Longitudinal Analysis of Wildcard Certificates in the WebPKI

David Hasselquist[1,2], Ludvig Bolin[1], Emil Carlsson[1], Adam Hylander[1], Martin Larsson[1], Erik Voldstad[1], Niklas Carlsson[1]

[1] Linköping University, Sweden
[2] Sectra Communications, Sweden

Proc. IFIP Networking, Barcelona, Spain, June 2023

Certificates

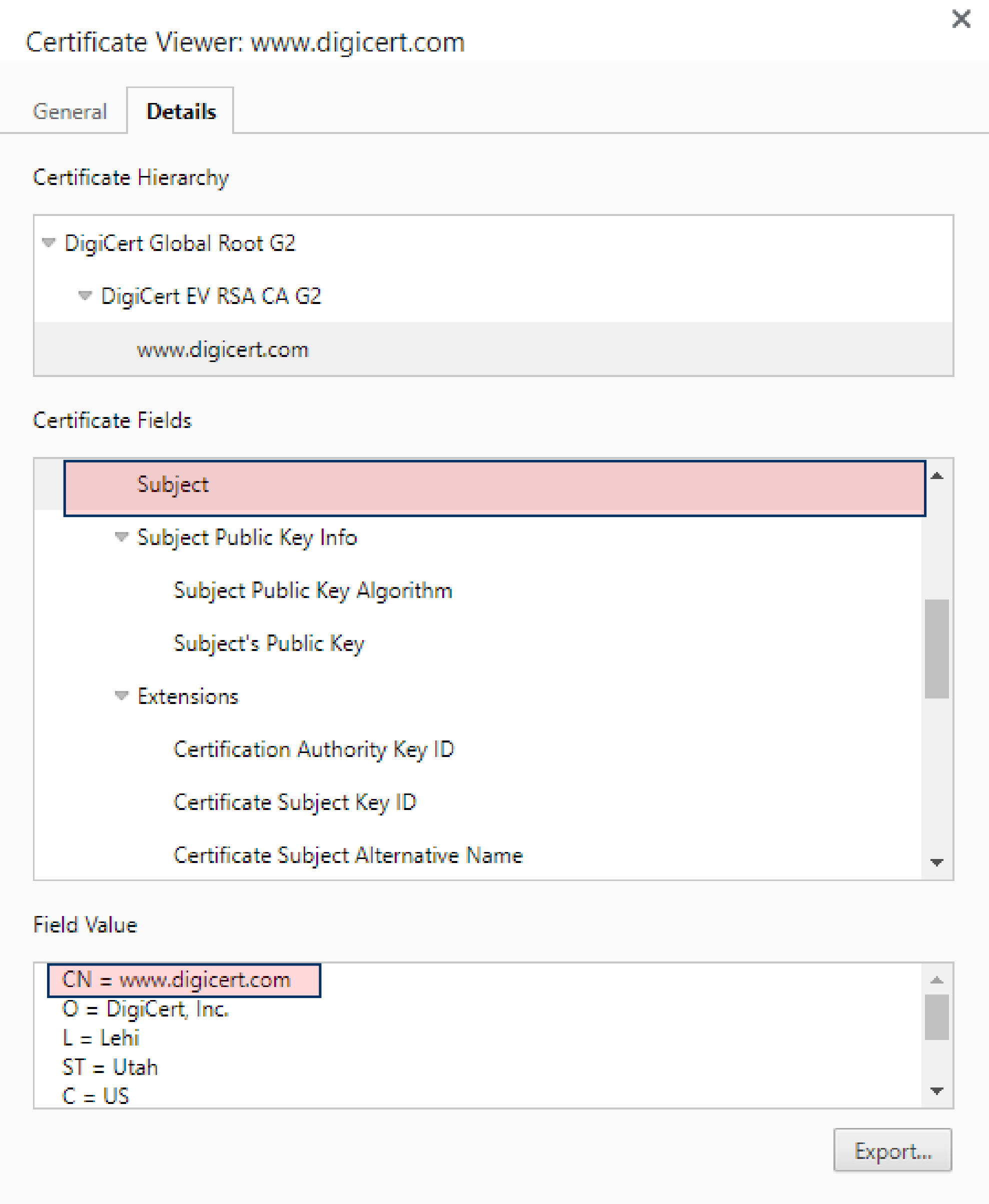
- Certificates are issued by a Certificate Authority (CA) for webservers to prove authenticity

Certificates

- Certificates are issued by a Certificate Authority (CA) for webservers to prove authenticity

- Wildcard certificates
  - *.example.com

- Multi-domain certificates
  - Subject alternative name (SAN) extension



Certificate Viewer: www.digicert.com

Certificate Hierarchy

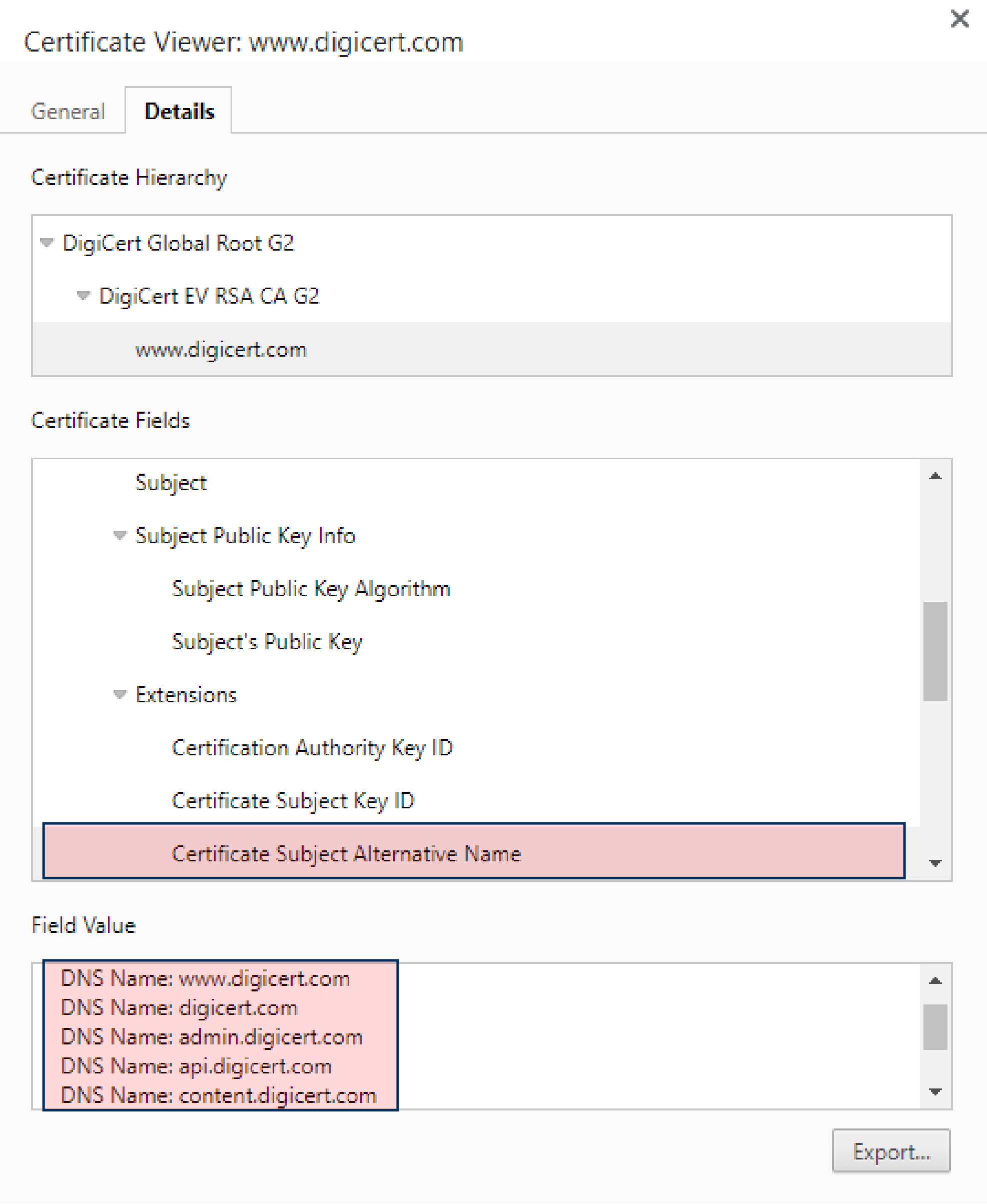
- DigiCert Global Root G2
  - DigiCert EV RSA CA G2
    - www.digicert.com

