Global Co-Operation in the New Millennium
The 9th European Conference on Information Systems
Bled, Slovenia, June 27-29, 2001

THE FUNDAMENTAL INADEQUACIES OF CONVENTIONAL PUBLIC KEY INFRASTRUCTURE

Roger Clarke

Principal, Xamax Consultancy Pty Ltd, Canberra,
Visiting Fellow, Department of Computer Science, Australian National University
Roger.Clarke@xamax.com.au

ABSTRACT

It has been conventional wisdom that, for e-commerce to fulfil its potential, each party to a transaction must be confident about the identity of the others. Digital signature technology, based on public key cryptography, has been claimed as appropriate means to achieve this aim. Digital signatures do little, however, unless a substantial 'public key infrastructure' (PKI) is in place, such that parties know what is being authenticated, and what level of assurance is provided. Conventional PKI, built around the ISO standard X.509, has been, and will continue to be, a substantial failure. This paper examines conventional X.509v3-based PKI architecture, and identifies key deficiencies including its inherently hierarchical and authoritarian nature, its unreasonable presumptions about the security of private keys, a range of other technical and implementation defects, confusions about what it is that a certificate actually provides assurance about, and its inherent privacy-invasiveness. A model is presented that explains the naiveté of identity authentication, and how e-commerce needs to be baaed on 'nyms' rather than 'identifiers'. Alternatives to conventional PKI are identified.

1. INTRODUCTION

There has been a popular perception that the adoption of e-commerce has been significantly slowed because, in cyberspace, buyers don't trust unidentifiable sellers. Digital signatures, and the mechanism that supports them, Public Key Infrastructure (PKI), have been touted as the solution to the problem. Despite well over a decade of development, however, very limited progress has been made, and each step forward with PKI seems to create a set of new sub-problems.

Meanwhile, a range of other impediments to net-consumer trust of cyberspace merchants has been identified (Clarke 1999c), and PKI has been criticised on both technical grounds (e.g. Davis 1996, Ellison and Schneier 2000, Schneier 2000), privacy grounds (e.g. Greenleaf & Clarke 1997) and commercial effectiveness (e.g. Winn 2001). This paper consolidates the critiques of PKI, with the intention of working towards digital authentication mechanisms that are more attuned to what the Information Society really needs.

The paper commences by stating the trust problem as it was originally perceived, and describing the currently conventional technology that has been applied in an endeavour to solve it. Major problems with that solution are then identified, in the areas of its hierarchical nature, insecurity of the private key, technical and implementation deficiencies, its failure to provide useful assurances to net-users, and its privacy-invasiveness. The paper concludes with an explanation of the critical nature of 'nyms', and a brisk

assessment of alternative approaches to achieving trust which offer better prospects for meeting the real needs of the Information Society.

2. THE PERCEIVED NEED

The commercial potential of the Internet became apparent only in the mid-1990s. Wired Magazine, launched in October 1994, claimed that its Hotwired venture was the first commercial web-site (Clarke 1999c), although Pizza Hut has also staked a claim to that mantle (Hobbes 1990-).

From an early stage, the conventional wisdom was that e-commerce, in comparison with purchasing in a physical location like a shop, lacks the important comfort factors of seeing who you're dealing with, or at least being able to see the merchant's physical 'foot-print', and check the physical attributes of the value being transferred. It was therefore postulated that successful commerce on public networks would be dependent on some other means of establishing trust.

A leap was then made to the conclusion that trust would need to be based on a mechanism for the identification of parties who deal on the net, supplemented by authentication mechanisms to test the assertions of identity. A recent expression of this is that "Fundamentally, electronic commerce involves the use of remote communications and therefore necessitates all parties involved to authenticate one another ... [because] the parties will not at the time of transacting have face to face dialogue" (McCullagh A. & Caelli, 2000).

Moreover, the demand for identity was presumed to be two-sided, i.e. not only would the merchant or service-provider identify themselves to the consumer but consumers would also identify themselves to sellers. It is unclear whether this was a conscious assumption, and if so whether it was based on an analysis of merchant behaviour, or was merely a pretext for the creation of exploitable trails of consumer behaviour. Either way, it represents a significant compromise to the freedom of consumers who have hitherto conducted most of their purchases anonymously.

3. CONVENTIONAL TECHNOLOGY

This section provides a brief overview of the key technologies that have enabled engineers to address the perceived problem described above.

