The sorry state of X.509
from certification weaknesses to Man-in-the-middle detection

Ralph Holz

Network Architectures and Services
Technische Universität München

17 September 2012
About the speaker

- PhD student at Technische Universität München, Germany
- PKI background - measurement and analysis of X.509 and OpenPGP
- Also been involved in protocol design and P2P security
Agenda

- The SSL Landscape
- Proposals to enhance or replace X.509
- Crossbear: Detecting and Localising the MitM
The SSL Landscape
Significance of SSL/TLS and X.509

**SSL/TLS**
- The backbone protocols for securing the WWW
- Authentication, confidentiality, integrity
- Public-key cryptography

**X.509: Public Key Infrastructure standard**
- Certification Authorities (CAs) certify Web sites
- Non-forgeable signature on \((identity, public\ \text{key})\)
This is probably not the site you are looking for!

You attempted to reach www.symantec.de, but instead you actually reached a server identifying itself as symantec.com. This may be caused by a misconfiguration on the server or by something more serious. An attacker on your network could be trying to get you to visit a fake (and potentially harmful) version of www.symantec.de.

You should not proceed, especially if you have never seen this warning before for this site.

Proceed anyway  Back to safety

Help me understand
Basic Idea of PKI

CA1 → I1 → E1
CA1 → I4 → I5 → E3
CA1 → I1 → E2

CA2 → I2 → I3 → I6 → E3 → E4
CA2 → I2 → E7

CA3 → R3

Root Store

Ralph Holz: The sorry state of X.509
Basic Idea of PKI

Root Store

Root Certificates

Ralph Holz: The sorry state of X.509
Basic Idea of PKI

- CA
- I1
- I4
- R1
- R2
- CA1
- CA2
- E1
- E2
- E3
- E4
- E5
- E6
- E7
- Root Store
- CA3
- Host Certificates

Ralph Holz: The sorry state of X.509
Basic Idea of PKI

CA

I

R

E

Root Store

Intermediate Certificates
Basic Idea of PKI

CA_1 \rightarrow I_1 \rightarrow E_1 \rightarrow I_6 \rightarrow E_3

CA_2 \rightarrow I_4 \rightarrow E_7

CA_3 \rightarrow R_3 \rightarrow E_7

Root Store

Cross-signing
Basic Idea of PKI

One CA, two different trees

Ralph Holz: The sorry state of X.509
Basic Idea of PKI

CAs in Root Store

CA 1

I 1

E 1

E 2

I 6

E 3

E 4

CA 2

I 4

I 5

I 2

E 5

E 6

R 1

R 2

Root Store

Root Store

CA 3

R 3

E 7

CA not in Root Store
Basic Idea of PKI

Root certificate not in Root Store

CA1 → I1 → E1
CA2 → I4 → E2
I5
E3
I6
E4

Root Store

CA3 → E7
R3

Ralph Holz: The sorry state of X.509
## An X.509 Certificate

### X509v3 Certificate

<table>
<thead>
<tr>
<th>Version</th>
<th>Serial no.</th>
<th>Sig. algo.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Issuer**

**Validity**

<table>
<thead>
<tr>
<th>Not Before</th>
<th>Not After</th>
</tr>
</thead>
</table>

**Subject**

**Subject Public Key Info**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Public Key</th>
</tr>
</thead>
</table>

**X509 v3 Extensions**

| CA Flag, EV, CRL, etc. |

**Signature**
How This Got Our Interest

PKI weaknesses since 2008

- Early December 2008:
  - ‘Error’ in Comodo CA: no identity check
  - Whitehack hacks StartSSL CA
- February 2009
  - ‘Easy’ attack on MD5: fake CA certificate
- March 2011: Comodo CA hacked
  - Blacklisting of ≈ 10 certificates
- July 2011: DigiNotar CA hacked
  - 531 fake certificates *in the wild*
How This Got Our Interest

PKI weaknesses since 2008

- Early December 2008:
  - ‘Error’ in Comodo CA: no identity check
  - Whitehack hacks StartSSL CA
- February 2009
  - ‘Easy’ attack on MD5: fake CA certificate
- March 2011: Comodo CA hacked
  - Blacklisting of ≈ 10 certificates
- July 2011: DigiNotar CA hacked
  - 531 fake certificates in the wild
PKI weaknesses since 2008

- Early December 2008:
  - ‘Error’ in Comodo CA: no identity check
  - Whitehack hacks StartSSL CA
- February 2009
  - ‘Easy’ attack on MD5: fake CA certificate
- March 2011: Comodo CA hacked
  - Blacklisting of ≈ 10 certificates
- July 2011: DigiNotar CA hacked
  - 531 fake certificates in the wild
How This Got Our Interest

