Real-World Problems of PKI Hierarchy

Daniel Cvrcek
cvrcek@vabo.cz

Department of Computer Science and Engineering
Brno University of Technology
Brno, Czech Republic

Abstract

Public key cryptography has become very popular recently. To use it securely, it is necessary to employ systems offering at least a basic set of functions associated with public key cryptosystems containing generation of keys, secure generation of certificates, verification of signatures, dissemination of revoked certificates, and so on. This paper demonstrates that the use of public key cryptography is not easy and it is very difficult to perform the above-mentioned actions in a way satisfying some predefined security level. The reason is in the complexity of the problem. We are not generally able to ensure the usage of systems able to work with all existing algorithms, key lengths, and formats in the whole public key infrastructure. Also, some people are of the opinion that X.509 as a basis for PKI is obsolete.

Keywords: PKI, X.509, hierarchy, CRL.

1. Introduction

Public cryptography is attracting much attention through various kinds of implementation of public key infrastructures (PKI). This popularity has lead, in the Czech Republic, to the adoption of the Electronic Signature Law. The problem is that public key cryptography is hard to understand without basic knowledge of its mathematical background and a general security model. This paucity of knowledge results in public displays by uneducated persons, during witch they introduce basic concepts of the technology but ignore fundamental security risks that make usage of the technology vulnerable and dangerous for its users. The goal of this paper is to highlight problems that are not so obvious but force experts to think over the basic concepts and its possible modifications.

1.1 Current problem areas

During the time we were working with the X.509-based standard, some flaws and problems gradually appeared. The problems were associated mostly with the technology. As the project advanced, new problems related to administration appeared. In short, three basic problem areas related to the PKI have been identified.

- Technology – the problems related to the principles on which the whole idea of PKI is based.
- Administration – the problems resulting from application of the technology.
- General security –application of the technology seemed to violate some basic security requirements.

Most of the negatives are related to the complexity of the PKI implementations. While attending any security course, you are told that public cryptography is very useful for exchanging secrets necessary for private communication satisfying privacy. This statement is true, but only if you know with whom you are communicating. At the moment you try to build versatile system (identification of users is the hardest problem), it becomes apparent that the problems to be solved are comparable (because of their complexity) to applications of symmetric cryptography.

Let us assume the basic concept of the X.500 family of standards – a global hierarchy of domain names. The impossibility of reaching this goal has been one of major reasons that defeated X.400 as an e-mail standard. One could say that changes in the premises imply at least rethinking the model built on them, but it has never taken a place.

1.2 Solutions

Are there any easy solutions that would solve problems associated with the standard? It does not seem so. People are being assured by vendors and security consultants that today's PKI solutions solve all their security problems and needs. We are told that PKIs may be used everywhere and by anyone.

The opposite is proved by new emerging technologies, based on different assumptions and promising solutions to some problems of X.509 technology. However, any new system has to be compatible with the existing one because of interoperability. It is a difficult problem to introduce packages preserving communication with existing public key infrastructures while offering new technology. An ideal security product cannot be seen by users, and it is hard to understand and even harder to explain why the new solution is better when nothing can be seen on the display.

The rest of this paper consists of three chapters devoted to the problem areas already introduced. One chapter also describes some other approaches and technologies.

2. Technology

The basic problem of the technology is the need for trusted third parties, but there is no solution. The second-worst problem is the distribution of information about revoked certificates.

The traditional method to distribute certificate-status information involves a certification authority (CA) periodically issuing a CRL that is posted to repositories. Each CRL includes all revoked and still unexpired certificates issued by the CA. One of the pieces of information included is the **nextUpdate** field, which specifies the time the next CRL will be issued. It is clear that the size of the CRL may be very large, depending on the number of certificates issued, the environment, and the period of the certificates' validity and users' activity. The paper by David A. Cooper [1] examines the load on the repository. His work determines the peak of load as being on about 35 requests per second for 300.000 relying parties and 24 hour-validity of CRL. I consider it an unacceptable number regarding the size of CRL at about 10 kB and a connection through a line of 100 kb/s.

The rate of CRL requests has an exponential probability and allows one to offer some solutions to make the rate more balanced.

