

Show Me Your Public Key and I Will Tell Who You Are

L I S E A R N E B E R G

A short introduction to digital certificates and some thoughts on their significance to the digital economy.

Internet and Security – Two Contradictory Terms?

It has been repeated again and again – the lack of security mechanisms on the Internet slows down the development of the new economy. Is it true? Hard to say, really. It is a fact that there are no global trust mechanisms on the Internet infrastructure, you cannot be really sure of whom you are talking to. The only global identification mechanism is the network address, and network addresses may be forged rather easily.

On the other hand, you can build as much security as you like into one specific application. You may issue usernames and passwords and even equip your users with password calculators or one-time passwords to ensure strong authentication between the user and your application. And you may encrypt user sessions by SSL and build crypto based cookie mechanisms to obtain confidentiality and preserve session integrity.

What is the problem then? A number of issues of course. Key lengths is one, secure storage is another. But in my opinion, the lack of common authentication mechanisms is one key issue. The consequences differ slightly depending on which market segment is addressed, the Business to Business (B2B) or the Business to Consumer (B2C) segment.

In the B2B segment, the relation to the customer is often established through other channels than the Internet. For a long time relation, you can afford a (partly) off-line registration procedure and you might take the trouble of managing usernames and passwords for all your users. So, the authentication problem can be coped with. As I see it, the problem is on the customer side. How many different passwords, userids and tokens can you possibly handle? If your daily work involves using five or six web-based services, delivered by as many companies, you might have to remember five or six passwords or handle five or six password tokens. And all of them issued to you in the same role in the same company. A general authentication mechanism would leave the user with one token, valid as authentication mechanism on any service.

The problem in the B2C segment is slightly different because you often do business with someone you do not know in advance. From the business perspective you may have no option but to trust that the name, address and credit card number supplied are actually genuine. It is a fact however, that Internet shops make quite an effort to check for false orders, false credit cards and false shipment addresses, but may still experience losses or fraud.

As a consumer, you may have experienced that it is a lot of work to remember username and password for a number of different Internet shops or services. It would be nice if I did not have to. And it would also be nice if I had a way of checking the authenticity of the service at the other end. Is this really my bank that I am talking to?

Whether the need for more global authentication mechanisms comes from practicality reasons, productivity reasons or genuine lack of trust, the answer to the authentication problem may be digital certificates. As web-based services involves more and more valuable transactions, the issue of non-repudiation and legally valid signatures on documents arises. These services may be built on top of an infrastructure based on digital certificates.

The rest of this paper is an introduction to digital certificates, the infrastructure to make them function and some examples of use.

The General Idea of Digital Certificates

A digital certificate is really an ID card for the digital world. As an analogy, consider an old-fashioned, paper-based ID card. What it does is really to establish a binding between a name and a picture. If you resemble the picture I accept that the name belongs to you. In the same way, a digital certificate establishes a binding between a name and a cryptographic key. If you can prove that you possess that key I accept that the name belongs to you.

Lise Arneberg (40) is Senior Consultant at Telenor Business Systems, in the department of e-commerce. She holds a Cand.Scient. degree in Mathematics/Algebraic Geometry from the University of Oslo, 1985. She has long experience from security and cryptography related issues at Alcatel (research department, telecom and defence communications), and at Scandinavia Online (electronic commerce).

lise.arneberg@telenor.com

In both cases, my trust depends on the issuer of the ID card or the digital certificate. If I recognise the issuer as trustworthy, the ID card has a value. Passports and ID cards issued by banks or the postal service are normally recognised as trustworthy anywhere. While the ones issued by a school or an employer are not generally accepted outside that school or company. The same mechanism applies for digital IDs. If you present a digital certificate and I do not trust or recognise the issuer, it would be of little value to me.

Issuers of digital certificates are normally referred to as Trusted Third Parties (TTPs). In Norway, we have currently three actors on this scene: Bankenes Betalingssentral (BBS), Posten SDS and Telenor. All these three are companies or institutions we are used to trusting. They are all currently running digital certificate services.