PKI weaknesses since 2008

- Early December 2008:
  - ‘Error’ in Comodo CA: no identity check
  - Whitehack hacks StartSSL CA
- February 2009
  - ‘Easy’ attack on MD5: fake CA certificate
- March 2011: Comodo CA hacked
  - Blacklisting of \( \approx 10 \) certificates
- July 2011: DigiNotar CA hacked
  - 531 fake certificates *in the wild*
How This Got Our Interest

PKI weaknesses since 2008

- Early December 2008:
  - ‘Error’ in Comodo CA: no identity check
  - Whitehack hacks StartSSL CA
- February 2009
  - ‘Easy’ attack on MD5: fake CA certificate
- March 2011: Comodo CA hacked
  - Blacklisting of ≈ 10 certificates
- July 2011: DigiNotar CA hacked
  - 531 fake certificates in the wild
CAs in Root Store

Root Stores Contain CA Certificates

Root Store

CA 1

Root Store

CA 2

CA 3

Ralph Holz: The sorry state of X.509
Your browser chooses the ‘trusted CAs’. Not you.

Any CA may issue a certificate for any domain.

This means the weakest CA determines the strength of the whole PKI.
Development of Mozilla Root Store

Up to 150 trustworthy Root Certificates

* Number dropping again
Can we Assess the Quality of this PKI?

A good PKI should

- allow HTTPS on all WWW hosts
- contain only valid certificates
- offer good cryptographic security
  - Long keys, only strong hash algorithms, ...
- have a sensible setup
  - Short validity periods (1 year)
  - Short certificate chains (but use intermediate certificates)
  - Number of issuers should be reasonable (weakest link!)
Acquiring Our Data Sets

Active scans to measure *deployed* PKI

- Scan hosts on Alexa Top 1 million Web sites
- Nov 2009 – Apr 2011: scanned 8 times from Germany
- March 2011: scans from 8 hosts around the globe

Passive monitoring to measure *user-encountered* PKI

- Munich Research Network, monitored all SSL/TLS traffic
- Two 2-week runs in Sep 2010 and Apr 2011

EFF scan of IPv4 space in 2010

- Scan of 2-3 months, no *domain* information
## Our Data Sets

### Active Scans

<table>
<thead>
<tr>
<th>Location</th>
<th>Time (run)</th>
<th>Type</th>
<th>Certificates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuebingen, DE</td>
<td>November 2009</td>
<td>Active scan</td>
<td>833,661</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>December 2009</td>
<td>Active scan</td>
<td>819,488</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>January 2010</td>
<td>Active scan</td>
<td>816,517</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>April 2010</td>
<td>Active scan</td>
<td>816,605</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>September 2010</td>
<td>Active scan</td>
<td>829,232</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>November 2010</td>
<td>Active scan</td>
<td>827,366</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Active scan</td>
<td>829,707</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Active scan with SNI</td>
<td>826,098</td>
</tr>
<tr>
<td>Shanghai, CN</td>
<td>April 2011</td>
<td>Active scan</td>
<td>798,976</td>
</tr>
<tr>
<td>Beijing, CN</td>
<td>April 2011</td>
<td>Active scan</td>
<td>797,046</td>
</tr>
<tr>
<td>Melbourne, AU</td>
<td>April 2011</td>
<td>Active scan</td>
<td>833,571</td>
</tr>
<tr>
<td>İzmir, TR</td>
<td>April 2011</td>
<td>Active scan</td>
<td>825,555</td>
</tr>
<tr>
<td>São Paulo, BR</td>
<td>April 2011</td>
<td>Active scan</td>
<td>833,246</td>
</tr>
<tr>
<td>Moscow, RU</td>
<td>April 2011</td>
<td>Active scan</td>
<td>830,765</td>
</tr>
<tr>
<td>Santa Barbara, US</td>
<td>April 2011</td>
<td>Active scan</td>
<td>834,173</td>
</tr>
<tr>
<td>Boston, US</td>
<td>April 2011</td>
<td>Active scan</td>
<td>834,054</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>September 2010</td>
<td>Passive monitoring</td>
<td>183,208</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Passive monitoring</td>
<td>989,040</td>
</tr>
<tr>
<td>EFF servers</td>
<td>March–June 2010</td>
<td>Active IPv4 scan</td>
<td>11,349,678</td>
</tr>
</tbody>
</table>

