Journal of Information, Law and Technology

Signature Stripping: A Digital Dilemma

Adrian McCullagh,
Ph.D Candidate, Information Security Research Centre,
Queensland University of Technology and
Director, Electronic Commerce, Gadens Lawyers.

William Caelli,
Head of School of Data Communication,
Queensland University of Technology

Peter Little
Professor of Business Law,
Queensland University of Technology

This is a **refereed** article published on: 28 February 2001

Citation: McCullagh A et al, 'Signature Stripping: A Digital Dilemma', Refereed article, 2001 (1) *The Journal of Information, Law and Technology (JILT)*.
<http://elj.warwick.ac.uk/jilt/01-1/mccullagh.html>

Abstract

There have been a number of papers that identify the various differences and flaws as between a digital signature and a traditional signature. A primary risk in digital signature technology is the so-called 'binding' factor of the public/private key pair and the identity of the rightful holder of the key pair. This paper does not analyse this issue but instead concentrates on the issue of digital signature stripping. A digital document is simply a series of bits that can be interpreted by a computer and thus be transformed into human readable form. A digitally signed document is no different to that of digital document except that it has appended to it another series of bits that can be used to identify the signer and verify the integrity of the digital document. This paper investigates the theoretical though fundamental flaw of digital signature integrity and the right of attribution. In doing so the paper compares digital signatures to traditional hand written signatures that are pervasive in the paper-based environment. The paper also provides a solution, which involves the use of a trusted third party and the three-way communications hand-shake, which is a common communications protocol..

Keywords: Digital Signatures, Encryption, Right of Attribution, Signature Integrity, Signature Stripping, Traditional Signatures, and Trusted Third Parties.

1. Introduction

Traditional signatures are mainly used for evidential purposes and are, in longitudinal contracts as opposed to instantiated contracts, verified only when a dispute arises. Hence, the value of a traditional signature is primarily for forensic purposes and not as is commonly believed a dependent function that must be verified at the time of contract formation. This forensic value is highly important and as such commerce has developed a number of specific rules governing the evidential value of, and trust bestowed in traditional signatures.

In 1976 Diffie and Hellman in their seminal paper proposed that it was possible through the use of public key cryptography for a person to digitally sign a digital document. Diffie and Hellman had independently rediscovered the groundbreaking concept of public key cryptography. Public key cryptography was not only capable of solving the key distribution problem that had plagued conventional cryptographers for centuries but it could also be used to digitally sign electronic documents. Instead of encrypting a message with the public key of the intended recipient and thus securing confidentiality, Diffie and Hellman proposed that a plain text message could be encrypted using the sender's private key. The recipient of such an encrypted message who possessed the sender's corresponding public key would then be in a position to authenticate the signer.

Therefore, as the concept went, the receiver would allegedly be in the same position as a person receiving a physical document duly signed by the sender. In this paper we will show that a digital signature and a traditional signature are two very different constructs and as such need to be treated differently. In particular we will show that, in the physical environment, a signed document possesses not only integrity of the contents of that document but that the signature itself also possesses the element of integrity. The current

commercial implementations of digital signature and electronic signature technology do not address the integrity of the electronic signature itself. We propose a solution that should be legally acceptable to the courts and to law enforcement agencies. Despite this issue being primarily of theoretical value, due to the fact that most jurisdictions do not impose a restriction on the use of cryptography for commercial confidentiality purposes, the authors submit that if the current liberal situation changed then the solution contained in this paper should be considered by legislators as a valid mechanism to provide the necessary integrity of digital signatures.

2. Prologue

In paper-based commercial transactions the traditional signature has established itself as a cornerstone for effecting such transactions. A set of well-defined rules governing the use of traditional signatures has developed over a substantial period of time. These rules are the foundation for the current established commercial legal infrastructure, but the use of electronic signatures, it is suggested, will challenge many of these well-established rules. It is proposed in this paper that new rules will be developed. The judiciary and commerce will only accept these new rules if electronic signatures possess characteristics that are no less than the minimum-security features possessed by traditional signatures. If an electronic signature does not possess these minimum features then it is likely that it will not enjoy the trust of commerce and the judiciary.

