Digital Signatures, Certificates and Electronic Commerce

Brian Gladman[1], Carl Ellison[2] and Nicholas Bohm[3]

1. Introduction

Cryptography as practised for 3500 years before the mid-1970s used one secret, known by both the enciphering and deciphering parties. Anyone possessing that secret (called the ***key***) was able to do either operation. In public key cryptography there are two different keys, one used to encipher data and the other used to decipher it. While one of these keys is kept private, the other is made public, and the system is designed so that knowledge of the public key does not allow the private key value to be determined. Normally it is assumed that the private key is controlled by just one person who will be referred to here as the ***keyholder***.

Public key cryptography can be used to achieve either ***confidentiality*** or ***digital signatures***. Confidentiality is provided when the deciphering key is kept private and the enciphering key is made public, so that anyone with the public key can encipher a message that only the keyholder can decipher. A digital signature is provided when the enciphering key is kept private and the deciphering key is made public, so that anyone with the public key can decipher a message that only the keyholder could have enciphered. With a correctly verified digital signature, we know that the signed object has not been modified since the signature was made and that this was done with a specific private key.

2. Digital Signatures

For digital signatures there is a key pair: a secret ***digital signature key***, which is used for signing data items, and a public ***digital signature verification key***, which is used by others to check if a signature was made with the related signature key.

Although the term 'signature' is used for both, in practice digital signatures and hand-written signatures have quite different characteristics. If a digitally signed document is tampered with in any way, its digital signature will not verify. However, the private key that generated the signature can be used by anyone who can gain access to its value, making it more like a seal. On the other hand, a written signature is linked in a biometric way to its owner but its use on a document does not prevent the latter's subsequent, undetected modification.

A written signature is evidence that the person acted on a document whereas a digital signature is evidence that a given private key acted on a document. Usefully, a digital signature also tells us that the document has not been changed since the signature was made (the way digital signatures work provides this confidence). However, without further information, a digital signature provides no evidence about the participation of any particular person.

If the secret signing key is kept under the control of a single person, then one can assume that the key acts as an agent for that person and one might assume that documents signed by the key were actually signed by the person. This is analogous to a check-writing machine. As long as the machine is kept well guarded and access to the machine is only by authorised signers, the machine signature can stand for the human signature it replaces. However, establishing that a signing key was under the control of a particular person at the time of a particular signature and that the person actually knew what he was signing with that key is a very complicated process.

2.1. Identifying the Keyholder

If a person we know and trust hands us their verification key, we can then associate this key with them provided we are sure that they alone have access to the related secret key value. Under these conditions we can link their name (or some other data that identifies

[1] *Brian Gladman, Information Security Consultant, United Kingdom.*

[2] *Carl Ellison, Intel Incorporated, United States.*

[3] *Nicholas Bohm, Solicitor, United Kingdom.*

them for our purposes) to their verification key to remind us of this association. In so doing, we have provided extra information with the verification key that will subsequently allow us to associate a verified signature with our friend, the keyholder. We have thus linked information to the verification key that describes a property of the related secret key, in this case, that it belongs to our friend. We will see later that we can link a wide range of different properties to secret signing keys in this way.

If a stranger hands us a verification key that they claim is for their signature, we know much less. We can give them something to sign and, if the resulting signature is valid, we know that they have access to the related secret key. However, we don't know if other people also have this key, so we cannot be sure that they alone can use it. Moreover, since we don't know them, they can claim to be anyone they please and we will be none the wiser (unless they are unlucky and choose someone we know). In general, therefore, unless we have obtained signature verification keys directly from people we know and trust, it is difficult to be sure about the identities of their owners.

3. Digital Certificates

As suggested earlier, if a friend hands us their signature verification key, we might associate their name with this key to remind ourselves that they have the related signing key. Our friend can then send us documents with signatures that we can check using the verification key that we associate with them. We can also pass this key, linked with our friend's name, to our other friends so that they can also check this signature. In addition, in order to protect the association of this name with this key from tampering, and to record its origin, we can digitally sign this (key, name) combination. In essence, we are asserting that the named person is the keyholder for the key in the (key, name) combination that we have signed. Provided that our friends associate the name we use with the right person[4], they can then check this signature even though they have not obtained the verification key directly.

A digitally signed combination of a signature verification key with other data that states a property of the related secret key is known as a ***certificate*** [1]. The person who signs such a certificate is asserting the stated property of the signing key associated with the verification key in the certificate. When we attach our friend's name to their verification key and sign the combination, we are stating that we use the name in the certificate to identify the related signing key. This is an example of an ***identity certificate***. The person receiving and using this certificate might also want to know that the named keyholder has sole control over the corresponding private signing key, but in general we as issuer of the ID certificate have no control over the protection and use of that private key and can therefore make no such assertion.

It is particularly important to distinguish between trust in a signature and trust in the owner of a signature. Under the right conditions digital signatures can provide confidence that a person (or an entity) has signed a data item but still say nothing about the trustworthiness of the person concerned. As with ordinary signatures it is possible to contemplate situations in which a signature is trusted but its owner is not.

