December 2003

Authentication Re-visited: How Public Key Infrastructure Could Yet Prosper

Roger Clarke
XaMax Consultancy Pty Ltd, Visiting Professor, Baker Cyberspace Law & Policy Centre

16th Bled eCommerce Conference

eTransformation

Bled, Slovenia, June 9-11, 2003

Authentication Re-visited: How Public Key Infrastructure Could Yet Prosper

Roger Clarke

Xamax Consultancy Pty Ltd, Australia
Visiting Professor, Baker Cyberspace Law & Policy Centre, U.N.S.W., Australia
Visiting Fellow, Department of Computer Science, A.N.U., Australia
Roger.Clarke@Xamax.com.au

Abstract

What are the nails for which public key technologies are supposed to be the hammer? This paper examines the kinds of assertions that e-business needs to be authenticated, and evaluates conventional and alternative public key infrastructures (PKI) against those requirements. It concludes that the root cause of the limited success enjoyed by public key technologies has been inadequate requirements analysis, and proposes how PKI can be re-conceived in order to meet the real needs of e-business..

1. Introduction

Digital signatures and the public key infrastructure (PKI) that supports them were portrayed by purveyors of the dot.com myths as the means whereby trust was to be achieved in e-commerce and other applications of the Internet. Yet a web-century has passed since the seminal article, and public key cryptography has yet to fulfil its promise.

This paper argues that the reason is that inadequate care was taken with the definition of requirements. One pervasive myth needs to be debunked at the outset. The term 'authentication' does not necessarily have anything to do with identity. **Authentication** refers to a process whereby a degree of confidence is established in an assertion. The assertion might relate to identity, but in many cases it does not.

The primary purpose of this paper is to present a taxonomy of assertions whose authentication is relevant to e-business. By **eBusiness** is meant here the application of telecommunications-based tools to the business of people, corporations and government agencies. It encompasses all segments of activity, including B2C, B2B, C2C, e-government and electronic services delivery.

The categories of assertion considered in this paper relate to human identities and entities, organisational identities and entities, artefact identities and entities, attributes of each of them, and value. The analysis depends on terminology that is variously new and more carefully defined than has been the case in the past. Key terms are shown in bold-face type at the point in the text where they are defined.

The later sections of the paper consider the extent to which, for all their inadequacies, X.509 certificates and the conventional PKI that is associated with them may have something to offer. Alternative certificate-formats and approaches to PKI are identified that may address the needs of e-business more effectively. This paper is a short-form presentation of analysis undertaken in two much more detailed papers, Clarke (2001b) and Clarke (2001c). Those two papers provide references to a more substantial underlying literature.

2. Assertions Important to eBusiness

A fundamental problem with PKI is that most of the research underlying it has been applied in orientation rather than instrumentalist research. That is to say that it has not sought to solve a problem, but to find something to do with a tool. The concept was created by Diffie & Helman (1976), as a response to a perceived need, and hence the origins were instrumentalist. Since Diffie & Helman, however, the vast majority of the work undertaken has comprised attempts to apply the technology. Inadequacies in PKI arise from a failure to refine the original conception. These inadequacies can only be overcome by articulating the needs of e-business, and then conceiving PKI that will address those needs.

The caption to Steiner's 1993 cartoon, "On the Internet, nobody knows you're a dog", is usually interpreted to mean that a serious problem arises from the lack of identity authentication. There are some circumstances in which this is a problem. But in most circumstances, people can get on with e-business perfectly well, with or without reliable knowledge of the other parties' identities by achieving other means for achieving trust (Clarke 2000). This is very fortunate, because as the following sub-section will argue, an assertion of human identity is very challenging to authenticate.

2.1 Assertions About Human Entities and Identities

The notion of human identity has been inadequately addressed in the literature. The term '**human identity**' is used in this paper to refer to a particular presentation of a **human entity**. Individual people perform various social, economic and political functions, in roles such as citizen, consumer, sole trader, and member of partnerships and unincorporated associations. A person may present the same persona for every role, or different personae for each of them, or a few personae each of which is used in multiple contexts.

It is useful to have a term available that encompasses both identities and the entities that underlie them. In this paper, the term '**(id)entity**' is used for that purpose.

Organisations construct models of relevant (id)entities, by capturing data into data structures within information systems. In particular:

- an (id)entity is represented by data that is stored in a **record**;
- an attribute of an (id)entity is represented by a **data-item or field** within a record; and
- an event involving an (id)entity is represented by a **transaction**.

Within an organisation's information systems, a real-world (id)entity is operationalised as some sub-set of the data that describes it, and that differentiates it from other, similar identities. For example, a car may be differentiated by its accessories, its paint-scheme, a particular pattern of dents and scratches, a particular grinding sound when changing

gears, and its peculiar cornering characteristics. More formally, an **(id)entifier** is one or more data-items concerning an (id)entity that are sufficient to distinguish it from other instances of its particular class, and that can therefore be used to signify that (id)entity.