The origins of traditional signature recognition, usage and meaning are rarely considered, yet these origins appear to offer guidance for the digital age. According to one eminent authority:

> *'The obscure origins of the art of writing must be regarded as dating back to the picture writing…. Before these pictograms could be regarded as real writing, however, it was necessary that they should pass through three well-defined stages of development. In the first place, the pictures had become conventionalised so that they always had the same appearance and designated the same object.*
>
> *'It was necessary that they should not only refer to a concrete object but also become the symbols of abstract conceptions. Finally, it was essential that these conventionalised symbols should pass into that stage with a combined representation of an abstract conception and the sound of the human voice. The last stage itself went through a number of developments.*

It is not unreasonable to expect that those entering into commercial transactions via electronic means will require assurances that the electronic signing of electronic documents will be no less secure than traditional processes. Without such assurances it is likely that people will not contract as freely electronically as they would otherwise do by conventional means in which they can repose their trust through the actual or perceived security of that process.

We have previously visited the topic of comparing traditional signatures to electronic

signatures. We concluded in our previous work that the trusted mechanism developed in the physical environment namely witnessing was not a mechanism that could be readily transposed into the electronic environment. This paper advances on our previous work by identifying that digital signatures that result from current commercial products are very insecure as compared to hand written signatures from an integrity and legal perspective. We propose an electronic solution to the lack of integrity of electronic signatures. Our solution does not prevent fraudulent false attribution but it does provide a mechanism to identify documents where an original electronic signature has been stripped from a document and substituted with a new electronic signature. It is suggested that our solution is capable of assisting organisations that wish to build electronic lodgment facilities such as the US Patents office or Land Title agencies that may desire to implement an online conveyancing register for land transactions.

3. Traditional Signatures

In order to understand better the deficiencies of digital signatures as compared to traditional signatures it is necessary to appreciate the attributes of traditional signatures. Surprisingly, there has been very little published research on what a traditional signature is from a legal perspective. There has been some historical research dealing with the concept of witnessing and in particular notaries, but by and large it is a subject that appears to have been taken for granted or assumed.

The locus classicus in Australian judicial statements on the meaning of a signature appears in R v. Moore, Ex Parte Myers, where Higginbotham made the following comment:

> *'It was observed by Patterson J in Lobb v Stanley, that the object of all Statutes which require a particular document to be signed by a particular person is to authenticate the genuineness of the document. A signature is only a mark, and where a Statute merely requires a document shall be signed, the Statue is satisfied by proof of the making of the mark upon the document by or by the authority of the signatory. ... In like manner, where the Statute does not require that the signature shall be an autograph, the printed name of the party who is required to sign the document is enough,... ; or the signature may be impressed upon the document by a stamp engraved with a fac-simile of the ordinary signature of the person signing ... . But proof in these cases must be given that the name printed on the stamp was affixed by the person signing, or that such signature has been recognised and brought home to him as having been done by his authority so as to appropriate it to the particular instrument'.*

This case raises three important issues, namely:

(a) the recognition by the Australian judiciary that a signature of a person in order to be bound by the contents of the document need not be the physical action of pen to paper but can be achieved via an agent or through the used of some mechanical means, such as an impress stamp bearing a facsimile of the persons signature. It is unclear whether the printed name of the ostensible

signatory is sufficient especially if the document is signed pursuant to a statutory obligation. For a company because it is an artificial legal entity and as such can only act through a human agent, there should, it is submitted, be a general requirement for the affixer of such a stamp to identify himself/herself not only by printed name but also by an accompanying mark.