Whereas a conventional signature is quite strongly associated with its owner and less so with a signed document, a cryptographic signature has the opposite properties. In isolation, a digital signature is only weakly associated with its owner since it is easy to publish a signature verification key with a claim that it belongs to someone else. . It is also possible for the computer that employs the signing key to be used by someone other than the key's reputed owner or for the software in that computer to be hostile to the interests of the owner.[5]

---

[4] The naming issue is not trivial. One needs to use names that are both unique and meaningful to the viewer of the name. In the days of small communities, common names served this purpose. In today's large communities, common names are not unique, but we use common names and even smaller personal nicknames so only those names would be meaningful. In other words, there is no possible name choice that is both unique and meaningful, at present, and any new name structure fitting those requirements would require a change in social conventions that might take centuries to accomplish.

[5] As would be the case with computer viruses or applications that have hidden execution modes.

Certificates offer a way of overcoming the first weakness by requiring that verification keys are 'countersigned' using a digital signature that we know is trustworthy. As an example, we might only trust keys that we have obtained directly from people we know and trust. However, we might extend this by not only accepting their keys but also any keys that they have signed. This sort of arrangement is often called a 'web of trust'. But this does not fully resolve the difficulty since we might still receive a key countersigned by someone we cannot trust because we don't know them. We could insist that countersignatures are themselves signed but this process has to stop somewhere and there can only be an unsigned verification key at the end of the chain. This last key is often referred to as the root key and another way has to be found to ensure that it is trustworthy (this will be discussed later).

Generally, a digital certificate is a collection of data that has been signed with a digital signature key. The keyholder of the signature on the certificate is asserting the validity of the contents of the certificate and, where there is an enclosed signature verification key, more specifically that they assert the stated property of the secret signing key associated with this verification key.

In practice, the vulnerabilities of digital signatures (and the systems in which they are used) will mean that certificates will sometimes fail and this has to be considered in designing the systems that make use of them.

One use of certificates that we have already discussed is that of certifying a name for a given key. Here a certificate will contain a verification key linked with information that allows the keyholder to be identified. The person signing such a certificate may be confirming that the identified individual has access to the secret key associated with the public verification key in the certificate. The signer might also claim that the identified person alone has access to the private key. Details about what is claimed by a certificate depend on information outside the certificate (typically called a Certification Practices Statement).

In order to check for a valid digital signature it is necessary to have a trustworthy verification key. Such a key may already be available in which case the signature can be verified immediately. If, however, the verification key is provided in the form of a certificate then the signature on the certificate must also be checked. At this point the process may repeat itself since we may need another certificate for this signature. Eventually, however, we must reach a key that we know we can trust or one that is not countersigned. In the latter case, one way of checking the key is to look up its value in a widely available printed book. Recently a publication – The Global Trust Register [2] – has been established for this purpose.

4. Types of Digital Certificate

4.1. Identity Certificates

An Identity Certificate is one that contains a signature verification key combined with sufficient information to identify (hopefully uniquely) the keyholder. This type of certificate is much subtler than might first be imagined and will be considered in more detail later.

4.2. Accreditation Certificates

This is a certificate that identifies the keyholder as a member of a specified group or organisation without necessarily identifying them. For example, such a certificate could indicate that the keyholder is a medical doctor or a lawyer. In many circumstances, a particular signature is needed to authorise a transaction but the identity of the keyholder is not relevant. For example, pharmacists might need to ensure that medical prescriptions are signed by doctors but they do not need to know the specific identities of the doctors involved.

Here the certificate states in effect that the keyholder, whoever they are, has 'permission to write medical prescriptions'. Accreditation certificates can also be viewed as authorisation (or permission) certificates. It might be thought that a doctor's key without identity would undermine the ability to audit the issue of medical prescriptions. However, while such certificate might not contain keyholder identity data, the certificate issuer will know this so such requirements can be met if necessary.

4.3. Authorisation and Permission Certificates

In these forms of certificate, the certificate signing authority delegates some form of authority to the key being signed. For example, a Bank will issue an authorisation certificate to its customers saying 'the key in this certificate can be used to authorise the withdrawal of money from account number 271828'.

In general, the owner of any resource that involves electronic access can use an authorisation certificate to control access to it. Other examples include control of access to secure computing facilities and to World Wide Web pages.

Although authorisation certificates might contain identity information this will not generally be necessary and may often be undesirable. For example, in the use of digital certificates to implement 'electronic cash' individual transactions need to be anonymous if they are to mimic normal money transactions and this means that the certificates involved will be authorisation certificates ***without*** identity information.

In banking an identity certificate might be used to set up an account but the authorisation certificate for the account will not itself contain identity data. To identify the owner of a certificate a bank will typically look up the link between account numbers and owners in its internal databases. Placing such information in an authorisation certificate is actually undesirable since it could expose the bank or its customers to additional risks.

For example, if such information were to be included in authorisation certificates, this could allow a bank's competitors to compile lists of its clients as certificates pass through the banking system. It would also allow anyone handling the certificate to use the identity information for unintended purposes that might put the client at risk. Moreover, privacy laws could make such practices unlawful since the issuing bank may have revealed personal client information to third parties.

While it might be argued that access to such information could be controlled, by far the best means of control is not to include it if it is not needed. This illustrates an important general principle for certificates: they should be designed carefully for specific purposes and the information they contain should be only that required for this and no more. This suggests that general certificates with wide availability and wide use are likely to be less useful than those designed for specific purposes.

