Certification Infrastructure Needs For Electronic Commerce And Personal Use

Carl M. Ellison[1]

16 July 1997

Introduction

A great deal of effort is being put into the question of building a Public Key Infrastructure, probably global, to enable electronic commerce and enable secure communication among citizens of the Net. In the process, a number of previously good assumptions are being used without being questioned. This paper questions some of those, then looks at the real needs for certification on the Internet and suggests that current plans for a PKI may need to be changed radically.

History of Certification

Diffie and Hellman, in their paper introducing the concept of public key cryptography[2], noted that public key cryptography had the chance to do away with couriers of secret keys between correspondents. "Given a system of this kind, the problem of key distribution is vastly simplified. Each user generates a pair of inverse transformations, E and D, at his terminal. The deciphering transformation, D, must be kept secret but need never be communicated on any channel. The enciphering key, E, can be made public by placing it in a public directory along with the user's name and address. Anyone can then encrypt messages and send them to the user, but no one else can decipher messages intended for him."[3]

Loren Kohnfelder, in his bachelor's thesis in electrical engineering from MIT[4], noted: "Public-key communication works best when the encryption functions can reliably be shared among the communicants (by direct contact if possible). Yet when such a reliable exchange of functions is impossible the next best thing is to trust a third party. Diffie and Hellman introduce a central authority known as the Public File."[5]

Kohnfelder then noted, "Each individual has a name in the system by which he is referenced in the Public File. Once two communicants have gotten each others' keys from the Public File then can securely communicate. The Public File digitally signs all of

[1] `cme@acm.org`

[2] Whitfield Diffie and Martin Hellman, "New Directions in Cryptography", IEEE Transactions on Information Theory, November 1976, pp. 644-654

[3] ibid., p. 648, c. 1

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

[5] Ibid., p.15

its transmission so that enemy impersonation of the Public File is precluded."[6] He then noted problems in this operation, including the need to make and break connections to this Public File, the potential performance bottleneck, the complexity of maintaining such a large system, the control of access into the huge database, etc. He then introduces the concept of a **certificate** – a signed entry, like the response from the Public File, but able to be communicated without direct connection to a centralized database.

Kohnfelder had a clever mechanism for forming this certificate, but the invention of secure hashes made his clever construction unnecessary. About ten years later, the form of certificate to which we are more accustomed was introduced as part of the X.500 plan.

The X.500 dream

The X.500 dream was that there was to be a global directory of every individual, computer or anything else connected to communications. The design of this worldwide master directory foresaw the problems of scaling and started out as a distributed database, maintained by multiple people, held in multiple locations.

To specify who had authority to change which portion of this distributed database, the X.509 certificate was designed, linking a public signature key to a node in the distributed database.

By Web speed, a 10-year-old proposal like X.500, which has never achieved its global implementation, is effectively dead. There are many reasons for this, but one is that corporations were expected to host and contribute large portions of this global database by listing their employees. Corporations, however, have a tendency to consider their employee lists confidential.

X.509, on the other hand, has continued to live – not as a mechanism for proving permission to modify a node in the X.500 data structure but rather as an **identity certificate**, the kind envisioned by Kohnfelder but grown considerably from his very simple design.

Use of X.509 Certificates

An X.509 certificate is expected to be used the way Diffie, Hellman and Kohnfelder described – to map from an individual in the physical world to his or her public key. As shown in Figure 1, Alice has some relationship with Bob. Bob is not available for direct communication of his public key, so Alice instead looks up Bob's **distinguished name** in an X.500 database, finding the X.509 certificate for Bob's key.

The path Alice follows has two links. The first is assumed to be common knowledge, from 3D space to a name space NS. The second is an X.509 certificate, mapping from the name space, NS, into the space of keys, KS.

As people have discussed certification over the past twenty years, this model has remained. Concern has been focussed on the strength of algorithms, the ability to form

[6] Ibid., p.39

hierarchies of certification authorities for issuing X.509 certificates, the Certification Practice Statement for each of the CAs involved, etc.