(a) the object of a signature affixed to a document is to authenticate the genuineness of the document; and

(b) a person in order to be bound by the contents of a document, must put his/her mind to the act of signing the document as opposed to simply providing an autograph.

When an autograph is affixed to a document there is on the part of the signatory a lack of intention to be bound by the contents of the document or to be associated with the contents of that document. Hence, a mark on a document will not be regarded as a signature unless there is the necessary intention to be bound by the contents of the document or to be associated with the document. The document also must be authenticated.

The Macquarie Dictionary defines 'authenticate' as 'to make authoritative or valid; to establish as genuine'. Further the term 'valid' is defined in part as 'legally sound, effective, or binding; having legal force; sustainable at law'.

Hence, a signature, if properly executed, and there being no surrounding circumstances that legally vitiate the signature such as fraud, undue influence or unconscionable conduct, will bind the signatory to the contents of the document even if the signatory has not read the document.

A signature is capable of performing a number of functions, namely it can:

(a) identify the signatory;

(b) provide certainty as to the personal involvement of a particular person in the act of signing;

(c) associate a particular person with the contents of the document.

(d) attest to the intention of a person to be bound by the contents of the document.

(e) attest to authorship of the document by the signatory;

(f) attest some written agreement which may have been written by some third party who is not a party to the binding agreement.

The general physical characteristics of traditional signatures are that they:

(a) can be easily produced and reproduced by the same person;

(a) are easily recognised by third parties;

(b) are relatively difficult to forge by third parties;

(c) become bound to the document such that the physical object and its contents and the signature become one composite physical thing;

(d) involve a physical process (ink to paper);

(e) are comparatively standard for all documents signed by the same person; and

(f) are relatively difficult to remove without leaving some trace.

The general legal characteristics of a traditional signature are that:

(a) any kind of mark is acceptable provided it was affixed by the person or by some person authorised by the person intended to be bound,

(b) unless there is some specific legislative requirement, the mark can be affixed by some mechanical means;

(c) the mark can be highly insecure such as a mark that has been effected by pencil;

(d) at the time of affixing the mark, the signatory must have the necessary intent to be bound by the contents of the document or, in the case of being a witness, the necessary intention to be associated with the document as a witness; and

(e) the mark can be located anywhere within the document and does not have to be at the foot of the document unless there is some legislative requirement as to form as in a will.

Hence, an electronic signature should be capable of acceptance as a signature; if:

(a) the electronic signature is accepted as being a mark which is capable of being affixed by the person or by some other person authorised by the person to be bound;

(b) the electronic signature is accepted by the receiver as being affixed via mechanical means, as can a traditional signature;

(c) at the time of affixing the electronic signature, the signer has the necessary intention to be bound;

(d) similar to traditional signatures, an electronic signature can be located anywhere on the document and does not have to be at the foot of the document unless there is some legislative requirement as to the form of the document as in a will, which usually requires each page to be signed by the testator and each witness. In this regard the mechanism used to affix the electronic signature should ensure that the electronic signature is attached to the document such that any detachment of the electronic signature from the document will cause a total failure in the mechanism used to verify the electronic signature;

(e) the electronic signature has the same integrity qualities as a traditional signature;

(a) an electronic signature can be affixed in such a manner that if the electronic signature is removed and substituted by the signature of a third party then the receiver of the falsely signed message will be in a position to identify the fraudulent action and thus be put on a train of enquiry.

It is elements (d), (e) and (f) above that are the most problematical. As will be discussed more fully a digital signature is simply an extra set of bits that are attached to or logically associated with some digital document. As such they can be removed without trace of their prior existence. If the reader is not interested in understanding how a digital signature is formed then the reader can skip Section 4 and continue at Section 5.

4. Electronic Signatures

4.1 Synopsis

This part of the paper describes what an electronic signature means and that there exists a sub category known as digital signatures. This paper concentrates on digital signatures and identifies that a digital signature is simply a series of bits that are appended to or logically associated to a digital document, which is also simply a series of bits. The issue is that the digital signature can not be attached to a digital document in the same manner as a traditional signature.