5. The Parties to a Digital Certificate

In principle there are three different interests associated with a digital certificate:

- **The Requesting Party** – the party who needs the certificate and will offer it for use by others – they will generally provide some or all of the information it contains;
- **The Issuing Party** – the party that digitally signs the certificate after creating the information in the certificate or checking its correctness;
- **The Verifying Party (or Parties)** – parties that validate the signature on the certificate and then rely on its contents for some purpose.

For example, a person – the requesting party – might present paper documents giving proof of identity to a government agency – the issuing party – who will then provide an identity certificate that could then be used by a bank – the verifying party – when the requesting party opens a bank account.

The term 'relying party' is sometimes uses instead of 'verifying party' but this can be misleading since the real purpose is to identify a party who checks the certificate before relying on it. In a credit card transaction many parties might handle a certificate and hence rely on it in some way but only a few of these might actually check the validity of the certificate. Hence a 'verifying party' is a party that checks and then relies on the contents of a certificate, not just one that depends on it without checking its validity.

The actual parties involved in using a certificate will vary depending on the type of certificate. The following table shows some typical examples.

Type of Certificate	Requesting Party	Issuing Party	Verifying Party
Identity	The person concerned	The appropriate government agency	Anyone undertaking an identity check
Accreditation	A qualified member of a profession	The professional body	A user of the services offered by the member
Authorisation	A customer wishing to access a resource	The resource owner	The resource owner

Normally there will be one requesting party and one issuing party but, possibly, many verifying parties. In some applications, these parties need not be distinct – in the last example above the issuing party and the verifying party are the same (although this might not always be true in such uses). This often applies for authorisation certificates.

6. Open and Closed Digital Certificates

In many circumstances, the parties associated with a digital certificate will use it within the context of a wider contractual relationship established between them. For example, when a customer opens a bank account, there will be a contract that sets out the conditions under which the customer and the bank will operate the account.

When a digital certificate is used in such situations, the contract that exists between the parties can be used to set out the conditions for the use of this certificate, including any liability issues if things go wrong. This is an example of a closed digital certificate – one for which all parties with an interest in it are bound by a contractual relationship.

In contrast, an open digital certificate is one in which one of the interested parties (usually a relying party) is not a party to any agreement covering its use. For example, a doctor and the professional medical body that issues his or her accreditation certificate will probably have an agreement governing their relationship. A patient, however, may not be a party to this agreement, although they could be at risk if such a certificate was false. This is an example of a situation where one party that relies on a certificate could be at risk but has no agreed basis for seeking recompense if things go wrong.

A party verifying an open digital certificate has several difficulties. First, they may have no immediate way of knowing whether it is trustworthy or for what particular purpose its use is allowed. While they may know the identity of the organisation that signed the certificate, this might not tell them much about the care that they took in its issue or the specific meaning attached to the certificate and its various fields.

Second, because they are not a party to any contract governing its use, they have no way of knowing whether this places any constraints on its operation or provides for compensation if things go wrong. Although some of these issues might be mentioned in the certificate, it is unlikely that their implications will be easy to understand without considerable research. It will also be important to understand whether the provisions established in the domain (jurisdiction) of issue will apply in the domain (jurisdiction) of use.

If these issues are not tackled uniformly, it is likely that the value of open digital certificates will be severely limited. Certificate Authorities (CA) aim to overcome such problems by providing a uniform and widely understood framework for the use of open digital certificates. However, this is a new and somewhat unfamiliar concept that may take a considerable time to become established. Moreover, there is now a growing realisation that open digital certificates are not necessary for electronic commerce, the majority of which can be undertaken using certificates of closed form.

7. Trust and Confidence in Digital Signatures and Certificates

7.1. Trust and Confidence in Certificates

It is important to distinguish between the validity of a certificate and the validity of the signed data it contains. The fact that there is a signature on a certificate should provide some confidence in the correctness of its content but this does not mean that this information is invariably correct.

First, a certificate signer might make a mistake or undertake insufficient checks on the data in the certificate. A digital identity certificate might be based on a Passport but the latter might be forged and there might be large variations in the care with which this is checked.

Second, certificate-issuing mechanisms may be faulty. They could simply have errors in them or they might have been infiltrated in order to issue apparently valid certificates with false contents.

Third, it is easy to forget that where a certificate includes a signature verification key, confidence in this key depends not only on confidence in the certificate but also confidence in the way the associated secret signing key is stored and used.

For identity certificates there are subtle problems of what constitutes identity and how this can be established. In most current identity certificates, a key will be linked to a person or an object by using a name – a string of characters – and this involves many pitfalls. If such names are not unique, then the fact that a certificate is valid will not necessarily mean that the associated signature key is valid because it might be associated with the wrong person.

In practice, the certificate issuer and a verifier will each know and use a set of names but there is no guarantee that these names are associated with the same people. The 'John Smith' known to the issuer might not be the same 'John Smith' that the verifier knows.

Worse still, the extra data that the issuer uses to construct unique names – for example, a name augmented with a date of birth – may not be available to the verifying party. Thus when a verifying party requests certificates to verify John Smith's signature they may receive a number of completely trustworthy certificates, none of which they can trust because they have no way of knowing which one to use! If they then proceed to use the one that they deem most likely to be correct, they have injected a human guess into what was described and relied upon as a cryptographically strong process.

