Global Information Assurance Certification Paper

Putting PKI in Context

Ian Christofis

25 March 2002

Summary

A number of writers have recently argued that PKI and digital signatures are not the saviours of electronic commerce that their proponents claim. Some argue for abandoning conventional PKI altogether. Whilst there are many problems with conventional PKI, some of these criticisms are not justified and amount to "throwing the baby out with the bathwater".

This paper argues that whilst changes are needed, what is most needed is for the context for PKI to be better understood, and for people to have more realistic expectations about what the technology can offer. Digital signatures and the PKI that supports them are simply another cryptographic security measure that provides a level of assurance, not absolute guarantees. In commercial transactions they can provide useful but refutable evidence.

1. Introduction

A number of writers have raised serious concerns about conventional Public Key Infrastructure (PKI) and digital signatures. They argue that conventional PKI and digital signatures are not the saviours of electronic commerce that their proponents claim, and should not be used as the primary means of online authentication. Some have proposed changes to conventional PKI. Some have proposed alternative PKI approaches. Some even argue for abandoning PKI altogether. Some examples of this dissent are:

> "Conventional PKI, built around the ISO standard X.509, has been, and will continue to be, a substantial failure." (Clarke (a), Abstract.)
>
> "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." (Clarke (a), Conclusion.)
>
> "There is mounting evidence that trying to use asymmetric cryptography as a signature on a contract is like trying to fit a square peg into a round hole, and the effort to get that square peg into that round hole has created a phenomenal sink hole into which countless individuals and organizations have poured vast resources with few tangible payoffs in sight." (Winn, Introduction)

These authors generally acknowledge that asymmetric cryptography has a place, but that conventional PKI, digital certificates and digital signatures are flawed in conception and implementation.

This paper outlines some of the problems with conventional PKI, and looks at some of the alternatives proposed. Whilst agreeing that changes to usage of PKI are needed, it argues that some of the criticisms are too harsh and that usage of PKI will evolve into a useful set of norms that will support electronic commerce and its legal underpinnings.

2. Defining "conventional PKI"

In general "conventional PKI" it is typically at the minimum used to mean PKI based on;

- X.509 certificate formats, and
- a hierarchical certification authority (CA) infrastructure.

It may also be taken to include

- use of personal names in certificates (as opposed to pseudonyms or anonymous certificates);
- other requirements of the X.509 standard, such as specific requirements for determining Distinguished Names;
- other technical standards (such as the PKCS[1] and PKIX[2] standards);

[1] See (RSA Laboratories).
[2] PKIX is an IETF initiative to adapt X.509-based PKI to the Internet. See (IETF).

- current usage conventions (eg poor practice such as a) using the same key for confidentiality and singing, b) generation of signing keys other than by the end-user, c) not using hardware devices to protect private keys, d) expecting users to have only one active signing key, etc);
- Government policy (eg Australia's Gatekeeper policy which sets various rules about what sorts of certificates Commonwealth Government agencies will take as acceptable); and
- legal environments & implications (which depends mainly on whether the jurisdiction has Utah/Malaysia-style digital signature type law, or UNCITRAL/Massachusetts/Australia/US-federal-style technology neutral electronic commerce type law).

"Conventional PKI" is not a tightly defined term. That is in itself a problem with some of the criticisms because they do not clearly distinguish between the criticisms in each of these areas. Whilst there are problems in each of these areas, debate about improving them is not assisted by vaguely directed criticisms at "conventional PKI".

3. Alternatives to "conventional PKI"

The main alternatives to conventional PKI are:

- SPKI/SDSI (Ellison), a merging of the Simple PKI (SPKI) and Simple Distributed Security Infrastructure (SDSI) initiatives; and
- Pretty Good Privacy (PGP), originally freeware developed by Phil Zimmerman, since both commercialised and taken up by the Open Source community.

Clarke also lists some others (Clarke (b)).

This paper is not intended to be a critique of the alternatives. These approaches are valuable and their development has contributed to, and continues to contribute to, a healthy debate about how best to apply asymmetric cryptography to business and personal processes.