Over-issued CRLs solve the problem that CRL cached by relying parties expires at one moment. This method spreads out requests for CRLs by ensuring that cached information expires at the same time only for a part of the relying parties. The way it is done is simple – there are several updates of CRL during their validity. For example, 24 hour-validity CRLs are generated every three hours. It means that ideally only one eighth of the relying parties need to update their copy of CRL at the same moment.

Segmented CRLs solve the problem by dividing certificate owners into smaller groups. Each group is then assigned to its own CRLs. This decreases sizes of revocation lists, and frees communication channels a bit and decreases response time of the repositories. The reverse side of the solution is that relying parties probably need to download more CRLs.

On-line status protocols spread the requests steadily over time but bring in more communication sessions. With ten verifications per day (assumed so far) it makes nearly 35 requests per second (300,000 relying parties) and does not solve the problem at all. Furthermore, it decreases the security of the system as a whole, because the private key of the CA has to be used automatically (it is not possible to have a human operating 35 requests per second throughout the day).

The calculations introduced so far suppose that the frequency of verification of digital signatures is uniform over time. It is not, however, true in the real life. An average clerk starts his or her computer at 8:00AM, and the first application to appear is an e-mail client. It means the repository needs to answer many more requests at one moment (assuming the new CRL is published sometime between 4:00PM and 8:00AM).

There are also other problems we have not mentioned that are connected with the proposed solutions. Over-issuing of revocation lists means that one certificate appears in each CRL at a different time, and it takes time for the CRLs' validity to appear everywhere. This raises two questions. Is the signature valid when identification of the appropriate certificate appears in for example, three CRLs but does not appear in five others? Who is responsible for the possible losses?

The segmented CRL is a good idea (especially for CAs because it allows keeping size of the revocation lists small, which is very useful for implementation), but we need to identify the correct CRL, and it adds a requirement to signature-verification process.

We have not mentioned the usage of delta-CRL that is directly defined in X.509. When using this method, complete CRLs are issued periodically and, meanwhile, only changes in the given CRL are issued and distributed. The target is similar to the segmented CRLs. We want to reduce communication costs. The reverse side of this method is the necessity to check the base and all newer delta CRLs during the verification process by the relying party. It looks satisfactory now, but there is a problem. There has be a restriction on the length of the delta-CRLs chain to verify. When the limit is reached, all the relying parties need to obtain a new base CRL. You can see that it does not result in any improvement.

A modification **sliding window delta-CRL** has been proposed [2]. It is a combination of the delta-CRL principle and over-issuing. A new base CRL is then generated with each delta CRL. The advantages combine results of over-issuing (decreased rate) and delta-CRL (decreased size of transmitted data). But the complexity of the system to issue CRLs makes this a costly option.

Speaking of technology, there is another critical difficulty. The secure usage of asymmetric cryptography requires the usage of secure hardware – at least smart cards – that is able to perform cryptographic operations. The problem is that this cryptography is rather complicated and requirements on smart cards are therefore quite demanding. There are not many vendors offering such crypto cards, and their price is relatively high. The complexity is due to two reasons: the complexity of the mathematics itself and verification of the certification path. The second function becomes problematic with high certification trees.

3. Administration

Real implementation of PKI also causes problems associated with administration of all the parts of the system. What we see certification policies as being very problematic. Standard X.509v3 [3] has made it possible to define the particular policy used for issuing certificates and to insert information about that policy into each certificate.

There is a good reason for the idea. It is possible to discern certificates according to their *quality*. The reverse side of that is the implementation of the idea.

- It is necessary to express policy used for the certificate.
- When verifying the certificate, we have to monitor fulfillment of some quality requirements.
- There is no common root, and we have to define rules to be able to apply certificates from different CAs.

The X.509v3 standard allows the expression of policy-mapping rules in certificates, but it also makes it more complex to use them securely. To make the situation even more complex, there is also a concept for an *attribute certificate* that is being greatly extended in version 4 of the standard [4].

To solve the problem of cooperation among several certification domains is the main objective of *Federal Bridge Certification Authority Initiative* [5]. One of the basic terms defined "trust domain", is a portion of the Federal PKI (FPKI) operating under the management of one policy-management body. An FPKI policy authority that governs collaboration on how best to ensure interoperation of existing agencies has also been

formed. The result of this effort is the Federal Bridge Certification Authority. FPKIPA approves Certification Practice Statement (CPS) for the FBCA and determines the assurance levels at which agency principal CAs may interoperate through the FBCA. To determine the assurance level, it is necessary to compare certification policies, CPSs, and other submitted materials (really nothing for a smart card).