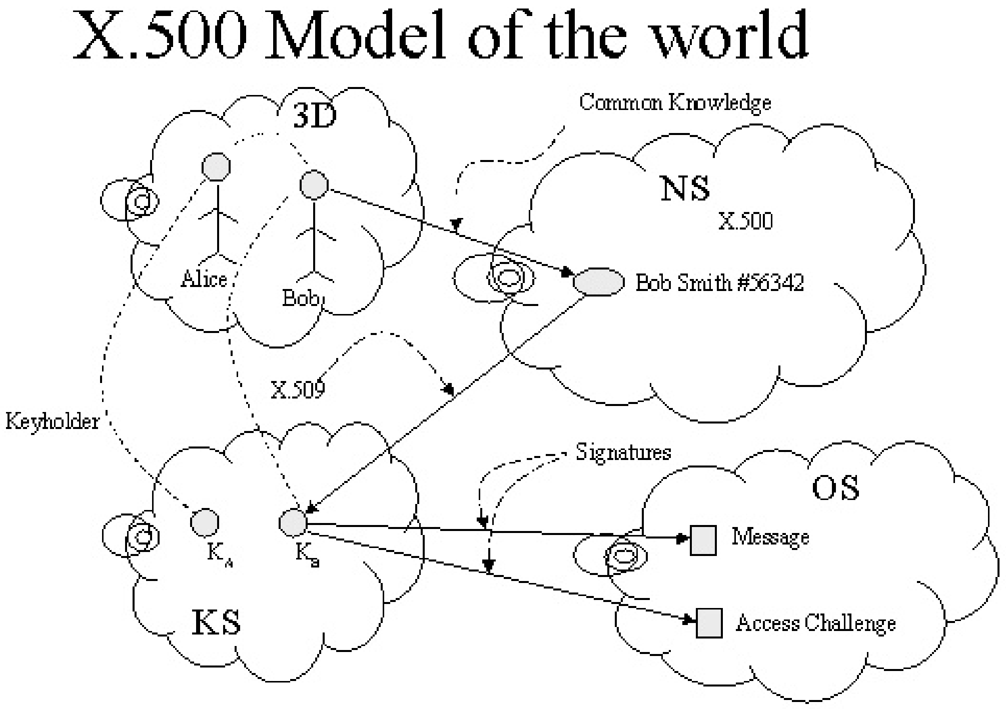

Figure 1

However, the process shown in Figure 1 is broken by at least two unanticipated consequences of the Internet, aside from the fact that there is reason to believe that a global X.500 database will never come into existence.

If we assume the existence of the global X.500 database, for the sake of argument, the first problem with the scenario in Figure 1 is that it doesn't scale. The name space scales. X.509 scales. The problem is that the "common knowledge" link from 3D space to the name space doesn't scale. For Alice to be able to follow this path, she needs to know Bob's name in that name space.

The problem here is that the original metaphor of a telephone book is wrong. If Alice met someone named Bob Smith from Baltimore MD several years ago at summer camp and finds herself passing through Baltimore today, she might be tempted to look up Bob Smith in the telephone book. She would find 166 entries under Robert, Robt, Rob or Bob. At current prices, she could put $41.50 into a payphone and try them all, not sure of course whether her old camp buddy Bob still lived in Baltimore or had moved on. Alice might find her old friend Bob this way and if she doesn't then it's just a matter of a small amount of money and time spent. However, the PKI assumption was that Alice had a confidential message for her old friend Bob's eyes only. If this were a Baltimore PKI, should Alice encrypt this sensitive message for each of the 166 possible Bob Smiths?

The telephone book does not promise to list Alice's old friends. That is the purpose of a personal address book, not a city, national or global directory.

An X.500-style PKI does use unique names (under any one root for the hierarchy), but for Alice to get Bob's unique name, she will have to establish a secure channel to Bob to learn his PKI name. If she has such a secure channel, then she can learn Bob's key directly rather than Bob's PKI name, and have the most secure knowledge of his key. Or, put the other way, to build a secure channel to her old friend Bob in order to learn his PKI name, Alice needs to know Bob's public key first. To learn Bob's public key, Alice needs to know his PKI name first.