25 million certificates to evaluate.
## Our Data Sets

### Active Scans

<table>
<thead>
<tr>
<th>Location</th>
<th>Time (run)</th>
<th>Type</th>
<th>Certificates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuebingen, DE</td>
<td>November 2009</td>
<td>Active scan</td>
<td>839,661</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>December 2009</td>
<td>Active scan</td>
<td>819,488</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>January 2010</td>
<td>Active scan</td>
<td>816,517</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>April 2010</td>
<td>Active scan</td>
<td>816,605</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>September 2010</td>
<td>Active scan</td>
<td>829,232</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>November 2010</td>
<td>Active scan</td>
<td>827,366</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Active scan</td>
<td>829,707</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Active scan with SNI</td>
<td>826,098</td>
</tr>
<tr>
<td>Shanghai, CN</td>
<td>April 2011</td>
<td>Active scan</td>
<td>798,976</td>
</tr>
<tr>
<td>Beijing, CN</td>
<td>April 2011</td>
<td>Active scan</td>
<td>797,046</td>
</tr>
<tr>
<td>Melbourne, AU</td>
<td>April 2011</td>
<td>Active scan</td>
<td>833,571</td>
</tr>
<tr>
<td>Ízmir, TR</td>
<td>April 2011</td>
<td>Active scan</td>
<td>825,555</td>
</tr>
<tr>
<td>São Paulo, BR</td>
<td>April 2011</td>
<td>Active scan</td>
<td>833,246</td>
</tr>
<tr>
<td>Moscow, RU</td>
<td>April 2011</td>
<td>Active scan</td>
<td>830,765</td>
</tr>
<tr>
<td>Santa Barbara, US</td>
<td>April 2011</td>
<td>Active scan</td>
<td>834,173</td>
</tr>
<tr>
<td>Boston, US</td>
<td>April 2011</td>
<td>Active scan</td>
<td>834,054</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>September 2010</td>
<td>Passive monitoring</td>
<td>183,208</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Passive monitoring</td>
<td>989,040</td>
</tr>
<tr>
<td>EFF servers</td>
<td>March–June 2010</td>
<td>Active IPv4 scan</td>
<td>11,349,678</td>
</tr>
</tbody>
</table>

25 million certificates to evaluate.

Ralph Holz: The sorry state of X.509
## Our Data Sets

### Active Scans — Passive Monitoring — EFF IPv4 scan

<table>
<thead>
<tr>
<th>Location</th>
<th>Time (run)</th>
<th>Type</th>
<th>Certificates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuebingen, DE</td>
<td>November 2009</td>
<td>Active scan</td>
<td>839,661</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>December 2009</td>
<td>Active scan</td>
<td>819,488</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>January 2010</td>
<td>Active scan</td>
<td>816,517</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>April 2010</td>
<td>Active scan</td>
<td>816,605</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>September 2010</td>
<td>Active scan</td>
<td>829,232</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>November 2010</td>
<td>Active scan</td>
<td>827,366</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Active scan</td>
<td>829,707</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Active scan with SNI</td>
<td>826,098</td>
</tr>
<tr>
<td>Shanghai, CN</td>
<td>April 2011</td>
<td>Active scan</td>
<td>798,976</td>
</tr>
<tr>
<td>Beijing, CN</td>
<td>April 2011</td>
<td>Active scan</td>
<td>797,046</td>
</tr>
<tr>
<td>Melbourne, AU</td>
<td>April 2011</td>
<td>Active scan</td>
<td>833,571</td>
</tr>
<tr>
<td>Ízmir, TR</td>
<td>April 2011</td>
<td>Active scan</td>
<td>825,555</td>
</tr>
<tr>
<td>São Paulo, BR</td>
<td>April 2011</td>
<td>Active scan</td>
<td>833,246</td>
</tr>
<tr>
<td>Moscow, RU</td>
<td>April 2011</td>
<td>Active scan</td>
<td>830,765</td>
</tr>
<tr>
<td>Santa Barbara, US</td>
<td>April 2011</td>
<td>Active scan</td>
<td>834,173</td>
</tr>
<tr>
<td>Boston, US</td>
<td>April 2011</td>
<td>Active scan</td>
<td>834,054</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>September 2010</td>
<td>Passive monitoring</td>
<td>183,208</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Passive monitoring</td>
<td>989,040</td>
</tr>
</tbody>
</table>