FBCA acts as a *bridge* among agencies' principal (root) CAs by issuing cross-certificates with them. The cross-certificates contain mapping of certification policies. The verification of the certificate issued in another trusted domain is done by verification of the whole certification path up to the principal CA of that trusted domain. The second step is transformation of policies expressed in the certificate by policy-mapping rules. We have not mentioned checking of lists of revoked certificates, which also has to be done. Progress is being made, but it is not easy.

We are using the notion of a trusted third party (TTP), but we have not mentioned what the word *trusted* means. Well, that is because it is very difficult. It is even more difficult when the TTP (for example TTP_1) has to trust some other TTP_2. TTP_2 is not as trusted as TTP_1, but is it trustworthy more or less? And how much more or less trustworthy is it?

There are some general answers designed for the purpose. We mean standards for evaluation of computer security [6, 7, 8]. The problem is that there are too few subjects able to really evaluate the security of complex information systems, and CA is a complex system. The Czech electronic signature law solves the problem with the statement *ability to by certified.*

Another aspect comprises registration processes. This is one of the most important parts of the PKI activities. The quality of the registration process determines trust in the PKI as a whole. The organizations running CAs are able to ensure security of CA's secret keys by secure hardware tokens, locks, and electronic security systems, but it is impossible (for organizational, financial and technical reasons) to ensure the same level of trust for registration authorities' (RA) secret keys. It means that it is very difficult to attack PKI as a whole, but it is much less difficult to attack one half, one fifth, or one tenth of the system, depending on the number of RAs attacked.

Our last remarks are about revocation. It seems that requirements on certificate revocations are strongly connected to the purpose (application) of the certificate used. This is implied by the fact that there is not generally a secure method for revocation. There are two extremes: frequent revocation and secure usage of secret keys needed for revocation. We always have to make some kind of compromise. We do not discuss the realistic probability that users are able to discover abuse of their secret keys, although they are fully responsible for them.

4. General Security

What we are going to discuss in this chapter is a more general conception of security related to PKI. We want to depict the impact of PKI on privacy [9]. We may talk about secret and private keys.

- Secret key generation – must be performed entirely under the holder's control but also in a certifiably secure manner. We have to ensure that the secret key is never outside the possession of the holder.
- Secret key storage and backup – has to be secure to ensure uninterrupted possession by the holder.
- Secret key escrow – is contradictory to the very concept of PKI.
- Secret key access – by the holder has to be uninterrupted (as stated by the law). Therefore it should be exempt from court orders and search warrants.
- Certification identification requirements – may involve intrusive demands for documents and possibly biometrics, perhaps exceeding the needs of the police.
- Registers of public keys and/or certificates – are necessarily public (according to Czech law). It is important to realize that they create a serious risk for the certificates' owners as a multi-purpose identification database.

The root of most problems is the tendency to preserve absolute identification, which is common in real world but does not take into account the increased risks associated with full identification in cyber space. The result of this position may lead to the death of private space.

The main benefit spoken of in the media is the possibility of easy communication between citizens and the state apparatus. Only a few know that an electronic signature, as defined by the law, may serve only as an instrument to secure communication, but cannot be used as a mechanism for proving the originator of stored documents. One may sign a tax return and a minute after that may revoke its certificate. I would not want to be the judge resolving a dispute about whether the certificate holder and alleged signer signed it or not. You may say that most people would not even be aware of that possibility, but there are a small number of people who could find flaws in the system.

We have included ten, slightly changed, risks of PKI identified by Schneier and Ellison [10] at the end of the chapter. The article states ten questions that may be very interesting to answer before applying PKI by your organization but also if you are already using it.

Whom do we trust, and for what?	Do you know, what information you certify?
Who is using my secret key?	Is Registration Authority as secure as CA?
How secure is the verifying computer?	How did the CA identify the certificate holder?
I know the name, but who is it?	How secure are the certificate practices?
Is the CA an authority?	Why are we using the CA process, anyway?