During the 1980s, public key (or 'asymmetric') cryptography had emerged. Public key cryptography involves two related keys, referred to as a 'key-pair', one of which only the owner needs to know (the 'private key') and the other which anyone can know (the 'public key'). Because only one party needs to know the private key, it does not need to be transmitted between parties, and hence it need never be exposed to the risk of interception. Knowledge of the public key by a third party, on the other hand, does not compromise the security of message transmissions (Diffie & Hellman 1976, Schneier 1996). For a tutorial treatment, see Clarke (1996), and for a short history see Ellison, in RFC2963 (1999).

The following sub-sections introduce firstly the application of public key cryptography to 'digital signatures', and then the infrastructure on which they depend. The dominant standard is then outlined and interpreted.

3.1. Digital Signatures

Digital signatures are a particular application of public key cryptography. A digital signature is a block of data that is generated from a message prior to its despatch, and is appended to it. The block is prepared by a two-step process:

- a 'message digest' is created by processing the actual message using a pre-agreed one-way hash algorithm; and
- this message digest is encrypted with the sender's private key.

The recipient re-creates the message digest from the message that they receive, uses the sender's public key to decrypt the digital signature that they received appended to the message itself, and compares the two results. If they are identical, then, so cryptographers argue:

- the content of the message received must be the same as that which was sent (assuring message content integrity);
- the message can only have been sent by a sender that had access to the private key (providing a means of authentication); and
- the sender cannot credibly deny that they sent it (addressing the need for non-repudiability of messages).

This paper concerns itself with only the second of these, the use of a digital signature to authenticate something about the message-sender.

Digital signatures were naively presumed by many people to provide unqualified assurance. In practice, however, the effectiveness of the mechanism is dependent on a number of conditions, in particular:

- a third party must have checked that the private key is in the possession of the appropriate party;
- that third party must be trustworthy;
- the private key must be subject to strong security measures, such that no other party can ever gain access to it or invoke it;
- the public key used must be the appropriate one, and not one provided by an imposter;
- a significant number of infrastructural elements must all be in place and functioning effectively, and their security not compromised;
- means must be established of discovering when a private key has been compromised, of issuing notices revoking keys and associated certificates, and ensuring that revocation is rapidly and reliably transmitted to all who need to know about it, without generating vast network traffic, access contention and slow service; and
- legal infrastructure exists, that appropriately distributes risks, and protects data.

As this paper explains, those conditions are generally not fulfilled by conventional PKI.

3.2 Public Key Infrastructure

Digital signature schemes depend on the public key of the message-sender being available to the recipient. The most practicable methods of achieving this are:

- senders can include their public keys in each message;
- senders can store them on a site of their own that is readily accessible (e.g. using FTP or HTTP); or
- public keys may be stored in one or more centrally managed directories, enabling each party to an exchange to look up the public key of the other party.

All of these approaches are subject to 'spoofing', i.e. an imposter can send a message that includes a public key, or store a public key in a readily accessible directory, and thereby fool the other party into thinking the message came from a particular person or organisation.

To address this risk, the concept was created of a 'certificate' that attests to the fact that the particular public key is associated with a particular party. (The technical literature uses the term 'is bound to' rather than 'is associated with'. Many readers would infer from that term a far stronger form of association than the technique actually warrants).

More precisely, a 'certificate' is a digitally signed, structured message that asserts an association between specific data and a particular public key. An 'identity certificate' is then a particular class of certificate that associates a particular identifier with a particular public key. (It will be argued later in this paper that the term 'identifier' should really be replaced by 'nym'). Regrettably, most of the literature uses the term

'certificate' ambiguously, to refer to both certificates generally and identity certificates in particular, despite the fact that the differences are extremely important.

According to conventional thinking, a certificate needs to be created by a trusted 'public key certification authority' (CA). A CA digitally signs each certificate using its own private key. In most schemes, the certificate is provided to the party that claims the particular key to be its own. That party then includes it in the messages that they send. A message with a CA's certificate attached therefore functions in a manner analogous to a letter applying for a job being accompanied by a letter from a referee attesting to something about the applicant, such as their identity, their good character, their experience, or their qualifications.