EFF servers March–June 2010 Active IPv4 scan 11,349,678

25 million certificates to evaluate.
## Our Data Sets

### Active Scans — Passive Monitoring — EFF IPv4 scan

<table>
<thead>
<tr>
<th>Location</th>
<th>Time (run)</th>
<th>Type</th>
<th>Certificates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuebingen, DE</td>
<td>November 2009</td>
<td>Active scan</td>
<td>833,661</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>December 2009</td>
<td>Active scan</td>
<td>819,488</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>January 2010</td>
<td>Active scan</td>
<td>816,517</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>April 2010</td>
<td>Active scan</td>
<td>816,605</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>September 2010</td>
<td>Active scan</td>
<td>829,232</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>November 2010</td>
<td>Active scan</td>
<td>827,366</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Active scan</td>
<td>829,707</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Active scan with SNI</td>
<td>826,098</td>
</tr>
<tr>
<td>Shanghai, CN</td>
<td>April 2011</td>
<td>Active scan</td>
<td>798,976</td>
</tr>
<tr>
<td>Beijing, CN</td>
<td>April 2011</td>
<td>Active scan</td>
<td>797,046</td>
</tr>
<tr>
<td>Melbourne, AU</td>
<td>April 2011</td>
<td>Active scan</td>
<td>833,571</td>
</tr>
<tr>
<td>İzmir, TR</td>
<td>April 2011</td>
<td>Active scan</td>
<td>825,555</td>
</tr>
<tr>
<td>São Paulo, BR</td>
<td>April 2011</td>
<td>Active scan</td>
<td>833,246</td>
</tr>
<tr>
<td>Moscow, RU</td>
<td>April 2011</td>
<td>Active scan</td>
<td>830,765</td>
</tr>
<tr>
<td>Santa Barbara, US</td>
<td>April 2011</td>
<td>Active scan</td>
<td>834,173</td>
</tr>
<tr>
<td>Boston, US</td>
<td>April 2011</td>
<td>Active scan</td>
<td>834,054</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>September 2010</td>
<td>Passive monitoring</td>
<td>183,208</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Passive monitoring</td>
<td>989,040</td>
</tr>
<tr>
<td>EFF servers</td>
<td>March–June 2010</td>
<td>Active IPv4 scan</td>
<td>11,349,678</td>
</tr>
</tbody>
</table>

25 million certificates to evaluate.
## Our Data Sets

### Active Scans — Passive Monitoring — EFF IPv4 scan

<table>
<thead>
<tr>
<th>Location</th>
<th>Time (run)</th>
<th>Type</th>
<th>Certificates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuebingen, DE</td>
<td>November 2009</td>
<td>Active scan</td>
<td>833,661</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>December 2009</td>
<td>Active scan</td>
<td>819,488</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>January 2010</td>
<td>Active scan</td>
<td>816,517</td>
</tr>
<tr>
<td>Tuebingen, DE</td>
<td>April 2010</td>
<td>Active scan</td>
<td>816,605</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>September 2010</td>
<td>Active scan</td>
<td>829,232</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>November 2010</td>
<td>Active scan</td>
<td>827,366</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Active scan</td>
<td>829,707</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Active scan with SNI</td>
<td>826,098</td>
</tr>
<tr>
<td>Shanghai, CN</td>
<td>April 2011</td>
<td>Active scan</td>
<td>798,976</td>
</tr>
<tr>
<td>Beijing, CN</td>
<td>April 2011</td>
<td>Active scan</td>
<td>797,046</td>
</tr>
<tr>
<td>Melbourne, AU</td>
<td>April 2011</td>
<td>Active scan</td>
<td>833,571</td>
</tr>
<tr>
<td>İzmir, TR</td>
<td>April 2011</td>
<td>Active scan</td>
<td>825,555</td>
</tr>
<tr>
<td>São Paulo, BR</td>
<td>April 2011</td>
<td>Active scan</td>
<td>833,246</td>
</tr>
<tr>
<td>Moscow, RU</td>
<td>April 2011</td>
<td>Active scan</td>
<td>830,765</td>
</tr>
<tr>
<td>Santa Barbara, US</td>
<td>April 2011</td>
<td>Active scan</td>
<td>834,173</td>
</tr>
<tr>
<td>Boston, US</td>
<td>April 2011</td>
<td>Active scan</td>
<td>834,054</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>September 2010</td>
<td>Passive monitoring</td>
<td>183,208</td>
</tr>
<tr>
<td>Munich, DE</td>
<td>April 2011</td>
<td>Passive monitoring</td>
<td>989,040</td>
</tr>
<tr>
<td>EFF servers</td>
<td>March–June 2010</td>
<td>Active IPv4 scan</td>
<td>11,349,678</td>
</tr>
</tbody>
</table>