The sub-sections below provide brief overviews of SPKI/SDSI and PGP.

3.1. SPKI/SDSI

Proponents of SPKI/SDSI argue that conventional PKI is "both excessively complex and incomplete" (Rivest & Lampson, abstract).

> "[The] complexity arises from a dependence on global name spaces and an attempt to formalize too many things. [The] incompleteness can be immediately perceived if one tries to define a security policy (e.g. write a ACL) based on the scheme. "(Rivest & Lampson, abstract)

The SPKI/SDSI approach is ``key-centric'' rather than focussed around an "individual" (e.g. person, corporation, process, or machine) whilst recognising that individuals will actually control the associated private keys. The public/private keys can be viewed as "proxies" for those individuals." (Rivest & Lampson)

The main design goals of SDSI are (Rivest & Lampson):

- Principals are public keys
- Egalitarian design – no global hierarchy necessary
- Each principal is a 'certification authority'
- Local name spaces
- Simple data structures
- Flexible signatures
- Identity certificates have human-readable content
- Manual process for creating identity certificates
- Certificates also give name/value bindings and assert membership
- On-line Internet orientation
- Linked local name spaces

3.2. PGP

PGP uses a different technical approach and format for what are essentially the equivalent of certificates, and is based on a "web of trust" model rather than a hierarchical certification authority (CA) approach. Trust is established by referrals from one user to another, or by direct exchange of public keys between users. Users digitally sign the public keys of others, so in effect all users can act as CAs.

4. The main problems with "conventional PKI"

Even the original 1988 version of the X.500 standard clearly acknowledges and lists a range of potential threats that are specific to *strong authentication*[3] itself. (CCITT, X.509, Annex E) It lists the threats as:

- Compromise of the user's secret key
- Compromise of the CA's secret key
- Misleading CA into producing an invalid certificate
- Collusion between a rogue CA and a user
- Forging a certificate
- Forging a token
- Replay of a token
- Attack on the cryptographic system

Clarke (Clarke (b), Conclusion) summarises the problems as related to:

- the insecurity of key-pair generation;

[3] *Strong authentication* is the X.509 term for authentication based on asymmetric cryptography as opposed to *simple authentication* based on a password.

- the insecurity of private-key storage; and
- insecurity arising from timing difficulties, especially key-pair revocation.
- Lack of assurance about whether:
    - the private key was originally available to other (id)entities as well as the (id)entity to which it purports to be associated;
    - the private key is now available to other (id)entities as well as the (id)entity to 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.

Ellison and Schneier outline ten major risks. (Ellison & Schneier):

- "Who do we trust, and for what?" There's a risk from an imprecise use of the word "trust." A CA is often defined as "trusted."
- "Who is using my key?"
- "How secure is the verifying computer?"
- "How secure is the verifying computer?"
- "Which John Robinson is he?"
- "Is the CA an authority?"
- "Is the user part of the security design?"
- "Was it one CA or a CA plus a Registration Authority?"
- "How did the CA identify the certificate holder?"
- "How secure are the certificate practices?"
- "Why are we using the CA process, anyway?"

Clarke and others have also argued that personal data privacy is inadequately addressed.

> "Regrettably, four years into the PKI development process, policy-makers and technology-providers have still failed abysmally to appreciate the privacy risks inherent in PKI. Public acceptance of public key cryptography will be seriously undermined by this failure." (Clarke (c), Conclusions)

Gereck provides a detailed critique of many of these issues. (Gereck)

Winn argues that the specific set of underlying assumptions that are wrong is "that contracts will be formed over the Internet among parties with no prior relationships through reliance on digital signature certificates issued by trusted third parties establish the identity of the parties" (Winn, Introduction)

Note that alternative approaches are not immune to criticism. For example Gerck (Gerck, PGP Section) raises a number of issues with PGP:

> "Because there is no entity responsible if (or when) something goes wrong – not even the user – the use of PGP in a commercial situation is difficult and may not adequately protect the business interests involved, which usually need to be guaranteed in well-defined contracts with loss responsibilities and fines. Furthermore, PGP does not scale well in size (because of the aforementioned asynchronous maintenance difficulties of the web of trust) or time (because of the same maintenance problems reflected in the so-called certificate revocation certificates, a CRL for PGP certificates). But again, within a circle of close friends this is not important." (Gerck, PGP Section)

5. Missing context

This paper does not intend to trivialise or dispute the concerns outlined above. There are many serious issues with how PKI is being implemented. What it aims to do however is place currently implemented PKI in more context.