A CA needs to undertake some form of authentication process in order to satisfy itself that the claimed association actually exists. A conventional approach is to depend on the services of a Registration Authority (RA), such as a Post Office. A comprehensive process would require the person with whom the key is to be associated to undertake all of the following:

- present themselves at the RA's premises;
- provide physical evidence of the characteristic claimed. For an identity certificate, this would typically involve 'photo-id', and documentary evidence of (for example) age, qualifications and/or professional membership, supported by a documentary trail evidencing the use of the relevant identifier(s) over a period of time (including, for example, marriage certificate or deeds poll);
- provide the public-key;
- provide evidence that they are the holder of the private-key (e.g. by signing a message in the presence of the RA);
- provide evidence that they have the private-key secure;
- nominate a contact-point; and
- nominate a delivery-point for the certificate.

Such procedures are highly inconvenient, intrusive and expensive, and the load falls on the individuals who are required by organisations to participate. As a result of the costs and difficulties, all existing schemes make very significant concessions to practicality, and thereby undermine their integrity.

The security of private keys is vital to the whole process, but is capable of being compromised. When, not if, a private key is compromised, the certificate must be revoked very quickly. An efficient and effective mechanism is therefore required to record and provide access to revocations of certificates and the associated key-pairs.

3.3 The X.509v3 Standard

The dominant standard used at present as the foundation for PKI is the family of CCITT X.500 standards, in particular X.509 (X.509 1988, 1997, and Housley et al. 1999). The current version of X.509 is number 3, usually referred to as X.509v3, which was finalised in 1997. A set of standards, dubbed PKIX, enables use of X.509 approaches within the web-context (W3C 2000). Guidance has been provided by texts such as Ford & Baum (1997), Adams & Lloyd (1999) and Austin (2000).

Ellison (1997) describes the history this way: "the X.500 proposal was published [in the late 1980s]. It was to be a global directory of named entities. To tie a public key to some node or sub-directory of that structure, the X.509 certificate was defined. The Subject of such a certificate was a path name indicating a node in the X.500 database – a so-called 'Distinguished Name'. The X.500 dream has effectively died but the X.509 certificate has lived on. The distinguished name took the place of a person's name and the certificate was called an 'identity certificate', assumed to bind an identity to a public key ...". In short, X.509 was the hammer that came to hand when the nail was discovered.

All forms of PKI necessarily involve some degree of intrusiveness, in order that sufficient quality can be achieved. Conventional PKI, built around X.509v3 certificates, is especially severe. Implementations commonly have many of the following features:

- a single key-pair per person;
- a 'distinguished name' that is unique across a name-space that is in principle vast, and in practice large, and that denies the opportunity for pseudonyms;
- a certificate that expressly claims to 'bind' the key to a person;
- little or no choice in the manner in which the key-pair is generated;
- in many cases, generation of the key-pair outside the control of the person concerned, with the result that the private key is *ab initio* in someone else's possession;
- issuer-ownership of the key-pair, with individuals merely licensed to use it;
- little or no choice as to what token (such as a diskette or chip-card) is used to store and carry the private-key and certificates;
- little or no choice as to who issues the token;
- issuer-ownership of the token, with individuals merely licensed to use it; and/or
- little or no choice in the organisation from which the individual acquires a certificate.

Current X.509v3 certificates go so far as to permit an agent of an organisation to protect their personal identity through the use of a role-title, but they actually preclude an individual (referred to as a 'residential person') from having that capability. Moreover, some implementations may preclude a residential person from possessing multiple personal key-pairs, even though the same person is permitted to possess multiple key-pairs for organisations that they represent.

Some schemes involve the key-pair generation process being compulsorily performed by some organisation on behalf of individuals, and perhaps even compulsory storage of the private key for the benefit of parties other than the individual concerned (commonly referred to as 'escrow').

X.509v3 certificates provide a limited means for communicating attributes, within the primary certificate or through the creation of secondary certificates which may attest to one or more characteristics of the individual. But the attributes are inherently linked to and dependent on the primary certificate, which bears the individual's identifier. Hence anonymity and even pseudonymity are still precluded.

The issuing of notice that a key-pair and certificate(s) have been revoked is supported by an inefficient download mechanism called Certificate Revocation Lists (CRLs - X509, 1988, 1997 and Housley et al. 1999). A more recent specification for an on-request look-up is Online Certificate Status Protocol (OCSP – Myers et al. 1999).

This paper uses the terms ‘conventional PKI‘ and ‘X.509-based PKI‘ to refer to public key infrastructure based on X.509v3 certificates, including its Internet variant, PKIX, .

4. DEFICIENCIES IN CONVENTIONAL PKI

This section presents a catalogue of problems with X.509-based PKI.