25 million certificates to evaluate.
Validity of End-Hosts Certificates

Root Store

CA1

R1

I1

E1

E2

I4

E5

E6

I5

I6

E3

E4

I2

R2

CA2

I3

E7

R3

CA3

Host Certificates
Validation of Certificate Chains

Just check chains, not host names

- Chain valid
- Expired
- Self-signed end host certificate
- Root certificate not in root store
- No root certificate found

% of all certificates

0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%

Ralph Holz: The sorry state of X.509
Correct Domain Name in Certificate

Now also check host names

- Look in Common Name (CN) and Subject Alternative Name (SAN)
- Munich, April 2011, only valid chains:
  - 12.2% correct CN
  - 5.9% correct SAN

Only 18% of certificates are fully verifiable

- Positive ‘trend’: from 14.9% in 2009 to 18% in 2011
- Addendum: recent scans show this is increasing (+ 0.5%)
Host Names in Self-signed Certificates

Active scan

- **2.2%** correct Common Name (CN)
- **0.5%** correct Subject Alternative Name

Top 3 most frequent CNs account for > 50%

- **plesk** or similar in 27.3%
- **localhost** or similar in 25.4% – standard installations?
Certificate Occurrences

Many certificates valid for more than one domain

- Domains served by same IP
- Some certificates issued for dozens of domains
- Certificate reuse on multiple machines increases options for attacker

Often found on hosters

- E.g. *.blogger.com, *.wordpress.com
How often does a certificate occur on $X$ hosts?

- Number of hosts per certificate $= X$
- $\Pr[\text{#hosts} > X]$
  - $1e^{-5}$
  - $1e^{-4}$
  - 0.001
  - 0.01
  - 0.1

- All certificates
- Only valid certificates
CDF of validity periods, active scans

Validity (years)

Pr[X < validity]

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Nov2009
Apr2011
Public Key Properties

Key types

- RSA: 99.98% (rest is DSA)
- About 50% have length 1,024 bit
- About 45% have length 2,048 bit
- Clear trend from 1,024 to 2,048 bit

Weird encounters

- 1,504 distinct certificates that share another certificate’s key
- Many traced to a handful of hosting companies
- Nadiah Henninger’s work: Embedded devices, poor entropy!
- www.factorable.net
MD5 is being phased out

![Graph showing the prevalence of signature algorithms over time.](image)

**Signature Algorithms**

**MD5 is being phased out**

- Active scans from Germany
- Monitoring
- Active scans from China

<table>
<thead>
<tr>
<th>Date</th>
<th>RSA/SHA1</th>
<th>RSA/MD5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1/2010</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>1/1/2011</td>
<td>90%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Ralph Holz: The sorry state of X.509
Certificate Quality

We defined 3 categories

- ‘Good’:
  - Correct chains, correct host name
  - Chain $\leq 2$
  - No MD5, strong key of $> 1024$ bit
  - Validity $\leq 13$ months

- ‘Acceptable’
  - Chain $\leq 3$, validity $\leq 25$ months
  - Rest as above

- ‘Poor’: the remainder
Validity correlates with rank

- Share of ‘poor’ certificates higher among high-ranking sites

Ralph Holz: The sorry state of X.509
Proposals to enhance or replace X.509
What to do about these problems?

No silver bullet known

- Part of the problem: SSL meant to protect stuff like credit card numbers
- But state-scale attacks were not in scope back in the 1990s

Several proposals:

- Extended Validation, Base Line Requirements
- Pinning Information
- Keys in DNSSEC (DANE)
- Perspectives/Convergence
- Public Logs: Sovereign Keys, Certificate Transparency
Extended Validation

- CAs to require state-issued documents before certification
- More expensive
- Rarely bought by customers

Base Line Requirements

- CA/Browser forum standard
- Absolute minimum requirements for validation
- Audit-based, rules for audits
Idea

- Browser stores last-seen public key of a site
- Alternatively: store issuing CA
- Recognise again upon next visit

Discussion

- Does not help against attack on first contact
- False alarms when certificates change (not rare!)
- How many certs to store?
Keys in DNSSEC: DANE

Idea

- DNSSEC already a hierarchical **state-level** PKI
- Verification from Root Server down to end-host
- New Resource Record in DNSSEC: public key of site

Discussion

- Straight-forward and strong
- Performance? Caching? DJB says it’s poor.
- Countries control their own TLDs. Think bit.ly!
- Defence against country-level attack?
Idea: Notaries
Perspectives and Convergence

Reconfirm with notaries
Discussion of Notary System

Advantages

- Works well against MitM in the network
- Reinforcement or replacement of CA system?

