Everything you Never Wanted to Know about PKI but were Forced to Find Out

> Peter Gutmann University of Auckland

What is Public Key Infrastructure

Public-key encryption is used for encryption and digital signatures

The public key is a string of bits

- Whose bits are they?
- What can they be used for?
- Are they still valid?
- Examples
 - Is this really the key for foo.com?
 - Was the key used to sign this valid at the time of signing?
 - Fetch me the key of Alfredo Garcia

The purpose of a PKI is to answer these questions (and more)

Certificate History

To understand the X.509 PKI, it's necessary to understand the history behind it

Why does X.509 do otherwise straightforward things in such a weird way?

[The] standards have been written by little green monsters from outer space in order to confuse normal human beings and prepare them for the big invasion — comp.std.internat

- Someone tried to explain public-key-based authentication to aliens. Their universal translators were broken and they had to gesture a lot
- They were created by the e-commerce division of the Ministry of Silly Walks

Certificate History (ctd)

Original paper on public-key encryption proposed the Public File

- Public-key white pages
- Key present \rightarrow key valid
- Communications with users were protected by a signature from the Public File

A very sensible, straightforward approach...

- ... today
- Not so good in 1976





Certificate History (ctd)

Concerns about misuse of the directory

- Companies don't like making their internal structure public

 Directory for corporate headhunters
- Privacy concerns
 - Directory of single women
 - Directory of teenage children

X.500 proposed various access control mechanisms

- Passwords
- · Hashed passwords
- Digital signatures



Certificate History (ctd)

X.509v1 clearly shows these origins

- Issuer and subject DN to place a cert in the directory
- Validity period
- Public key

No indication of...

- CA vs. end entity certs
 - Implicit from position in directory
- Key usage
 - Only one usage, directory authentication
- Cert policy
 - Only one policy, directory authentication
- Any of the other X.509v3 paraphernalia

Certificate History (ctd)

No directories of this type were ever seriously deployed

• We've had to live with the legacy of this approach ever since

This model turns certificates into capabilities

- Tickets which can be used for authorisation/access control purposes
- Capabilities can be passed around freely
- Revocation is very hard

X.500 tried to address revocation with...

- Replacing the cert with a new one
- Notifying the owner "by some off-line procedure"
- Certificate revocation lists (CRLs), a blacklist of revoked certs
- Assorted handwaving





Identity in Certificates

No-one understands X.500 DNs

- Locality? Organisational unit? Administrative domain?
- Don't fit any real-world domain
- Require extensive versing in X.500 theology to comprehend

Solution: Users cram anything they feel like into the DN

- DN becomes a meaningless blob
- In some cases privacy requirements (e.g. FERPA) result in the creation of DNs containing random noise (a full DN reveals too much info)
- Only useful components are the common name (CN) + email address, or server URL

Other PKI designs use a more pragmatic approach

Identity in Certificates (ctd) PGP: Used for email encryption

- Identity is name + email address
- SPKI: Used for authorisation/access control
 - Identity is a name meaningful within the domain of application
 - Account name on a server
 - Credit card number
 - Merchant ID

PGP and SPKI also use the public key as a unique ID

Certificate Revocation

Revocation is managed with a certificate revocation list (CRL), a form of anti-certificate which cancels a certificate

- Equivalent to 1970s-era credit card blacklist booklets
 - These were based on even earlier cheque blacklists
- Relying parties are expected to check CRLs before using a certificate
 - "This certificate is valid unless you hear somewhere that it isn't"

CRL Problems

CRLs don't work

- Violate the cardinal rule of data-driven programming "Once you have emitted a datum you can't take it back"
- In transaction processing terms, viewing a certificate as a PREPARE and a revocation as a COMMIT
 - No action can be taken between the two without destroying the ACID properties of the transaction
 - Allowing for other operations between PREPARE and COMMIT results in nondeterministic behaviour
- Blacklist approach was abandoned by credit card vendors 20 years ago because it didn't work properly

CRL Problems (ctd)

CRLs mirror credit card blacklist problems

- Not issued frequently enough to be effective against an attacker
- Expensive to distribute
- Vulnerable to simple DOS attacks
 - Attacker can prevent revocation by blocking CRL delivery