Certificate Fields

```
    Subject
    Subject Public Key Info
        Subject Public Key Algorithm
        Subject's Public Key
    Extensions
        Certification Authority Key ID
        Certificate Subject Key ID
        Certificate Subject Alternative Name
```

Field Value

```
CN = www.digicert.com
O = DigiCert, Inc.
L = Lehi
ST = Utah
C = US
```



Certificate Viewer: www.digicert.com

Certificate Hierarchy

- DigiCert Global Root G2
  - DigiCert EV RSA CA G2
    - www.digicert.com

Certificate Fields

```
    Subject
    Subject Public Key Info
        Subject Public Key Algorithm
        Subject's Public Key
    Extensions
        Certification Authority Key ID
        Certificate Subject Key ID
        Certificate Subject Alternative Name
```

Field Value

```
DNS Name: www.digicert.com
DNS Name: digicert.com
DNS Name: admin.digicert.com
DNS Name: api.digicert.com
DNS Name: content.digicert.com
```

Example 1

- One private key to validate several (sub)domains
- One compromised key can lead to several domains being compromised

Example 2

- A recent study identified 343,336 unique wildcard domains with a wildcard followed by a TLD
  - *.com-deals.online
  - *.com.example.com
- Each such wildcard certificate can be used to target any domain with the matching TLD
  - amazon.com-deals.online
  - apple.com.example.com

R. Roberts et al., **You are who you appear to be: A longitudinal study of domain impersonation in TLS certificates**, in *Proc. ACM CCS*, 2019.

Example 3

- A recent study show that TLS certificates shared by multiple domains enables HTTPS hijacking attacks, despite state-of-the-art security policies and countermeasures
- HTTPS MITM attacks based on the shared TLS certificates
- Hijack ongoing HTTPS connection
  - Misconfigured HTTP response headers
  - Rerouting encrypted traffic to another flawed server sharing the TLS certificate
- 25% subdomains of Alexa Top 500 websites are affected by these issues

M. Zhang et al., **Talking with familiar strangers: An empirical study on https context confusion attacks**, in *Proc. ACM CCS*, 2020.

Contributions

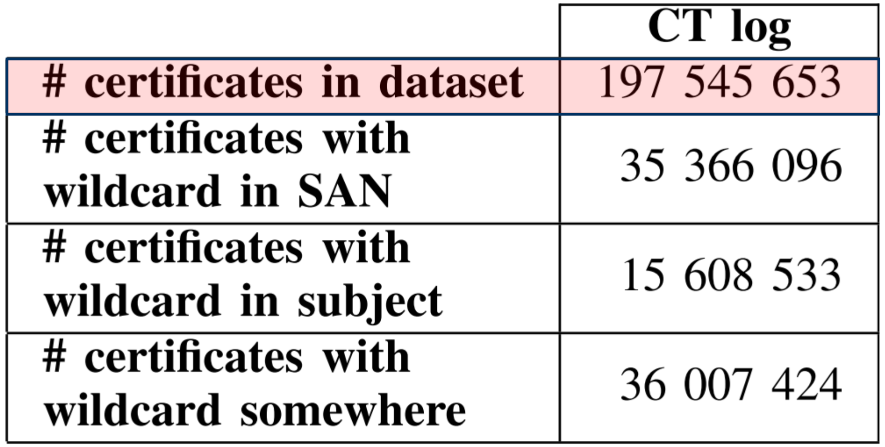
- 10-year longitudinal analysis of the wildcard certificates and multi-domain certificates usage on the internet
- High-level analysis of three large certificate datasets
- Capture and highlight substantial differences in the heterogenous wildcard and multi-domain certificate practices, by studying certificates along five dimension:
  - (1) domain popularity
  - (2) certificate authority
  - (3) certificate type
  - (4) certificate validity period
  - (5) certificate key type

Dataset

- Certificate transparency (CT) logs
- Crt.sh
- Rapid7

Dataset

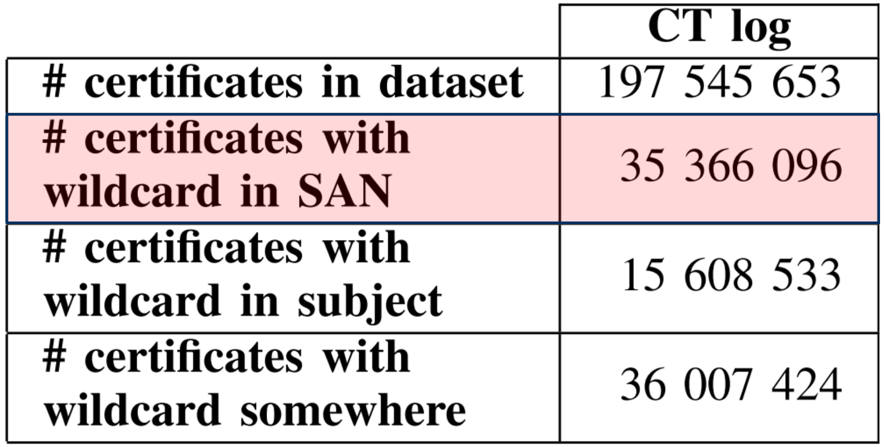
- Certificate transparency (CT) logs
- Crt.sh
- Rapid7



	CT log
# certificates in dataset	197 545 653
# certificates with wildcard in SAN	35 366 096
# certificates with wildcard in subject	15 608 533
# certificates with wildcard somewhere	36 007 424

Dataset

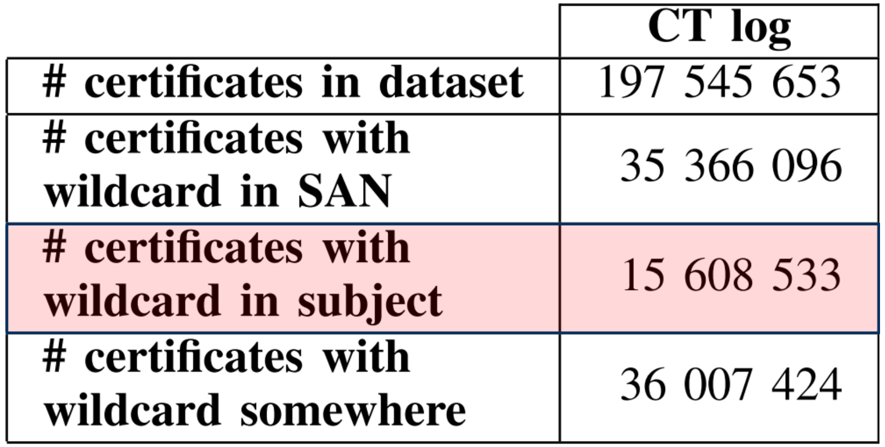
- Certificate transparency (CT) logs
- Crt.sh
- Rapid7