The preceding paragraph defined both an identifier for an identity, and a new term, '**entifier**', which is the signifier for an entity. The distinction is important: a name or code may be an identifier, but not an entifier. An entifier for a human being, because it must distinguish the physical person from other individuals, is of necessity some form of biometric.

(Id)entification is the process whereby data is associated with a particular (id)entity. It is performed through the acquisition of data that constitutes an (id)entifier for that (id)entity. An organisation's purpose in performing an (id)entification process is to establish that an (id)entity presenting to it is either:

- a previously known real-world (id)entity – in which case it will be possible to associate the new transaction with an existing record in the relevant information system; or
- a previously unknown real-world (id)entity – in which case it will be appropriate for a new organisational (id)entifier to be established for that party, and a new record created in the relevant information system.

The process of (id)entification is a search for the one among many data records that corresponds to the presenting (id)entity. For a comprehensive treatment of human identity in information systems, see Clarke (1994).

One further refinement of existing language is needed, in order to reflect the realities confronting e-business. The relationship between an identity and an underlying entity may or may not be known to a record-keeper, and indeed may or may not be knowable. The term **'nym'** is used here to refer to one or more data-items relating to an identity that are sufficient to distinguish it from other instances of its particular class, but without enabling association with a specific entity. That the concept is commonly recognised is evidenced by the wide range of synonyms, including also-known-as, aka, alias, avatar, handle, nickname, nick, nom de guerre, nom de plume, moniker, persona, personality, profile, pseudonym, pseudo-identifier, sobriquet, and stage-name. The term 'nym' is to be preferred, because it is gaining currency, it is derived from a relevant Greek root, and it carries little semantic baggage with it.

A nym enables an individual to act without disclosing which entity they are. There are two important cases. **Anonymity** is a characteristic of data, such that it cannot be associated with a particular human entity from the data itself, nor by combining it with other data. **Pseudonymity** is a characteristic of data, such that it cannot, in the normal course of events, be associated with a particular human entity. In most cases, this is achieved through the use of some form of pseudo-identifier, with the index that relates the identifier to the underlying entity being inaccessible, and effectively protected. For a comprehensive treatment of nymity, see Clarke (1999).

It is proposed that meaningful discussions about the authentication of human (id)entity are not possible unless a model is available of at least the richness depicted in Exhibit 1.

Some key aspects of this model that differentiate it from the conventional wisdom are that:

- entities underlie identities;
- the term 'entifier' is necessary, to distinguish a signifier of an entity. An entity may have multiple entifiers, but an entifier relates to precisely one entity;

- any entity may have multiple identities;
- multiple entities might present to other parties using the same identity;
- an identity may have multiple identifiers;
- data may not be able to be reliably associated with a particular entity, even though it can be related to an identity. The term 'nym' is useful to signify an identifier for such an identity;
- it may or may not be apparent that what appears to be an identifier may be only a nym.

On the basis of this model, it is possible to propose definitions for the first of a series of assertion-types relevant to e-business.

Identity authentication is the process whereby an organisation establishes its degree of confidence in an assertion that a party is who they purport to be. More laboriously, but more precisely, it is a process designed to cross-check against additional evidence the identity that is asserted or inferred by an identifier acquired during an identification process. An item of evidence is usefully referred to as an '**authenticator**' (such as the nomination of additional identifier, or demonstrated knowledge), or a '**credential**' (such as a document purporting to be associated with the person, or a token such as an 'identity card' or 'photo-id').