CRLs add further problems of their own

- Can contain retroactive invalidity dates
- CRL issued right now can indicate that a cert was invalid last week
 - Checking that something was valid at time *t* isn't sufficient to establish validity
 - Back-dated CRL can appear at any point in the future
- Destroys the entire concept of nonrepudiation

CRL Problems (ctd)

Revoking self-signed certificates is hairy

- Cert revokes itself
- Applications may
 - Accept the CRL as valid and revoke the certificate
 - Reject the CRL as invalid since it was signed with a revoked certificate
 - Crash
- Computer version of Epimenides paradoxon "All Cretans are liars"
 - Crashing is an appropriate response

CRL Problems (ctd)

CRL Distribution Problems

- CRLs have a fixed validity period
 Valid from *issue date* to *expiry date*
- At *expiry date*, all relying parties connect to the CA to fetch the new CRL
 - Massive peak loads when a CRL expires (DDOS attack)
- Issuing CRLs to provide timely revocation exacerbates the problem
 - 10M clients download a 1MB CRL issued once a minute = ~150GB/s traffic
 - Even per-minute CRLs aren't timely enough for high-value transactions with interest calculated by the minute

CRL Problems (ctd)

- Clients are allowed to cache CRLs for efficiency purposes
 - CA issues a CRL with a 1-hour expiry time
 - Urgent revocation arrives, CA issues an (unscheduled) forced CRL before the expiry time
 - Clients that re-fetch the CRL each time will recognise the cert as expired
 - Clients that cache CRLs won't
 - Users must choose between huge bandwidth consumption/ processing delays or missed revocations

Certificate Revocation (ctd)

Many applications require prompt revocation

- CAs (and X.509) don't really support this
- CAs are inherently an offline operation

Requirements for online checks

- Should return a simple boolean value "Certificate is valid/not valid right now"
- Can return additional information such as "Not valid because ..."
- Historical query support is also useful, "Was valid at the time the signature was generated"
- Should be lightweight (c.f. CRLs, which can require fetching and parsing a 10,000 entry CRL to check the status of a single certificate)



Online Status Checking (ctd)

OCSP acts as a selective CRL protocol

- Standard CRL process: "Send me a CRL for everything you've got"
- OCSP process: "Send me a pseudo-CRL/OCSP response for only these certs"
 - Lightweight pseudo-CRL avoids CRL size problems
- Reply is created on the spot in response to the request
 - Ephemeral pseudo-CRL avoids CRL validity period problems
 - Requires a signing operation for every query

Online Status Checking (ctd)

- Returned status values are non-orthogonal
 - Status = "good", "revoked", or "unknown"
 - "Not revoked" doesn't necessarily mean "good"
 - "Unknown" could be anything from "Certificate was never issued" to "It was issued but I can't find a CRL for it"
- If asked "Is this a valid cert" and fed...
 - A freshly-issued cert, can't say "Yes"
 - An MPEG of a cat, can't say "No"
- Compare this with the credit card authorisation model
 - Response is "Authorised" or "Declined" (with optional reasons)

Online Status Checking (ctd)

- Problems arise to some extent from the CRL-based origins of OCSP
 - CRL can only report a negative result
 - "Not revoked" doesn't mean a cert was ever issued
 - Some OCSP implementations will report "I can't find a CRL" as "Good"
 - Some relying party implementations will assume "revoked"
 ⇒ "not good", so any other status = "good"
 - Much debate among implementers about OCSP semantics

Cost of Revocation Checking

CAs charge fees to issue a certificate

• Most expensive collection of bits in the world

Revocation checks are expected to be free

- CA can't tell how often or how many checks will be made
- CRLs require
 - Processor time
 - Multiple servers (many clients can fetch them)
 - Network bandwidth (CRLs can get large)
- Active disincentive for CAs to provide real revocation checking capabilities

Cost of Revocation Checking (ctd)

Example: ActiveX

- Relatively cheap cert can sign huge numbers of ActiveX controls
- Controls are deployed across hundreds of millions of Windows machines
- Any kind of useful revocation checking would be astronomically expensive

Example: email certificate

- Must be made cheap (or free) or users won't use them
- Revocation handling isn't financially feasible

Cost of Revocation Checking (ctd)