	CT log
# certificates in dataset	197 545 653
# certificates with wildcard in SAN	35 366 096
# certificates with wildcard in subject	15 608 533
# certificates with wildcard somewhere	36 007 424

Dataset

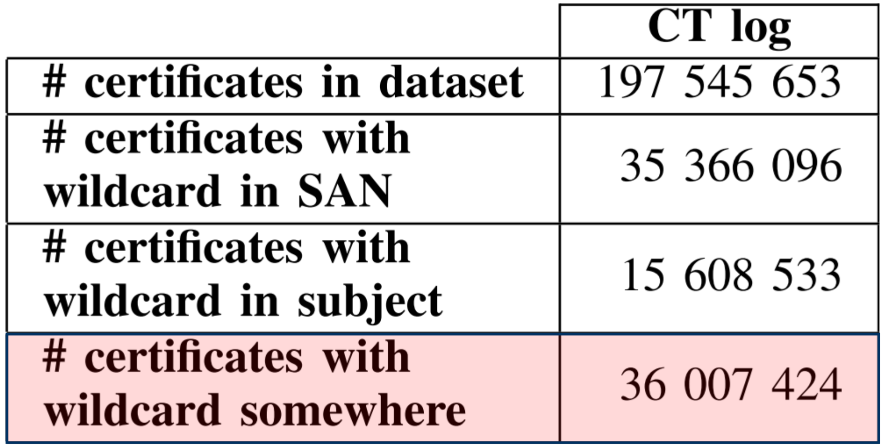
- Certificate transparency (CT) logs
- Crt.sh
- Rapid7



	CT log
# certificates in dataset	197 545 653
# certificates with wildcard in SAN	35 366 096
# certificates with wildcard in subject	15 608 533
# certificates with wildcard somewhere	36 007 424

Dataset

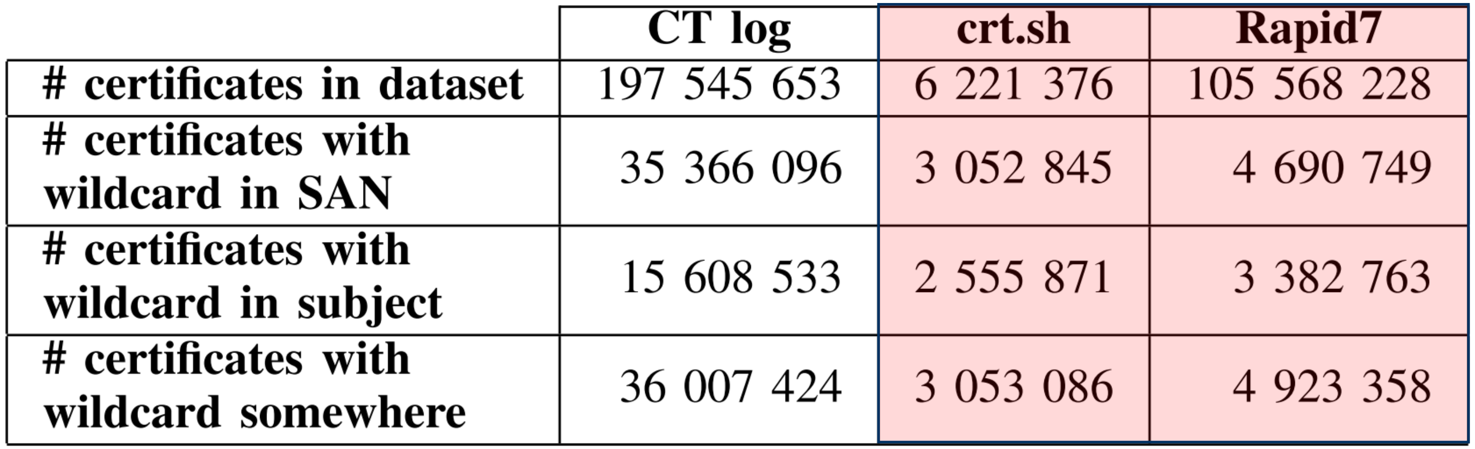
- Certificate transparency (CT) logs
- Crt.sh
- Rapid7



	CT log
# certificates in dataset	197 545 653
# certificates with wildcard in SAN	35 366 096
# certificates with wildcard in subject	15 608 533
# certificates with wildcard somewhere	36 007 424

Dataset

- » Certificate transparency (CT) logs
- » Crt.sh
- » Rapid7



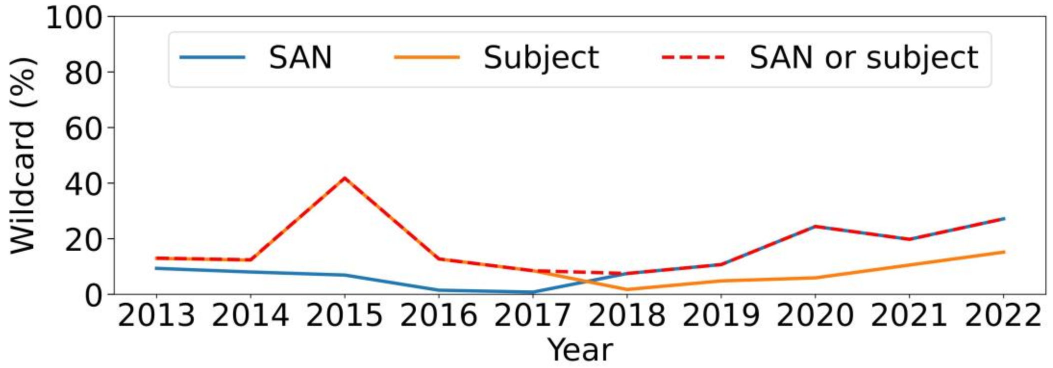
	CT log	crt.sh	Rapid7
# certificates in dataset	197 545 653	6 221 376	105 568 228
# certificates with wildcard in SAN	35 366 096	3 052 845	4 690 749
# certificates with wildcard in subject	15 608 533	2 555 871	3 382 763
# certificates with wildcard somewhere	36 007 424	3 053 086	4 923 358

