefore, if the signature will ordinarily go unchallenged, and if ordinarily the signature will only receive a visual examination, then the simple tracing of the signature would be sufficient. However, for the small portion of high value cases where a more detailed examination would be made, the availability of the recorded velocity, direction, acceleration, and perhaps pressure would make it possible to compare the signature against a pre-enrolled template with considerably greater accuracy.

---

63. At the present time, the FBI's database of fingerprints is the only reasonably complete collection of biometric data in the U.S., and the largest in the world. It contains approximately 219 million sets of fingerprints records, most collected since 1924, including over 132 million criminal cards and almost 87 million civil cards, representing approximately 75 million individuals in total, including 36 million individuals who have been arrested and/or convicted of a criminal offense in the United States. In addition to fingerprints taken of those persons who are arrested or convicted, approximately 25% of the fingerprint cards received each year come from the Immigration and Naturalization Service, with additional cards coming from the military, licensing and regulatory agencies, banking institutions, segments of the securities industry, registered futures associations, nuclear power plant personnel, child-care workers, educators, foster care providers, and others. *Fingerprint Identification and Related Information Services*, Statement of Dennis G. Kurre, Deputy Assistant Director, Criminal Justice Information Services Division, Federal Bureau of Investigation, before the Subcommittee on Immigration and Claims, Committee on the Judiciary, United States House of Representatives, Washington, D.C. (April 30, 1997) <http://www.fbi.gov/congress/cjis/kurre.htm>.

Another example of a biometric identifier is "voice print analysis"—the recorded sample of a person's voice that might be subjected to sophisticated speaker recognition algorithms that are based on analysis of the frequency spectrum (relative loudness at each frequency) while speaking certain words. The frequency spectrum resulting from one speaker saying the same word as another speaker, even with the same pitch and approximate intonation, will be quite different.[64] As in the case of the signature dynamics, there is a dynamic aspect to speech, so that the frequency spectrum, in addition to having a particular distribution at any moment in time, also changes in a reasonably consistent fashion for any given speaker sounding a given word.

### C. False Positives and False Negatives

Biometric measurements are almost always subject to some degree of imprecision, not because the devices themselves are inaccurate, but because the human body varies and so does the interaction between the body and the measurement device. Someone pressing down on a fingerprint platen may press harder and thereby flatten the fingerprint more than he did the last time he used the device, and the image may be rotated or mis-positioned as well. As a result, biometric devices are subject to two types of errors: false positives and false negatives.

A "false negative" occurs when the device does not recognize the biometric indicia of a legitimate user. This is a nuisance to the user, because it means that he must repeat the process until a good sample is obtained, but it is relatively harmless to the system, as it is fail safe. A "false positive" is much more serious, because it means that someone who is *not* a legitimate user is accepted as though he were. Most biometric devices can be adjusted across a range of sensitivities, and can trade off between an annoyingly high rate of false negatives and an unacceptable rate of false positives, depending on the application.

For most applications a false negative rate of 10% is considered tolerable, at least if the rejection occurs immediately so the user can try again. The rate of false positives varies with the technology, as some are more accurate than others. In most cases, a false positive rate of 1% would be considered quite unacceptable, and false positive rates of 0.1% to 0.001% would be considered more or less normal. A few technologies, especially fingerprint readers, may claim false

---

64. The sounds one makes while sounding a vowel, primarily phonation, are largely determined by the vocal cords and the resonant cavities of the mouth, nose and nasal sinuses, the pharynx, and even chest cavity; while the sounds of the fricative and plosive consonants, primarily articulation, are determined by the lips, tongue, soft palate, and sometimes the teeth. Arthur C. Guyton, TEXTBOOK OF MEDICAL PHYSIOLOGY (7th ed. 1986). Although mimics and humorists are often quite convincing when impersonating a famous person, they succeed primarily by duplicating or even exaggerating the speaker's mannerisms, drawl, and regional pronunciations, such as John F. Kennedy's famous "Cuber," and "Haavahd," rather than by exactly duplicating their acoustic speech patterns.

positive rates of one in a million or even higher based on theoretical arguments, but such claims should be viewed with considerable scepticism until the devices have actually been used by 10 million or more people.