Revocation checking in these cases is, quite literally, worthless

• Leave an infrequently-issued CRL at some semi-documented location and hope few people find it

Charge for revocation checks

- Allows certain guarantees to be associated with the check
- Identrus charges for every revocation check (i.e. certificate use)
- GSA cost was $40 \notin ...$ \$1.20 each time a certificate was used

Rev./Status Checking in the Real World

CA key compromise: Everyone finds out

• Sun handled revocation of their CA key via posts to mailing lists and newsgroups

SSL server key compromise: No-one finds out

- Stealing the keys from a typical poorly-secured server isn't hard (c.f. web page defacements)
- Revocation isn't necessary since certificates are included in the SSL handshake

- Just install a new certificate

email key compromise: Who cares?

• If necessary, send a copy of your new certificate to everyone in your address book



software

- Serves no real purpose, and slows everything down a lot
- CRLs are useful in special-case situations where there exists a statutory or contractual obligation to use them
 - Relying party needs to be able to claim CRL use for due diligence purposes or to avoid liability

Alternative: Use Online Authorisation Check

Simple Public Key Infrastructure (SPKI)

- Prefers online authorisation/validation checks
 - This is a true online authorisation check, not the OCSP silly-walk
- Positive assertions are more tractable than negative ones
 - Compare "Aliens exist" vs. "Aliens don't exist"
- Cert renewal interval is based on risk analysis of potential losses
 - X.509 renewal interval is usually one year, motivated by billing concerns
 - Treated like a domain name: Once a year, re-certify the same key
- Provides for one-time renewal
 - Cert is valid for a single transaction



- Ties keys to accounts
- Revocation is handled by removing the key/closing the account





Cross-Certification

Original X.500-based scheme envisaged a strict hierarchy rooted at the directory root

• PEM tried (and failed) to apply this to the Internet

Later work had large numbers of hierarchies

- Many, many flat hierarchies
- Every CA has a set of root certificates used to sign other certificates in relatively flat trees

What happens when you're in hierarchy A and your trading partner is in hierarchy B?

Solution: CAs cross-certify each other

- A signs B's certificate
- B signs A's certificate

Cross-Certification (ctd)

Problem: Each certificate now has two issuers

- All of X.509 is based on the fact that there's a unique issuer
- Toto, I don't think we're in X.509 any more

With further cross-certification, re-parenting, subordination of one CA to another, revocation and re-issuance/ replacement, the hierarchy of trust...





Cross-Certification (ctd)

Different CAs and paths have different validity periods, constraints, etc etc

- Certificate paths can contain loops
- Certificate semantics can change on different iterations through the loop
- Are certificate paths Turing-complete?
- No software in existence can handle these situations

Cross-certification is the black hole of PKI

- All existing laws break down
- No-one knows what it's like on the other side

Cross-Certification (ctd)

The theory: A well-managed PKI will never end up like this

• "If it does occur, we can handle it via nameConstraints, policyConstraints, etc etc"

The practice: If you give them the means, they will build it

- Allow cross-certification and it's only a matter of time before the situation will collapse into chaos
- c.f. CA vs. EE certificates
 - There are at least 5 different ways to differentiate the two
 - Only one of these was ever envisaged by X.509
- Support for name and policy constraints is dubious to nonexistant
 - Playing Russian roulette with your security

Cross-Certification in Browsers

Hard-coded into browsers

- Implicitly trusted
- Totally unknown CAs
 - CA keys have been on-sold to third parties when the original CA went out of business
- Moribund web sites
- 512-bit keys
- 40-year cert lifetime (!!)
- How much would you trust a "NO LIABILITY ACCEPTED" CA?