4.1.1 Electronic Signatures

An electronic signature is any electronic mark that indicates the identity of some person who is being attributed as being the signatory. The difficulty with many electronic signature mechanisms is that they do not possess an adequately secure mechanism of authentication. This authentication not only relates to the authentication of the document but should also relate to the authentication of the person to be charged with being bound by the contents of the document. In non-face-to-face transactions the process of authentication is highly important. When a transaction is being effected in a face-to-face situation both parties are in a good position to identify each other and if, for example the transaction is a cash transaction then the seller of the goods is satisfied as soon as he/she has possession of the money. The buyer will or should at the time of the transaction also

possess the goods or if there is a time delay in delivery of the goods the buyer will know where to attend if the goods are not forthcoming or are later discovered to be defective. Hence there is a simple mechanism of authentication of the parties to the transaction in a face-to face encounter. That is, the buyer is at the time of effecting the transaction in a position to identify the seller and the seller's location. In a non-face-to-face transaction neither party will be in a position to properly identify the other unless the electronic signature mechanism being used securely authenticates each party to the other. In credit card transactions, for example, credit card companies assign greater risk to 'card not present' transactions, such as phone transactions and internet transactions, than that given to 'over the counter transactions', as indicated above. The maturest technology that provides the necessary security to achieve authentication is digital signature technology. The remainder of this paper will discuss the issue of affixing digital signatures and how to implement an integrity mechanism for the entire message which must include the digital signature, in an equivalent manner to that found in the paper based environment.

4.1.2 Digital Signatures

Diffie and Hellman coined the phrase 'unforgeable digital signatures and receipts are needed' within an electronic messaging system. This was apparently the first use of the term 'digital signatures' in the public literature. Diffie and Hellman stated that a public key cryptosystem comprises two keys:

$$\{E_k\}\ k \square \{K\}, \text{ and } \{D_k\}\ k \square \{K\}$$

which are mathematically related but it is computationally infeasible to determine one key from the other. Their work centered upon one-way functions that incorporated a trap door. The primary characteristics of a public key crypto-system are that for the message space {M} there exists:

(a) for every k□{K}, Ek is the inverse of Dk;

for every k□{K} and M□{M}, the algorithms Ek and Dk are easy to compute;

for almost every k□{K}, each easily computed algorithm equivalent to Dk is computationally infeasible to derive Ek; and

for every k□{K}, it is feasible to compute inverse pairs Ek and Dk from {K}.

Element (a) above provides that within the key space {K} there exists a pair of keys Ek and Dk such that they are the inverse of each other. If Ek is used to encrypt a message to give the ciphertext then Dk being the inverse of Ek must be capable of decrypting the ciphertext to reveal the original message in 'plaintext'.

Element (b) provides that the whole process must be computationally feasible by being easy to compute. That is, for every key (k) that is within the key space {K}

> and for every (M) that is within the message space {M} the encryption and decryption algorithm must be easy.

As stated by Diffie and Hellman element (c) is the property that allows one of the user's key pair to be made public without compromising the secrecy of the other key. Therefore, the signer can disclose the verification key to the world with the knowledge that such disclosure does not compromise the secrecy of the signing key. This is a primal element of digital signature technology. Provided only one person has the relevant private key then anyone who receives a message that has been digitally signed through the use of that private key will use the corresponding public key to verify that that particular private key was used to digitally sign the message. Ipso facto, the receiver is then in a position to identify the signer.

The fourth element (d) ensures that there must be a way to feasibly compute a key pair related to inverse transformations when no constraint is placed on either what the enciphering or deciphering mechanism is to be.