4.1 The Hierarchical Model of Trust

X.509-based PKI is inherently hierarchical. This is because trust in the CA is not automatic, and each layer of CAs needs to be attested to by some superior layer. Conventional PKI therefore depends on a third party that is partly but not entirely trusted, which in turns depends on another such partly but not entirely trusted third party, which needs to be attested to by some further superior layer. This results in an unholy spiral up to some mythical authority in which everyone is assumed to have ultimate trust. Trust in the real world has never worked like that, and trust in cyberspace won't either.

Such schemes can also be readily argued to be authoritarian in nature (Clarke 1994b). For example, there is an intrinsic assumption that every party that acquires a certificate is required to disclose their identity, even if the only functional need is to communicate eligibility (e.g. their age, qualifications, or agency relationship

with a principal). RAs have considerable power, to the extent that they are able to deny a person a digital identity.

4.2 The Identifier Associated with a Key-Pair

X.509-based PKI makes the assumption that the 'distinguished name' has to be unique within the 'name-space'. This precludes the second and subsequent individuals who seek to use a particular name (Clarke 2000b) from using their own name without some kind of qualifier. It also provides no basis for individuals to use alternative identifiers, and implicitly denies individuals the capability to have and use multiple key-pairs, and multiple certificates. The engineers who created the X.509 standard appear to have been blithely unaware that multiple identities per person are entirely legal in many jurisdictions, particularly those whose legal systems derive from that of the United Kingdom (Clarke 1994c).

4.3 Private Key Insecurity

Underlying digital signatures and PKI is the assumption that the holder of a private key will be able to ensure its security. During the 1999-2000 period, **corporate servers** have been subject to a rash of electronic break-ins. The ease with which many of these ‘hacks‘ have been performed has demonstrated the serious inadequacy of the precautions taken by organisations of all kinds and all sizes. Standards have been issued by governments (e.g. TCSEC 1985, ITSEC 1991, Common Criteria 1998), and guidance provided by text-books (e.g. Garfinkel & Spafford 1997), but the degree to which organisations have applied the principles and guidance is embarrassingly low.

Conventional PKI also assumes that consumers and citizens will have, and will need to use, private keys. There are many ways in which malicious software (malware) can be applied to discover, copy or invoke private keys, in memory or on disk. The hardware and systems software of commodity workstations, particularly mainstream Windows and MacOS machines, currently provide very little in the way of security features. Moreover, few products are available that enable consumers to graft such security features on to their work-and-play facilities, and such products as exist require considerable expertise to install and configure (Kaiser 2000). Private keys on **'commodity workstations'** that are connected to the Internet via commercial Internet Access Providers therefore remain highly susceptible to a wide array of risks, both of capture, and of invocation without the authority of, or even knowledge of, the consumer/citizen.

4.4 Technical and Implementation Weaknesses

A range of problems have been identified with the technical design of X.509-based PKI and with its implementation in real-world applications (Ellison & Schneier 2000). These include problems with the assumption that a single global **name-space** exists, the difficulty of detecting that a private key has been subject to **compromise**, many difficulties in implementing an effective **revocation process**, and the onerousness and demeaning nature of **registration processes**.

Even where solutions exist, they are commonly ignored or flouted. Hierarchical schemes are undermined by reliance on **'self-signed' certificates** by CAs, i.e. blind trust by other parties in the CA, its intentions, and its procedures. Most schemes fail to implement effective **revocation procedures**, using either the CRL or OCSP specifications. The major implementations of X.509-based PKI, such as that based on the Verisign certificates embedded in commercially-available web-browsers, are at best 'relaxed' applications of formal X.509 standards, and hence the current PKI is even less meaningful than that which would be feasible if it was applied as intended. CAs find it necessary to deflect attention from the critical weaknesses of their services by drawing attention to the physical and electronic security of the facilities that they use to generate the certificate.

In addition to all this, the X.509 standards are long, rich, complex and imprecise, with the result that **interpretations of the standard** are required, and many variants, commonly termed 'profiles', exist (see, for

example, Gutmann 2000). Commercial applications are clumsy to implement, and considerable **difficulties and delays** are experienced, even by skilled technicians, in relation to the generation of keys, the acquisition of certificates, and the management of certificates. All schemes compromise the theoretical requirements, and thereby undermine their purpose.