5. Alternatives

We do not want to say that there is a general alternative to PKI, just as we do not say that PKI is generally useful. First, someone should consider what the requirements on the system are. Such a statement should be then used for deciding on devices that are able to satisfy them.

A very good example is in banking. There is a bank that has realized that the most important task is authentication of clients and that the relation is 1:n. All clients are to communicate with just one subject – the bank. With this in mind, they have decided to use authentication calculators that keep the system simple, from both the customer's point of view and the bank's. Another bank has decided on PKI. It resulted in a more complicated system not only for the bank itself but also for its customers. One has to visit the bank twice. The first time one obtains software necessary for the generation of keys, and the second time (with 3½" disk) the certificate is generated. The first bank uses just symmetric cryptography; the second bank has to use both the symmetric and asymmetric (public key) types.

The opposite alternative to PKI is symmetric cryptography. To secure communication with a one-time pad (it offers perfect secrecy), it is necessary to use 1 CD for 650 MB of data, and it is largely data for correspondence. It is the simplest and the most secure way to ensure secure communication among a few of people. One may also use the CD as a source of encryption keys. When changing the key (128 bits) twice a day, one may do it, for 50,000 years. (This is only a theoretical computation.)

5.1 PGP

Pretty Good Privacy uses the concept *web of trust* instead of hierarchy, which is used by X.509. This system does not use any trusted third party, but each user issues its own *certificate*. When you want to communicate with a friend, you simply give him your certificate in as secure a way, as you believe necessary. Users of the system must be aware of the security risks, and the resulting security is fully in their hands.

By the way, is it a bad idea to print a fingerprint of the public key on business cards?

5.2 SPKI/SDSI

There are two concepts; these are Simple Public Key Infrastructure and Simple Distributed Security Infrastructure, which are being harmonized [12, 13]. SDSI has introduced the concept of local names. The basic idea is that it is useless to have globally unique name beyond what we know and use as everyday local names. We know many people only by their first name; perhaps we know their whole name, but how many of your friends do you know by their address? When one says ”Peter”, he knows who it is. There is also a significant difference from X.509, in that it is possible to use names for groups of persons. Each member of the groups has his or her own keys, but the name is the same.

X.509 based systems use CRLs as the mechanism to distribute information about revoked certificates. SPKI knows three types of validity tests: CRL, re-validation, and one-time. The original CRL was modeled on print books used as blacklists for account numbers and credit card numbers. However, there may be problems with timing when it is not possible to obtain a new CRL, and some attacks may be based on the fact.

The positive version of CRL (revalidation) *contains* one or more certificates valid for another interval. The third possibility is an on-line check of the certificate’s validity. This is similar to the OCSP protocol defined for X.509 certificates.

It is also very important to use very simple syntax for certificates and other data, so that no special and complicated parsers are needed. This prevents much of the complexity of the resulting software solutions. In fact, BNF is used for definition of SPKI certificates.

5.3 Private Credentials

Stefan Brand has proposed a different conception of digital certificates. They are called *private credentials* and their validity can be checked without revealing the identity of their holder. They also protect certificate holders from linking separate actions [14].

The prime goal of private credentials is to preserve the privacy of their holders. The holders are the subjects that decide what information is to be disclosed to the recipient. Although the certificate may contain a set of attributes, the holder may choose a subset of them to be accessible by the verifier of the certificate.

6. Conclusions

PKI solutions are enthusiastically advertised by a number of vendors. The most important reason is money. One has to create the infrastructure, but the process of issuing a certificate costs nearly nothing and each customer needs to ask for at least one certificate a year, on a regular base.

PKI as understood today is a solution for environments where identification is needed. Its usage for communication between a citizen and a civil servant is understandable, but the necessity to have a list of certificates publicly available is not good news for certificates holders’ privacy. Moreover identification of the holder of the certificate has to be unique.

Most people do not realize that secure usage of public key cryptography is not just about strong cryptography and secure algorithms. The system is much more complicated and includes secure viewer-to-display information that is to be signed, a secure environment for storing and using private keys (usually secure hardware), and signature verification that must be done in a secure environment because it is as important as signing. When hearing statements like ”a digital signature is 99.9999999% secure”, one has to reassure that it is the security expert’s word for it. The security of the signature is only as secure as the lock on your door or smart card in your pocket. The security of a system is as good as the weakest part of the system.