High-level analysis

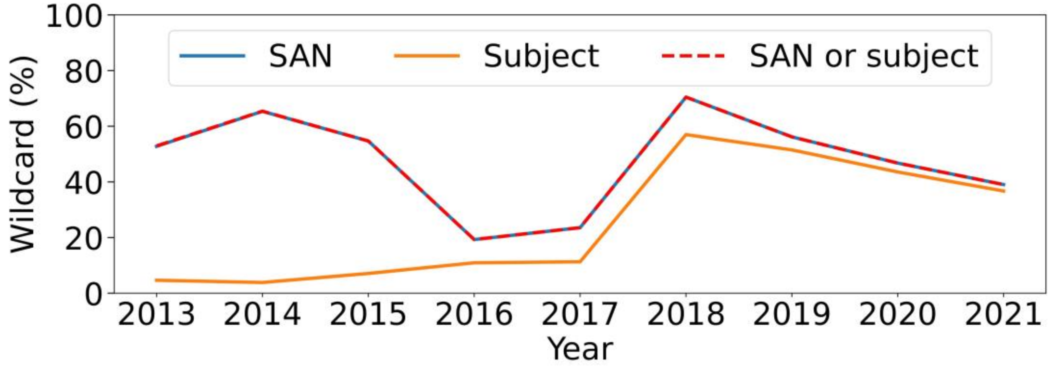
- » Yearly percentage of wildcard usage

- » Chrome removed support for validating against the subject in 2017

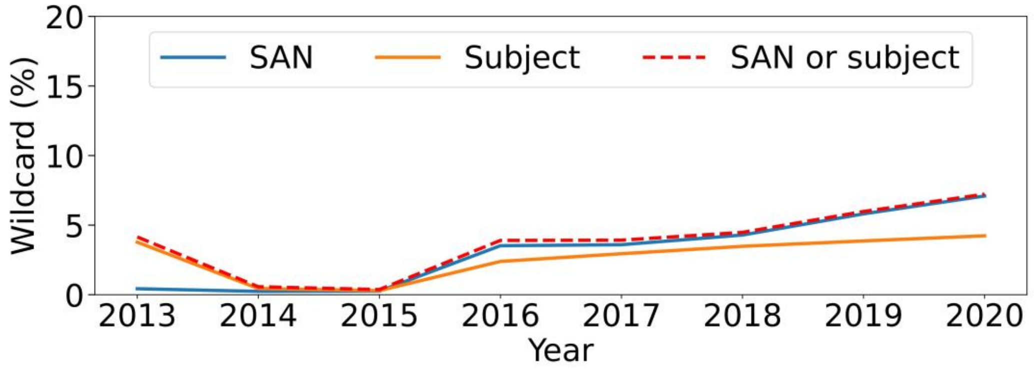
CT log



Crt.sh

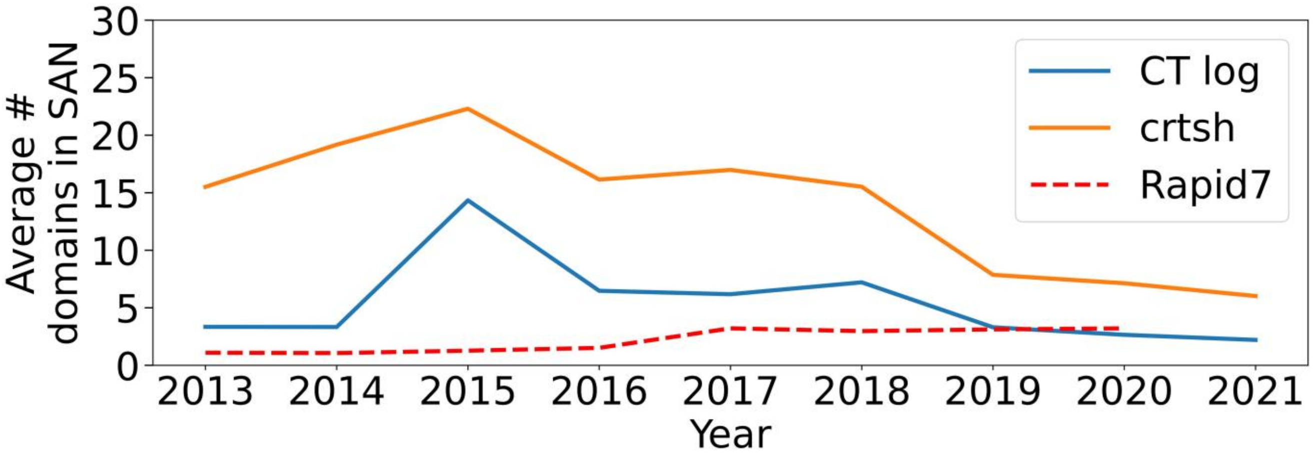


Rapid7



High-level analysis

- Yearly average number of domans in SAN



Impact of factors

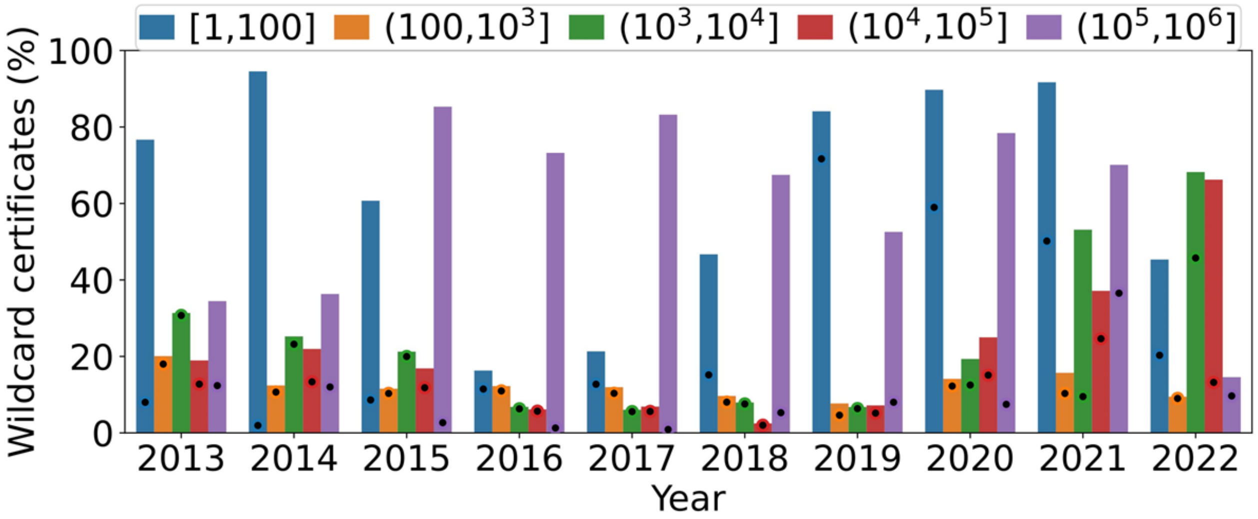
- (1) domain popularity
- (2) certificate authority
- (3) certificate type
- (4) certificate validity period
- (5) certificate key type

Impact of factors

- (1) domain popularity
- (2) certificate authority
- (3) certificate type
- (4) certificate validity period
- (5) certificate key type