SDSI Solution to Certification

Ron Rivest and Butler Lampson in the design of SDSI (Simple Distributed Security Infrastructure) addressed the problem of inability to disambiguate names in a large name space[7]. They created small name spaces, just large enough to be meaningful to the user.

Whether one accepts the certificate structure proposed by Rivest and Lampson or not, the lesson and solution are clear. If Alice has a relationship with Bob in the 3D world and wants to map down to his public key through a name space, she needs to know that name space. The only way to guarantee that she knows the name space is for it to be her own name space. She must learn Bob's key by some secure channel, not a PKI, before she adds Bob's key to her local SDSI certificate space – but she will form such certificates only for those with whom she has a 3D space relationship.

SDSI provides each user (each public signature key) with a name space. It also defines a method of linking those name spaces securely. That is, Alice's key might be K_A and Bob's key might be K_B. Alice would define a mapping, just as she does for nicknames or aliases in her mail program, from her name for Bob (which might be "Bob" or "Red" or "Stinky" or whatever) to K_B. This process is shown in Figure 2, which bears a strong resemblance to Figure 1.

Under this model, Alice forms the linkage for personal contacts of hers (members of her business, personal friends, etc.) via SDSI certificates. However, SDSI also allows Alice to refer to keys of people not in her personal address book, by reference. For example, the name (K_A Red) maps to K_B, because Alice has defined it that way. The name (K_A Red Mother) might map to Bob's mother's key, by indirection through Bob's name space. That is, (K_A Red Mother) reduces to (K_B Mother) and Bob defines the SDSI certificate to map that name to a public key.

This model of naming fits the real world and avoids the scaling problem that plagues the X.500 model.

There is an analogy between SDSI and Einstein's Relativity. Einstein showed that you can do better physics if you drop the fantasy of a global space-time, give each observer

[7] See http://theory.lcs.mit.edu/~rivest/publications.html for information on SDSI.

his own space-time and provide the rules for mapping from one observer's space-time to another's. SDSI shows that you can do better security if you drop the notion of a global name space and give each participant his own name space and provide the rules for mapping from one participant's name space to another's.