Diffie and Hellman did not provide a feasible solution to the public key problem though they did give a simple example based upon calculating the inverse of a matrix. The computational power to calculate the result of a vector through matrix transformation is N^2 operations where N is the size of the matrix. If the matrix has an inverse then the computational power to compute the inverse is N^3 operations. As Diffie and Hellman observed this power differential is too small. If the computational power to calculate the inverse were N^{100} operations then it would be possible to say that it is computationally infeasible to calculate one key of the pair from simply knowing the other key of that pair.

Rivest et al advanced the Diffie Hellman proposition by developing the first publicly accepted and working public key cryptosystem. Rivest et al proposed as regards to signatures that:

> *If electronic mail systems are to replace the existing paper mail system for business transactions, 'signing' an electronic message must be possible. The recipient of a signed message has proof that the message originated from the sender. This quality is stronger than authentication (where the recipient can verify that the message came from the sender); the recipient can convince a 'judge' that he did not forge the message himself! In an authentication problem the recipient does not worry about this possibility, since he only wants to satisfy himself that the message came from the sender (our emphasis)*

Since Rivest et al there have been innumerable papers and articles concerning digital signatures and digital signature schemes. To date, the vast majority of this material has concentrated on the technical aspects and not the legal, social or economic aspects of digital signatures.

In simple terms a digital signature is a cryptographic conversion of some data such that the encrypted data can be allegedly attributed to the holder of the private key used to perform that cryptographic conversion where the private key forms part of a public key

pair. The basic assumption behind digital signatures is that if only one person has access to the private key and an encrypted document can only be decrypted using the corresponding public key then the encryption process must have occurred through the use of such private key, which means that is was the holder of the private key who encrypted the document. Consequently, the identity of the signatory of the electronic document has been revealed. As will be discussed further in this paper this does not mean that attribution has been achieved.

A major issue in using public key crypto-systems is that they are substantially slower than symmetric key crypto-systems. When Diffie and Hellman first proposed the use of public key crypto-systems as a method of digitally signing documents they were anticipating the encryption of the entire document through the use of the private key. Essentially the entire digital document acts as the digital signature. In doing so each message is broken up into manageable blocks that are equal in size to that of the signing key. This is a novel proposition in that a traditional signature is affixed at some place on the document. The document does not become converted into the signature. To overcome the slowness of public key cryptography for signing purposes, Rivest proposed a mechanism known as hashing.

4.2 Hashing Algorithms

The purpose of the hashing algorithm is to produce a message digest of the document. A message digest is sometimes called the fingerprint of a document and is much smaller in size to that of the original document. There are a number of one-way hash algorithms currently being used such as MD5 and SHA-1. MD5 produces a message digest that is 128bits long whereas SHA-1 produces a message digest that is 160 bits long.

All hash algorithms have the following characteristics:

- A message M of arbitrary length must be able to be converted into a message digest h(M) of fixed length;
- The algorithm must be one-way, i.e. even knowing h(m) and the algorithm , it must be computationally infeasible to determine M;
- h(M) must be ‘collision free’. It must be computationally infeasible for messages say M1 and M2 to result in the same hash. That is, it must be computationally infeasible for h(M1) = h(M2).

This issue of collision free is very important because as will be detailed below it is the hash that is operated on by the public key crypto-system. If two documents possessed that same hash then the digital signature of one document could be uplifted and inserted into the other document without affecting the integrity of the document. Furthermore, the false digital signature would be verified as a valid digital signature. Dobbertine identified a flaw in the hashing algorithm MD4 which was a precursor to MD5 as not being collision free.

In fact Dobbertine was able to show two contracts that were simply different as regards to the amount to be paid under them. In contract ‘A’ the amount was $176,495 and in contract ‘B’ the amount was $276,495. There were no other difference specified in the contract. As can be seen the difference in liability is $100,000. The digital signatures were the same because the hash of each document was the same.

Figure 1: Message using the Hash Algorithm

The entire exercise of affixing a digital signature is as follows:

Figure 2: Affixing a digital signature

The digital signature is then either appended to the file containing the DOC1 or can be sent separately to the recipient. Therefore, a digitally signed document comprises the message and the digital signature which is in essence is a series of bits known as the digital signature concatenated to another series of bits known as the message.