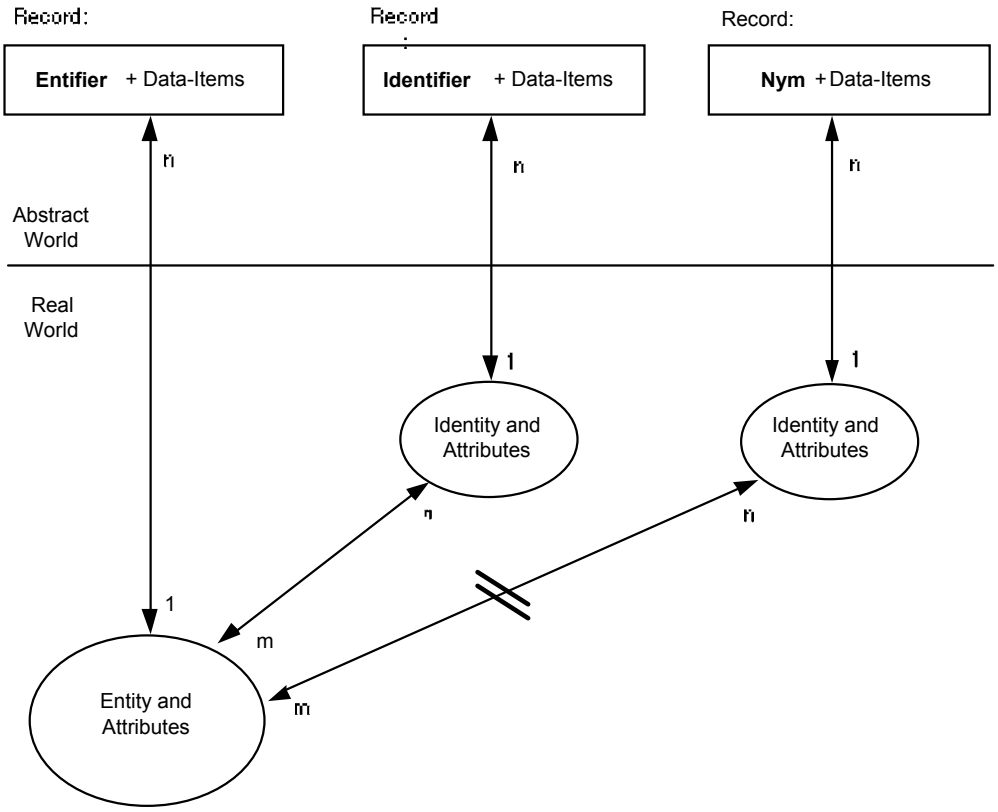

Exhibit 1: *A Model of Human (Id)entity*

The term **entity authentication**, on the other hand, refers to the process whereby an organisation establishes its degree of confidence in an assertion that a party is a specific instance of the species *homo sapiens*. The entification of a human entity depends on the

gathering of an entifier of the person, i.e. a biometric. The authentication process involves a cross-check of the entifier against a reference measure. Despite the confidence expressed by marketers of biometrics technology, a considerable degree of difficulty is involved in implementing effective schemes, and they are highly privacy-invasive (Clarke 2002).

The quality of (id)entification, and of the authentication of (id)entity, depends on many factors, and the challenges involved result in a substantial incidence of false inclusions and false exclusions. Because of the inevitability of quality shortfalls, a system designer needs to carefully consider the following:

- the **repudiability** of an assertion. This includes such questions as how an (id)entity contests information stored on another (id)entity's records;
- the **onus of proof**. This involves establishing on which (id)entity the responsibility lies to establish that data is or is not of appropriate quality; and
- the **allocation of costs**. This determines which party bears the cost and inconvenience that arise in the (id)entification and authentication processes, and where the quality of data is contested.

This sub-section has demonstrated that the authentication of human identity and entity is far more complex and challenging than it is assumed to be by the purveyors of conventional PKI, and is in most circumstances probably a forlorn hope. The following sub-sections consider the extent to which, even where human (id)entification and authentication are feasible, it is a sufficient, and an appropriate, means of enabling e-business.

2.2 Assertions about Organisational Entities and Identities

A huge amount of e-business involves organisations, including business enterprises, government agencies and associations (also known as non-profit and not-for-profit organisations). An assertion that a message has originated from, or been sent to, a particular organisation, would appear to be a very important category of assertion to authenticate.

The concept of **'organisation'** requires consideration. In order to mobilise resources, the concept of 'incorporation' was created. The original 'bodies corporate' took the form of 'joint stock companies'. The idea has been applied in many other circumstances as well, in order to create entities distinct from the people who make them up.

Identifiers of corporations include their names, and the codes assigned to them by registration bodies and other organisations. On the other hand, corporations evidence many identities, such as business units, business names and brands, which may not be distinguished for the relevant legal purposes, and which may or may not have reliable identifiers.

Government agencies are even more problematical. Many cannot be formally distinguished from the 'body politic' of which they are a part, and many that do have an independent legal existence have uncertain names, and no registration codes.

Moreover, there has to be doubt that the concept of an entifier for an organisation has any meaning. Organisations have no physical existence, and are merely legal fictions: there can be no equivalent to a human biometric. The authentication of an assertion of organisational entity is therefore seriously problematical, in both the worlds of conventional business and of e-business.

The authentication of an assertion of organisational identity is also difficult. Commonly-used authenticators include the affixing of a company seal, letterhead, and callback to a telephone-number acquired from another source. These provide only a modest degree of assurance. Equivalents in the electronic world are similarly difficult to contrive. To seriously suggest that an entity that has no real-world existence can possess a private key, and can invoke it in order to sign messages, is to drift towards a dangerous fantasy-land.

This and the previous sub-section have demonstrated that assertions that relate to humans and to organisations are very difficult to authenticate. This suggests that it might be an appropriate time to examine other kinds of assertions and establishe whether they are more readily authenticated.

2.3 Assertions about Artefact Entities and Identities

The term **artefact** is used in this paper to refer to devices such as workstations, smart cards and robots, together with software agents. These exhibit more or less intelligent behaviour, with more or less independence from individuals, and are perhaps gradually tending towards sentience (Clarke 1993-94). Artefacts are substantially involved in e-business, and there are many circumstances in which it is appropriate to check the likelihood that the artefact that originated a message is as it appears to be, or is asserted to be.

Entifiers for hardware artefacts include processor-ids, and network interface card (NIC) ids. They are somewhat more challenging to define for software artefacts. Artefact identities, on the other hand, can be signified using smartcard segment-IDs, process-ids such as sockets and web-server ids, and web-page URLs and email-addresses. An IP(v4)-address says nothing about which artefact was using it at the time, and is therefore at best a proxy for an artefact entifier.