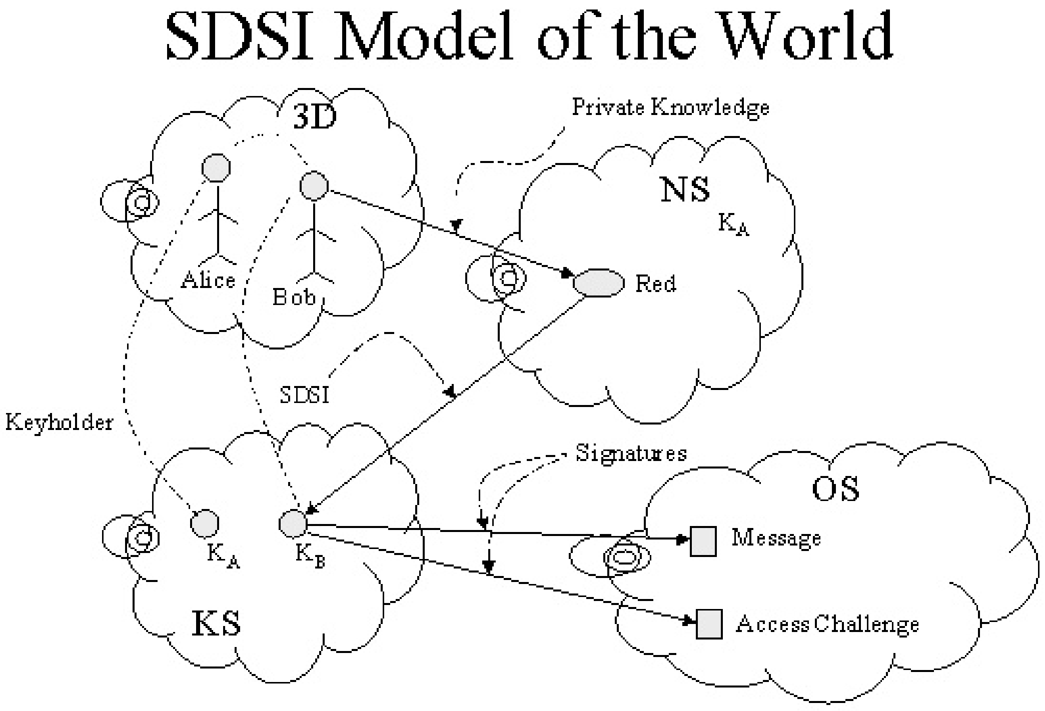

Figure 2

Difference between SDSI and PGP

The lack of hierarchy in SDSI reminds some of PGP. PGP is closer to X.500/X.509 than to SDSI, because PGP uses a global name space. It therefore suffers from the same problem when Alice tries to disambiguate various Bob Smiths. Its saving grace is that the global name used (an e-mail address) actually is implemented and is useful, but for mapping from 3D world to keys, it lacks the security of SDSI.

The other flaw

In 1976 when Diffie and Hellman were writing, in 1978 when Kohnfelder was writing and even in the late 1980's when X.500/X.509 was put forth, "cyberspace" was a science fiction term. The reality was that all relationships of interest (friendships, employment, banking, commerce...) were formed in the 3D world and the job was to map from entities in 3D space to keys in cyberspace, securely.

In 1997, relationships form in cyberspace between people who have never met in 3D space and might never meet. For such people and such relationships, a mechanism designed to map from 3D space to cyberspace is irrelevant at best.

In his July 1, 1997 address to the nation on electronic commerce and the future of the Internet, President Clinton opened with a story of his recent trip back to Arkansas for his great uncle's funeral. At his cousin's house, he talked with her son in his mid-30's "who lives in another small town in Arkansas, who after we talked for 5 minutes, proceeded to tell me that he played golf on the Internet several times a month, from this small town in Arkansas, with an elderly man in Australia, who unfailingly beat him. [laughter] An unheard-of experience just a few years ago. [chuckle] He knows this guy. He explained to me how he finds this man." The President was describing a relationship that formed in cyberspace and was likely to remain confined to cyberspace. This relationship did not need a PKI to certify **true names** of the participants.

In the same press conference, Macdara MacCole spoke of parents who don't have time to make relationships with their physical neighbors and who, instead, find common interest groups on the Internet and engage in their "over the back fence" relationship building there. There are intentionally anonymous groups that meet on the Internet – 12-step groups, groups of survivors of abuse, etc. – and for these groups, as for MacCole's parents, true names are of little interest or even a detriment. However, there **are** characteristics of interest and they call for certificates.

The mapping from 3D space to cyberspace keys, which was assumed in 1976, is sometimes needed (e.g., for e-mail to old friends), but it is not enough. Especially for the world of electronic commerce, where each merchant hopes to have a worldwide customer base, we must assume that relationships are formed within cyberspace as well. This reality is reflected in the model shown in Figure 3.

SPKI Capability Certificates[8]

A man named Bruce started a company recently to do consulting. He liked the writings of another person, named John, which he encountered from time to time on the net. So, he hired John as his first employee. John worked for Bruce for six months before they met in person for the first time.

If this had been a few years in the future, with all electronic mail digitally signed and access to networked file systems controlled by digitally signed random challenges, Bruce would not have needed John's <name, key> certificate to form and conduct this relationship. John's true name had no importance to Bruce. Bruce hired John for his writings. Assuming all John's writings were digitally signed, John's public key would be all Bruce needed to identify future writings from this person.

Bruce could, in this future world, give John access to space on the company file system and maybe permission to login on the company computer. These permissions would be passed directly from Bruce's key to John's. They would not pass through 3D space since the two men would not have met yet in 3D space. (In Figure 3, the arrow between K_A

[8] See http://world.std.com/~cme/html/spki.html for information on SPKI.

and K_B indicates this permission or **capability certificate**.) [A SET (Secure Electronic Transactions) cardholder certificate is such a mapping, since it contains no name. Instead of a name, it holds the permission that is being passed from the issuing key to the subject key.]