5.1. Paper signatures

It is instructive to consider how we would critique hand-written paper signatures if they were new technology that was being introduced today. We would sensibly argue that they provide so little assurance that it would be foolish to adopt them. Some of the obvious problems are:

- They are comparatively easy to forge. Even handwriting experts cannot categorically attribute a signature to a particular person. We often even accept faxed and photocopied signed documents. By cutting and pasting we can easily copy a paper signature to a photocopy or faxed document.
- Information on a signed page can readily be altered after the page was signed.
- We usually have no way of verifying if a signature is that of the purported signatory.

Paper signatures cannot withstand the sort of scrutiny some writers are currently applying to PKI and digital signatures, but we use them on a daily basis. This does not mean we are foolish to do so, it means that we understand the limitations of paper signatures and have developed business practices and a legal environment to cope with their imperfections and inadequacies.

We need to get to a similar point of understanding of the inadequacies of digital signatures so that we understand their limitations and use them sensibly as a risk-managed security measure.

Digital signatures will normally also be surrounded by various other evidence. Court cases very rarely hinge on a single signature. They take a whole range of available evidence and make assessments based on that. The same will be true of digital signatures.

5.2. An imperfect world

The world is full of imperfect solutions. Striving for improvement is laudable, but we should not be halted in our tracks by imperfections in available technology or poor practice that has become entrenched. Expecting that perfect solutions will become available is probably not even realistic.

We should rather expect to live with imperfect solutions and incrementally fix them. This sort of "organic" approach is fundamentally how the Internet has developed, not to mention the law.

The fact that digital signatures and PKI are imperfect in their implementation or even in their underlying premises should not stop us using the technology in the proper context.

Information security measures provide varying levels of assurance, they do not provides absolute guarantees. Our legal systems are used to dealing with these sorts of imprecise mechanisms.

5.3. Time horizons should be lengthened

An underlying assumption by many people advocating digital signatures is that widespread adoption of the technology in open environments is feasible within short timeframes e.g. a few months or years. Many of the criticisms of conventional PKI also have this as an underlying premise.

These timeframes are unrealistically short. Adoption of these technologies should be viewed as a part of a radical change to society that is arguably a more significant shift than the industrial revolution.

Changing from a paper-centric to a very electronic-focussed approach to information handling is a major paradigm shift for the technically-minded, let alone the population at large. The normal reaction to most people in dealing with their personal affairs is to keep paper documents if they need long-term access to the document or are likely to require proof of a transaction at a later time. Even in business, filing documents electronically rather than filing paper copies is still the exception. Many people feel that the electronic copy of a document is ephemeral and feel more comfortable with a "real" copy on paper. The practice of printing emails is a good example of this. Even when documents arrive electronically, people often print paper copies to file.

People very comfortable with the technology and with ready access to computing resources may, in contrast, see electronic storage as better for long-term, dependable, accessible storage than paper, even for their personal information. This is due to the ability to protect electronic documents from loss with off-site back-ups, the ability to more easily find documents and search within them, to store large volumes of information in very small physical space, and to transport the information easily. In contrast they typically view printed copies as a disposable temporary instance of a document, merely for convenience of reading or mark-up. Over time as the technology becomes even more ubiquitous and widely available, it is possible that the population at large will migrate closer to this view of the permanence of electronic copies compared to paper copies.

It is clear to anyone familiar with digital technology that electronic documents and other records are trivially easy to alter. There is clearly a need for effective data integrity and authentication if we are going to place reliance on electronic documents. Whilst digital signatures are not the only method for achieving this, for all their problems they would appear to still be the best technology choice. However, whilst the general population remains very paper-focussed, the need for digital signatures is not strong and will take many years to become compelling.