Figure 3: Components of a digitally signed document - the original message and the digital signature

This is very different to what occurs in the physical environment where the traditional signature and the document become one indivisible artefact. In the physical environment, in order to remove a signature and replace it by another signature, the signed document would have to be mutilated by physically cutting the document so as to separate the traditional signature from the document and then have the traditional signature affixed by a paper clip to the document or for the traditional signature to be affixed on a separate

piece of paper and for that separate piece of paper to be attached by a paper clip to the document. A traditional signature has all the characteristics of a classical one-way function in that it is easy to process in one direction but very difficult to reverse the process, i.e. easy to affix, difficult to remove. The analogy to a paper clip being used to affix a digital signature is apt because the clip can be removed without leaving a trace in the same manner that a digital signature can be removed from a digital document without leaving a trace. This clearly does not occur in the physical word. It is absurd to think that business would rely upon paper clip technology as a basis for securing a signature to a document. But this is what is being suggested through the use of digital signatures.

The process to verify the digital signature is as follows:

Figure 4: Verifying the digital signature

A comparison of Message Digest (a) and Message Digest (b) will reveal two factors associated with the document. If the Message Digests are the same then the recipient knows that:

(i) the document sent by the sender has not been altered; and

(ii) the document has been digitally signed using the private key that corresponds to the public key that was used in the verification process.

Provided the sender is the only person who has access to the private key then it must have been the sender who encrypted the associated Message Digest(a). This is because the public key used by the receiver to decrypt the digital signature could only correctly decrypt a digital signature that had been encrypted using the private key associated with the key pair. The main assumption here is that the sender has not compromised the private key.

As can be seen from the above exercise a digital signature will be different for each document because the private key is used to operate on the message digest which in turn results from the operation of the hashing algorithm on the original document.

Hence the general physical characteristics of digital signatures are that they:

(a) are simply a series of bits that are attached to, or logically associated with another set of bits known as the digital document;

(a) can be easily produced by the same person provided they have access to a computer, appropriate software, their signing (private key) and a digital document;

(a) are easily verified by document receivers and independent third parties;

(a) are very difficult to forge by third parties unless the forger has through some means obtained access to the private key;

(a) are not permanently bound to the document such that the contents of the digital document and the digital signature become one indivisible series of bits;

(a) involve a mathematical process;

(a) are exactly the same for all documents that are exactly the same when signed by the same person; and

(a) are very easy to remove without leaving some trace.

In summary, public key crypto-systems are basically block ciphers that break up a message into manageable blocks, which are individually operated on through the use of one of the key pairs. If the exercise is to digitally sign a document then the private key is used in the operation. Because public key crypto-systems are substantially slower than symmetric key crypto-systems, the message is for digitally signing purposes first operated on by a one way hash function, which will calculate the fingerprint of the document. The fingerprint will normally be either 128 bits long or 160 bits long depending on the algorithm used. Further this is all possible because every message is expressed within a computer as a binary string and as such each character can be concatenated so as to produce a large bit string. It is this bit string that is broken up into blocks and operated on

as if it were a series of numbers and not a character. Each of the mathematical algorithms above process whole numbers for encryption purposes.

5. The Right of Attribution

5.1 Synopsis

This part describes a new protocol that will place a recipient of a digitally signed document in the same position as the recipient of a physical document, which has been altered by changing the signature on the document. In the paper based environment the recipient will usually be able to identify if the signature to the document has been changed. This protocol is to provide a functional equivalent conclusion for the benefit of the recipient of a digitally signed document.

5.1.1 Signature Stripping

As stated above a digital signature is simply a set of bits that are attached to or logically associated with a digital message. They can very easily be removed by a fraudulent third party without leaving a trace. The issue is: 'why would someone want false attribution to a message'?