4.5 The Limited Assurance Actually Provided

A critical feature of PKI schemes is the warranties and indemnities provided by the CA to accompany the assurance. It would be expected that the CA would incur financial liability if the assurance that the sender was who the sender purported to be transpires to be incorrect, and a party's reasonable dependence on the assurance results in economic cost. The wording provided by web-browsers suggests considerable protection, e.g. "The signer of the Certificate promises you that the holder of this Certificate is who they say they are" (Macintosh Netscape Navigator 4.08).

Such bold assurances are, however, subject to a great deal of qualification. CAs commonly describe their procedures for associating persons with online identities in 'Certification Practice Statements', and express the commercial aspects in 'Certificate Policy Statements'. These are often phrased in ways that obscure rather than clarify. Moreover, "The certification authority may establish different classes of certificates with different prices and different degrees of scrutiny applied in reviewing the application" (Winn 1998), and the conditions are generally phrased so that they minimise the CA‘s exposure to liabilities.

In any case, the concept of 'authentication' has been seriously misunderstood by the designers of X.509-based PKI. Authentication is a process whereby a degree of confidence is established in the truth of an assertion. There are many kinds of assertions that can be the subject of authentication processes. Among them are assertions of the form 'this artefact has a value equivalent to so much of a particular currency', and 'the sender of this message has a credential that attests to their having a particular attribute, or their eligibility to perform a particular function'.

In order to discuss the real meaning of a certificate, some definitions of terms are needed:

- an **entity** is a real-world thing. A pallet, a package and a widget are examples of real-world entities that are generally not relevant in the current context; whereas a person, an organisation, and an artefact that is capable of taking a relevant action (e.g. a hardware-server, a software-server, a hardware-client, a software-client, a software agent) are of direct relevance;
- an **identity** is a defined and specific instance of a specific entity (e.g. a particular person, organisation, computer or software process, performing a particular function, at a particular point in time). At any given time, each entity has precisely one identity (but, because its experience and hence behaviour are cumulative, its identity changes over time). Neither an entity nor an identity is capable of being directly expressed as data;
- a **digital persona** is a group of data items that together form a simplified representation of an identity (Clarke 1994a); and
- an **identifier** is a data-item or group of data-items that reliably distinguishes the identity of an entity. An identity may have many identifiers, i.e. the mapping is 1:n. (Note that, in most of the literature relating to digital signatures, certificates and PKI, the term 'name' is used variously to refer to a specific kind of identifier and to refer to identifiers generally).

The kind of assertion that certificates are supposed to provide assurance about is 'the sender of this message is the entity that uses a particular identifier'. A certificate does **not**, however, attest to that. What it does attest to is that:

- a particular message was generated by an artefact that had available to it a particular private key; and
- the CA that provided the certificate has, at some time in the past, had grounds for believing that that private key was associated with a particular entity.

Depending on the registration process that was applied, a certificate may also attest that:

- the CA that provided the certificate has, at some time in the past, had grounds for believing that the entity had some kind of right to use that identifier, or had used that identifier in the past; and
- the CA that provided the certificate has, at some time in the past, had grounds for believing that the entity had access to the appropriate private key.

A certificate provides no assurance, however, about whether:

- the private key was originally available to other entities as well as the entity to which it purports to be 'bound';
- the private key is now available to other entities as well as the entity to which it purports to be 'bound';
- the private key invocation that gave rise to a particular message was performed by the entity; and
- the private key invocation that gave rise to a particular message was performed with the entity's free and informed consent.

Moreover, such assurance as a certificate provides is qualified by the terms of the CA's Certificate Policy Statement, as dictated by the CA's lawyers; and very limited recourse is available should the assurance be wrong. A relying party appears to have little or no legal protection, not just if the CA was wrong, but even if the CA was negligent (e.g. Sneddon 2000).

4.6 Privacy-Invasiveness

The previous sections have focussed mainly on technical inadequacies, but mentioned privacy in passing. Greenleaf & Clarke (1997) considered the privacy impact of conventional digital signatures and PKI. That paper categorised the wide range of threats into those involving private keys (variously during generation, storage and backup, escrow, access and revocation), and risks arising in relation to public keys (including identification requirements, registers and revocation logs). In addition, there are consequential implications such as further increases in expectations of identification, imposition of requirements to carry a token such as chip-card, and the gross privacy imposition of biometrics as a security mechanism for the private key.

Some of these problems are features of conventional PKI schemes that could be avoided or designed around. Many, however, are direct implications of the nature of the X.509 architecture and certificate design.