All CA certs are trusted equally

• Implicit universal cross-certification







Closing the Circle (ctd)

All we really care about is the key

- Issuer/subject DN, etc are historical artifacts/baggage
- "Bring me the key of Alfredo Garcia"
- This operation is currently performed locally when the key is fetched from a certificate store/Windows registry/flat file
- Moving from a local to a remote query allows centralised administration





Finding a Workable Business Model (ctd)

A PKI is not just another IT project

- Requires a combined organisational, procedural, and legal approach
- Staffing requires a skilled, multidisciplinary team
- Complexity is enormous
 - Initial PKI efforts vastly underestimated the amount of work involved
 - Current work is concentrating on small-scale pilots to avoid this issue

To be accepted, a PKI must provide perceived value

- Failure to do so is what killed SET
- No-one has really figured out a PKI business model yet

CA Business Model

Free email certs

- No-one will pay for them
- Clown suit certificates

SSL certificates run as a protection racket

- Buy our certs at US\$500/kB/year or your customers will be scared away
- Actual CA advertising:

If you fail to renew your Server ID prior to the expiration date, operating your Web site will become far riskier than normal [...] your Web site visitors will encounter multiple, intimidating warning messages when trying to conduct secure transactions with your site. This will likely impact customer trust and could result in lost business for your site.

CA consulting services

Getting your CA Key into Browsers

Total cost: \$0.5M per browser

- Netscape: Hand over the cash and a floppy
- MSIE: No special charge, but you must pass an SAS70 electronic data security audit
 - US CPA Statement on Auditing Standards 70
 - Lengthy (up to 6 months), expensive, and painful
 - Infrastructure, policy, staff, and auditing costs run to \$0.5M

CA keys are bought and sold on the secondary market

- Equifax's certificates are actually owned by Geotrust
- Cheaper to buy another CAs HSM than to have your own key added

Problems with X.509

Most of the required infrastructure doesn't exist

- Users use an undefined certification request protocol to obtain a certificate which is published in an unclear location in a nonexistent directory with no real means to revoke it
- Various workarounds are used to hide the problems
 - Details of certificate requests are kludged together via web pages
 - Complete certificate chains are included in messages wherever they're needed
 - Revocation is either handled in an ad hoc manner or ignored entirely

Standards groups are working on protocols to fix this

• Progress is extremely slow

Certificates are based on owner identities, not keys

- Owner identities don't work very well as certificate ID's
 - Real people change affiliations, email addresses, even names
 - An owner will typically have multiple certificates, all with the same ID
- Owner identity is rarely of security interest (authorisation/ capabilities are what count)
 - When you check into a hotel, buy goods in a store, you're asked for a payment instrument, not a passport
- Revoking a key requires revoking the identity of the owner
- Renewal/replacement of identity certificates is nontrivial

Problems with X.509 (ctd)

Authentication and confidentiality certificates are treated the same way for certification purposes

• X.509v1 and v2 couldn't even distinguish between the two

Users should have certified authentication keys and use these to certify their own confidentiality keys

- No real need to have a CA to certify confidentiality keys
- New confidentiality keys can be created at any time
- Doesn't require the cooperation of a CA to replace keys

 Will never fly for exactly that reason
- PGP uses this model

Aggregation of attributes shortens the overall certificate lifetime

- Steve's Rule of Revocation: Frequency of certificate change is proportional to the square of the number of attributes
- Inflexibility of certificate conflicts with real-world IDs
 - Can get a haircut, switch to contact lenses, get a suntan, shave off a moustache, go on a diet, without invalidating your passport
 - Changing a single bit in a certificate requires getting a new one
 - Steve's certificate is for an organisation which no longer exists























In a closed system (SWIFT, Identrus, ACH)

- Members sign up to the rules of the club
- Only members who will play by the rules and can carry the risk are admitted
- Members are contractually obliged to follow the rules, including obligations for signatures made with their private key
- Design can be frozen at some point when members sign off on it
 - Continuous flow of standards, drafts, modifications, and proposals is impossible to keep up with
 - PKIX has become a standing committee that will standardise anything with an ASN.1 syntax
 from ietf-pkix

Problems with X.509 (ctd)

In an open system

- Parties have no previously established network of contracts covering private key use on which they can rely
 - On what basis do you sue someone when they repudiate a signature?
 - Have they published a legally binding promise to the world to stand behind that signature?
 - Do they owe a duty of care, actionable in the case of negligence?

