Chair of Network Architectures and Services
Department of Informatics
Technical University of Munich

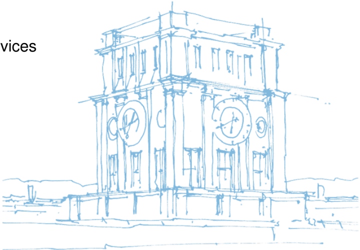
In Log We Trust: Revealing Poor Security Practices with Certificate Transparency Logs and Internet Measurements

Oliver Gasser, Benjamin Hof, Max Helm, Maciej Korczynski, Ralph Holz, Georg Carle

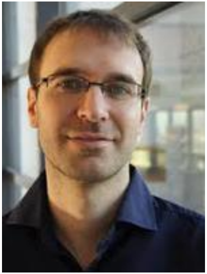
Tuesday 27th March, 2018

Passive and Active Measurement Conference 2018
Berlin, Germany

Chair of Network Architectures and Services
Department of Informatics
Technical University of Munich



Joint work



Why should I care about CT?

What is Certificate Transparency (CT) in a nutshell?

- CT provides a repository of certificates to make misissuance detectable
- Pushed by Google and others
- RFC 6962

Why should I care about CT?

What is Certificate Transparency (CT) in a nutshell?

- CT provides a repository of certificates to make misissuance detectable
- Pushed by Google and others
- RFC 6962

Why is CT interesting for researchers?

- CT offers a timeline of issued certificates
- Allows to analyze current state and evolution of certificate ecosystem

Why should I care about CT?

What is Certificate Transparency (CT) in a nutshell?

- CT provides a repository of certificates to make misissuance detectable
- Pushed by Google and others
- RFC 6962

Why is CT interesting for researchers?

- CT offers a timeline of issued certificates
- Allows to analyze current state and evolution of certificate ecosystem

Why do we need measurements?

- Not all certificates are in CT (yet)
- Find discrepancies between certificates in CT and certificates deployed in the wild

Why should I care about CT?

What is Certificate Transparency (CT) in a nutshell?

- CT provides a repository of certificates to make misissuance detectable
- Pushed by Google and others
- RFC 6962

Why is CT interesting for researchers?

- CT offers a timeline of issued certificates
- Allows to analyze current state and evolution of certificate ecosystem

Why do we need measurements?

- Not all certificates are in CT (yet)
- Find discrepancies between certificates in CT and certificates deployed in the wild

What if I don’t care about security at all?

- Wait for the bonus slide at the end

Certificate Transparency

What problem is CT trying to solve?

- Misissued certificates pose a threat to TLS security
  - Example: DigiNotar hack in 2011 resulted in unauthorized certificate issuance
- Timely detection of misissued certificates is hard
  - Domain owner or CA might not be aware of misissuance
  - CA might not go public about misissuance
- Idea: All CAs publish a list of issued certificates
  - Others can then look at those lists and detect misissued certificates

Certificate Transparency

What problem is CT trying to solve?

- Misissued certificates pose a threat to TLS security
  - Example: DigiNotar hack in 2011 resulted in unauthorized certificate issuance
- Timely detection of misissued certificates is hard
  - Domain owner or CA might not be aware of misissuance
  - CA might not go public about misissuance
- Idea: All CAs publish a list of issued certificates
  - Others can then look at those lists and detect misissued certificates

Involved parties in CT

- Log: Public, untrusted, append-only certificate store
- Monitor: Service evaluating certificates found in logs
- Auditor

Certificate Transparency

What problem is CT trying to solve?

- Misissued certificates pose a threat to TLS security
  - Example: DigiNotar hack in 2011 resulted in unauthorized certificate issuance
- Timely detection of misissued certificates is hard
  - Domain owner or CA might not be aware of misissuance
  - CA might not go public about misissuance
- Idea: All CAs publish a list of issued certificates
  - Others can then look at those lists and detect misissued certificates

Involved parties in CT

- Log: Public, untrusted, append-only certificate store → data source for this work
- Monitor: Service evaluating certificates found in logs
- Auditor

Certificate Transparency

What problem is CT trying to solve?