The authentication of assertions about artefacts is challenging, but less seriously difficult than the authentication of assertions about people and about organisations. Moreover, confidence arising from authentication of artefacts can make valuable contributions to trust in e-business. But before lowering our sights too far, we need to consider what other categories of assertion are relevant.

2.4 Assertions about Attributes, Agency and Location

(Id)entities have attributes. Attributes of human entities relevant to e-business include age-range, association membership, and educational or other qualification. Organisations have attributes such as registered health care provider, and pre-qualified tenderer. Artefacts may have a particular configuration, or a particular capability such as a being able to display or print data.

Assertions of the possession of an attribute can be subjected to **attribute authentication** through the inspection of a credential that attests to that (id)entity possessing that attribute. Many circumstances exist in which the credential identifies the person, but this is not actually necessary. All that is needed is some means whereby the credential is reliably associated with the (id)entity presenting the credential. For example, a series of challenges for information may be sufficient to establish that a person qualifies for entry to premises, without even knowing their (id)entity let alone authenticating it.

Moreover, even where the process of attribute authentication involves the provision of an (id)entifier, there may be no need to record anything more than the fact that authentication was performed. In this way, the transaction ceases to be identified. An example of this is the inspection of so-called 'photo-id', without recording the (id)entifier

displayed on the card. Smartcard-based schemes can be readily devised such that identity and even entity authentication can be performed, but without yielding up the (id)entifier for recording in an information system.

In addition to the kinds of attributes discussed above, two particular sub-categories are of great significance in e-business. One is the legal authority to act on behalf of another (id)entity, generally referred to as **agency**. The representative is referred to as an agent, and the party being represented is called the principal.

Humans appoint agents, with various terms being used in various contexts, such as authorised signatory and attorney (as in 'power of attorney', rather than 'district attorney'). Organisations lack corporeal form, and therefore have no ability to act in either the physical or the electronic worlds. In order to enter into contracts, place orders, receive deliveries, instigate payments, accept orders, and initiate the fulfilment of orders, they have no option but to delegate to agents. In many cases, organisations delegate to other organisations, but, eventually, for any action to be taken, the last organisation in the chain has to depend on a human, or possibly an artefact that has the capacity to act in the real world.

In this context, as in many other instances discussed so far, an assertion about agency may or may not include an assertion about (id)entity, either of the agent or the principal. For example, many auctions permit nymous offerors and/or nymous bidders. Credentials and processes designed to assist in authentication need to support identity-less agency relationships.

When conducting **agency authentication**, care is needed to ensure that the relationship between principal and agent exists at the relevant time, that it encompasses the kind of transaction being conducted, and that the transaction does not exceed any limitations on the agent's power to act on behalf of the principal, and to bind the principal in contract. A further complication is that an agent may act for multiple principals, and a principal may be represented by multiple agents. This results in multiple credentials, and scope for conflicts of interest to arise that need to be managed.

Analogous arrangements have been envisaged for the electronic context, applying cryptographic techniques. One approach that might be used is to authenticate the (id)entity of the individual and/or body corporate (as discussed in the preceding sub-sections), and then check some kind of register of (id)entities authorised to act on behalf of the relevant body. The register might even be implemented in distributed fashion, by setting an indicator within the person's own digital signature chip-card.

Another approach is direct authentication of an authorisation. For example, a body corporate's private key could be used to digitally sign a particular kind of instrument, which a recipient could confirm (using the body corporate's widely available public key). This would be a more direct mechanism, and would avoid unnecessary declaration and authentication of the (id)entity of the agent. It would, on the other hand, involve risk of appropriation or theft of what amounts to a bearer instrument.

Another important kind of attribute is **location**. An assertion might be of the form 'the (id)entity that originated this message did so from, or in respect of, a particular location, within some tolerance range'. **Location authentication** might involve location and tracking technologies such as the triangulation of cell-phone signals or the use of global positioning systems (GPS).

Location authentication has potential applicability to a variety of contexts, such as distributed order fulfilment, mobile commerce (e.g. fleet management, motor vehicle hire, driver assistance, road-tolling, breakdown services, and insurance), and legal constraints (e.g. censorship and on-line gambling). The justification advanced for imposing tracking capabilities on cellular phones has been search-and-rescue. As is the case with all other

kinds of attributes, mechanisms are needed that support location authentication with and without (id)entity.

2.5 Assertions About Value

Every instance of authentication considered above as a basis for trust in e-business is challenging, some of them extremely so. This makes it especially important that a further relevant form not be overlooked: value authentication.

A party commonly seeks assurance that the consideration offered by another party delivers the value it purports to. In most cases, 'value' is best understood in terms of fungibility (or convertibility to cash); but value may also be represented by vouchers such as certificates and tickets; and value can be imputed by the recipient of goods, services or information.

Examples of **value authentication** include the checking of a banknote for forgery-resistant features like metal wires or holograms. In the electronic context, they include the seeking of pre-authorisation of credit-card payments; messages stating that funds have been transferred from the sender's account to an account nominated by the receiver; and messages that contain the electronic equivalent of a coin of a particular value in a particular currency.