If you can imagine an online lodgment service where legal rights are created or granted upon the lodgment of documents then by deleting the rightful signature and replacing it with another signature the person identified by the replaced signature will be attributed as the rightful beneficiary of such legal rights. Such a realistic situation arises with the online lodgment of inventions with the Patents Office. The inventor or the rightful owner of the invention is required to sign the application form. Provided the invention meets the legal requirement of registration, the patent rights will be granted to the party noted as the applicant. That is the party who signed the application form. The same could also be said for online land transactions. If an agency established an online lodgment facility then the noted purchaser will be registered as being the rightful proprietor of the land title.

The problem is that in the paper based environment there exists through the physicality of the signature and the document an automatic integrity mechanism relating to the signature itself. As stated above it is very difficult for a signature to be removed without leaving a trace. It is in essence a one way function. If on the face of a document there is evidence of tampering then the recipient can not simply rely upon the document but has a duty to make enquiries as to the authenticity of the document and in this case, the signature. Therefore the issue in the digital environment is:

> 'Is it possible to provide an integrity mechanism such that if there is a substitution of digital signatures then the recipient can determine when such substitution takes place and thereby be put in train a series of enquiries to ascertain the true situation?'

Hence, the recipient of a digitally signed document must at a minimum be placed in the

position of being able to determine whether or not a digitally signed document has a substituted digital signature attached to it. If the recipient can make such a determination, then, if a substitution has occurred, the entire veracity of the document must be questioned.

5.2 Digital Signature Integrity

One of the benefits of a digital signature is that it provides an integrity mechanism for the contents of the document but it does not provide an integrity mechanism for the digital signature itself, when considered in relation to conventional documents whereby the signature and document become one indivisible whole.

Stinson proposes that the digitally signed document could be encrypted using the intended recipient's public key. That is given the message x, the sender would compute the digital signature ds = sig_{sender}(message), and then encrypts both the message and ds using the receiver's public key; Z = Epr(ds, message). The difficulty with this solution is:

(a) it will be slow;

(a) if the El Gamal system is used to encrypt the entire message then the length of the encrypted message is doubled;

(a) In some jurisdictions it is for policy reasons illegal for citizens to encrypt messages or to even possess encryption technology for confidentiality purposes.

Points (a) and (b) can easily be overcome by encrypting the message M and ds using a symmetric key function and then encrypting the symmetric key using the receiver's public key. That is the sender selects a symmetric key (k) and encrypts M and ds obtaining Z=Ek (M,ds). The sender also encrypts the symmetric key using the public key of the recipient and obtains Y=Epr(k). The sender forwards to the recipient both Z and Y. The recipient first decrypts the symmetric key using his/her own private key and then proceeds to verify the digital signature.

If encryption technology is illegal then what mechanism can be used that can give the required integrity over the entire document which must include the digital signature? This is not a trivial task. For example the following protocol at first analysis appears to be a method of providing integrity of digitally signed documents inclusive of the digital signature.

h(ds,M,) →X

h(M)→Y

(where: X is the hash of the digital signature together with the message;.
ds is the digital signature attached to or logically associated to the document;
M is the message).

By hashing the aggregate of the message and the digital signature the sender has captured the desired fingerprint in X.

The sender then forwards a new message comprising (Y, X, S, R) to a trusted third party who time stamps (Y, X, S, R) and places (Y, X, S, R, T) in a secure public directory that has read capability to all but restricted write, amend and append capabilities.

$$S \rightarrow TTP : (Y,X,S,R)$$

(where: Y is the hash of the unsigned message. X is the hash of ds and M. T is a time stamp appended by the TTP).

S then sends M and ds to R. R will upon receipt of M and ds calculate:

$$J=h(M)$$

$$I=h(M,ds_1)$$

(where: M is the unsigned message actually received by the recipient; ds_1 is the digital signature actually received by the recipient; J is the hash of the message M; I is the hash of the digitally signed message actually received by the recipient).