Certificates linking keys with biometric data can overcome some of these difficulties but the infrastructure required to generate and check such data is considerable and this means that the widespread use of these approaches is still some way away.

It turns out, therefore, that handling identity is much more difficult than might first be thought. However, careful analysis shows that it is often not needed in many circumstances where we might expect it to be required.

Of course, one aim of certificates is to make such associations stronger but, as indicated earlier, certificates alone cannot accomplish this task. Moreover, if we are faced with a key of unknown ownership, we may have to access a certification infrastructure to find out more about its ownership. But once we have done this we not only have to trust things that are local to us but also the design, the implementation and the operation of this infrastructure.

Moving from trust in the owner alone to trust in a possibly large infrastructure as well is a very considerable step and one that introduces significant new security vulnerabilities. It is hardly surprising, therefore, that experts have been looking in detail at applications in order to discover whether these can be engineered in a way that avoids the need for this large infrastructure. There is now a growing belief that many digital signature applications can be successfully engineered without such facilities.

7.2. Trust and Confidence in Secret Signature Keys

With all the interest in Certificate Authorities (CAs) and Trusted Third Parties (TTPs) it is easy to forget that digital signatures involve much more than certificates alone. Certificates deal only with confidence in the public components of public/private key pairs and give no immediate basis for confidence or trust in the other end of the chain – the secret signing key. It is possible to have a perfect CA infrastructure but achieve nothing of value because there are no equivalent guarantees about the environment in which the associated secret signature keys are being stored or used.

In order to trust a digital signature we must not only be confident about the ownership of the verification key but also confident that this owner is the only person who can use the related secret signing key. There are a number of concerns here:

- if an owner is careless with their signing key they might reveal its value to others;
- their computer may not be very secure and their key might hence be stolen without their knowledge even if they have been careful;
- the computer might include software that will sign with the user's securely held key, but under direction from some other entity and without the user's knowledge;
- the computer might be accessed physically by some person other than the intended user, after the user had enabled the signing key;
- a user might sign something and then regret this later; they could then invalidate their own signature by deliberately publishing their secret signing key. This cannot be distinguished from an honest case of theft of the secret signing key and the user should not be punished for something he cannot have prevented;
- digital signatures might be forgeable without access to secret signing keys.

These requirements in combination present enormous challenges that many consider well beyond the current 'state of the art'.

Software based mechanisms for key storage will not offer the level of protection needed and this means that hardware tokens of some form will be necessary [3]. Although smart-cards spring to mind, these are not yet powerful enough to support key generation 'on the card' as well as being vulnerable to attacks that can reveal secret data value. If the key values are generated elsewhere and loaded onto such cards, the very fact that these values exist in more than one place adds opportunities for illicit duplication and signature forgery.

In an ideal situation, a signature key pair would be generated on a hardware card and the public component would be made available externally through the card interface. The secret signature key would remain in protected storage on the card and never ever leave it so that not even the card owner knows its value. When a signature is required, the data to be signed would then be passed to the card for signature in order to avoid the need for the secret key to leave the card.

Unfortunately this means that the card processor needs to be quite powerful and it also requires a card interface that supports a high data rate so that signatures on large data items can be achieved without undue delay. This combination of features is too challenging for current smart cards. Alternative formats, for example, PCMCIA cards, can support these requirements but they are relatively expensive, especially when the whole infrastructure requires modification in order to employ them.

Moreover, while liability issues and legislation in respect of Certificate Authorities has received much attention, the equivalent issues in respect of the risks and vulnerabilities of the related signature mechanisms have received little if any coverage. This is surprising since many believe these risks are more likely to occur and more difficult to overcome.

Preventing users from repudiating their own keys is an especially challenging requirement that is not likely to be possible for some time. In consequence, practical applications of digital signatures must remain effective in the presence of signature repudiation.

The loss of non-repudiation does not imply that digital signatures are useless. Credit cards offer their users the chance to repudiate line items on the monthly bill but that does not mean that credit cards are useless. In addition, house keys and car keys carry no notion of non-repudiation but remain quite useful as access control devices.

7.3. Trust and Confidence in Signing

In a conventional signature a person sees the document they are signing and uses a pen to make their signature. In the case of digital signatures, the process is much more complicated. A user will have their signature capability in some form, for example, a

smart card and will have to apply this to digital data of some kind. They might be shown an electronic document on a screen and then be asked to insert their card so that this document can be signed.

But how do they know that the system used their signature on the document on the screen and not on some other document? How do they known that their secret signing key was not stored so that their signature could be forged later? And, if this is a 'point of sale' terminal, how do the customer or the trader know that the terminal has not been designed or modified to defraud either or both of them?

The process of signing something digitally is enormously complex compared to a conventional written signature and this means that confidence in the process will be a great deal more difficult to achieve. The same applies to proving validity of the process in a court of law, should the occasion ever arise.

The checking of signatures on certificates is an important part of any signature verification process and this means that 'signature verification terminals' will need to have access to a number of signature verification keys. As already pointed out, there is always a key at the end of a set of certificates that cannot be signed and hence at least one key that has to be made trustworthy in some other way. All signature verification stations will hence generally contain at least one trusted key that is used to complete the last links in certificate chains. For example, a 'point of sale' terminal may incorporate master signature verification keys for the credit card companies whose cards it recognises.