A further important mechanism is **value escrow**. This involves a third party receiving value from both parties to a transaction, authenticating it, holding it temporarily in trust, and releasing it to the respective parties only after it has evidence that both have fulfilled their obligations. The interpolation of such an intermediary brings advantages; but it also incurs additional costs, and creates additional risks, such as malperformance, fraud and insolvency of the escrow agent.

In a great deal of conventional commerce, value authentication without identity is a primary means whereby trust is achieved. In e-commerce, however, an aberration has arisen: in its few short years to date, the sole practical payment mechanism has been through the transmission of credit card details, which carry an identifier of the cardholder. Payment mechanisms that do not have an identifier associated with them have been conceived, designed, prototyped, implemented, and trialled, but have not yet been widely adopted. The deployment of value authentication without disclosure of identity represents a real opportunity to unlock the potential of e-commerce.

The model presented in this section has identified 15 kinds of assertions whose authentication is relevant to e-business. They are summarised in Appendix 1. The following section considers public key technology, as a prelude to evaluating conventional and alternative approaches to applying it to support authentication of the various assertion-types.

3. Public Key Infrastructure

Public key technologies assure a message-recipient that the artefact that originated the message had access to a particular private key. Making the assumption that the sender was therefore one particular arefact is inadvisable, however, unless a range of risks is satisfactorily addressed. These include the possibilities that the private key might be available to other artefacts as well, that the signature-generation process might be able to be invoked by other artefacts as well, and that the public key used to check the signature might have been provided by an imposter.

Managing those risks requires infrastructure to support the bare technology. In particular, the key-pair needs to be associated with something in the real world. (The term commonly used in the computer security literature is the 'binding' of the key to something in the real world – and the something is almost always presumed to be an (id)entity, and almost always the (id)entity of a human being. But the term 'binding' implies a much tighter form of association than is actually feasible, and is therefore avoided in this paper).

Various infrastructure designs and processes have been proposed, mostly based on directory entries and/or signed copies of directory entries conventionally referred to as 'digital certificates'. Depending on the degree of control exercised over artefact manufacture, the network, and connections to it, moderate degrees of confidence can be established in relation to 2 of the 15 categories of assertion: those relating to artefact (id)entity.

Authentication of the other 13 categories of assertion cannot be satisfied by public key technology alone, because they involve entities that are outside the networked world. The association of the key-pair with something in the real world needs to be pre-authenticated.

The term **'public key infrastructure' (PKI)** is subject to a variety of interpretations, and all too commonly its meaning is left defined. (See, for example, Webopedia and FOLDOC). It is used in this paper to refer to the comprehensive set of measures needed to enable public key technologies to support the authentication of assertions. Appendix 2 summarises the elements that make up such a PKI. Business and public policy requirements are examined in greater detail in Clarke (2001c).

Fundamental requirements are that the PKI must:

- encompass the authentication of all of the 15 forms of assertion that are relevant to e-business, and not just a sub-set of them;
- provide comprehensive protections against the risks involved, rather than addressing only some arbitrarily-selected sub-set of them;
- support varying strengths of authentication, from unauthenticated, via weakly authenticated and moderately authenticated, to strongly authenticated;
- be practicable;
- be economic; and
- reflect the interests of all stakeholders, rather than just those of the sponsors. Public policy requirements have to date generally been overlooked.

4. Is There a Role for Conventional, X.509-Based PKI?

The vast majority of designs that have been proposed to date, and that have been implemented to date, depend upon digital certificates of a particular format, specified in the CCITT X.509 standard. This was designed in a very different context from the open public Internet that has emerged since the mid-1990s; but X.509 was the hammer that came to hand when the nail was discovered. X.509 primarily defines a certificate-format, but implies some requirements of the infrastructure to support their use.

As a product of a lengthy consultative process, the X.509 standard embodies very substantial flexibility and even looseness. See Gutmann (2000). Reflecting its origins as an element of the X.500 family of directory standards, it is heavily oriented towards identity, to the extent that attribute certificates are tied to an identity (as 'children' of a

'parent' certificate). Fixities of definition present serious challenges to the design of anonymity and pseudonymity, and even to the availability of multiple keys and certificates for members of the public.

Conventional infrastructure built to apply the X.509 standard depends on organisations called Certification Authorities (CAs) that provide cryptographically secure 'certification' that purports to offer some kind of assurance about the association between the public key and something in the real world. CAs in turn depend on so-called Registration Authorities (RAs) to perform pre-authentication of that association, by means of procedures conducted in the real world, such as comparison of a person's appearance against 'photo-id' that they present, inspection of documents, and perhaps checking of the possession of the private key. For a comprehensive treatment of the inadequacies of conventional PKI, see Clarke (2001a), and several other references provided in that paper.

In an attempt to stimulate uptake of conventional PKI, information technology providers have productised CA and RA services. This has had the effect of imposing a fixed form of the providers' limited vision on all potential users: instead of application-specific choices in relation to the type of assertion to be authenticated, the processes to be used, and the strength of authentication to be achieved, CAs have offered a take-it-or-leave-it approach. It has achieved very slow take-up.