### D. Latent Images and Spoofing Attacks

If a biometric technique is used in the presence of a human observer, *e.g.*, a clerk processing a charge sale, the risk of a successful spoofing attack is presumably quite low. But if the biometric device is used in an unattended environment, *e.g.*, to control access to an unattended secure facility, the risk could be much higher.

One of the threats to a biometric technique is a "latent image" attack, such as "lifting" a fingerprint from some object that a user touched, dusting it with fingerprint powder or the like, and then transferring that image to a fingerprint reader and have it accepted as the user's. To combat this attack, fingerprint readers include a variety of false finger detection mechanisms . One technique is to scan the finger using three different colors of light, one of which is in the near-infrared region. Not only does this technique tend to reject a "finger" that is not the same color as the enrolled sample, but the infrared light actually penetrates the skin. As it happens, oxygenated hemoglobin has a characteristic absorption spectrum in the near infrared, and so the fingerprint reader can actually discern the subcutaneous capillary pattern. Presumably, this prevents the use of a wax finger or even a detached or dead person's finger.[65]

### E. Capture and Replay Attacks

To illustrate another threat to a biometric technique, assume a signature dynamics system used at a point of sale terminal, such as are already appearing in many stores. A dishonest merchant could install software on a terminal that would capture the electrical impulses generated when the consumer signs her name, and those captured signals could then be replayed whenever it was desired.

Even if the signature dynamics system itself contains sophisticated proprietary processing to authenticate the electrical impulses produced while using the system against a pre-enrolled sample and then attaches that "electronic signature" to a document using a message digest technique, the system could still be spoof ed. By disconnecting the signature dynamics device from the computer that performs the processing, and then plugging in another computer, the previously

---

65. Fingerprints are not the only kind of latent images that might be used in such an attack. In a sense, we leave a latent image of ourselves every time we walk by one of the increasingly ubiquitous but relatively low resolution video surveillance cameras. In the case of an iris scanner, for example, one of the threats would presumably be the use of a high resolution photograph of an authorized person's face and eyes—a very high quality Halloween mask, as it were.

recorded electrical impulses of the user's real signature could be played back and the processing software could not tell the difference.[66] To protect against such attacks, manufacturers of such devices must incorporate both the signature recording pad and the processing capability within one tamper-proof, sealed box; and then protect that box from unauthorized substitution through some form of emblem of authenticity that would be difficult and expensive to forge—perhaps some kind of a holographic, three dimensional image which would be visibly altered if tampered with.

### F. Binding the Biometric Indica to the Document

In addition to capturing the biometric indicia at the source, there is another form of attack that could be carried out against an electronically signed document while it is being transmitted or stored. In order to identify the originator of the document, the value of the biometric indicia has to be bound to the document in some manner. However, if the biometric indicia are transmitted with the document, it would be a simple matter to copy them and then bind them to some other document. For example, in the case of an electronic fingerprint pattern, the same electronic pattern could be appended to any document and used to authenticate it.

The most reasonable way to prevent this kind of a substitution and replay attack would be to use a digital signature to do the binding. In other words, a message digest would be created that covers not only the text of the document, but also the digital pattern of the recorded fingerprint. That would make it impossible to change either the document or the biometric indicia without detection, so long as the attacker did not have access to the private key.

Unfortunately, the justification for using the biometric identifier has now become somewhat circular. The primary reason to use a biometric technique was due to concern that an unauthorized person might gain access to the user's private key, and now we are using the private key to authenticate the biometrics.

One approach to resolving this problem might be to use a different key, one that is provided by the vendor of the biometric package and somehow kept secret. In effect, the legitimate user would be authenticating himself to a trusted agent—the vendor's software package—and the vendor's private key would be used to sign the biometric indicia and bind it to the document. Assuming that the problem of spoofing the system by recording the real signature and replaying it at the electrical connection can be solved, this approach would seem to be workable,

66. This same threat exists with automated teller machines. It would be possible to build a fake ATM that would accept a person's credit card and PIN, recording them for future use, and then report that an error had occurred in processing and close the machine. The disappointed customer would go away, not realizing that her credit card number and PIN had just been stolen.

although it now involves having the vendor testify as the reliability of their system if the electronic signature is contested.