Possible problems

- Privacy
- False positives: some sites change certificates frequently
- Content Distribution Networks?
Store information
Retrieve information
Public Log Schemes

Idea: store information publicly and append-only

- **Sovereign Keys**
  - Sites stores authoritative key to cross-sign its certificates
  - Goal: cross-certification and cross-validation of keys

- **Certificate Transparency**
  - CAs and others store info about who is certified by whom
  - Goal: detect rogue CA issuing key for a site

Schemes are very new - end of 2011
Sites store information on < 30 timeline servers

<table>
<thead>
<tr>
<th>timestamp</th>
<th>name</th>
<th>key</th>
<th>protocols</th>
<th>evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1322736203</td>
<td>A</td>
<td>0x427E8A</td>
<td>https, smtps</td>
<td>$\text{Sig}_{CA}(A, \ldots)$</td>
</tr>
<tr>
<td>1323254603</td>
<td>B</td>
<td>0x7389FB</td>
<td>https:8080</td>
<td>$\text{Sig}_{B}(B, \ldots)$</td>
</tr>
<tr>
<td>1323657143</td>
<td>C</td>
<td>0x49212A</td>
<td>imaps</td>
<td>$\text{Sig}_{C}(C, \ldots)$</td>
</tr>
<tr>
<td>1413787143</td>
<td>A</td>
<td>0x427E8A</td>
<td>https, smtps</td>
<td>$\text{Sig}_{CA}(A, \ldots)$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Work-in-progress (alive)

- Timeline is auditable by clients
- Mirrors proposed
- https://www.eff.org/sovereign-keys
Sovereign Keys: Discussion

Pros

- Does not need CA support
- Evidence can be based on DANE DNSSEC, CAs, ...
- Performance and bandwidth?

Cons

- Continuous monitoring of timeline server needed
- Maintain list of timeline servers
- Entries are not space-efficient
- Privacy (suggested remedy: TOR-like proxying)
- Key loss
Certificate Transparency (Google)

Store certification proof on public servers

<table>
<thead>
<tr>
<th>timestamp</th>
<th>name</th>
<th>cert</th>
<th>evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1322736203</td>
<td>A</td>
<td>Cert chain by Verisign</td>
<td>$MSig$($hashes$)</td>
</tr>
<tr>
<td>1323254603</td>
<td>B</td>
<td>Self-signed cert</td>
<td>$MSig$($hashes$)</td>
</tr>
<tr>
<td>1323657143</td>
<td>C</td>
<td>Cert by CACert</td>
<td>$MSig$($hashes$)</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Work-in-progress (alive)

- Timeline consistency can be monitored
- Roles: clients, auditors, monitors (on-behalf)
Figure: Log is a Merkle tree, $d_i$ are new certificate chains.
Discussion of Certificate Transparency

Pros

- Protects against rogue/hacked CAs
- Efficient data structure
- Has Google campaigning for it

Cons

- Requires continuous monitoring of logs
- Monitors need full log at all times, act on behalf of others
- Proofs are $O(\log n)$, but storage is linear
Crossbear: Detecting and Localising the MitM
Case study 1: Syria vs. Facebook?

A Syrian Man-In-The-Middle Attack against Facebook

UPDATE: If you are in Syria and your browser shows you this certificate warning on Facebook, it is not safe to login to Facebook. You may wish to use Tor to connect to Facebook, or use proxies outside of Syria.

Certificate:
Data:
  Issuer: C=US, ST=California, L=Alto Palo, O=Facebook, Inc., OU=Facebook, CN=s.static.ak.facebook.com
  Subject: C=US, ST=California, L=Alto Palo, O=Facebook, Inc., OU=Facebook, CN=s.static.ak.facebook.com
[...] I spent the night in a hotel in Warsaw, Poland. I bought access to the Internet WiFi. [...]

For the SSL connection to imap.googlemail.com (also known as imap.gmail.com) at port 993, Thunderbird warned me about a certificate error. The certificate presented by IP address 74.125.115.16 was issued to
Subject: C=US, ST=California, L=Mountain View, O=Google Inc, CN=imap.googlemail.com
and it was issued by
Issuer: C=US, ST=California, L=Sunnyvale, O=Fortinet, OU=Certificate Authority, CN=FortiGate CA/emailAddress=support@fortinet.com

Actually attaches a traceroute.
Case study 3: DigiNotar vs. Iran?

Invalid Server Certificate

You attempted to reach www.google.com, but the server presented an invalid certificate.