Widespread adoption of PKI and digital signatures, to the point where they are a normal aspect of our way of handling information, will therefore probably take something of the order of 50 to 100 years. From this perspective we are still in the early stages of adoption of the technology and there are likely to be significant changes in how the technology is used before it becomes ubiquitous.

One of the problems with digital signatures is that the analogy to paper signatures is misleading. In time we will get to a point where the general public more clearly understands what a digital signature is, much like the way people now understand intrinsically what a debit card and PIN are. They understand the risks to a reasonable degree, without resorting to earlier analogies like "your PIN is your signature". People generally understand that if they provide a friend or an assailant with their PIN and their card they have provided access to banks accounts controlled by that card – quite different from a paper signature in their minds.

5.4. Redefining "non-repudiation"

One key area of debate is the ability of conventional PKI to provide effective non-repudiation services.

Digital signatures provide persistent evidence of intent, making fraudulent repudiation more difficult. Proof of intent after the fact with most other approaches relies on showing that sound processes were followed at the time and that there is little doubt that these were compromised at the time in question. This is usually strong enough evidence, but addresses the whole issue of non-repudiation in a less direct and generally weaker manner.

The term "non-repudiation" is somewhat misleading. Use of "non-repudiation" data security services does not preclude repudiation of that evidence by the purported signer. Nor does it imply an automatic reversal of the onus of proof.

"Non-repudiation" is just a convenient name for a data security service. Data security services provide varying degrees of assurance, not absolute guarantees. You can have weak and strong non-repudiation services and anywhere in between. Digital signatures are the strongest technology currently available, but they are not the only technical mechanism that can be used to provide a "non-repudiation" service. Most other mechanisms rely on a 3rd party handling the message and acting as an arbitrator in the event of dispute, or just rely on trying to show that normal practices were followed in the case in question, so the evidence is somewhat indirect and not straight-forward to check. Digital signatures also have the advantage that they provide persistent evidence (see below).

One purpose of digital signatures is to provide evidence, but such evidence is refutable. Under Digital Signature legislation (such as Utah, Malaysia, etc) there is a prima facie acceptance of digital signatures, but there is still the possibility of repudiation. Under the various Australian Federal and state Electronic Transaction Acts and similar laws (eg USA federal E-SIGN law) there is a deliberate avoidance of anointing any particular technology or giving anything prima facia weight.

There is never a case where use of a "non-repudiation" service actually means that the sender of a message is disbarred from repudiating the message. There are all sorts of reasons, such as someone else getting control of the signing private key (e.g. by stealing a

smartcard and finding out the associated PIN), why a digital signature might be repudiated.

5.5. Persistent vs Transient Data Security Services

When a data security service is persistent, that security aspect can be verified at any time after the communication has taken place. Digital signatures, like paper signatures, provide persistent authentication. The data integrity and non-repudiation aspects of digital signatures are also persistent.

Many other techniques for authentication and data integrity are transient i.e. the protection ceases at the end of the communication. For example SSL (with or without client certificates) provides transient data confidentiality and data integrity services, and with client-certificates also provides transient authentication services. User ID and password similarly provides a transient authentication service.

Transient data security services can be quite appropriate, but relying on them means that, in the event of a dispute, there will be a need to show that the processes surrounding the transaction were appropriate, uncompromised and working correctly at the time.

One of the key strengths of digital signature techniques is their ability to provide persistent authentication, data integrity and non-repudiation. The digital signature can be stored, archived and transported without losing its effect.

6. Conclusion

Conventional PKI does need fixing, but it is not as broken as some claim. Many of the criticisms of "conventional PKI" fail to be sufficiently specific in their targets. "Conventional PKI" is conveniently vague thing to criticize. Criticisms should be directed to specific issues for them to be meaningful.

Energy is best directed at improving the way PKI is implemented rather than arguing that PKI and digital signatures are not the right approach. The emphasis should be on fixing the problems rather than arguing that PKI should be abandoned.