4.7 Conclusions

Conventional PKI involves enormous complexity, effort and expense, in return for insecure protections, very weak evidence, and very limited recourse. Both corporations and individuals, including consumers, citizens, employees and contractors (especially those in sensitive circumstances) should have serious doubts about schemes of this nature being inflicted upon them.

5. THE CRITICAL NEED FOR NYMS

The previous section argued that PKI's impacts on individuals are severe. If e-trust schemes are to serve the needs of the Information Society, the focus must be moved away from identities of individuals, and mechanisms must be at least tolerant, and even actively supportive, of anonymity and pseudonymity (Clarke 1993, 1994 and 1999). Application of these concepts is critical to ensure that the advent of cyberspace does not mean the death of private space.

The following related needs exist:

- both people and incorporated organisations need to be able to have acts performed on their behalf by human and software agents;
- people need to be able to perform actions without necessarily declaring their identity;

- people need to be able to communicate that they have particular attributes, without being forced to declare their identity in the process; and
- persistent relationships need to be enabled even if either or both parties are unidentified.

These objectives can be achieved through the application of the concept of a 'nym'. This is the pseudo-identity that arises from anonymous and pseudonymous dealings (McCullagh D. 1996-, Clarke 1999b).

An earlier section offered definitions for the terms 'entity', 'identity', 'digital persona', and 'identifier'. Three further terms require explanation:

- a **role** is a particular presentation of an entity. An entity may have many roles; and a role may be associated with more than one entity, i.e. the mapping of role to entity is m:n. A role is not capable of being directly expressed as data;
- an **agent** is a particular role performed by an entity, with the delegation of, and on behalf of, another entity, which entity is referred to as a 'principal'. An agent may act for multiple principals, and a principal may have multiple agents, i.e. the mapping is once again m:n. An agent may be a human or an artefact;
- a **nym** is a data-item or group of data-items that reliably distinguishes a role. However, because a role is not reliably related to an entity, there is no reliable mapping between a nym and the underlying entity or entities, i.e. the mapping is m:n, and is not fully determinable. Moreover, there may be a chain of nyms, because an agent may perform an act on behalf of an agent, which performs on behalf of a principal.

This gives rise to the web of concepts depicted in Exhibit 1.

Nyms are not mere imagination: technologies exist that enable them. See EPIC (1997-) and Clarke (1999a). Moreover, it is critical to the future of e-commerce that the information infrastructure supports nyms, and that people adjust to their existence and nature. As Ellison (1997) argued: "The [U.S. House Hearing] asked 'Do you know who you are doing business with?'. Before answering that question, one should really answer the two questions: 'Do you need to know who you are doing business with?', and 'Can you know who you are doing business with?'".

Nyms are in practice replacing identifiers. Services and protocols such as IRC, MUDDs and ICQ expressly support them. So do several of the alternatives to conventional PKI that are discussed below. Any approach to inculcating trust in marketspaces will need to implement persistent nyms at least for the consumer side of transactions.

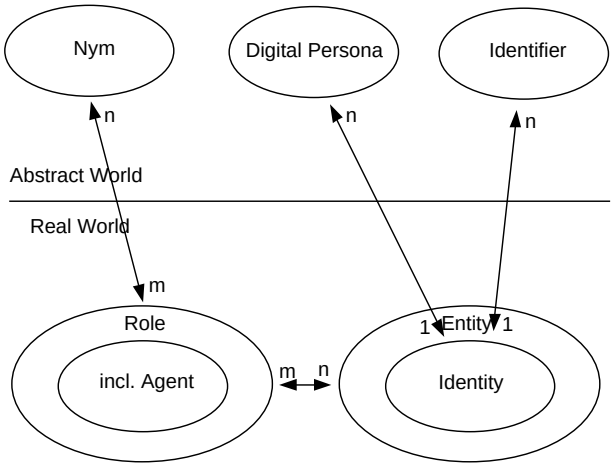

Exhibit 1: *A Realistic Model of Entities in e-Commerce*

6. ALTERNATIVE MODELS OF TRUST

Conventional PKI are ineffectual and privacy-invasive. Fortunately, there are other ways to address the need for trust in marketspaces. Their discovery depends in part on re-definition of the problem. This section briefly scans some alternative approaches that may provide better fit to the needs of the Information Society.