Another more serious objection might be the question of how the vendor's private key would be distributed and protected. If the system is implemented in software, it would presumably be necessary to embed the private key in the software modules, using some form of copy protection and obfuscation technique to hide the key. However, the software vendor would apparently have to do one of two things. Either the same key would be included in every software package, which would mean that once one system was compromised they all would be compromised, or the vendor would have to incorporate a different key for every customer's package. That would at least limit the extent of a compromise to that one customer, but unfortunately that would also make it quite easy to compare two different versions to see what was different, and thereby expose the obfuscation technique.

In any case, anyone who wanted to do so badly enough could use what is called an in-circuit emulator to step the program through its operation one step at a time, until the obfuscation technique was revealed and the private key compromised.

### G. Smart Card-based Biometrics

The preceding analysis casts some doubt on the viability of end-to-end biometric techniques as a primary means of identifying the originator of an electronic document, although there is no doubt that they have real utility for access control and similar uses. However, biometrics might be usable in conjunction with a smart card to control when the card can be used. As the processing power of smart cards and other tokens continues to improve, it might be possible to put the pattern matching and comparison function on the smart card itself. Ideally, even the biometric input device would be on the card, and a number of card manufacturers are reportedly exploring various schemes for putting at least a low resolution fingerprint reader on the card itself. But it is too early to tell if such attempts will be successful or economically viable, especially since smart cards without such capabilities are still considered too expensive to be deployed in mass quantities as yet, at least within the United States.

If such an approach should eventually prove successful, and if such smart cards could be so certified by some kind of an accrediting agency, and if CAs were to issue certificates which identify such signatures as having originated with such a smart card, then relying parties would have a considerable degree of confidence that only the authorized user could have originated the digital signature in question, even if they could not independently validate the biometric identification at the time of receipt.

## H. Dynamic Biometric Approaches

The attacks based on copying the biometric indicia and replaying them are based on an assumption of relatively unvarying indicia. If, however, the biometrics were to involve something that changed every time they were used, and if that change could be included in the electronic document that was being sign ed, the replay attack could be defeated.

Consider a signature dynamics system. If in addition to the user's signature the user also wrote down the date and time of the transaction and perhaps a portion of the message digest of the text that is being signed, then a simple replay attack would not work. Instead, the attacker would have to break up the signature into two pieces, the signature itself and the date/time information, then delete the date/time, and substitute a forged date/time and message digest that would be suitable.

In the case of a handwriting system, this would be relatively easy. Numbers are not particularly distinctive in the way that they are written, and it is doubtful that the velocity and acceleration techniques could be applied to validate the origin of such short, disconnected fragments. Hence, the forger could record all of the various numbers, strip off the user's numbers and insert his own in the recorded biometrics, perhaps artificially varying them slightly so that they wouldn't be precisely the same each time.

However, a voice-print based biometrics system might be able to deal with this problem. Consider a system which creates a digitally signed document, message or transaction, signing it with the user's private key which is stored on a smart card that is access-controlled by the use of a password or PIN. The certificate that binds the user's public key to his identity might also contain two additional attributes that are confirmed and signed by the CA, a color photograph of the individual and a voice print template of the user's speech. Now, in addition to the user's presumed exclusive control over his private key, biometric evidence such as an audio or audio/video recording could be used to authenticate the signer's volitional act. For example, the user might identify himself, give the data and time of the transaction, and read the message digest of the transaction that is being signed. ("I, John Doe, hereby approve this document as my Last Will and Testament, and am so indicating by affixing my digital signature to it, as of [today's date and time]. The message digest for this document is 1a47 3bc9, 8745 ef9c, 97da 0156, ebb3 2995, 58f2 888d.") The digital audio/video recording would be protected against undetected alteration by being digitally signed by the user's key, in the same manner that the electronic version of the text document is signed.

The user's identity could then be automatically validated by the relying party by comparing the signer's voice print (taken from the recorded message) against the pre-enrolled voice print template that was contained in the certificate issued by the CA. In addition, the date/time and message digest information could be

manually compared to the same information which is contained in the document itself and might conceivably be handled automatically by advanced speech recognition techniques. Finally, the validity of the document and the signature could further be confirmed, even after death, by a comparison of the video recording of the signer to the color photograph, and a lip reader could be used to confirm that the video image was in fact saying what the audio portion contained.