Digital signatures and supporting PKI are very important and useful technologies provided they are understood and used in the right context. They can provide valuable evidence provided they are seen for what they are, namely a risk-managed security approach.

References

CCITT. Data Communications Networks Directory Recommendations X.500 – X.521 Geneva: International Telecommunications Union, 1988. 77 – 78.

Clarke, Roger (a). "The Fundamental Inadequacies of Conventional Public Key Infrastructure." 3 May 2001.
URL: http://www.anu.edu.au/people/Roger.Clarke/II/ECIS2001.html (27 February 2002)

Clarke, Roger (b). "Why Do We Need PKI? Authentication Re-visited." Review Draft of 28 January 2002.
URL: http://www.anu.edu.au/people/Roger.Clarke/EC/PKIRW02.html (25 March 2002)

Clarke, Roger (c). "Privacy Requirements of Public Key Infrastructure." Internet Law Bulletin 3, 1 (April 2000) 2-6. Republished in 'Global Electronic Commerce', published by the World Markets Research Centre in collaboration with the UN/ECE's e-Commerce Forum on 'Electronic Commerce for Transition Economies in the Digital Age', 19-20 June 2000.
URL: http://www.anu.edu.au/people/Roger.Clarke/DV/PKI2000.html. (27 February 2002)

Ellison, Carl. "SPKI/SDSI Certificates" (2 February 2002)
URL: http://world.std.com/~cme/html/spki.html (27 February 2002)

Ellison, Carl. & Schneier Bruce. "Ten Risks of PKI: What You're Not Being Told About Public Key Infrastructure." Computer Security Journal, v 16, n 1, 2000, pp. 1-7.
URL: http://www.counterpane.com/pki-risks.html (27 February 2002)

Gutmann, Peter (a). "Re: problem with the death of X.509 PKI" ansi-epay list, Thread: "Re: problem with the death of X.509 PKI (forwarded)"
URL: http://lists.commerce.net/archives/ansi-epay/200106/msg00007.html (25 March 2002)

Gutmann, Peter (b). "X.509 Style Guide." October 2000
http://www.cs.auckland.ac.nz/~pgut001/pubs/x509guide.txt (25 March 2002)

Gerck, Ed. "Overview of Certification Systems: X.509, PKIX, CA, PGP & SKIP. Do you understand digital certificates? Do you know what they warrant?" 18 July 2000.
URL: http://www.mcg.org.br/certover.pdf (25 March 2002)

IETF. "Public-Key Infrastructure (X.509) (pkix)" 11 January 2002
URL: http://www.ietf.org/html.charters/pkix-charter.html (27 February 2002)

Leibowitz, Wendy. "E-Signatures: Where's the Beef?" Wendy Tech Articles. 14 November 2000.
URL: http://www.wendytech.com/articlesesignature.htm (27 February 2002)

Rundgren, Anders. "Re: use of digital signatures and PKI." ansi-epay list, Thread: "use of digital signatures and PKI"
URL: http://lists.commerce.net/archives/ansi-epay/200106/msg00003.html (25 March 2002)

The Internet Society. "X.509 Internet Public Key Infrastructure, Online Certificate Status Protocol – OCSP." RFC2560 1999
URL: http://www.ietf.org/rfc/rfc2560.txt?number=2560 (27 February 2002)

Rivest, Ronald & Lampson, Butler. "SDSI - A Simple Distributed Security Infrastructure." 2 October 1996.
URL: http://theory.lcs.mit.edu/~rivest/sdsi11.html (27 February 2002)

RSA Laboratories. "Public-Key Cryptography Standards"
URL: http://www.rsasecurity.com/rsalabs/pkcs/ (27 February 2002)

Schneier, Bruce. "Why Digital Signatures Are Not Signatures" Crypto-Gram 15 November 2000 (2000)
URL: http://www.counterpane.com/crypto-gram-0011.html#1 (27 February 2002)

Winn, Jane. "The Emperor's New Clothes: The Shocking Truth About Digital Signatures and Internet Commerce." Revised Draft - March 9, 2001.
URL: http://faculty.smu.edu/jwinn/shocking-truth.htm (25 March 2002)