Back

Help me understand

When you connect to a secure website, the server hosting that site presents your browser with something called a server certificate. This certificate contains important information, such as the address of the website, which is verified by a third party. This third party is called a certificate authority (CA). When you connect to a secure website, the server presents its certificate to your browser, which checks that the address in the certificate matches the address of the website. If it does, your browser can be confident that you are connecting to the website you intended, and not a third party (such as an attacker on your network).

In this case, the server certificate or an intermediate CA certificate presented to your browser is invalid. This certificate is not trusted by your browser because it is not part of its trusted list of CAs, or it is part of a list ofRevoked CAs. This certificate is not valid, contains invalid fields, or is not supported.
This is *not* a proposal to strengthen X.509.

Crossbear: a tool to gather *hard data*.

- Raise reliable data about MitM *in the wild*
- *How often* do MitM occur?
- *Where* are the attackers located?
- *Who* are the attackers?
- Are we jumping at shadows?

Method: combine notary principle, tracing and centralised reporting and analysis.
Alice is surfing...
Man-in-the-middle
Alice queries Crossbear
Crossbear checks the server
Crossbear reports result
Alice traceroutes to server
Alice reports to Crossbear
Distribute hunting tasks
Bob goes hunting
Bob reports
There are many Bobs
Crossbear ecosystem

Implemented and running

Server under MitM attack (V)
Victim client Alice (A)
Client Charlie (C)
Certificate database
Crossbear server (S)
Observation database
HuntingTask database
Attacker Mallory (M)
Client Bob (B)
Client Dave (D)

Ralph Holz: The sorry state of X.509
Components

Server: store and analyse

- Crossbear server at TU München, Germany
- Uses Convergence project’s notaries for diversity
- Server cert hard-coded into client!

Detection and localisation

- Clients as Firefox add-on (detection and localisation)
- 150 stand-alone hunters on stand-by on PlanetLab (localisation)
Verification request

User

HTTP REQUEST

CertVerifyRequest
   - Observed certificate chain
   - Host from which that chain was observed

HTTP RESPONSE

CertVerifyResult
   - Report and score for the certificate

HuntingTask [optional]
   - Hashes of already known certificate chains
   - Target

PublicIPNotification [optional]
   - Public IP of the user
   - HMAC on that public IP

ServerTime [optional]
   - Timestamp of the server’s current local time

Crossbear server

NB: SSL-secured connection, server cert hard-coded
Hunter reply

NB: SSL-secured connection, server cert hard-coded

Ralph Holz: The sorry state of X.509
Actually, we also determine on server-side:

- CAs used in certificate chain (→ continuity)
- AS number of hosts in traceroute (→ frequent reports?)
- Geo data: location of hosts in traceroute (→ traversed countries)
- WHOIS info

Firefox add-on

- For savvy users
- Score-based, several factors
- UI → see code on github
Attacker model

Chosen based on MitM reports we have

Attacker behaviour

- Non-selective: MitM all attached ‘client’ systems
- Selective: MitM only some of attached ‘client’ systems

Attacker position

- Towards periphery, close to victim client
- Towards periphery, close to victim server
- Central location in network (important AS, ...)

Ralph Holz: The sorry state of X.509
Non-selective, close to victim client
Non-selective, state-level attacker
Analysis: non-selective attacker

Detection
- Attack is detected if $\geq 1$ reports
- Attacker can only drop connections to Crossbear server

Lends itself well to localisation
- Get $\geq 1$ traceroute from victim, $\geq 1$ from unpoisoned hunter
- The more, the better
- The closer to intersection point, the better
- Success depends on the number of hunters
- An estimate can be given
Assumptions

Possible to give a closed-form model, at both router and AS-level.

- Non-selective attacker
- Routing symmetric (OK if path lengths not much different)
- Routing based only on destination address (hot-potato routing etc. rare)
- Probability for a node (router, AS) as location for hunter is evenly distributed (*)
- Probability that traffic is forwarded to specific neighbour is evenly distributed (*)
Closed-form model

Ideas

- First, treat path length victim ↔ server as fixed length
- Probability that randomly placed hunter covers a node is function of node degree
- Probability that one hunter is victim and another just escapes MitM is function of node degree
- Aggregate: sum over all possible path lengths, and all possible locations of attacker on path
- Model only depends on node degrees and path lengths
Input data

Router level

- Node degrees: Rocketfuel
- Path lengths: number of IP hops: traceroutes to 30k random hosts

AS level

- Node degree and path lengths: RouteViews archive
Number of hunters vs. success

![Graph showing the relationship between number of hunters and attacker localisation probability.](image)