In closed networks, tight control may be able to be exercised over artefacts, and conventional PKI may be effective for authenticating artefact (id)entity and perhaps also artefact attributes. It may be feasible to apply such techniques to artefacts in open networks. It is not clear, however, that certificates and CAs offer a great deal more than a simpler scheme based on private keys embedded within devices at the time of manufacture, and relying on directory-entries rather than meaningless certificates.

In summary, it appears that the X.509v3 standard supports authentication of artefact (id)entities, yet is marketed primarily as a means of authentication of human identities. X.509 certificates essentially preclude anonymity. They might conceivably be used to support pseudonymity, but only if the policies, procedures and practices within RAs and CAs are designed to do so, and if technical, organisational and legal protections exist for the records that relate the name in the certificate to the (id)entity of the 'subject'.

Hence conventional X.509-based PKI offer inferior solutions, and for only perhaps 4 or 5 of the 15 needs identified earlier in this paper, namely:

- identity authentication for humans, provided that highly privacy-intrusive processes are successfully imposed;
- (id)entity authentication for artefacts; and
- attribute authentication, for humans and for artefacts, but only with identity.

Their very substantial bias towards identification, and away from anonymity and pseudonymity, work in favour of surveillance and privacy-invasiveness, and against public acceptability. Applications of PKI that inherit these inadequacies include SSL/TLS, PKIX, S/MIME and W3C XML-Signature (aka XMLDSig). A recent application, so-called 'Qualified Certificates' (RFC3039, 2001) specifies the kind of electronic national ID mechanism of which totalitarian dreams and nightmares are made.

The deficiencies in X.509 are intrinsic to the standard. X.509 derived from the simplistic and threatening X.500 notion of a centralised world directory. It was an inappropriate foundation for PKI to support e-business across the open public Internet. The prospect of further revisions to X.509 overcoming its deficiencies seems unlikely. If progress is to be made in PKI, consideration needs to be given to other approaches.

5. Alternative Public Key Technologies

This section considers alternative approaches to the use of public key technologies for authentication in the e-business context. For example, certificate formats other than X.509 can be used as a basis for a PKI. Several exist that, unlike X.509, were designed for the specific purpose. See Gerck (1997-2000) and Ellison (2000). These include:

- PKCS#6 (which is a specification for a super-set of X.509) (RSA 1993);
- Pretty Good Privacy (PGP – Zimmerman 1995);
- SPKI/SDSI (Wang 1998, Ellison et al. 1999, Rivest & Lampson 1996); and
- Brandsian Private Credentials (Brands 2000).

Another approach is to move the focus away from certificates, and avoid unjustified assumptions that are inherent in conventional PKI. PGP, SPKI/SDSI and Brandsian technologies each move in that direction. Additional technologies that avoid certificates include:

- Blazian Trust Management Systems (Blaze 1999); and
- Account Authority Digital Signature Model (AADS – Wheeler 1998, Wheeler & Wheeler 1998).

AADS is an example of a scheme that depends on independent or 'out of band' pre-authentication of the association between a key-pair and an (id)entity. This may be achieved through the existence of a prior relationship between the parties (or an 'account' held by one with the other), or, more generally, by leveraging off an existing community of interest. It is, however, primarily targetted at the authentication of the identity of humans and organisations, and attributes with identity, and possibly agency with identity, for a total of perhaps 4 of the 15 categories that this paper concluded were relevant to e-business.

PGP uses certificates, but enables anyone to issue a certificate. It thereby places the onus on the relying party to make its own decisions as to the level of confidence it places in the key-pair actually being associated with whatever real-world thing it purports to. (This 'web of trust' notion has drifted back into fashion in the X.509 world under the alternative rubric of 'mesh architecture'). PGP is capable of being applied so as to support perhaps 7 of the 15 categories of authentication, including identity, attribute and agency.

The SPKI/SDSI and Blazian approaches regard the association between key-pair and identity as a separate matter unaddressed by public key technology. Blazian trust management, in common with all authorisation technologies, focusses primarily on privileges and restrictions. Attributes are associated with public keys, not with (id)entities. It would appear capable of being applied to perhaps as many as 10 of the 15 categories.

SPKI/SDSI uses local rather than global names, and hence supports both pseudonyms and multiple identities per entity. It appears to be applicable to about 10 of the categories of assertion that require authentication, with the major exceptions of value authentication and probably entity authentication.

The most revolutionary and complete alternative appears to be Brandsian Private Credentials. These use a refined form of cryptography and certificate, 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. Brandsian private credentials are fundamentally anonymous, but implementations can be devised to achieve pseudonymity or identification. They are claimed by their originator to encompass conventional X.509v3 digital certificates as a special case. This approach appears capable of being applied to virtually all of the 15 categories, including value authentication, although (in common with every approach except RFC3039 'Qualified Certificates') it is not targetted at entity authentication.