Last Updated: July 24th, 2017

Upcoming Training

SANS Prague 2017	Prague, Czech Republic	Aug 07, 2017 - Aug 12, 2017	Live Event
SANS Boston 2017	Boston, MA	Aug 07, 2017 - Aug 12, 2017	Live Event
SANS New York City 2017	New York City, NY	Aug 14, 2017 - Aug 19, 2017	Live Event
SANS Salt Lake City 2017	Salt Lake City, UT	Aug 14, 2017 - Aug 19, 2017	Live Event
Community SANS Omaha SEC401*	Omaha, NE	Aug 14, 2017 - Aug 19, 2017	Community SANS
Virginia Beach 2017 - SEC401: Security Essentials Bootcamp Style	Virginia Beach, VA	Aug 21, 2017 - Aug 26, 2017	vLive
SANS Chicago 2017	Chicago, IL	Aug 21, 2017 - Aug 26, 2017	Live Event
SANS Virginia Beach 2017	Virginia Beach, VA	Aug 21, 2017 - Sep 01, 2017	Live Event
SANS Adelaide 2017	Adelaide, Australia	Aug 21, 2017 - Aug 26, 2017	Live Event
Community SANS Pasadena SEC401 @ NASA	Pasadena, CA	Aug 23, 2017 - Aug 30, 2017	Community SANS
Mentor Session - SEC401	Minneapolis, MN	Aug 29, 2017 - Oct 10, 2017	Mentor
SANS San Francisco Fall 2017	San Francisco, CA	Sep 05, 2017 - Sep 10, 2017	Live Event
SANS Tampa - Clearwater 2017	Clearwater, FL	Sep 05, 2017 - Sep 10, 2017	Live Event
Mentor Session - SEC401	Edmonton, AB	Sep 06, 2017 - Oct 18, 2017	Mentor
SANS Network Security 2017	Las Vegas, NV	Sep 10, 2017 - Sep 17, 2017	Live Event
Mentor Session - SEC401	Ventura, CA	Sep 11, 2017 - Oct 12, 2017	Mentor
Community SANS Albany SEC401	Albany, NY	Sep 11, 2017 - Sep 16, 2017	Community SANS
Community SANS Columbia SEC401	Columbia, MD	Sep 18, 2017 - Sep 23, 2017	Community SANS
Community SANS Dallas SEC401	Dallas, TX	Sep 18, 2017 - Sep 23, 2017	Community SANS
SANS London September 2017	London, United Kingdom	Sep 25, 2017 - Sep 30, 2017	Live Event
SANS Baltimore Fall 2017	Baltimore, MD	Sep 25, 2017 - Sep 30, 2017	Live Event
SANS Copenhagen 2017	Copenhagen, Denmark	Sep 25, 2017 - Sep 30, 2017	Live Event
Community SANS Boise SEC401	Boise, ID	Sep 25, 2017 - Sep 30, 2017	Community SANS
Baltimore Fall 2017 - SEC401: Security Essentials Bootcamp Style	Baltimore, MD	Sep 25, 2017 - Sep 30, 2017	vLive
Community SANS New York SEC401	New York, NY	Sep 25, 2017 - Sep 30, 2017	Community SANS
Rocky Mountain Fall 2017	Denver, CO	Sep 25, 2017 - Sep 30, 2017	Live Event
Community SANS Sacramento SEC401	Sacramento, CA	Oct 02, 2017 - Oct 07, 2017	Community SANS
SANS DFIR Prague 2017	Prague, Czech Republic	Oct 02, 2017 - Oct 08, 2017	Live Event
Community SANS Charleston SEC401	Charleston, SC	Oct 02, 2017 - Oct 07, 2017	Community SANS
Mentor Session - SEC401	Arlington, VA	Oct 04, 2017 - Nov 15, 2017	Mentor
SANS Phoenix-Mesa 2017	Mesa, AZ	Oct 09, 2017 - Oct 14, 2017	Live Event