- **AS**
- **Router (uncertainty=2)**
- **Router (uncertainty=1)**
- **Router (exact)**

Ralph Holz: The sorry state of X.509
Selective attacker: close to victim
Selective attacker: in core

Ralph Holz: The sorry state of X.509
Selective attackers are a headache

Every attack report to be checked for plausibility

But attacker should leave some hints – cannot arbitrarily spoof IP addresses

Can be indistinguishable from non-selective attacks

- *Every* attack report to be checked for plausibility
- But attacker should leave some hints – cannot arbitrarily spoof IP addresses
Patterns to look for

**Attack seems to be restricted to few stub AS**

- Use BGP data to check traceroutes for plausibility
- Do MitM certificates share properties?
- Which AS in which countries involved? Some ‘known’ suspect?

**MitM reports from just a few companies?**

- Check traceroutes for traversed countries and AS
- Might be industrial espionage

**All of this is intensive manual work. But only localisation is affected, and it is better than no data all.**
Crossbear is an open system

- Malicious injection of data
  - Clients/hunters have no ID, no authentication
  - Attacker can eclipse real hunters in his network, too
  - Should results in clusters of suspicious reports, though

- Denial-of-service attacks

- It is an arms race

- Other detection systems are subject to same attacks
Our position

A first step towards gathering better data

- We do not advertise Crossbear as a silver bullet
- Best results can be expected against the non-selective attacker
- These are also the attackers we are most interested in

Crossbear is deployed and ready

- 150 hunters on PlanetLab
- 4,000 certificate reports – no MitM

Integration with OONI

- Tor’s Open Observatory of Network Interference
Thank you!

Contact

- Twitter: @crossbearteam
- WWW: https://pki.net.in.tum.de
- https://github.com/crossbear/Crossbear
Backup Slides
Errors in TLS Connection Setup

Scans from Germany, Nov 2009 and Apr 2011

% of all connections

- Other failure
- Unknown protocol
- Success

Ralph Holz: The sorry state of X.509
UNKNOWN PROTOCOL

- Rescanned those hosts and manual sampling
- Always plain HTTP...
- ... and always an index.html with HTML 2 ...
- Hypothesis: old servers, old configurations
- More likely to happen in the lower ranks
Unusual Host Names

**CN=plesk or similar**

- Found in 7.3% of certificates
- Verified: Plesk/Parallels panels

**CN=localhost**

- 4.7% of certificates
Symmetric Ciphers

Results from monitoring

(Mostly) in line with results from 2007 by Lee et al.

- Order of AES and RC4 has shifted, RC4-128 most popular
Debian Weak Keys

Weak randomness in key generation – serious bug of 2008

In line with findings of 2009 by Yilek et al.
CDF for RSA key lengths – linear Y axis
Public Key Lengths

CDF for RSA key lengths – double-log Y axis

Key length (bits)
Pr[X < length]

256 512 1024 2048 4096 8192

Tue–Nov2009
TUM–Sep2010
TUM–April2011

Ralph Holz: The sorry state of X.509
Certificate Occurrences

Most frequent Common Name occurrences

- *.blogger.com
- *.bluehost.com
- *.hostgator.com
- *.blogger.com
- www.snakeoil.dom
- *.hostmonster.com
- *.wordpress.com

Number of occurrences

<table>
<thead>
<tr>
<th>Domain</th>
<th>Number of Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>*.blogger.com</td>
<td>20000</td>
</tr>
<tr>
<td>*.bluehost.com</td>
<td>19000</td>
</tr>
<tr>
<td>*.hostgator.com</td>
<td>18000</td>
</tr>
<tr>
<td>*.blogger.com</td>
<td>16000</td>
</tr>
<tr>
<td><a href="http://www.snakeoil.dom">www.snakeoil.dom</a></td>
<td>12000</td>
</tr>
<tr>
<td>*.hostmonster.com</td>
<td>10000</td>
</tr>
<tr>
<td>*.wordpress.com</td>
<td>10000</td>
</tr>
</tbody>
</table>
Certificate Chains


certificate chains

Root Store
CA

Intermediate Certificates

CA

Ralph Holz: The sorry state of X.509
Certificate Chain Lengths

Finding more positive than negative:

- Trend to use intermediate certificates more often
- Allows to keep Root Certificates offline
- But chains still reasonably short
Certificate Issuers

Very few CAs account for > 50% of certificates

But there are 150+ Root Certificates in Mozilla.

Ralph Holz: The sorry state of X.509