- Possible ways to proceed
 - Claim a duty of care where negligence resulted in financial loss (generally negligence claims for pure financial loss won't support this)
 - Claim that publishing the key was a negligent misstatement (unlikely that this will work)
 - Go after the CA (CA won't suffer any loss if the keyholder is negligent, so they can't go after the keyholder)
- On the whiteboard:
 - "Alice does something magical/mathematical with Bob's key, and the judge says 'Obviously Bob is guilty""
- In practice: Would you like to be the test case?
 - Current digital signature legislation won't help

Problems with X.509 (ctd)

Certificates don't model standard authority delegation practices

- Manager can delegate authority/responsibility to an employee
 "You're in charge of purchasing"
- CA can issue a certificate to an employee, but can't delegate the responsibility which comes with it

Residential certificates are even more problematic

• No-one knows who has the authority to sign these things

Problems with Implementations

Relying parties must, by definition, be able to rely on the handling of certificates

Currently difficult to do because of

- Implementation bugs
- Different interpretations of standards by implementors
- Implementation of different parts of standards
- Implementation of different standards

Problems with Implementations (ctd) Examples of common problems rfc822Name has ambiguous definition/implementation (Assorted standards/implementations) Should be used as luser@aol.com Can often get away with President George W.Bush <luser@aol.com> Name constraints can be avoided through creative name encoding (Problem in standards) Multiple encodings for the same character, zero-width spaces, floating diacritics, etc Can make identical-appearing strings compare as different strings Can also evade name constraints by using altNames

Problems with Implementations (ctd)

- Software crashes when it encounters a Unicode or UTF-8 string (Netscape)
 - Some other software uses Unicode for any non-ASCII characters, guaranteeing a crash
 - At least one digital signature law requires the (unnecessary) use of Unicode for a mandatory certificate field
 - Standards committee must have had MS stockholders on it
- Software produces negative numeric values because the implementors forgot about the sign bit (Microsoft and a few others)
 - Everyone changed their code to be bug-compatible with MS
- Software hardcodes the certificate policy so that any policy is treated as if it were the Verisign one (Microsoft)



Problems with Implementations (ctd)

- CA flag in certificates is ignored (Microsoft, several Mozilladerived browsers)
 - Anyone can act as a CA
 - *You* (or Honest Al down at the diner) can issue Verisign certificates
- Software ignores the key usage flags and uses the first cert it finds for the purpose it needs (Microsoft)
 - If users have separate encryption and signing certs, the software will grab the first one it finds and use it for both purposes
 - CryptoAPI seems to mostly ignore usage constraints on keys
 - AT_KEYXECHANGE keys (with corresponding certificates) can be used for signing and signature verification without any trouble







Problems with Implementations (ctd)

- End entity certificates are encoded without the basicConstraints extension to indicate that the certificate is a non-CA cert (PKIX)
 - Some apps treat these certificates as CA certificates for X.509v1 compatibility
 - May be useful as a cryptographically strong RNG
 - Issue 128 certificates without basicConstraints
 - User other app's CA/non-CA interpretation as one bit of a key
 - Produces close to 128 bits of pure entropy
- CRL checking is broken (Microsoft)
 - Older versions of MSIE would grope around blindly for a minute or so, then time out and continue anyway
 - Some newer versions forget to perform certificate validity checks (e.g. cert expiry, CA certs) if CRL checking enabled





• The lunatic fringe: Certs from vendors like Deutsche Telekom/Telesec are so broken they would create a matter/antimatter reaction if placed in the same room as an X.509 spec

Interoperability considerations merely create uncertainty and don't serve any useful purpose. The market for digital signatures is at hand and it's possible to sell products without any interoperability

---Telesec project leader (translated) People will buy anything as long as you tell them it's X.509 (shorter translation)

Implementation Problem Redux

Certified for use with Windows

- Microsoft owns the trademark
- Submit software to Microsoft, who perform extensive testing
- Passing software can use the certification mark
- Reasonable (given the size of the deployed base) interoperability among tested products

S/MIME

- RSADSI owns (owned) the trademark
- Simple interoperability test for signing and encryption

 Anyone could participate, at no cost
- Passing software can use the certification mark
- Good interoperability among tested products

Implementation Problem Redux (ctd)

X.509

- No quality control
- You cannot build software so broken than it can't claim to be X.509v3

Problems with an X.509-style PKI

PKI will solve all your problems

- PKI will make your network secure
- PKI will allow single sign-on
- PKI solves privacy problems
- PKI will allow *<insert requirement which customer will pay money for>*
- PKI makes the sun shine and the grass grow and the birds sing