Such keys can be obtained direct from the credit card companies so we know that they are genuine but how do we know that someone has not modified their values since the terminal was manufactured? The terminal manufacturer could sign these keys and have a signature verification key in the terminal that is used to verify them, but we now need to know that their verification key has not been modified.

What this means is that there has to be at least one verification key on the platform that we can be confident has not been modified since manufacture and which provides an 'anchor' for trust in the terminal. Since this has to be very trustworthy, it has to be built into the terminal in a way that its value cannot be easily changed.

Even if we are confident in the terminals handling of signature verification, there is much more that we need to have confidence in. As already indicated, we need to be sure that what is on the screen is actually what is signed and we need to be sure that secret signing keys are not compromised in any way. There is, hence, a great deal more to the security of the terminal than just certificates. In particular we have to be sure that the hardware and software does exactly and only what we expect of it, and that it has not been modified after such a determination has been made.

We also need to know that the physical computer was being operated by the appropriate person, especially at the time of signature, and that its software was only the approved set and that every member of that set had been built in a trusted fashion and not been modified since.

Compared with 'pen on paper', digital signatures hence involve complex and highly opaque signing processes in which it will be much more difficult to develop trust. For this reason alone, it seems most unlikely that people will be prepared to commit rapidly to use such signatures at a personal or consumer level unless some other party underwrites the many risks involved.

8. Digital Signatures and Certificates in Electronic Commerce

The development of electronic commerce depends on many factors of which the security of the supporting infrastructure is just one. While, therefore, the provision of a secure approach is a necessary step in the development of this market, it is a far from sufficient one. However, the aim here is to look at this aspect in isolation, recognising that it is just one part of a much larger picture.

8.1. Consumer Concerns and Expectations

In large measure, consumer attitudes to security in electronic commerce are determined by coverage in the media. Electronic commerce depends on the use of electronic

networks such as the Internet and here nearly all publicity in respect of security is negative. This leaves consumers with the impression that they will be in great danger if they use it to make electronic transactions. The reality is very different and of the billions of transactions that take place each day all but a minute fraction complete with no security problems.

In practice by far the largest security risk in using the Internet for electronic commerce is the extent of the trust that can be placed in the other party (or parties) to any transaction. In comparison, risks arising from the use of intervening infrastructure are small and these can be removed by using modern cryptographic protocols that operate on an 'end-to-end' basis.

In the use of credit cards, traders do not have to worry too much about the signatures of consumers since their confidence in a transaction is built more on the use of the card than it is on the identity of the consumer. Although the consumer benefits from a guarantee against financial loss, they will often wish to trade with a known company and not another masquerading in their place. This means that certification of the trader's signature can be of interest to the consumer, although even here the primary concern is likely to be that of avoiding financial loss if things go wrong.

An important feature of digital certificates is that they state trust relationships and allow inferences to be made but they do not create or introduce trust where this does not already exist. For example, if I trust the British Medical Association (BMA) and the BMA trusts that 'X' is a doctor, then I can infer that I too should trust that 'X' is a doctor. However, my trust relationship with 'X' is not one that has introduced any new trust, it is simply a restatement of a chain of trust that already exists in a more extended form.

In practice, most trust relationships are made locally and the reason for this is evident since trust is a property of relationships between people and it is simply not sensible to trust someone who is completely unknown. Thus while standardisation bodies and governments have focussed on 'top down' certificate hierarchies to underpin the use of public key cryptography, private users have been much more inclined to collect the keys of their colleagues in 'personal address books'.

For these same reasons, most trust relationships that matter in commerce are forged in closed rather than open form. The relationships between banks and their clients are often seen as a model for trusted third party services but customers are more likely to see these as closed two-party arrangements in which banks provide them with banking facilities.

Although digital signatures may eventually generate new patterns of trust, experience with other technologies suggests that early exploitation will be in augmenting existing trust relationships and in improving their efficiency and effectiveness.

8.2. Trusted Parties

It is worth running through a typical electronic transaction using digital signatures in order to see what is involved.

If a client seeks a credit card, he or she will expect the card issuer to undertake a number of checks before they decide whether they will provide a card. If a card is provided, the client will be expected to sign the card but the company will not seek independent verification that this signature is truly that of the client in question. Card issuers don't do this for conventional signatures and there is no immediately obvious reason why they should behave differently when digital signatures are involved. Importantly, the relationship between the client and the card issuer is a closed, two-party one; it can hence be governed by a contract that includes the terms and conditions for the use of the card and any related digital signatures.

In the same way, the credit card company will also have a different, closed, two-party contract with traders specifying the terms and conditions for handling credit card payments. Again, therefore, there is a closed contract that can be used to set out how digital signatures are to be used.

This example shows that consumers and traders do not enter an open three party relationship with the credit card company but instead separate two-party relationships in

closed form. There is hence no need for an open CA infrastructure to support the use of digital signatures in this type of transaction since all the relationships involved are closed ones that can be managed by the credit card company or banks involved.

A major role of the credit card company here is to provide guarantees for both the consumer and the trader in respect of the financial obligations of the other. One way of operating this guarantee would be to use a digital certificate issued by the credit card company for the consumer's and the trader's digital signatures. When these parties relied on the digital signature of the other party, the certificate for this signature would give them confidence that their financial interests are being guaranteed. This is a use of a certificate for carrying a guarantee.