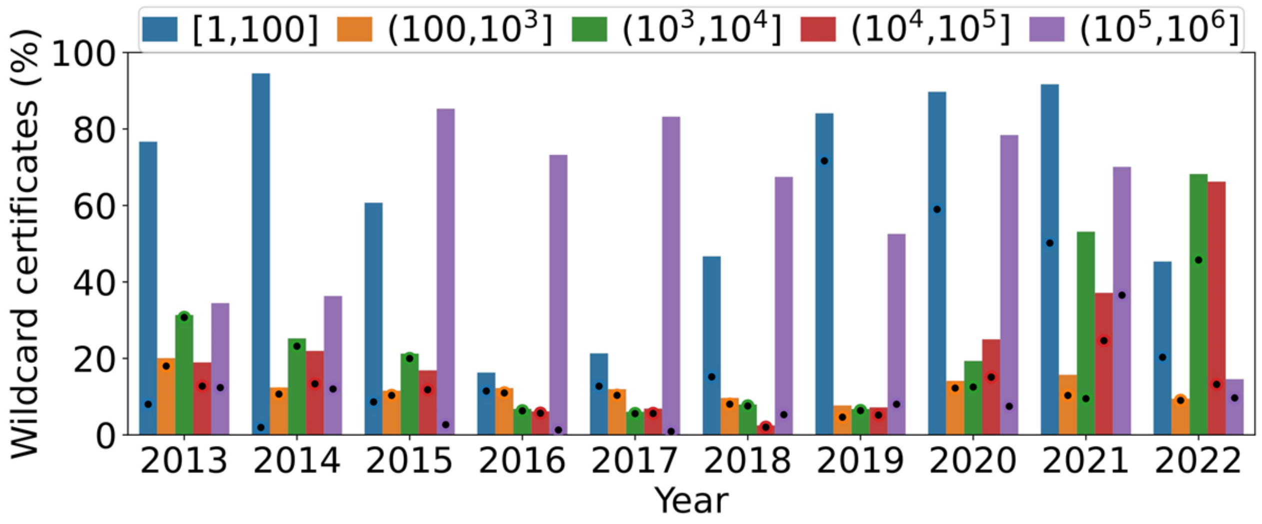
Domain popularity

- Yearly percentage of wildcard certificates issued for domains with different popularity
  - SAN (bars)
  - Subject (markers)

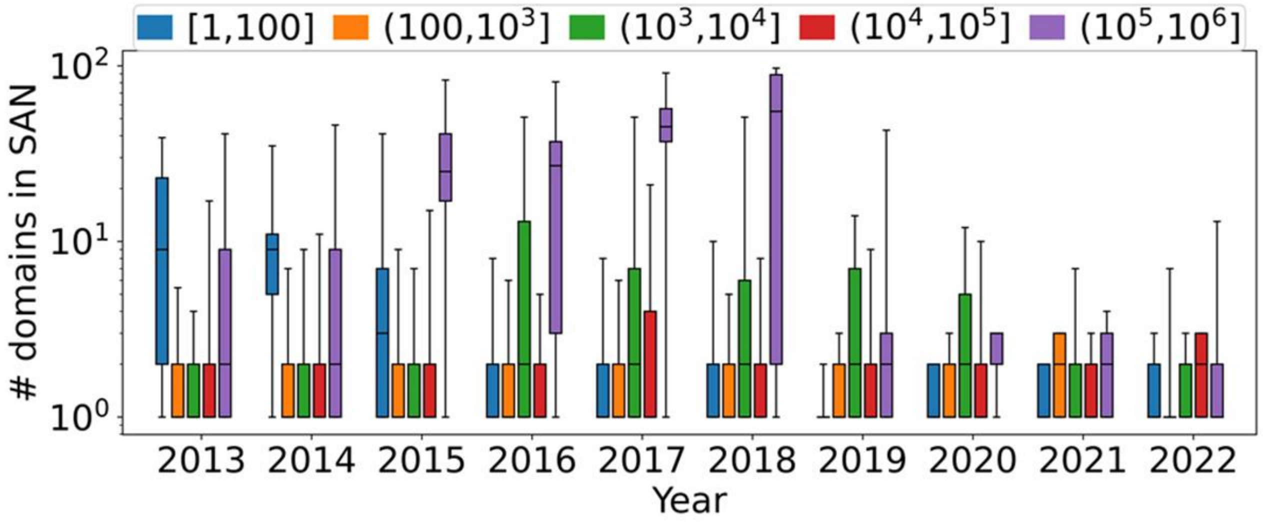


Domain popularity

- Yearly percentage of wildcard certificates issued for domains with different popularity
  - SAN (bars)
  - Subject (markers)