But if S's signed message M is captured and ds is deleted and substituted by ds' then there nothing to stop the interceptor C from altering the documents form but not it's substance. If C adds a space or some punctuation like a comma or a dash then the documents substance is maintained but h(M') will be substantially different to that of h (M). Hence the storage of h(M) for the benefit of R will be insufficient. S needs to send (S,R,h(M,ds)) to the trusted third party.

This should be done at a minimum using a three way handshake as follows:

Figure 5: Three way hand-shake

In this figure S has digitally signed a message and has attempted for forward it to R but C has intercepted the message and has stripped S's digital signature and substituted his/her own digital signature. The following three-way hand shake is recommended:

1: S →TTP : TTPk(S,R,h(M,ds,Na);

2: TTP→S : Sk(Na+1,Nb,T);

3 : S→TTP : TTPk(Nb+1).

(where: TTPk is the public key of TTP; Sk is the public key of S; T is the time stamp of the message received from S; S is the name of S; R is the name of R; H (M,ds) is the hash of the digitally signed document of S; Na is nonce issued by S, which will be used by S to satisfy itself on return that TTP actually received the message; Nb is a nonce issued by TTP, which will be used by TTP to satisfy itself on return that S actually received the message. It could be a registration number of the message relating to R and S and as such is stored in a public directory under that number. This will allow R an easy mechanism to search for h(M,ds) from S).

4 : TTP→R : Rk(S,R,Nb,Nc,T)

5. R→S : TTPk(Nc+1)

(where Rk is the public key of R; TTPk is the public key of TTP; Nb is nonce issued by TTP and can also be used as the registration number of message received by TTP from S; Nc is a nonce issued by TTP so that TTP can be satisfied that R received the message (S,R,Nb,Nc)).

The protocol detailed above can be classified as a active protocol as it requires TTP to notify R on receipt of a message from S. If R does not receive a message from S shortly thereafter then R should activate a train of enquiry so as determine the status of the missing message. Alternatively the protocol could be made into a passive protocol. In this situation when R receives a message from C, R will look up the TTP's public directory and extract all recent records that exists under his name. If no record is recovered that identifies S as having sent a message to R then R should initiate a train of enquiry as to why he/she has not received a message from S. R should request S to re-send the original digitally signed message M. The time stamping by the TTP will identify who was the first person to sign the correct document.

The above protocol does use encryption but it is not for confidentiality of plaintext document but for security of the hash of the aggregated message and digital signature. As

can be seen for jurisdictions where the private use of encryption for confidentiality purposes is illegal, the digital signature integrity protocol involves substantial overhead and the involvement of a trusted third party. Such trusted third parties could be certification authorities, which adds to their business operations and therefore their revenue mechanism.

Many of the electronic lodgment systems involve government agencies like patent offices and land title offices. Since the encryption is required between the agency and the private citizen then such communication can be encrypted because the agency can decrypt the digitally signed message. Between private citizens in jurisdictions where it is illegal to encrypt messages. We submit that a trusted third party system should be utilised. Of course the cost of implementing such a system may outweigh the benefit for all messages and there may only be value in business to business communications.

6. Conclusion

There is a clear and fundamental difference between a traditional signature and a digital signature. In this paper we have identified a fundamental flaw with digital signature technology in that the digital signature itself in not securely attached to a document so as to create one indivisible artefact. It is technically possible to better secure the attachment of a digital signature to a digital document by encrypting the digitally signed document but in some jurisdictions the use of encryption technology for confidentiality purposes is illegal.

We put forward that in order to have integrity of the entire message which includes the digital signature then the message and the digital signature should be concatenated and the hash of the concatenation should be lodged with an independent trusted third party who will time stamp the lodgment of the hash. This solution provides an added business opportunity for trusted third parties in providing digital signature integrity and emulates in some way the operation of a notary.