In other words, the identity of the user who is speaking would be confirmed by voice print analysis of the spoken text. The spoken text would contain both the date/time and the message digest of the document, providing a sufficiently long sample of speech for the voice print analysis to operate on and confirm. The inclusion of the date/time and message digest would also make it impossible to use a replay attack and apply the same biometrics to a different document. And the user's private key would be used to digitally sign the entire assembly, thereby preventing any undetected modification of the signed document while it is being transmitted or stored.

This approach assumes that voice-print analysis of spoken text other than previously enrolled phrases will be feasible and of sufficient accuracy to be useful. It also assumes that speech synthesis using the speaker's own recorded template to create the date/time and message digest portions will *not* be feasible, at least in the near future and/or at reasonable cost. Finally, if the video recording approach is also used, it is assumed that the frame by frame "morphing" of images that have recently been used in some of the blockbuster special effect movies and advertisements featuring long-dead performers such as Fred Astaire and John Wayne will remain so difficult and expensive, especially to achieve good lip-synch, as to be essentially prohibitive as a means of attacking medium-value electronic commerce transactions.

It could be argued that the user might be tricked into speaking the various numbers in some other context, and his voice recorded and later edited into the biometric data. One possible way to counter such a possibility would be to have the computer prompt the user with a list of obscure and even nonsense words. It is not likely that the user would have ever been recorded saying "obscurantist paleolithic infarcts ululate plaintively through the himmeldreck," especially not with the humorous inflection such a phrase would probably elicit.

Those are admittedly a number of assumptions that have to be validated before such a scheme could be implemented and shown to be reasonably secure, but the approach does appear to be promising.

## VI. STATUTORY REQUIREMENTS FOR SECURE ELECTRONIC SIGNATURES

Do electronic documents bearing digital signatures or some form of biometric identifier qualify for the enhanced evidentiary treatment specified by

the reform legislation discussed above, *i.e.*, should the use of a digital signature or a biometric identifier create a presumption that the originator of the message is the person indicated and that the content of the message has not been altered?

First of all, under most versions of this reform legislation, the parties are always free to agree in advance to the use a particular security procedure, and the presumptions arise if the security procedure indicates that one of the parties is the originator of the electronic document. However, there is the qualification that the security procedure be "commercially reasonable." Likewise, under the proposed Illinois Act, a security procedure must be both "commercially reasonable" and implemented in a "trustworthy manner" in order to qualify as a "secure electronic signature." Do digital signatures and biometric identifiers meet these standards?

In the case of a conventional digital signatures, there are reasonably objective criteria that can be used to measure the trustworthiness of the implementation, including the TRUSTED COMPUTER SECURITY EVALUATION CRITERIA (TCSEC)[67] together with the TRUSTED NETWORK INTERPRETATION,[68] and FIPS PUB 140-1, SECURITY REQUIREMENTS FOR CRYPTOGRAPHIC MODULES.[69] In addition to these U.S. criteria, efforts are underway within a number of countries to define a set of common criteria for computer security,[70] and it is understood that the British are developing a set of cryptographic standards that will be similar in scope to FIPS PUB 140-1. The recent German Digital Signature Law[71] and accompanying Ordinance[72] is particularly noteworthy for imposing technical requirements for CAs, and imposing a duty of instruction on CAs to inform persons who apply for a certificate as to which technical components satisfy the requirements for the generation and protection of keys and the creation and validation of digital signatures.

Given these objective measures, determining the commercial reasonableness of requiring a particular level of security is obviously a function of cost and

---

67. DEPARTMENT OF DEFENSE TRUSTED COMPUTER SYSTEM EVALUATION CRITERIA, U.S. DOD 5200.28-STD (U.S. Department of Defense, Dec. 1985).

68. NATIONAL COMPUTER SECURITY CENTER, TRUSTED NETWORK INTERPRETATION OF THE TRUSTED COMPUTER SYSTEM EVALUATION CRITERIA, NCSC-TG-005, Version-1, 31 July 1987.

69. SECURITY REQUIREMENTS FOR CRYPTOGRAPHIC MODULES, Federal Information Processing Standards Publication FIPS PUB 140-1, U.S. Department of Commerce/National Institute of Standards and Technology, (1994 January 11).