The **'web of trust'** approach is intrinsic to the longstanding alternative product Pretty Good Privacy (PGP) – (Zimmerman 1995, Garfinkel 1995, Bacard 1995, Stallings 1995). This avoids the need for professional CAs, because certificates can be issued by anyone. Fault-tolerance is achieved by depending on multiple certificates, probably with varying weightings assigned to them by the evaluator, on the basis of the degree of trust they place in the person who provided the certificate. PGP supports nyms. It depends on email-addresses, which are unique, because of the manner in which domain-names are allocated, and aliases and user-names are assigned. They are not formally linked to entities, however, and may have any of a 1:1 relationship with a single person, or 1:n (multiple people may share the same address), or n:1 (a person may have multiple addresses); or indeed m:n (multiple accounts may be used by multiple people). The practicality of PGP's specific implementation of the 'web of trust' notion has been criticised, but arguments have been pursued for the concept to be broadened and applied more generally (Grossman 2000).

Another standardisation process is that which grew out of Simple Public Key Infrastructure (SPKI) – (Ellison 1996, IETF 1997-, Wang 1998, Ellison 2000). The momentum has now shifted to a parallel initiative, the **Simple Distributed Security Infrastructure (SDSI)** – (Rivest & Lampson 1996, SDSI 1996, Ellison 2000), into whose current version SPKI features have been incorporated. The two approaches are in the process of being harmonised. The key element of SDSI is that the X.509 nirvana of a single, global name-space has been abandoned. With it, the presumption has been removed that 'name' (or, better expressed, 'identifier') is reliably bound to a particular entity. The certificate associates a public key (and hence a key-pair) to an entity that only the CA knows, and no warranties are provided to the recipient of the message by the CA as to who the keyholder is. It is up to the relying party to build up an image of the sender based on its successive interactions with the holder of that key. Attributes are associated with public keys, not with identities of real-world entities.. SPKI/SDSI supports nyms, because no identifier is reliably associable with a particular entity, and each entity may use multiple key-pairs.

Brands (2000) proposes a different conception and implementation of digital certificates, such that privacy is protected without sacrificing security. The validity of such certificates and their contents can be checked, but the identity of the certificate-holder cannot be extracted, and different actions by the same person cannot be linked. Certificate holders have control over what information is disclosed, and to whom. If they prove to be implementable, **Brandsian certificates** will be expressly anonymous.

Trust may be based on **reputation**, by which is meant 'generally held' positive opinion about an entity. There are several ways in which 'generally held' opinion can arise, including reputation based on experience, performance, or social networks. Marketing specialists have substituted image for substance, and manufactured **proxies for reputation**. An entity can use advertising and public relations techniques to establish or embellish a brand name, which it protects using the particular form of intellectual property law called trademarks. An entity can seek to engender trust in itself by using someone else's brand, such as a seal of approval from an organisation that projects advertising and public relations on behalf of its clients. I refer to such arrangements as ‘meta-brands‘ (Clarke 2001).

An approach that avoids and dissolves the problems with PKI rather than trying to solve them, is **trust-management systems** (Blaze 1999, Blaze et al. 1999a, Blaze et al. 1999b). These can be viewed as generalisations of longstanding access control techniques for achieving security of software processes and data. The trust management approach also offers ways of addressing privacy. This is because it focusses primarily on privileges and restrictions rather than the identification of individuals, and hence it can deal with nyms representing pseudonymous roles just as readily as with names that are associated with an identified entity.

7. CONCLUSIONS

The originally perceived need was that, for e-commerce to become mainstream, merchants needed to identify themselves, and to enable authentication of the identifiers they provided. Marketers sought schemes in which consumers also needed to identify themselves to the seller. This paper has cast grave doubt on the need for identification and authentication, particularly of consumers. It has drawn attention to the manifold failures of conventional PKI to deliver on its claims, and to its seriously privacy-invasive nature.

There remain a few contexts in which digital signatures can be effective. In particular, it can be applied internally by organisations that have structures that are strictly hierarchical and relatively stable. National defence agencies, and some kinds of large corporations, are arguably of that kind. In addition, a related approach can be applied on Extranets that link defined and bounded communities of organisations and individuals. Where the participants are well-known to one another from prior dealings, a scheme can be devised to leverage off the existing relationships in order to associate a key with a particular community-member. Winn (1998) refers to these as 'closed-bound communities'. Note that, in such circumstances, the conventional PKI is essentially irrelevant (Wheeler 1998, Wheeler & Wheeler 1998).

The technical orientation that has been adopted by the proponents of conventional, X.509-based PKI does not address the needs of the Information Society. The real requirement is for trust in e-interactions: consumers want security and convenience, but without surrendering personal data to sellers (and to others who may gain access to it, e.g. other merchants and government agencies).