8.3. Financial Guarantees

In order to promote electronic commerce a new form of guarantee could thus be provided if credit card companies were prepared to issue certificates for the digital signatures of their customers and thereby enable them to purchase goods with the benefit of the same guarantee of payment as is provided by a cheque guarantee card.

This superficially resembles an open "three party" relationship, but really involves two separate, closed two-party relationships. Moreover, the third party is not certifying digital signatures in an 'identity' sense but instead using them to enable keyholders to confer the certifier's guarantee.

It is important to note that the digital signatures used here are secondary to the trust relationships that are involved. They are obviously not essential to these relationships since the latter are precisely those used in existing credit card transactions. Digital signatures are hence being used to implement and extend an existing set of trust relationships rather than to introduce something entirely new.

8.4. Professional and Service Promises

A number of professional bodies are already considering the introduction of digital accreditation certificates for their members. Examples of regulated professional groups who might wish to do this include:

- The Medical Profession
- The Legal Profession
- The Accountancy Professions
- The Chartered Engineering Professions

In these cases, the certificates give their holders permission to practise as members of the profession concerned.

This form of certificate could also be used to underpin the sort of promises that some commercial organisations offer. Trading communities often subscribe to organisations that regulate their members in order to provide public confidence in their services (for example, the Association of British Travel Agents). Such bodies could use digital signature technology to provide certificates to their members to enable them to trade electronically under their terms and conditions.

Organisations considering the use of digital accreditation certificates will have the choice of operating their own certificate authorities or, alternatively, acquiring such services from certificate service providers.

Note again that certificates are being used to give keyholders permissions of some kind, for example, permission to practise as doctors or lawyers. In such cases the certifying body will normally be the professional body that regulates the profession or the activity in question. Again, therefore, digital signatures are being used to represent an existing set of trust relationships in electronic form.

8.5. Identity Certificates

Governments have always played a role in identifying their citizens using such documents as birth certificates, passports and drivers licences. It would hence be

realistic to expect governments to extend this traditional role into the electronic world through the issue of electronic identity certificates. For example such certificates, implemented in smart card form, might replace passports and could provide a much easier and more convenient way of implementing border controls.

It is often incorrectly assumed that identity certificates are needed to underpin electronic commerce. However, much of Electronic Commerce is about handling the transfer of money electronically and this does not depend on identity information in any direct way – what has to be certified are authorisations to transfer money between accounts. Although the financial institutions will have information on the presumed identity of account owners, the actual processes involved in an individual electronic transaction do not depend on access to this information.

This is not to say that identity information is unimportant in Electronic Commerce. For example, consumers will want to be confident about the identity of companies they are dealing with electronically before pursuing transactions with them. Moreover, the value of identity here will be that it brings with it a wider set of qualities (reputation) in which consumers have a strong interest.

8.6. The Characteristics of Digital Signature Certificates

All the above examples of digital signature use have important features in common.

First, in each of these applications, digital signatures are used to support pre-existing trust relationships. These relationships already exist, they are not new, and the value of digital signatures is in extending these relationships into cyberspace or improving the way in which they can be implemented.

Second, a certificate containing a signature verification key gives the keyholder permission to undertake action or obtain services of some kind:

- **financial**: permission to trade under the terms and conditions provided by the company that issues the certificate;
- **professional**: permission to provide the service regulated by the organisation that issues the certificate;
- **identity**: permission for the holder to receive the benefits provided by the UK government for UK citizens.

Third, and most important, in all cases the certifying agency provides something that relying parties see as valuable.

In the financial example, the certificate issuer is fulfilling a role analogous to that of a credit card company in providing a guarantee to each party that it will meet the responsibilities of the other party if they default.

In the professional and services example, the certificate issuing organisation is putting its professional reputation behind a member of the profession or the organisation concerned and may also be underwriting this with a promise of compensation if things go wrong.

In the identity example, this might be issued by one UK government agency and used by another. This will occur, for example, when a UK citizen seeks benefits from the UK government as a UK citizen using a mechanism based on digital signatures.

If another government were to rely on a ‘digital passport’ at a border check, they might then know that the holder is a citizen of a specific country with which they have an extradition treaty. They might hence have confidence that they can hold the person in question to account if it is found after they left that they had acted unlawfully. In practice, they cannot have complete confidence in this because the digital passport may be false; moreover, the citizen might not reside in their country of citizenship. In any event, it is very unlikely that the issuing country would offer any recourse if the passport proved to have been issued erroneously. However, established practice shows that most passports are valid, most governments co-operate in upholding the law, and this means that they can reasonably expect these things to be true although there are no absolute guarantees involved. A digital passport would also be an example of a certificate where the

requesting party and the relying party are the same since a person returning to their country of citizenship is relying on their passport for re-admission.

In these examples, the relying parties obtain somewhat different benefits from the certificate issuer. In the first, they obtain strong guarantees that they will not be put at risk in the digital signature based transactions involved. In the second, they have the confidence that they are dealing with someone who is backed by a professional or trade organisation, and one that might also offer a promise of compensation if things go wrong. In the third example – the 'digital passport' – the other government is not likely to be compensated if things go wrong but the passport is likely to be valid and they can reasonably expect any required extradition request will succeed. In this case, therefore, the value is not the compensation in the small number of cases when the digital certificate fails but rather in their expectations for the large percentage that are correct.