- Misissued certificates pose a threat to TLS security
  - Example: DigiNotar hack in 2011 resulted in unauthorized certificate issuance
- Timely detection of misissued certificates is hard
  - Domain owner or CA might not be aware of misissuance
  - CA might not go public about misissuance
- Idea: All CAs publish a list of issued certificates
  - Others can then look at those lists and detect misissued certificates

Involved parties in CT

- Log: Public, untrusted, append-only certificate store → data source for this work
- Monitor: Service evaluating certificates found in logs → us
- Auditor

Measurement methodology

Active measurements

- 600 M certificates downloaded from 30 CT logs
- Active HTTPS scans of more than 200 M IPv4 and IPv6 hosts
- Certificate Revocation List downloads resulting in 25 M entries

Measurement methodology

Active measurements

- 600 M certificates downloaded from 30 CT logs
- Active HTTPS scans of more than 200 M IPv4 and IPv6 hosts
- Certificate Revocation List downloads resulting in 25 M entries

Performing measurements in an ethical way

- Don’t annoy other people and take away their precious time
  - Limit query rate
  - Use incremental downloads for CT logs
  - Use conforming packets/requests
- Don’t hide your intentions
  - Use dedicated measurement machine
  - Informing rDNS name, WHOIS entry, web site explaining measurements

Primary analysis goals

What we wanted to find out:

1. Who are the issuers of certificates in CT logs?

Primary analysis goals

What we wanted to find out:

1. Who are the issuers of certificates in CT logs?
2. How secure are certificates in CT logs?

Primary analysis goals

What we wanted to find out:

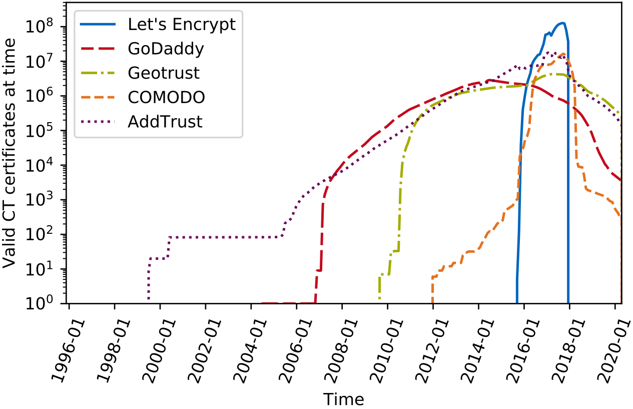
1. Who are the issuers of certificates in CT logs?
2. How secure are certificates in CT logs?
3. How do certificates in CT logs differ from those found in the wild?

Primary analysis goals

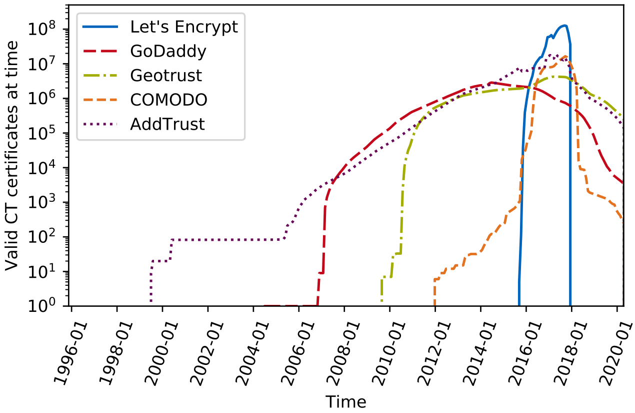
What we wanted to find out:

1. Who are the issuers of certificates in CT logs?
2. How secure are certificates in CT logs?
3. How do certificates in CT logs differ from those found in the wild?
4. Do we find old and non-HTTPS certificates in CT logs?

1. Who are the issuers of certificates in CT logs?



1. Who are the issuers of certificates in CT logs?



- Let’s Encrypt is the dominating issuer of CT log certificates
- Certificates in logs from before standardization of CT began

Insecure certificates

Baseline Requirements (BRs)

- Rules regarding certificates and issuing processes which CAs adhere to
- Devised within the CA/Browser Forum
- Each requirement has an enforcement date
- Example: RSA key size $\geq$ 2048 bits for certificates starting 2014

Insecure certificates

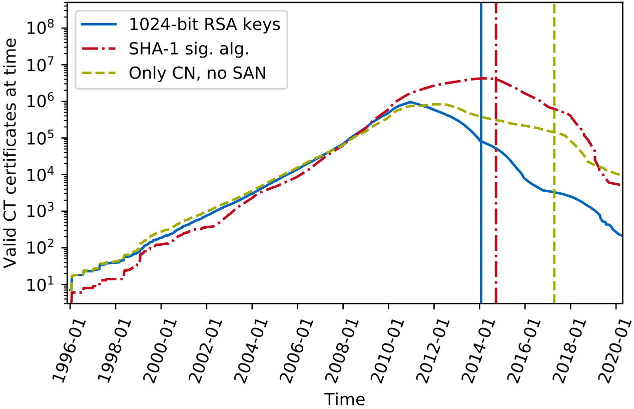
Baseline Requirements (BRs)

- Rules regarding certificates and issuing processes which CAs adhere to
- Devised within the CA/Browser Forum
- Each requirement has an enforcement date
- Example: RSA key size $\geq$ 2048 bits for certificates starting 2014