There are many ways to secure communication and to ensure authentication and privacy. It is foolhardy to say that a particular solution is suitable for all situations and all environments. When you ask an expert and he claims to have a universal solution, it is better that you find another expert.

“Real security is hard work. There is no cure-all, especially not PKI” [11].

7. References

[1] Cooper, D. A.: A Model of Certificate Revocation. Proceedings of the Fifteenth Annual Computer Security Applications Conference, pg. 256-264, December 1999.

[3] Cooper, D. A.: A More Efficient Use of Delta-CRLs. Proceedings of the 2000 IEEE Symposium on Security and Privacy, pg. 190-202, May 2000.

[4] ITU-T.: Recommendation X.509 (1997 E): Information Technology - Open Systems Interconnection – The Directory: Authentication Framework. June 1997

[5] ITU-T.: Draft revised ITU-T Recommendation X.509 (v4). 2000

[6] Electronic Mesasging Association.: Report of Federal Bridge Certification Authority Initiative and Demonstration, 2000.

[7] DoD.: Trusted Computer System Evaluation Criteria. December 1985

[8] SOG-IS.: Information Technology Security Evaluation Criteria v1.2. Department of Trade and Industry, London, 1991

[9] The Common Criteria Project Sponsoring Organizations.: Common Criteria for Information Technology Security Evaluation version 2.1. August 1999.

[10] Clark, R.: Conventional Public Key Infrastructure: An Artefact Ill-fitted to the Needs of the Information Society. submitted to the Euro. Conference in Information Systems 2001, Bled, Slovenia.

[11] Ellison, C., Schneier, B.: Ten Risks of PKI: What You're not Being Told about Public Key Infrastructure. Computer Security Journal, vol. XVI, November, 2000

[12] Ellison, C., Schneier, B.: Risks of PKI: Secure E-Mail. Inside Risks 115, Communications of the ACM, vol 43, n 1, January, 2000

[13] Ellison, C.: SPKI Requirments. RFC 2692, September 1999

[14] Ellison, C., Frantz, B., Lampson, B., Rivest, R., Thomas, B., Ylonen, T.: SPKI Certificate Theory. RFC 2693, September 1999

[15] Zero-knowledge Systems, Inc.: Private Credentials. November, 2000.

Tasks and Challenges of The National Cryptographic Environment Implementation in Small Countries

Karel Dolník
Karel.Dolnik@army.cz

Oldrich Pekárek
Oldrich.Pekarek@army.cz

Military Security Office
Prague, Czech Republic

Abstract

There are numerous types of equipment for cryptographic protection of information. This paper discusses reason for implementing a national cryptographic environment, and the accompanying, challenges and opportunities. Implementation of NCE is a strategic matter for all countries but for small ones it is very difficult to be self-sufficient.

Keywords: national cryptographic environment.

1. Introduction

One of the basic tasks of the Cryptographic Security Branch of the Military Security Office in Prague is to implement of the national cryptographic environment (NCE) in computer, information, and communication systems requiring cryptographic information protection.

NCE is a system of unique cryptographic methods and tools developed by a state administration authority to be used primarily within its own organization, possibly within the state administration of a respective country.

NCE is implemented mainly within the systems handling classified information to provide their cryptographic protection. This entails the protection of information transmitted by telephone, fax and radio as well as of information transmitted within information and communication systems at different layers of an OSI model (application, network, link) for various transmission technologies and protocols (TCP/IP, HDLC, X.25, ATM, ISDN, G.703, etc.) and for the cryptographic protection of the information processed and stored directly in computers. The above list shows the need for a very wide spectrum of national cryptographic algorithms and equipment, and therefore the national cryptographic environment is utilized primarily by economically and technologically developed subjects.

2. Motivation for implementation of NCE

At present, cryptography is becoming a public affair, the computer capabilities of particular subjects are increasing (either by using new and more capable or original, but significantly cheaper, technologies or by utilizing computer capacities of large networks – mainly the Internet). This increases the number of qualified attackers (students – hackers, virus creators, foreign intelligence, and international terrorist and extremist groups) as well as the motivations for attack (individual prestige, industrial espionage, and economic crime). Besides the visible attempts to paralyze the functioning of information and communication systems, it is possible to conduct hidden long-term attacks against confidentiality, integrity and authenticity