Their certificates may be designed for different purposes or services. For remote LAN access, for Internet banking, for security services within an organisation, for tax reporting or for other special applications. And the contents may differ. But generally the content is at least:

- Name of the owner;
- The public key;
- Name of the issuer;
- Validity period and expiry date;
- The issuer's signature.

The issuer's signature ensures the authenticity of the certificate. If any part of the content is altered, a signature check will reveal the fraud. And it is not possible for anybody else to issue false certificates, because you cannot forge such a signature. If we use the passport analogy, the issuer's signature is the seal of Norwegian authorities, as well as the special paper quality that makes a passport so difficult to forge.

Registration and Distribution – the Key Issues for Building Trust

If you have applied for a passport lately perhaps you remember that it is quite a tedious process. You are required to meet personally, show an ID and fill in a couple of forms. To receive the passport, you must once again meet at the police station and show your face and an ID card to verify that you are the correct owner of the passport. Getting a new credit card is much easier. You simply phone and tell that the previous card is broken. After a few days you receive a new one by mail.

These two examples represent different registration procedures and different distribution procedures. The result is that a passport has a higher

level of trust than a credit card (at least in the sense of ID cards). Once again, the analogy can be made to the world of digital certificates. If an authentication process is based on cryptographic keys, how sure can you be that those keys are in the hands of the right person? That depends pretty much on the checks performed during registration, as well as distribution procedures for the certificate and the cryptographic keys. If you have to show an ID card to apply for and receive the certificate, this is a high level of security. If you fill in a form on the Internet and receive the certificate by mail, the level of trust would be low. Requirements for registration and distribution processes is often called a certificate policy (CP). In a sense, the certificate policy defines the likeliness that the certificate is in the right hands.

Two different TTPs may agree to accept each other's digital certificates. They would do so only if the process of registration and distribution are similar and ensures the same level of security or trust. This is called cross certification and is a formal process between two TTPs. If your TTP accepts the certificates of my TTP and *vice versa*, then we can actually trust each other's certificates. And we may check each other's signatures.

And a Little Bit on the Cryptographic Protocols to Make it all Work

Digital certificates are based on public key techniques, mostly RSA. For the slightly rusty reader, I will just repeat that an RSA key pair consists of a modulus, a private key and a public key. If you sign a message with your private key, the signature may be checked using your public key. If someone encrypts a message with your public key, only the private key can decrypt the message.

The modulus and the public key are public knowledge. The private key must be kept secret. Key lengths these days are normally 1024 bits but sometimes the double. Encryption processes involves exponentiation modulo 1024 bits numbers and may be time consuming.

The beauty of public key encryption is that you can easily prove that you possess a key (the private) without exposing it.

Some definitions before we look into the authentication process:

- The Registration Authority (RA) receives certificate applications and verifies the applicant's identity according to what is specified in the Certificate Policy.

- The Certificate Authority (CA) issues certificates and distributes them to the user.
- The CA also publishes all issued certificates in a catalogue or directory, which is available via the Internet.
- When a certificate is revoked for some reason, the CA updates the Certificate Revocation List (CRL).

These are all services operated by the TTP. They are critical applications and are run in a physically secured environment by specially authorised personnel.

In the world of digital certificates, when I want to prove my identity to you, the authentication process would go like this:

1. I claim my identity by sending my digital certificate. Or you look up the certificate in a certificate directory.
2. You check the issuer's name. If you recognise the name as someone you trust, you do the signature check to verify authenticity of the certificate. If not, you look up the directory of your Trusted Third Party to check if this issuer is someone he trusts (i.e. is certified by him). If that is so, you receive the information you need to check the signature of my certificate. If not, you reject my certificate because you have no reason to trust its origin.
3. You check the revocation lists (CRLs) to see if this certificate is still valid.
4. If all is well so far, you send me a challenge.
5. And my response is the challenge signed by my private key.
6. You verify the signature, using the public key of my certificate.
7. If the signature is OK, you will trust that I am the genuine owner of the certificate and now you know who I am.