Problems with an X.509-style PKI (ctd)

Reality vs. hype

- Very little interoperability/compatibility
- Lack of expertise in deploying/using a PKI
- No manageability
- Huge up-front infrastructure requirements
 - Few organisations realise just how much time, money and resources will be required
- "PKI will get rid of passwords"
 - Current implementations = password + private key
 - Passwords with a vengeance
- Certificate revocation doesn't really work
 - Locating the certificate in the first place works even less



How Effective are Certificates Really? (ctd)

But

- Actual site it's being sent to is itn.net
- Company is located in Palo Alto, California
 - Who are these people?
 - Site contains links to the Amex web site
 - Anyone can add links to Amex site to their home page though
- Just for comparison
 - Singapore Airlines, British Airways, and Lufthansa have appropriate certificates
 - Air New Zealand also uses itn.net
 - American Airlines don't seem to use any security at all
 - Qantas don't even have a web site
 - They do if you spell their name Quantas (!!)



PKI Design Guidelines

Identity

- Use a locally meaningful identifier
 - User name
 - email address
 - Account number
- Don't try and do anything meaningful with DNs
 - Treat them as meaningless blobs

PKI Design Guidelines (ctd)

Revocation

- If possible, design your PKI so that revocation isn't required SET
 - AADS/X9.59
 - ssh
 - -SSL
- If that isn't possible, use a mechanism which provides freshness guarantees
- If that isn't possible, use an online status query mechanism
 - Valid/not valid responder
 - OCSP
- If the revocation is of no value, use CRLs

PKI Design Guidelines (ctd)

Application-specific PKIs

- PKIs designed to solve a particular problem are easier to work with than a one-size-(mis)fits all approach
- One-size-fits-all approach is only useful to verify well-known entities
 - amazon.com et al
 - Banks
 - Government departments
- Application-specific approaches work better for everything else
 - Use the same channels to verify Bob's key as you use to verify other transactions with Bob
- Third-party CAs merely get in the way

PKI Design Guidelines (ctd)

Application-specific PKIs

- SPKI
 - Binds a key to an authorisation
 - X.509 binds a key to an (often irrelevant) identity which must then somehow be mapped to an authorisation
- PGP
 - Designed to secure email
 - Laissez-faire key management tied to email address solves
 "Which directory" and "Which John Doe" problems

PKI Design Guidelines (ctd)

In many situations no PKI of any kind is needed

- Example: Authority-to-individual communications (e.g. tax filing)
 - The authority knows who its users/clients are; everyone knows who the authority is
 - Obvious solution: S/MIME or PGP
 - Practical solution: SSL web server with access control
 - Revocation = disable user access
 - Instantaneous
 - Consistently applied
 - Administered by the organisation involved, not some third party

PKI Design Guidelines (ctd)

- Example: AADS/X9.59
 - Ties keys to existing accounts
 - Handled via standard business mechanisms
 - Revocation = remove key/close account
 - (US) Business Records Exception allows standard business records to be treated as evidence (rather than hearsay) in court
 - Following standard legal precedent is easier than becoming a test case for PKI

PKI Design Guidelines (ctd)

- Example: Business transactions
 - Ask Citibank about certificate validity

Vs.

- Ask Citibank to authorise the transaction directly
 - \rightarrow Use an online authorisation
- Well-established mechanisms (and much legal precedent) for online authorisation
- Strong consumer protection via Reg.E and Reg.Z
 - Report loss within 2 days: No liability
 - Report loss within 2-60 days (time to get a bank statement): Liability of \$50
- Enacted when ATM/credit cards were introduced to keep the banks honest
 - Highly effective (c.f. UK banks' card security)

PKI Design Guidelines (ctd)

There's nothing which says you have to use X.509 as anything more than a complex bit-bagging scheme

- Provides broad toolkit and crypto token support without tying you to X.509 peculiarities
- If you have a cert management scheme which works, use it

Be careful about holding your business processes hostage to your PKI (or lack thereof)

Phew!

More information in part 2 of the godzilla crypto tutorial, http://www.cs.auckland.ac.nz/~pgut001/tutorial/index.html

"PKI: It's not dead, just resting", IEEE Computer, August 2002.