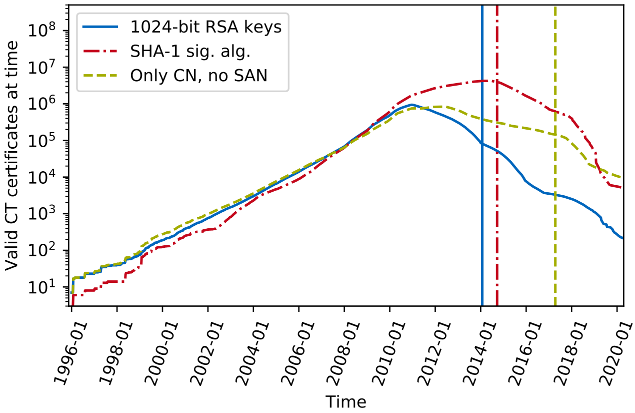
Analysis

- Analyze BR adherence of all collected certificates
- Use tool `cablint`
- Group violations into four categories
  - Identity (e.g. invalid domain in SAN)
  - Signature (e.g. SHA-1)
  - Keys (e.g. 1024 bit RSA key)
  - Time-validity (e.g. validity too long)

2. How secure are certificates in CT logs?

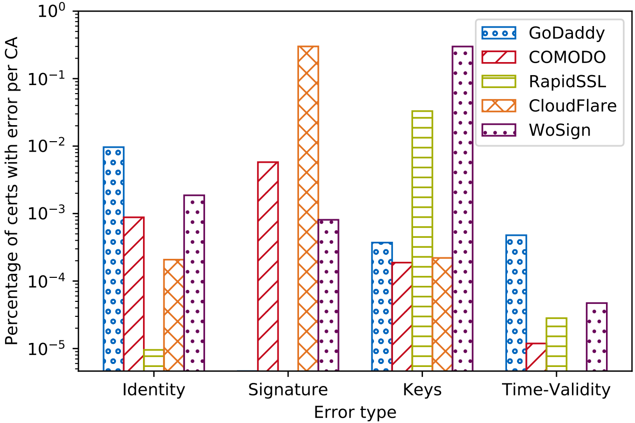


2. How secure are certificates in CT logs?

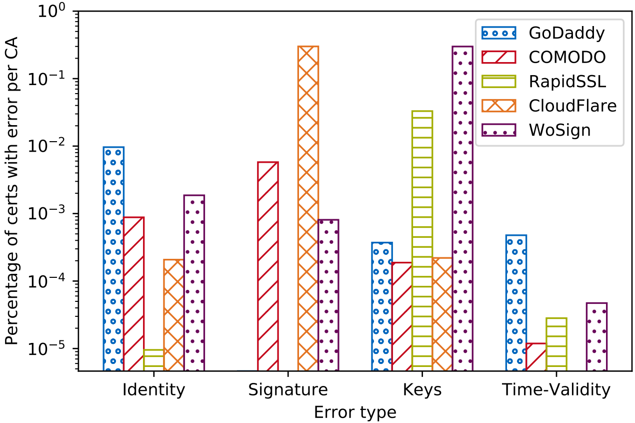


- Enforcement of stricter rules helps curb the number of insecure certificates
- But: Many insecure certificates remain in CT logs

BR violations per CA



BR violations per CA



- Some CAs with high violations in specific categories
- Let's Encrypt with no found violation

3. How do certificates in CT logs differ from those found in the wild?

Certificates

- CT logs: 216.8 M certificates
- Active scans: 128.1 M certificates

3. How do certificates in CT logs differ from those found in the wild?

Certificates

- CT logs: 216.8 M certificates
- Active scans: 128.1 M certificates

Overlap between CT logs and in the wild certificates

- 86 % of certificates in the wild are logged in CT
- Good milestone towards CT becoming mandatory in Chrome in April 2018

3. How do certificates in CT logs differ from those found in the wild?

Certificates

- CT logs: 216.8 M certificates
- Active scans: 128.1 M certificates

Overlap between CT logs and in the wild certificates

- 86 % of certificates in the wild are logged in CT
- Good milestone towards CT becoming mandatory in Chrome in April 2018

Baseline Requirements

- More adherence in CT logs (95 %) compared to in the wild (90 %)
- CT can help increase the security of certificates

4a. Do we find old certificates in CT logs?

Previous HTTPS scans

- Conducted between 2009 and 2015
- Targets: Alexa Top 1M and IPv4-wide

4a. Do we find old certificates in CT logs?