Thus, while some individual certificates might be incorrect, this does not diminish their overall value provided that a high percentage prove to be valid. Hence the value of certificates is not the result of their absolute correctness but rather the result of the a high probability that any certificate at random will be correct.

The concept of a digital certificate as endowing the keyholder of the embedded key with some form of permission is at the heart of digital signature applications in commerce and more widely in society. Although this is quite different to the original concept of associating identities and keys, it turns out to be much more powerful.

9. Third Party Digital Signature Certification Authorities

Some governments are advocating the concept of licensed 'third party digital signature certification authorities' – Certification Authorities (CA) for short – as essential enablers for Electronic Commerce. The idea is that such companies will offer their clients digital signatures that are certified by these companies as belonging to these clients.

The problem with this approach is that it is modelled on the use of digital signatures for 'identity' rather than 'empowerment' and this is not the most effective model for electronic commerce applications.

Considering professional organisations, and using certified doctors' digital signatures as an example, what would a third party CA contribute in comparison with the General Medical Council or the British Medical Association? They could never match the professional expertise of the latter bodies in a direct certification role and could not expect to certify doctors' digital signatures to the same level of professional assurance. Hence, while signature certification controlled directly by professional bodies might reduce the probability of certifying bogus doctors' signatures, it is hard to see this as the result of a third-party CA operation. Faced with using a medical prescription digitally signed with a signature certified by the General Medical Council or by 'The Acme Signature Company' it is not hard to see which of these most people would choose.

For certification to be worthwhile, the certificate issuer should have something of value to offer in signing a certificate. In the credit card example, the certificate issuer is providing a guarantee while in the professional/organisational example this is a promise of performance with, possibly, a promise of compensation for error.

To be commercially successful, third party CAs will have to provide 'value added' and about the only service so far considered here is that of assigning names to keyholders. In practice, however, it is far from obvious that this is a service that commerce requires since it has done quite nicely without this for a considerable time. The financial institutions that underpin modern commerce already have well-established ways of identifying their clients and it is far from obvious that they would wish to employ an outside party to perform (and charge for) such a service. Indeed if such services are worthwhile it seems far more likely that these institutions will offer such services rather than act to promote their emergence as another force in the market.

In effect this means that companies offering stand-alone digital signature certification services will have to become an additional trusted party to these already existing trust relationships without actually bringing anything of great value to the table. In practice, it

is hard to see this as an important service and certainly not one on which electronic commerce will depend.

A further concern here is that the focus of attention on this form of certification is creating uncertainties about the way forward in both legal and technical terms in respect of digital signature use. Because of this, it is proving more difficult for organisations to justify investments in digital signature applications in areas where they are more likely to be effective. In consequence, the focus on certification for identity based digital signatures seems more likely to be hindering rather than promoting the development of electronic commerce.

10. Technical Standards for Digital Signatures and Certificates

Considerable fragmentation and significant interoperability problems could result if many different technical mechanisms emerge for implementing digital signatures. However, attempts to standardise on particular algorithms or mechanisms will involve the danger that there could be a large investment in a specific solution that might later prove to have serious weaknesses.

Because of this, it is very important that any legislation for digital signatures does not depend in any way on the specific technology used for signature implementation. If such a separation is maintained it will then be possible for any technology that subsequently proves ineffective to be replaced without undermining the legislation concerned. This is also true of applications, which should be written so that dependency on particular signature mechanisms is avoided.

The importance of avoiding such technological dependence has been well illustrated by one recent event: the MD5 algorithm, which has been widely used in the creation of digital signatures, has been found to contain a flaw that can sometimes allow signatures to be forged. Consequently, this algorithm is no longer recommended but, had it been specified as a mandatory standard, the whole basis for digital signatures would have collapsed.

In fact, it is important that applications that include digital signature features offer users a choice of several different signature mechanisms. If only one such mechanism is available, and this subsequently proves to be faulty, then the application will fail completely. If, however, a number of choices are available, one of these alternatives can be bought into immediate use.

At first sight avoiding a requirement for the use of specified signature mechanisms might appear to create market fragmentation. In practice, however, openly available, high quality algorithms are very difficult to design and this means that only a relatively small number of algorithms will find widespread use.

Another reason for avoiding mandated algorithms is that different applications will often require different mechanisms. For example, implementing a digital signature mechanism in software or in a smart card will involve different considerations and this can lead to the use of different algorithms.

At the algorithm level, therefore, it is important to avoid the temptation to standardise on just one or even a small number of algorithms. Experience suggests that the dangers of too much diversification are unlikely to arise and the benefits of diversity are such that this is an important requirement in its own right. Although algorithm standardisation should be avoided, there are still requirements that should be met by any algorithms that are used to create digital signatures. For open digital signatures it is evidently important that all the parties using the signature should know the algorithm employed and this requires that it should be published in an openly available form. Although not an absolute requirement, it is also highly desirable that there are no royalties or licence fees involved in its use.

More importantly, open publication is essential in order that society as a whole can develop confidence in such mechanisms. The need to use published and widely available digital signature algorithms is important in order to ensure that they have been widely scrutinised. Cryptographic algorithms are very difficult to design and this means that even experts can make subtle mistakes in their design that are often very difficult to

discover. By ensuring that any algorithms used are subject to publication and open international review we can help to ensure that such weaknesses are identified and eliminated.