70. COMMON CRITERIA FOR INFORMATION TECHNOLOGY SECURITY EVALUATION, Version 1.0, (Jan. 31, 1996) (commonly referred to as the "ITSEC.").

71. INFORMATIONS UND KOMMUNIKATIONSDIENSTE-GESETZ- IUKDG VOM 22. JULI 1997 (BGBl. I S.1870) [Information and Communication Services Act], Article 3; SIGNATUR-GESETZ—SIGG [Digital Signature] <http://www.iid.de/rahmen/iukdgbt.html>, *translated at* <http://ww.iid.de/rahmen/ iukdgebt.html>.

72. VERORDNUNG ZUR DIGITALEN SIGNATUR (SIGNATURVERORDNUNG—SIGV) IN DER FASSUNG DES BESCHLUSSES DER BUNDESREGIERUNG VOM 8. OKTOBER 1997 [Digital Signature Ordinance] <http://www.iid.de/rahmen/sigv.html>, *translated at* <http://ourworld.compuserve.com/homepages/ckuner/verord04.html>.

availability of a suitable implementation versus the apparent risk.[73] For the average consumer transaction, requiring a C2-rated operating system and a FIPS 140-1 level 2 cryptographic implementation might be overkill, especially if the amount of money that might be at risk is limited by consumer protection legislation and regulation such as the $50 loss limit imposed by Reg. E.[74] On the other hand, consumers who wish to use their digital signature for purposes of securities trading, to buy or sell real property, or otherwise conduct economically significant business would be well advised to make use of at least those levels of computer security and cryptographic integrity. And vice versa, a merchant accepting an order for a $20 book probably need not worry too much about the security of the consumer's home computer or the extent to which the CA has attempted to identify the consumer, but if the consumer allegedly used the same computer system and the same low assurance certificate and private key to purchase a new Mercedes, a Lear jet, or an oil tanker, and the merchant accepted the order and suffered a loss, that would surely *not* be considered commercially reasonable reliance.

Unfortunately, to date no comparable objective, agreed-upon measures have been devised for biometric identification devices proposed for use in electronic commerce. Although the use of biometric devices could still qualify as a "secure electronic signature" if agreed to by the parties, there is less certainty as to how such devices could be determined to be "commercially reasonable" or implemented in a "trustworthy manner," in large part because few if any of the biometric approaches that have been proposed have been fully disclosed to the technical community, much less received the technical/scientific community's endorsement through recognized standards.

To the extent that biometric measures have not yet been standardized, the combination of digital signatures and biometrics obviously cannot be standardized either. Nonetheless, the use of biometrics, even though not yet standardized, might be useful in showing that a digital signature was implemented in a "trustworthy manner" in a high-value transaction where an abundance of caution would be considered prudent.

73. Under the Illinois Act, commercial reasonableness is determined by looking at the purposes of the security procedure and the commercial circumstances at the time it was used, including "the nature of the transaction, sophistication of the parties, volume of similar transactions engaged in by either or both of the parties, availability of alternatives offered to but rejected by either of the parties, cost of alternative procedures, and procedures in general use for similar types of transactions." ILLINOIS ELECTRONIC COMMERCE SECURITY ACT §303(a) (Jan. 16, 1998 Final Version) <http://www.mbc.com>.

74. 12 C.F.R. § 226.12(b) (1997).

Reform legislation giving legal recognition to electronic documents assumes that the use of a security procedure is sufficiently reliable to justify easing the normal rules about proving the originator and content of an electronic document. However, this paper has shown that even the highly reliable digital signature security procedure is potentially subject to compromise, primarily due to the fact that subscribers, especially consumers, may not take prudent precautions to guard access to the private key, thus making compromise of the key and impersonation of the subscriber possible. This paper has also shown that even the best biometric identification techniques, used alone, have potential security flaws that can best be overcome by using the digital signature to bind the biometric identifier to the electronic document in question. Finally, this paper has also established that digital signatures and biometric identifiers can be used in a complementary way, with the strengths of each security procedure offsetting potential weaknesses in the other. With the continued development of technology, both hardware and software, a combination of these security procedures may come very close to achieving a nonrepudiable method for identifying both the originator and content of an electronic document.