The reader may have noticed that the steps above require a bit of infrastructure. First of all, digital certificates must be issued by the CA and transported securely to their owners. Secondly, the user must know whom to trust, i.e. (s)he must have access to the public key (or rather the certificate) of the trusted TTP. Thirdly, certificates must be available in a directory. If the certificates are general purpose IDs, this directory must be available to the public. As some certificates inevitably will get lost, stolen or corrupted,

a blacklist or revocation list must be publicly available for everyone to check. And there must be someone who controls and updates the revocation lists.

There is a number of standards describing the details of these steps. Most important is the CCITT X.509 standard for digital certificates. Currently, version 3 of the standard is in use. The PKCS-7 standard describes how to create and verify signatures. The directory is an X.500 service, and the LDAP protocol specifies the interface to the directory. A number of PKCS protocols describe cross certification and other certificate exchange protocols.

On the Practical Side

Suppose your Internet provider or someone else has equipped you with a digital certificate. How do you receive it, keep it and how do you get to use it without knowing the details of the protocols above?

Keys can be kept in SW or in a smart card. Keys in SW must be protected by encryption and you will need a password to decrypt the keys and get access to cryptographic operations. In a smart card, the keys are protected by the operating system of the card and the private key will never leave the card. A PIN is required to open the card for use.

The certificate issuing process may take different forms. It may be done online with the CA and with keys generated locally. In this way, the private key never leaves "home". To secure the certificate generation process, the user is first equipped with two codes used for authentication.

The certificates may also be personalised offline with centralised key generation (mainly smart cards) and then shipped to the user according to policy.

If keys are in SW, there must be a "bridge" between the application (for instance a login script) and the cryptographic keys. Let us call this bridge a security module. It offers a specified interface (API) of security operations to any application on the PC. The application may be able to order a signature, perform the steps of an authentication and perhaps to check the certificate of the entity at the other end. The API may also specify a number of other services such as encryption mechanisms and exchange of encryption keys. With keys in a smart card, all cryptographic operations may be performed at the card, and the application (or the security module) may communicate with the card through the smart card reader.

A number of standard PC applications, such as mail clients and web browsers, is available today with plug-ins that are adapted to a specific security module. For the security module from one major supplier, Entrust, there is a wide range of applications, denoted "Entrust ready", that exploits the security services in their security module. This is all done seamlessly to the user. The user may sign or encrypt mail, or communicate securely with a web service, simply by clicking icons in her regular applications.

For more specific or proprietary applications, security services must be built in by utilising the API of the security module.

Digital certificates can be managed or unmanaged. With a managed certificate, the security module automatically communicates with the CA to handle necessary certificate updates, key backup for encryption keys and other services.

SET – Digital Certificates as Credit Cards

The Secure Electronic Transaction (SET) standard was perhaps the first example of a commercial use of digital certificates. It was developed as a cooperation between VISA and Eurocard. Their goal was to establish secure credit card transactions on the Internet. It is an open standard, based on digital certificates. It was published in 1997 and the first SET transactions in Norway were performed late 1997.

The certificate issuers or CAs of SET certificates are the credit card companies. Both card holders and shops will have certificates. Each credit card company will define their own procedures for registration and distribution of certificates. There is no cross certification between CAs, i.e. you cannot pay with a VISA credit card unless the shop (or merchant in SET terminology) has a VISA certificate too.

The payment process is based on a set of pre-defined messages between the card holder, the merchant and the acquirer payment gateway (also called payment gateway) – the interface to the credit card company. All messages in the protocol are encrypted and signed. The protocol ensures maximum anonymity. The merchant will only know that the card holder is authentic and that the amount is accepted by the credit card company; she will not know the credit card number of the card holder. The payment gateway will only know the amount and not what was purchased. The card holder can be sure that this is an authentic shop with a real VISA (or other) certificate and not just someone who set up a fraud service.

The protocol, it seems, took into account any experience with practical use of credit cards, as well as exploited the new possibilities that came with the new technology. The best of two worlds? Unfortunately, SET has not been a success so far, despite all its beautiful cryptography. Partly because it was early. But also because in its first implementation it had two practical weaknesses.