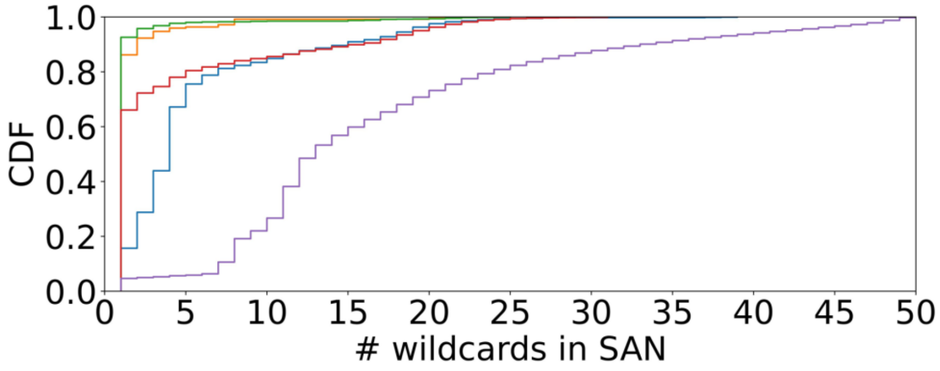
- Number of domains in SAN



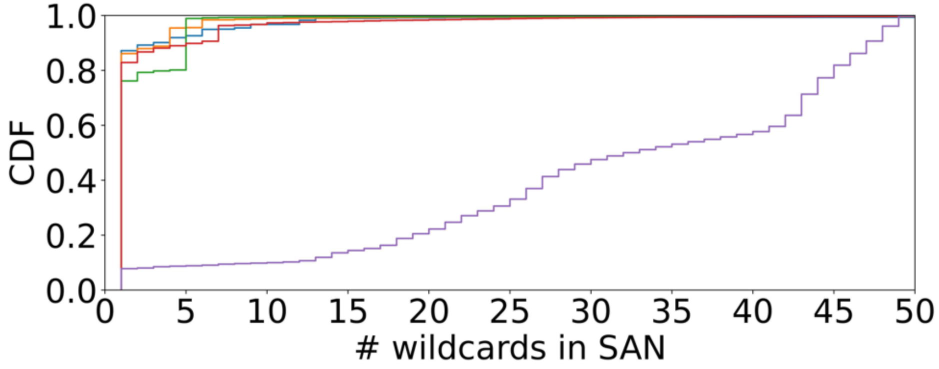
Domain popularity

» Number of wildcards in SAN

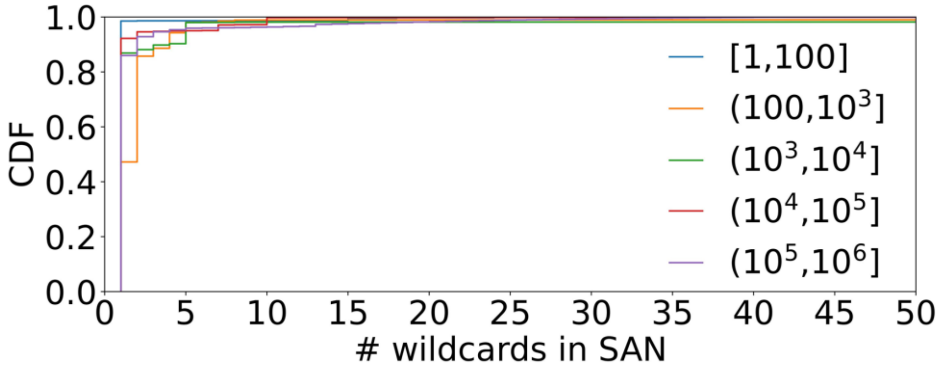
2013 -
2015



2016 -
2018



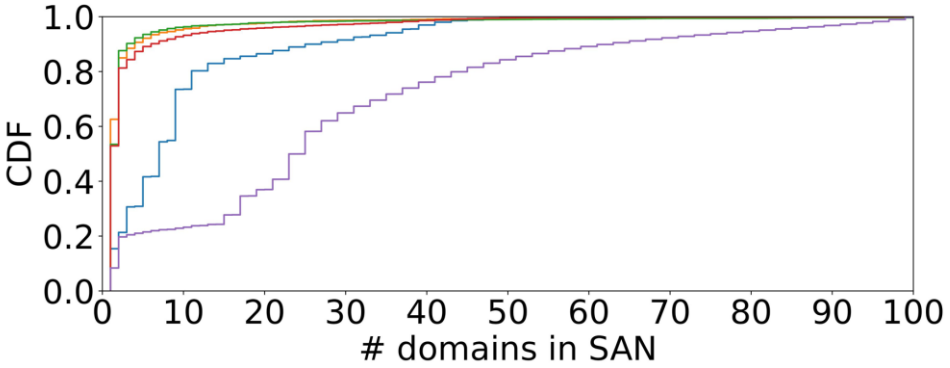
2019 -
2021



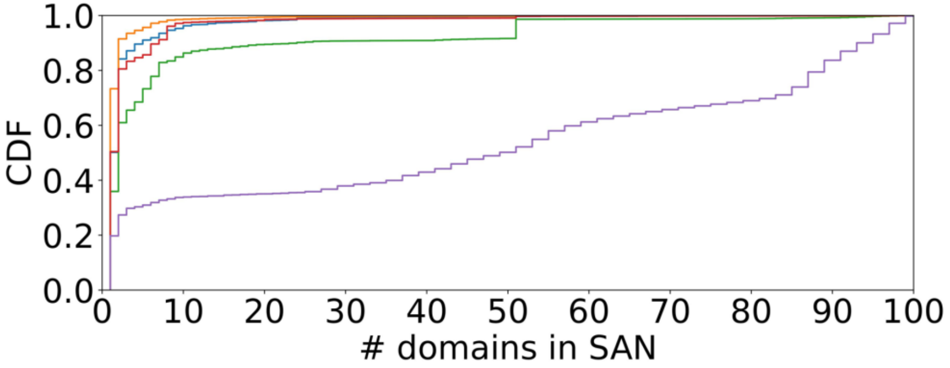
Domain popularity

- Number of domains in SAN

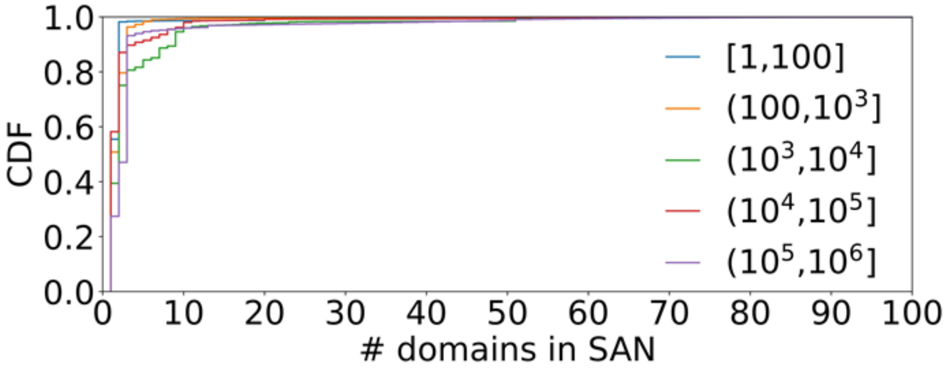
2013 - 2015



2016 - 2018

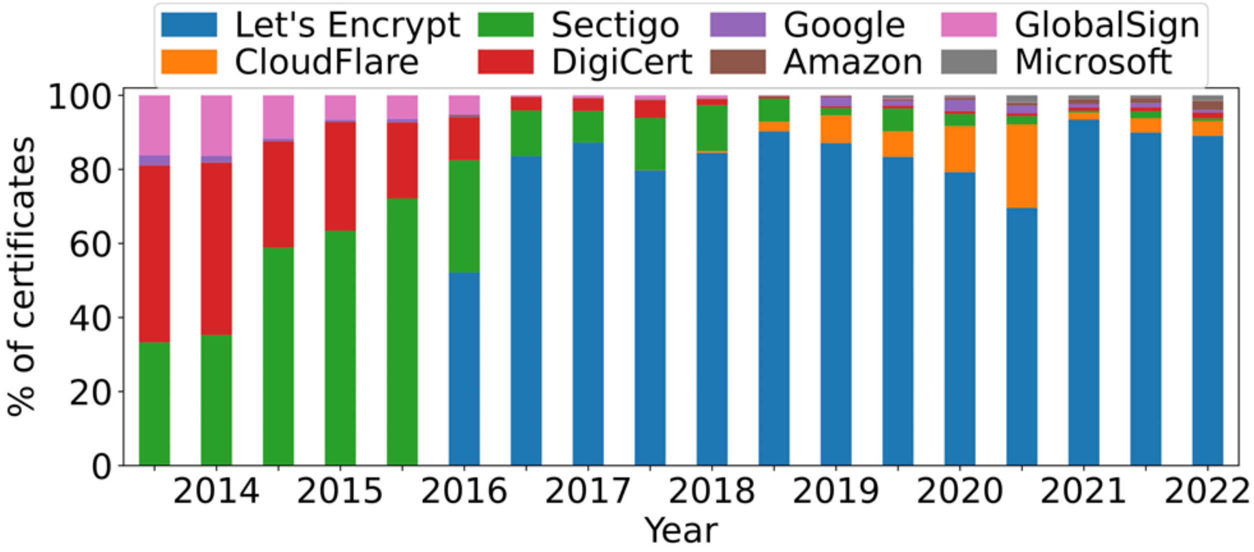


2019 - 2021



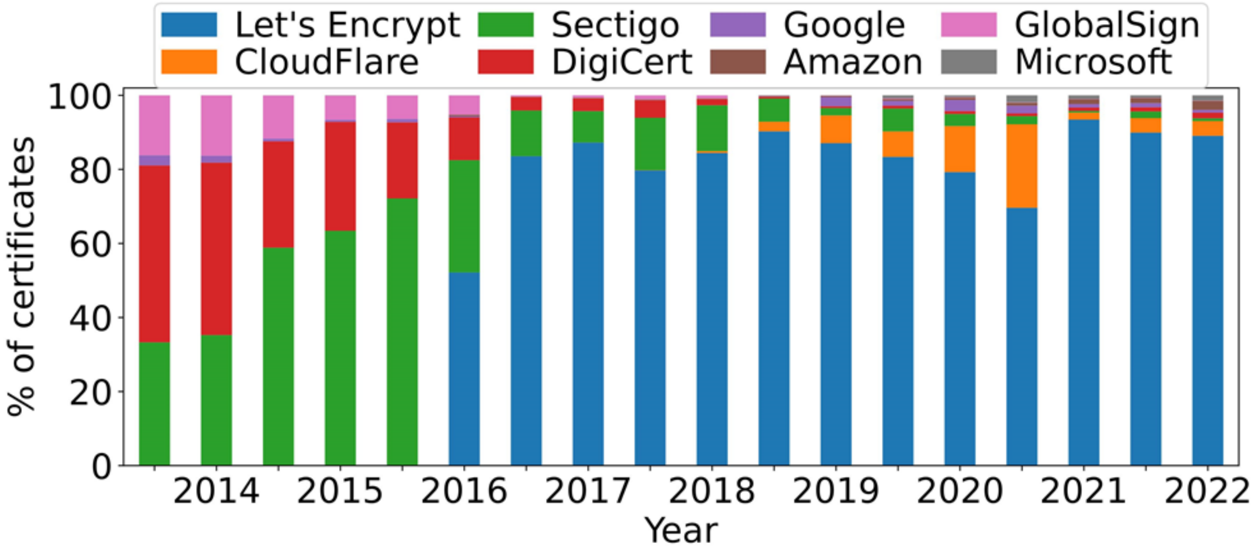
Certificate authority

- Relative certificate frequency per top-CA.