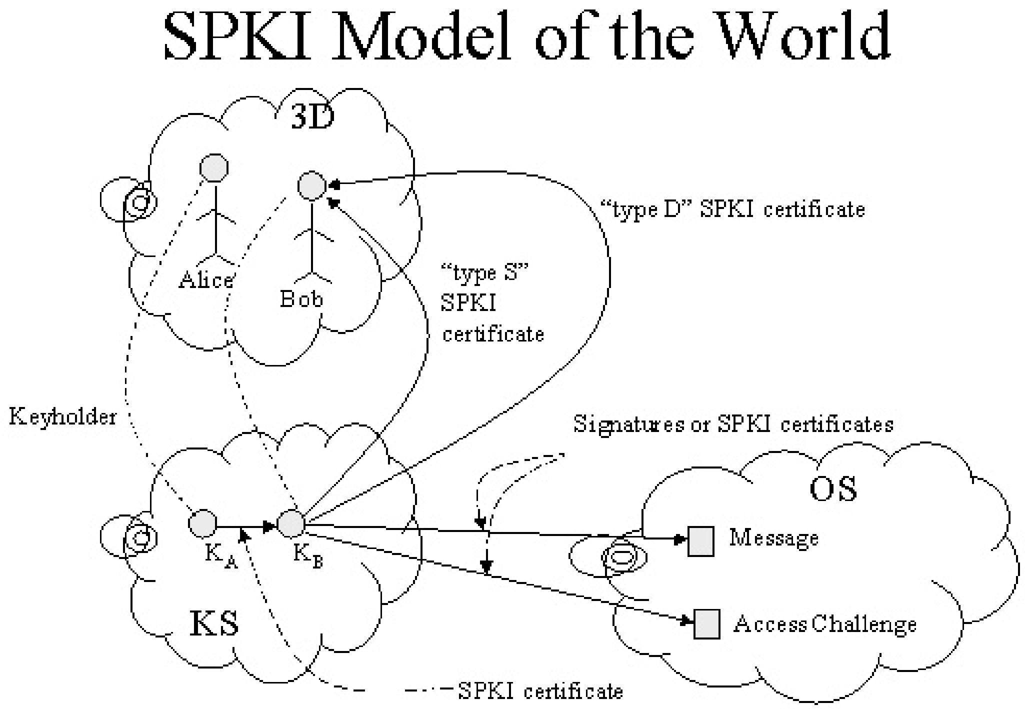

Figure 3

There is a need for other kinds of certificates as well. In particular, John needs to get paid. Therefore, there must be a certificate containing the name to put on John's paycheck and the postal address for the envelope used to mail that check. This certificate (called type "D" for "donation" in Figure 3) looks like an identity certificate, dating back to the Diffie-Hellman definition. It has a name and an address, bound to a public key and signed by a public key. However, there are two differences:

1. This certificate maps from KS to 3D, not from 3D to KS.
2. This certificate is best signed by John's key (the subject key of the certificate). No one else can be trusted more than John to provide the correct information.

Figure 3 shows another kind of certificate mapping from KS to 3D: type "S" for "subpoena".

Assume Alice and Bob sign a contract in which Bob is to deliver some work after a certain period of time. Alice needs the work done so she writes penalty clauses into the contract. Thinking ahead to the possibility that Bob might default on the contact, Alice demands a type S certificate – by which she might serve papers on Bob should he default. Alice doesn't know Bob. They met on the net and might never meet in person.

Clearly K_B should not sign the type S certificate. It needs to be issued by a third party. Is this a place for a national PKI of trusted third party CAs?