Conventional PKI suffers very serious inadequacies. The existence of an increasingly rich set of alternatives shows that the time has now come to recognise the inherent deficiencies of X.509 architectures, and abandon attempts to impose them on open, public systems.

REFERENCES

The following are key citations. The full list of 72 references is provided at:

http://www.anu.edu.au/people/Roger.Clarke/II/ECIS2001.html#Refs

Brands S.A. (2000) 'Rethinking Public Key Infrastructures and Digital Certificates: Building in Privacy' MIT Press, 2000

Clarke R. (1994c) 'Human Identification in Information Systems: Management Challenges and Public Policy Issues' Info. Technology & People 7,4 (December 1994). At http://www.anu.edu.au/people/Roger.Clarke/DV/HumanID.html

Davis D. (1996) 'Compliance Defects in Public-Key Cryptography' Proc. 6th Usenix Security Symp., San Jose CA, 1996, pp.171-178, at http://world.std.com/~dtd/compliance/compliance.pdf

Ellison C. (1996) 'Establishing Identity Without Certification Authorities', Proc. 6th USENIX Security Symposium, San Jose CA, July 22-25, 1996, at http://world.std.com/~cme/usenix.html

Ellison C. (2000b) 'SPKI/SDSI and the Web of Trust' September 2000, at http://world.std.com/~cme/html/web.html

Ellison C. & Schneier B. (2000a) 'Risks of PKI: Electronic Commerce' Inside Risks 116, Commun. ACM 43, 2 (February 2000), at http://www.counterpane.com/insiderisks5.html

Gerck E. (2000) 'Overview of Certification Systems: X.509, CA, PGP and SKIP', July 2000, at http://www.mcg.org.br/certover.pdf

Greenleaf G.W. & Clarke R. (1997) `Privacy Implications of Digital Signatures', IBC Conference on Digital Signatures, Sydney (March 1997), at http://www.anu.edu.au/people/Roger.Clarke/DV/DigSig.html

Housley R., Ford W., Polk W. and Solo D. (1999) 'Internet X.509 Public Key Infrastructure Certificate and CRL Profile', RFC 2459, January 1999, at http://www.ietf.org/rfc/rfc2459.txt

Khare R. & Rifkin A. (1997) 'Weaving a Web of Trust' Revised version of a paper World Wide Web Journal 2 3 (Summer 1997) 77-112, at http://www.cs.caltech.edu/~adam/local/trust.html

Myers M., Ankney R., Malpani A., Galperin S. & Adams C. (1999) 'X.509 Internet Public Key Infrastructure: Online Certificate Status Protocol - OCSP', IETF RFC2560, June 1999, at http://www.ietf.org/rfc/rfc2560.txt

RFC2692 (1999) 'SPKI Requirements' Internet Engineering Task Force of The Internet Society, September 1999, at ftp://ftp.isi.edu/in-notes/rfc2692.txt

Rivest R.L. & Lampson B. (1996) 'SDSI - A Simple Distributed Security Infrastructure', 15 Sep 1996, at http://theory.lcs.mit.edu/~rivest/sdsi10.html

Schneier B. (1996) 'Applied Cryptography' Wiley, 2nd Ed., 1996

Schneier B. (2000) 'Why Digital Signatures Are Not Signatures‘ Crypto-Gram 15 November 2000, at http://www.counterpane.com/crypto-gram-0011.html

SDSI (1996-) 'A Simple Distributed Security Infrastructure (SDSI)', 1996-, at http://theory.lcs.mit.edu/~cis/sdsi.html

W3C (2000) 'Public-Key Infrastructure (X.509) (pkix)', at http://www.ietf.org/html.charters/pkix-charter.html

Wheeler L. (1998) 'Account Authority Digital Signature Model (AADS)', at http://www.garlic.com/~lynn/aadsover.htm

Winn J.K. (1998) 'Open Systems, Free Markets,and Regulation of Internet Commerce' 72 Tulane L. Rev. 1177 (1998), at http://www.smu.edu/~jwinn/esig.html

Winn J.K. (2001) ‚The Emperor’s New Clothes: The Shocking Truth About Ditial Signatures and Internet Commerce‘ forthcoming, Idaho Law Review, 2001

X.509 (1988, 1997) 'The Directory - Authentication Framework', Volume VIII of CCITT Blue Book, pages 48-81, CCITT/ITU, 1988, 1997