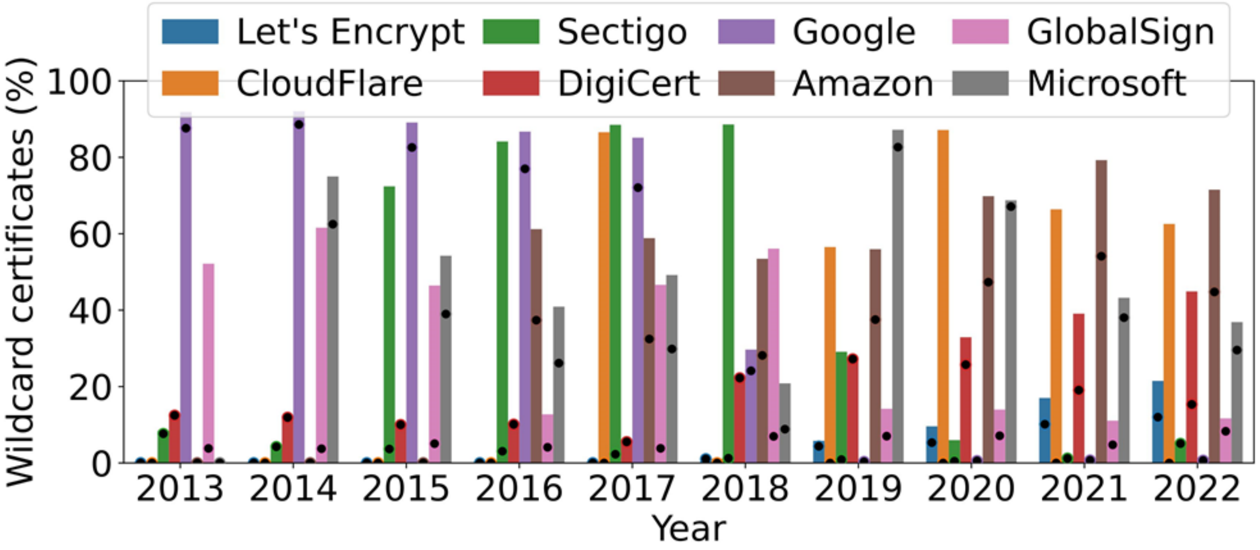


Certificate authority

- Relative certificate frequency per top-CA.

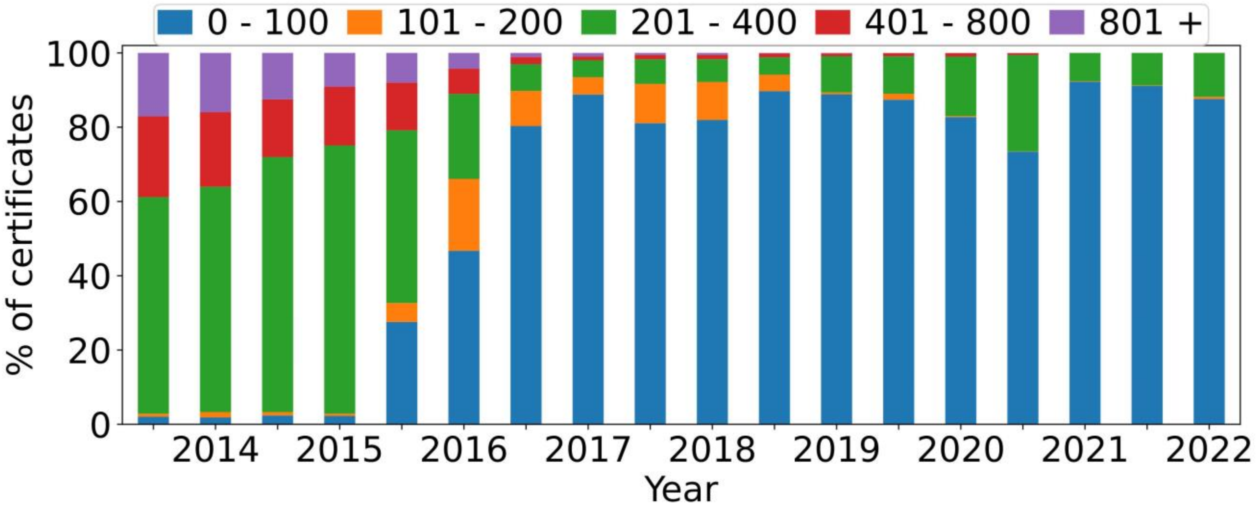


- Percentage of wildcard certificates in SAN (bars) and in subject field (markers) per CA.



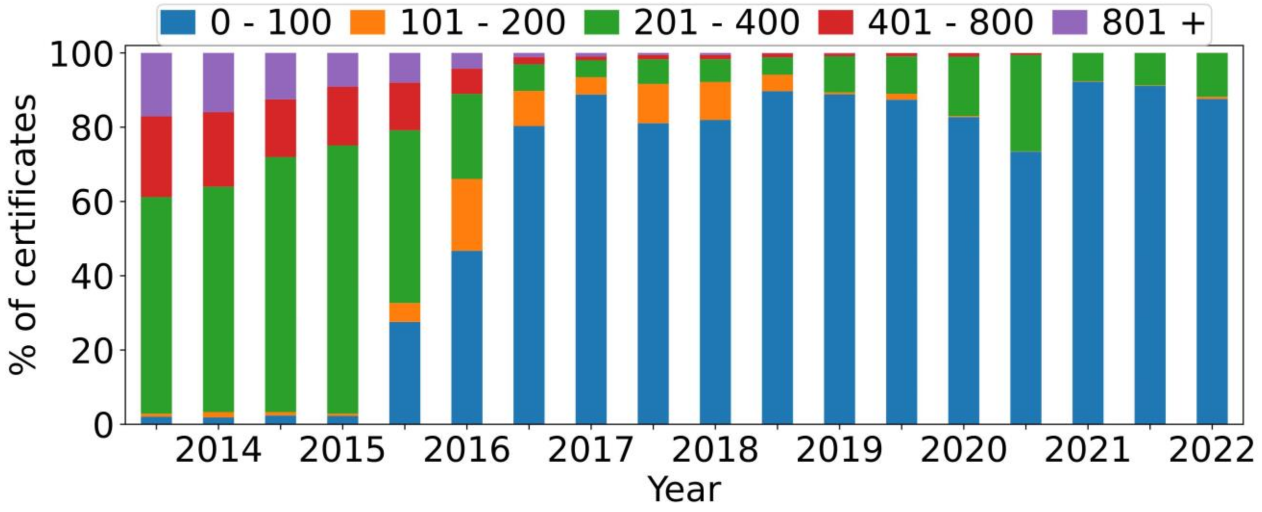
Certificate validity period

- Relative certificate frequency per validity period.