More likely, the type S certificate will be issued by Acme Process Servers, Inc. – a profit making company whose business it is to guarantee that should the **keyholder**[9] of K_B default, Acme will serve legal papers on him/her. Acme needs to know who that keyholder is, but as it turns out the type S certificate does not need to contain a name and address and Alice never needs to know who Bob is. In fact, it is in Acme's interest to have the certificate contain only a sequence number – an index into Acme's own files. That way, should Bob default on the contract, Alice is obliged to go to Acme to get the legal papers served (for a fee).

Note that this concentration on anonymity is not out of perversion but because, as with President Clinton's example, cyberspace permits relationships in which names and addresses in the physical world are of little importance. We need to be able to meet the real needs of cyberspace residents directly, not by attempting to resort to 3D world mechanisms that would break at the global scale of the Internet. We also need to satisfy the security requirements of organizations that refuse to release names of employees, especially when associated with their permission or authority, but need to authorize those employees' digital signatures.

A World of Certificates

The descriptions above have covered a number of different certificates, useful for different purposes.

- SDSI: covering people, groups or institutions with which you already have a relationship. This includes secure or signed e-mail as well as business-to-business EDI relationships except those in which the issuing company wishes anonymity.
- SPKI capability: covering the granting of permission when names are irrelevant, e.g., among people and institutions that have never met in 3D space or wish/need to remain anonymous. These should cover all remaining electronic commerce needs.
- Type "D" certificate: (called "auto-certificate" by SDSI) – a self-signed certificate noting information for which the subject key keyholder is the undisputed authority.
- Type "S" certificate: a new business service, waiting for companies to form to provide it.

There is at least one use for an identity certificate issued by a trusted third party:

- Online dating certificate: issued by a trusted dating service, this certificate would contain a picture, height and weight of some keyholder, perhaps with an anonymous e-mail address (provided by the dating service), bound to the keyholder's public key.

[9] "keyholder" is a coined word to mean the person (or other entity) with possession of the private key associated with a given public key. In this case, Bob. By using the word "keyholder", we can refer to Bob before we learn his identity.

> It would certify that the issuing agency took the picture and measured height and weight after obtaining a demonstration from the keyholder that he/she possessed the private key to match the public key in the certificate.

There will almost certainly develop other kinds of certificate, some of which will represent business opportunities.

Room for Legislation?

As President Clinton said, in introducing the story of his relative playing Internet golf, "I had two disparate experiences in the last few days that would convince a person of limited technological proficiency like myself that the world is changing rather dramatically."

The apparent motivation of digital signature laws, such as those in Utah and now in Germany, is to avoid the need to pass new laws and instead map all laws relating to physical signatures down cyberspace by somehow mapping physical people down to cyberspace. That can certainly be done. However, as this paper has attempted to show, there are new kinds of relationship which didn't exist in the physical world but which need to be handled by certificates. There are existing 3D relationships that can't be mapped to cyberspace using a global (or national) PKI because the name space would be too big to be useful.

The Internet has created a space in which people live and work. Relationships form. Things are sold. Fraud might be committed. Laws might be broken (e.g., with copyright violation). People fall in love. Companies form. People are hired. People have arguments and fall out of love. Partnerships dissolve. Disputes might arise.

Cyberspace is a society. It is not just a way of doing business within our traditional small Arkansas towns. It might be prudent to wait to see what needs arise in cyberspace before we consider attempting to map the 3D world onto it through legislation.

Meanwhile, we might benefit from the thought experiment of how one would live in a cyberspace in which the parties do not share any 3D-space connection – e.g., between Earth and some alien world, linked through some information-only wormhole in space. What cooperative, profit-making work could be done? What value would pass between the worlds? How would payment be made? Would there be a balance of payments? Could currencies on the two worlds have an exchange rate? How would disputes which cross between the worlds be resolved? Could someone hire a lawyer on the other world? Would lawyer-client privilege be respected? How would either party know? Would there be religious battles? How would we resolve them? Could there be war? How would we fight it? How would the situation change if physical goods could be shipped between worlds but the people could never meet (e.g., because the environment of one would kill the other)? What mechanisms, permissions, assurances – certificates – would be needed to facilitate the interaction between worlds?

Answers to these and similar questions might help prepare us for our future lives in cyberspace.