The certificates were in software and had to be installed on your PC. (Smart card readers were expensive in 1997, about 7 times the prices today, and were out of the question.) Do you know how to take care of a credit card residing on your PC? If it is not in your wallet, how do you keep control of who is using it? Should it be on your home PC or on your work PC? What happens if your children or your cleaning lady are not to be trusted? Now, after a few years, we have got used to the idea of SW certificates, but personally I am still not sure that credit cards should be in SW.

The second practical weakness occurred on the merchant side. A specialised SW called a SET Merchant is required to run the SET protocol. And in 1997, the SET Merchant applications were so expensive that they were out of the question for most Internet shops. Not to mention the requirements for infrastructure. A SET merchant with its certificates is not an application you would want to put on the server directly accessible on the Internet.

So, in spite of SET, most Internet transactions are still paid by old fashioned credit cards. Shops add a bit of security by use of SSL to avoid sending credit card numbers as cleartext on the Internet. But they have to put a lot of effort in manually handling all their credit card transactions. To help this situation, some credit card companies now develop further another option of the SET – without card holder certificates. It is called MOSET (modified SET?) and permits on-line handling of credit card transactions without card holder authentication.

What About the m-Business?

The "mobile Internet" is evolving just these days. There are two approaches to services on a GSM phone. One is the WAP – Internet browsing and Internet services adapted to the format of the cell phone screen and the bandwidth of the GSM network. The second approach is building explicit services into the SIM card of a regular GSM phone. Any way, a phone is much more personal than a PC. It is also more quickly enabled and services will seem more easily accessible than on a PC. So it should be expected that

services accessible by mobile phones will evolve quickly and offer a wide range of services.

With banking services, shopping and subscription services, the issue of authentication arises also in the m-world. Even though the GSM algorithms themselves provide some authentication, this is not automatically available to applications outside the GSM systems.

However, SIM cards are smart cards and as such ideal for storage of cryptographic keys and digital certificates. Telenor Mobil are planning to equip all their new GSM subscribers with a larger SIM card, containing a digital certificate, a security module and a set of application clients, thereby providing a new electronic world with a “global” authentication mechanism.

Paperless Business? Digital Certificates and Legally Valid Signatures

Although so much business communication is done over electronic channels, there is still no mechanism or infrastructure in place to substitute a legally valid, hand written signature on a paper document. What would it take to establish legally valid signatures on a document? In such a manner that none of the signing parties can deny that they have signed, what they have signed and when?

The answer to the “when” is a timestamp service – a trusted party that adds the time to the signed document and then uses its own certificate to sign it.

The answer to the “what” part may be a public notary – a trusted party that stores valuable documents and may verify signatures upon dispute, even after many years.

And the parties must have their digital certificates, containing keys that may be used for signing.

The issue of legally valid signatures still has not reached enough maturity to bring out the implementations. This is due to legislation, but also to infrastructure and trusted actors.

Conclusive Remark

Would it be nice to have only one digital id, one smart card valid in all situations? With credit cards, health information, access to entrance doors at work and whatever. I am not so sure. On the other hand, I may never have that option. So far, the structure or content of a digital certificate must reflect some information about its usage. Your certificate as an employee with a certain role in a certain organisation will be different from your SET certificate or from the cer-

tificate on your GSM phone, which only reflects you as a private person. So we will probably have to cope with a number of digital certificates – just as we cope with a number of magnetic cards. Let us just get used to the idea! And let us pray that someone invents the all-secure-PIN-storage-device – real soon now! (Based on fingerprint identification?)

Abbreviations

B2B	Business to Business
B2C	Business to Consumer
CA	Certificate Authority
CCITT	Comité Consultatif International Téléphonique et Télégraphique (Now: ITU-T)
CP	Certificate Policy
CRL	Certificate Revocation List
GSM	Global System for Mobile Communications
LAN	Local Area Network
LDAP	Lightweight Directory Access Protocol
PKCS	Public-Key Cryptography Standards
RA	Registration Authority
RSA	Rivest-Shamir-Adleman (algorithm)
SET	Secure Electronic Transaction http://www.europay.com/common/Index.html
SSL	Secure Sockets Layer
TTP	Trusted Third Party
WAP	Wireless Application Protocol