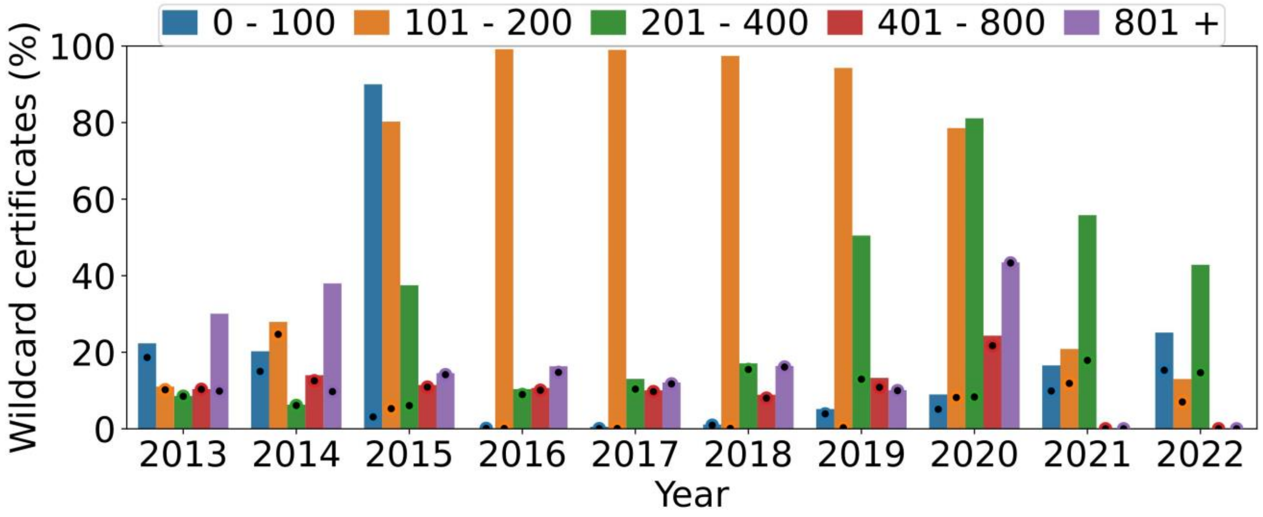


Certificate validity period

- » Relative certificate frequency per validity period.



- » Percentage of wildcard certificates in SAN (bars) and in subject field (markers) per validity period (in days).



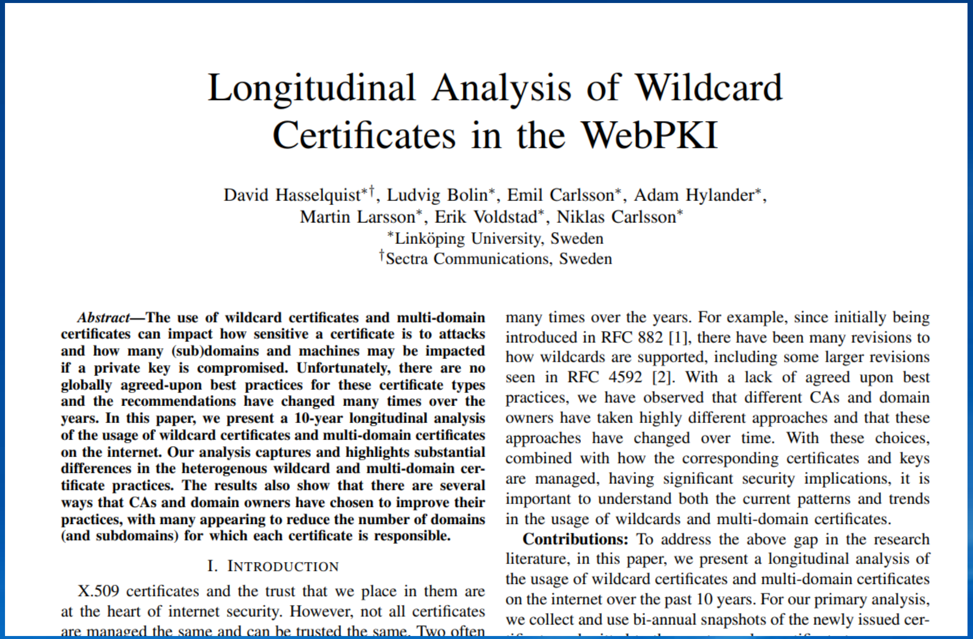
Conclusions

- 10-year longitudinal analysis of the wildcard certificates and multi-domain certificates usage on the internet
- High-level analysis of three large certificate datasets
- Capture and highlight substantial differences in the heterogenous wildcard and multi-domain certificate practices, by studying certificates along five dimension:
  - (1) domain popularity
  - (2) certificate authority
  - (3) certificate type
  - (4) certificate validity period
  - (5) certificate key type

Longitudinal Analysis of Wildcard Certificates in the WebPKI

David Hasselquist
Ludvig Bolin
Emil Carlsson
Adam Hylander
Martin Larsson
Erik Voldstad
Niklas Carlsson

Paper is online!



Longitudinal Analysis of Wildcard Certificates in the WebPKI

David Hasselquist[*][†], Ludvig Bolin[*], Emil Carlsson[*], Adam Hylander[*],
Martin Larsson[*], Erik Voldstad[*], Niklas Carlsson[*]
[*]Linköping University, Sweden
[†]Sectra Communications, Sweden

***Abstract*—The use of wildcard certificates and multi-domain certificates can impact how sensitive a certificate is to attacks and how many (sub)domains and machines may be impacted if a private key is compromised. Unfortunately, there are no globally agreed-upon best practices for these certificate types and the recommendations have changed many times over the years. In this paper, we present a 10-year longitudinal analysis of the usage of wildcard certificates and multi-domain certificates on the internet. Our analysis captures and highlights substantial differences in the heterogenous wildcard and multi-domain certificate practices. The results also show that there are several ways that CAs and domain owners have chosen to improve their practices, with many appearing to reduce the number of domains (and subdomains) for which each certificate is responsible.**

I. Introduction

X.509 certificates and the trust that we place in them are at the heart of internet security. However, not all certificates are managed the same and can be trusted the same. Two often

many times over the years. For example, since initially being introduced in RFC 882 [1], there have been many revisions to how wildcards are supported, including some larger revisions seen in RFC 4592 [2]. With a lack of agreed upon best practices, we have observed that different CAs and domain owners have taken highly different approaches and that these approaches have changed over time. With these choices, combined with how the corresponding certificates and keys are managed, having significant security implications, it is important to understand both the current patterns and trends in the usage of wildcards and multi-domain certificates.

Contributions: To address the above gap in the research literature, in this paper, we present a longitudinal analysis of the usage of wildcard certificates and multi-domain certificates on the internet over the past 10 years. For our primary analysis, we collect and use bi-annual snapshots of the newly issued cer-

David Hasselquist (david.hasselquist@liu.se)