6. Conclusions

Public key technology has some inherent problems in relation to:

- the insecurity of key-pair generation;
- the insecurity of private-key storage; and
- insecurity arising from timing difficulties, especially key-pair revocation.

Irrespective of the design of the PKI and of applications that use it, a certificate provides no assurance about whether:

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

The challenge is to deliver PKI that provides sufficiently convincing evidence to increase confidence in the categories of assertion identified earlier, enables a risk-managed approach to the inherent weaknesses of public key technology, and avoids creating any additional weaknesses.

Conventional X.509-based PKI embodies a large number of additional problems, and in any case addresses only 4 or 5 of the 15 categories of authentication actually needed to support e-business.

Alternative approaches exist, and have already been deployed in the field, which offer more effective authentication than does conventional PKI based on X.509 certificates. Of these, SPKI/SDSI and Brandsian Private Credentials offer particular promise as a basis for PKI that will satisfy the many conflicting interests of the many stakeholders in e-business. There is an urgent need for more, and more substantial, implementations of the alternative approaches.

References

Blaze M. (1999) 'Using the KeyNote Trust Management System', November 1999, at http://www.crypto.com/trustmgt/kn.html

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

Clarke R. (1993-94) 'Asimov's Laws of Robotics: Implications for Information Technology' IEEE Computer 26,12 (December 1993) pp.53-61 and 27,1 (January 1994), pp.57-66, at http://www.anu.edu.au/people/Roger.Clarke/SOS/Asimov.html

Clarke R. (1994) 'Human Identification in Information Systems: Management Challenges and Public Policy Issues', Information Technology & People 7,4 (December 1994) 6-37, at http://www.anu.edu.au/people/Roger.Clarke/DV/HumanID.html

Clarke R. (1999) 'Identified, Anonymous and Pseudonymous Transactions: The Spectrum of Choice' Proc. User Identification & Privacy Protection Conf., Stockholm, 14-15 June 1999, at http://www.anu.edu.au/people/Roger.Clarke/DV/UIPP99.html

Clarke R. (2000) 'Trust in the Context of e-Business ', July 2000, at http://www.anu.edu.au/people/Roger.Clarke/EC/Trust.html

Clarke R. (2001a) 'The Fundamental Inadequacies of Conventional Public Key Infrastructure' Proc. Conf. ECIS'2001, Bled, Slovenia, 27-29 June 2001, at http://www.anu.edu.au/people/Roger.Clarke/II/ECIS2001.html

Clarke R. (2001b) 'Authentication: A Sufficiently Rich Model to Enable e-Business', December 2001, at http://www.anu.edu.au/people/Roger.Clarke/EC/AuthModel.html

Clarke R. (2001c) 'The Re-Invention of Public Key Infrastructure', December 2001, at http://www.anu.edu.au/people/Roger.Clarke/EC/PKIReinv.html.html

Clarke R. (2002) 'Biometrics' Inadequacies and Threats, and the Need for Regulation' April 2002, at http://www.anu.edu.au/people/Roger.Clarke/DV/BiomThreats.html

Diffie W. & Hellman M. (1976) 'New directions in cryptography' IEEE Transactions on Information Theory IT-22 (November 1976) 644-654

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

Ellison C., Frantz B., Lampson B, Rivest R., Thomas B. & Ylonen T. (1999) 'Simple Public Key Certificate' The Internet Society, July 1999, at http://world.std.com/~cme/spki.txt

Gerck E. (1997-2000) 'Overview of Certification Systems: X.509, CA, PGP and SKIP', First published April 17, 1997, revisions to 18 July 2000, at http://www.mcg.org.br/certover.pdf

Gutmann P. (2000) 'X.509 Style Guide', at http://www.cs.auckland.ac.nz/~pgut001/pubs/x509guide.txt

RFC3039 (2001) 'Internet X.509 Public Key Infrastructure: Qualified Certificates Profile' Internet Engineering Task Force of The Internet Society, 2001, at ftp://ftp.isi.edu/in-notes/rfc3039.txt

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

RSA (1993) 'PKCS #6 - Extended-Certificate Syntax Standard' RSA Security Inc., November 1993, at http://www.rsasecurity.com/rsalabs/pkcs/pkcs-6/index.html

Wang Y. (1998) 'SPKI' December 1998, at http://www.hut.fi/~yuwang/publications/SPKI/SPKI.html

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

Wheeler A. & Wheeler L. (1998) 'PKI Account Authority Digital Signature Infrastructure', November 1998, at http://www.garlic.com/~lynn/draft-wheeler-ipki-aads-01.txt

Zimmermann P.R. (1995) 'PGP 5.0 User's Guide' MIT Press, 1995, at http://mitpress.mit.edu/book-home.tcl?isbn=0262740176

Appendix 1: 15 Kinds of Assertions Relevant to e-Business

The model of authentication for e-business presented in this paper distinguishes 15 kinds of assertions:

- Entity Authentication:
  - for humans, of the form 'the person who originated this message is the person who uses a particular entifier';
  - for organisations, of the form 'the organisation on whose behalf this message was originated is the organisation that uses a particular entifier';
  - for artefacts, of the form 'the artefact that originated this message is the artefact that uses a particular entifier'.
- **Identity Authentication:**
  - for humans, of the form 'the identity or role that originated this message is the one that uses a particular identifier';
  - for organisations, of the form 'the organisational role on whose behalf this message was originated is the organisational role that uses a particular identifier';
  - for artefacts, of the form 'the role within an artefact that originated this message is the role that uses a particular identifier'.
- **Attribute Authentication:**
  - with (id)entity., of the form 'the (id)entity that originated this message has both a particular (id)entifier, and a particular attribute';
  - without (id)entity, of the form 'the (id)entity that originated this message has a particular attribute', but no assurance exists as to who the (id)entity is that has that attribute.

- **Agency Authentication:**
  - with the (id)entity of both the principal and the agent, of the form 'the (id)entity that originated this message has a particular (id)entifier, and is the appointed agent for another specified (id)entity';
  - without the (id)entity of the principal, of the form 'the (id)entity that originated this message is the appointed agent for another unspecified (id)entity';
  - without the (id)entity of the agent, of the form 'the (id)entity that originated this message is the appointed agent for another specified (id)entity'.
- **Location Authentication:**
  - with (id)entity, of the form 'a specified (id)entity originated this message, and did so from, or in respect of, a particular location, within some tolerance range';
  - without (id)entity, of the form 'an (id)entity whose (id)entifier is not known originated this message from, or in respect of, a particular location, within some tolerance range'.
- **Value Authentication:**
  - with (id)entity, of the form 'this message was originated by a particular (id)entity, and its contents have a particular value to the recipient';
  - without (id)entity, of the form 'the contents of this message have a particular value to the recipient'.

Appendix 2: Elements of a PKI to Support Authentication

- **Standards and Protocols:**
  - standards for encryption algorithms and hashing algorithms;
  - protocols for agreeing parameters associated with encryption algorithms and hashing algorithms;
  - protocols for making public keys available to message-recipients;
  - protocols for making revocation notices available to message-recipients;
  - standards for the secure generation of secure key-pairs;
  - standards and protocols to support device-synchronisation and date-time-stamps with evidentiary value;
  - possibly, standards and protocols for notarisation services, to enhance the evidentiary value of transaction records.
- **Software:**
  - to generate key-pairs;
  - to store private keys;
  - to store public keys;
  - to make public keys available to message-recipients;
  - to generate MACs or message digests from messages;

  - to apply private keys to encrypt MACs or message digests;
  - to compose messages;
  - to acquire public keys;
  - to check public keys:
    - as to their validity;
    - as to their currency;
    - as to whether they have been revoked;
  - to apply senders' public keys to decrypt messages;
  - to apply senders' public keys to decrypt MACs or message digests;
  - to compare MACs or message digests in order to check digital signatures;
  - to acquire time-signals;
  - possibly to invoke, and to perform, notarisation services.
- **Protections for Private Keys**:
  - against unauthorised access while in storage;
  - against unauthorised access while in main memory;
  - against unauthorised invocation.
- **If a directory is used** as a repository for public keys:
  - protocols for inserting data into and maintaining data in the repository;
  - protocols for access to data in the repository.
- **If certificates are used** to communicate public keys and/or for communicating some kind of assurance about the entity that uses the corresponding private keys:
  - standards for certificate-formats;
  - profiles for application of the standards in particular contexts;
  - protocols for the communication of certificates to parties that need them;
  - means whereby recipients of messages can judge whether to check the digital signatures on certificates;
  - means whereby recipients of messages can check the digital signatures on the certificates;
  - means whereby recipients of messages can judge the extent to which various forms of assertion to have been authenticated;
  - if certificates are signed by 'certificate authorities' (CAs):
  - standards for CAs;
  - standards and procedures for registration and audit of CAs;
  - procedures for recourse by parties against CAs;
  - insurance arrangements for CAs.
- **A Legal Framework** to:
  - establish the responsibilities of the various parties;
  - establish the warranties provided by the various parties;

- assign residual risks;
- enable recourse by an injured party.

- **If pre-authentication of the association between the key-pair and something in the real world is defined to be part of the PKI** rather than an application-level issue, then the PKI must include means to establish the association of the key-pair with a device, person, legal entity, attribute, agency relationship and/or location. The elements involved are likely to comprise:
  - standards for authenticating the relationship claimed;
  - procedures for authenticating the relationship claimed;
  - means of communicating the relationship claimed and authenticated;
  - a sufficient set of operational registration authorities (RAs) providing pre-authentication services;
  - a statement as to the warranties provided by the RA and CA;
  - means of communicating the warranties provided.
- **Services**:
  - application design services;
  - software libraries;
  - security audit services;
  - possibly directory services for public keys, digital certificates and/or revocations;
  - possibly backup services for private keys;
  - if certificates are signed by CAs, then a sufficient set of operational CAs;
  - insurance services, especially for CAs and RAs;
  - time-services;
  - possibly notarisation services.