Previous HTTPS scans

- Conducted between 2009 and 2015
- Targets: Alexa Top 1M and IPv4-wide

Logged HTTPS certificates obtained from active scans over time

- 2009: 22 %
- 2015: 35 %

4a. Do we find old certificates in CT logs?

Previous HTTPS scans

- Conducted between 2009 and 2015
- Targets: Alexa Top 1M and IPv4-wide

Logged HTTPS certificates obtained from active scans over time

- 2009: 22 %
- 2015: 35 %
- 2017: 86 %
- Non-linear increase towards Google Chrome’s inclusion deadline

4b. Do we find non-HTTPS certificates in CT logs?

TLS scan focusing on messaging protocols

- Conducted in 2015
- TLS-enabled versions of SMTP, IMAP, POP3, FTP, XMPP, IRC

4b. Do we find non-HTTPS certificates in CT logs?

TLS scan focusing on messaging protocols

- Conducted in 2015
- TLS-enabled versions of SMTP, IMAP, POP3, FTP, XMPP, IRC

Non-HTTPS certificates in CT logs

- Overlap with certificates from HTTPS scan between 19 % (IRC) and 31 % (SMTP)
- Very low presence in CT logs
  - Highest: SMTP with 3.5 %
  - Lowest: XMPP with 2.0 %

4b. Do we find non-HTTPS certificates in CT logs?

TLS scan focusing on messaging protocols

- Conducted in 2015
- TLS-enabled versions of SMTP, IMAP, POP3, FTP, XMPP, IRC

Non-HTTPS certificates in CT logs

- Overlap with certificates from HTTPS scan between 19 % (IRC) and 31 % (SMTP)
- Very low presence in CT logs
  - Highest: SMTP with 3.5 %
  - Lowest: XMPP with 2.0 %
  - Much lower compared to 35 % of HTTPS
- CT focused on HTTPS certificates

Bonus slide: What if I am not interested in security at all?

CT logs as source for domains and IP addresses

- TUM's IPv6 hitlist available since 2016
- Extract domains from certificates in CT logs, resolve for IP addresses
- Adds 5.4 M IPv4 and 489 k IPv6 addresses
- Increase of 70 % of IPv6 addresses

Bonus slide: What if I am not interested in security at all?

CT logs as source for domains and IP addresses

- TUM's IPv6 hitlist available since 2016
- Extract domains from certificates in CT logs, resolve for IP addresses
- Adds 5.4 M IPv4 and 489 k IPv6 addresses
- Increase of 70 % of IPv6 addresses

We make our CT-extended IPv6 hitlist publicly available:

- `https://www.net.in.tum.de/pub/ipv6-hitlist/`
- Feel free to use it as a source for IPv6 addresses for your own research

Reproducible research

To encourage reproducibility in network measurement research we publish measurement tools, data, and analysis pipeline

- Data set: `https://mediatum.ub.tum.de/1422427`
- Source code: `https://github.com/tumi8/pam18-inlogwetrust`

Reproducible research

To encourage reproducibility in network measurement research we publish measurement tools, data, and analysis pipeline

- Data set: `https://mediatum.ub.tum.de/1422427`
- Source code: `https://github.com/tumi8/pam18-inlogwetrust`

Benefits

- Reproduce our results
- Conduct additional analyses on vast HTTPS data set
- Archive of the TUM University Library ensures long-term availability

150 Jahre
culture of
excellence

Conclusion

1. Who are the issuers of certificates in CT logs?
   - Let’s Encrypt issues most certificates found in CT logs
2. How secure are certificates in CT logs?
   - 900 k certificates violating Baseline Requirements, decreasing over time
3. How do certificates in CT logs differ from those found in the wild?
   - More adherence to BR of certificates in CT logs compared to active scans
4. Do we find old and non-HTTPS certificates in CT logs?
   - One fifth of certificates scanned in 2009 are in CT logs
   - Only a few percent of non-HTTPS certificates are logged

*** What if I am not interested in security at all?
   - Use our CT-extended hitlist for your IPv6 research

Conclusion

1. Who are the issuers of certificates in CT logs?
    - Let’s Encrypt issues most certificates found in CT logs
2. How secure are certificates in CT logs?
    - 900 k certificates violating Baseline Requirements, decreasing over time
3. How do certificates in CT logs differ from those found in the wild?
    - More adherence to BR of certificates in CT logs compared to active scans
4. Do we find old and non-HTTPS certificates in CT logs?
    - One fifth of certificates scanned in 2009 are in CT logs
    - Only a few percent of non-HTTPS certificates are logged

*** What if I am not interested in security at all?
    - Use our CT-extended hitlist for your IPv6 research

Oliver Gasser <gasser@net.in.tum.de>
`https://www.net.in.tum.de/~gasser/`