A further practical requirement is to use algorithms that are designed to provide digital signatures and not confidentiality6. The reason for this is both political and technical. In technical terms, signature and confidentiality uses require different algorithm properties. In policy terms, many governments impose restrictions on the provision of confidentiality and this will constrain the use of any digital signature technology that can also directly support confidentiality.

10.1. Digital Certificate Standards

X.509 Certificates

This standard is designed around a link between a digital signature key and a name in an X.500 directory that is hopefully sufficient to identify a person or an entity. It also embodies a capability for certificate extensions and these can be marked as critical or non-critical so that any extra features that they offer can be controlled. The aim here is to ensure that a verifying party will not accept a certificate as valid if there are critical extensions present that it cannot interpret. In contrast, a non-critical extension is not essential for validating the certificate and any inability to interpret it does not necessarily make the certificate invalid. More generally X.509 allows fields for which there is no universal definition of the semantics involved.

Although X.509 is widely used, it does have some potentially serious weaknesses including:

- the assumption that it is always necessary to link a key to an identified person or entity;
- the assumption that, even when such a link is appropriate, it is always possible to uniquely identify the person or entity involved.

The problem here is that X.500 presumes the existence of a global directory structure in which every entity that needs to be identified can be traced somewhere within its hierarchy. In the real world, however, things are not so simple and there are many situations where this will not be possible or even desirable – the Central Intelligence Agency, for example, is unlikely to add the names of its employees to an open X.500 directory hierarchy.

However, a more serious problem is that the link between names and keys is much less important than it at first appears. As discussed earlier, a certified digital signature is much better viewed as a permission for the keyholder to use the key for a specific purpose.

New Certificate Research

It is now being increasingly recognised that unique global names are not necessary to support certificates. For example, most people employ more localised ways of identifying those with whom they interact through direct meetings, through their immediate colleagues or using their (or their companies') address books.

In particular the improvements that result in considering certificates as empowerment mechanisms for digital signatures provide for direct links between keys and permissions without the complications that are involved in links to names and identities. As we have already seen in the examples given earlier, many real world examples map more easily onto certificates that empower keys rather than those that seek to link keys with names. Signatures represent the powers given to their owners to undertake actions so it very often makes sense to bind these actions to signature keys rather than infer the power of a signature owner through their identity.

For these reasons newer certification models such as SPKI and SDSI have been developed to be especially appropriate for access control applications, including

[6] In principle signature algorithms can be used for confidentiality but it is possible to make the incorrect use awkward and inefficient.

electronic commerce [4]. In addition, the problem of remote execution of complex policy is addressed by PolicyMaker [5] certificates and there is no reason to believe that research into the nature and form of certification has stopped with these developments.

11. Conclusions

Digital Signatures have potential uses in Electronic Commerce but attempts to apply them in ways that mirror written signatures are unlikely to be effective because this analogy is misleading.

'Trusted Third Parties (TTP)' as Certificate Authorities for digital signatures are often justified using an analogy with the role of the financial institutions in conventional commerce. In practice, however, this seems to be based on a misinterpretation of what these organisations provide since the trust relationships involved are only superficially three party ones. When analysed in more detail they are most often seen to be sequences of closed two-party relationships that combine to give the appearance of a relationship involving three-parties.

Because of this, it seems unlikely that open digital certificates have a significant role in Electronic Commerce. It also turns out that digital certificates are more effective as mechanisms for attaching permissions to digital signatures instead of names or identities (as the analogy with written signatures leads us to expect). And these properties in combination lead to uses of digital signatures, not as vehicles for identity, but rather as mechanisms that can represent the closed trust relationships on which commerce depends. Identity based digital signatures and the associated Certification Authorities have little immediate relevance in the development of Electronic Commerce. Put in the simplest terms, they are unnecessary for this purpose and seem more likely to delay the emergence of an electronic marketplace than they are to promote its development.

Certificates and Certificate Authorities – mechanisms for the management of the public key components of key pairs – have been the subject of much attention and debate. In contrast, the management of the private, signature key, components of key pairs has not been given the attention it deserves. This is surprising since many believe that the technical and legal issues involved are more severe. In particular, for written signatures, the relying party carries the burden of proof of authenticity. However, UK government proposals for licensed digital signatures propose to reverse this situation by placing the burden of rebuttal on keyholders. Such a step will bring very significant risks for consumers since the technology currently available for private key management is simply not capable of the security required to support such a profound change.

12. References

[1] Loren M. Kohnfelder, "Towards a Practical Public-key Cryptosystem", May 1978, p.15.

[2] Ross Anderson et al, "The Global Trust Register", Northgate Consultants Ltd, 10 Water End, Wrestlingsworth, Sandy, Bedfordshire Sg19 2HA, United Kingdom.

[3] Peter A. Loscocco, Stephen D. Smalley, Patrick A. Muckelbauer, Ruth C. Taylor, S. Jeff Turner, John F. Farrell, "The Inevitability of Failure: The Flawed Assumption of Security in Modern Computing Environments", National Security Agency.

[4] Carl M. Ellison, "Infrastructure Needs For Electronic Commerce and Personal Use", CyberCash Inc.

[5] Matt Blaze, Jane Feigenbaum and Jack Lacy, "Decentralised Trust Management", AT&T Research.