# The Rise of Certificate Transparency and Its Implications on the Internet Ecosystem

IETF 103, Bangkok, 2018

Quirin Scheitle *(TUM),* Oliver Gasser *(TUM),* Theodor Nolte *(HAW Hamburg),*
Johanna Amann *(ICSI/Corelight/LBNL),* Lexi Brent *(The University of Sydney),*
Georg Carle *(TUM),* Ralph Holz *(The University of Sydney),*
Thomas C. Schmidt *(HAW Hamburg),* Matthias Wählisch *(FU Berlin)*

# Owner of a name gets a certificate

www.ietf.org

# 3rd party incorrectly gets a certificate
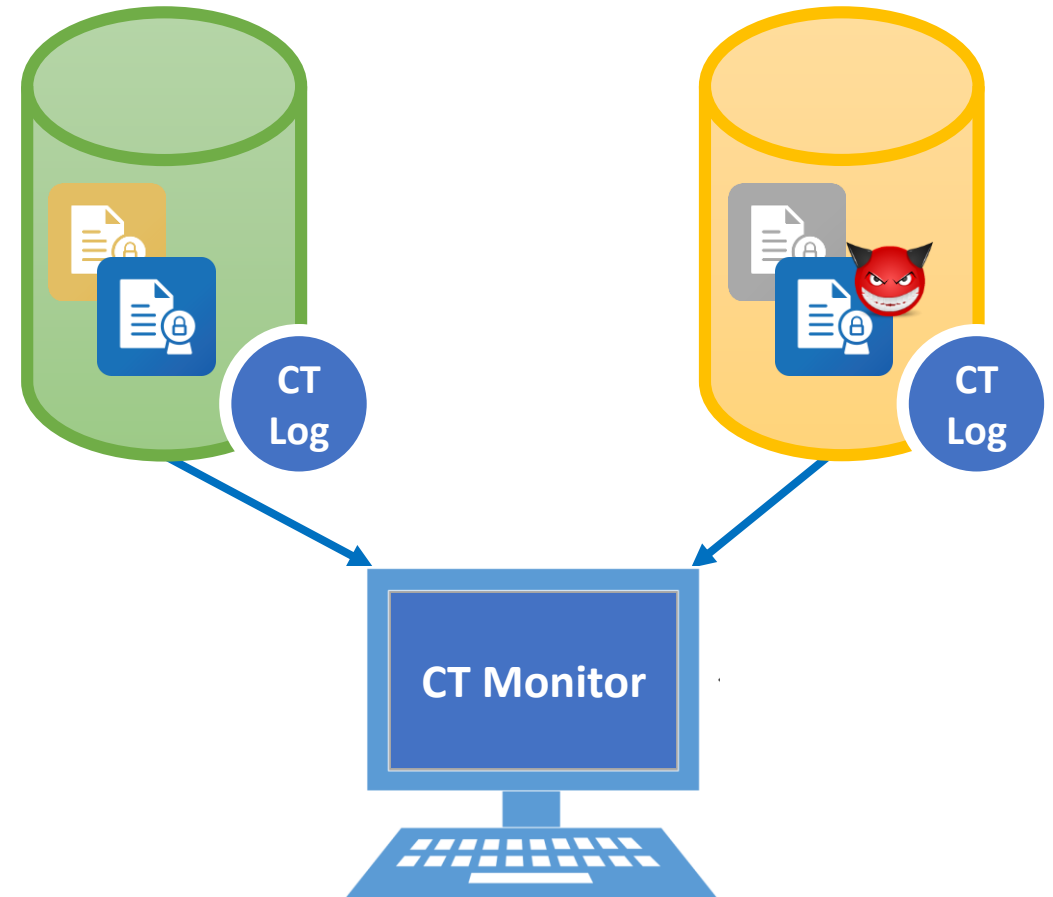Name owner cannot verify existence

www.ietf.org

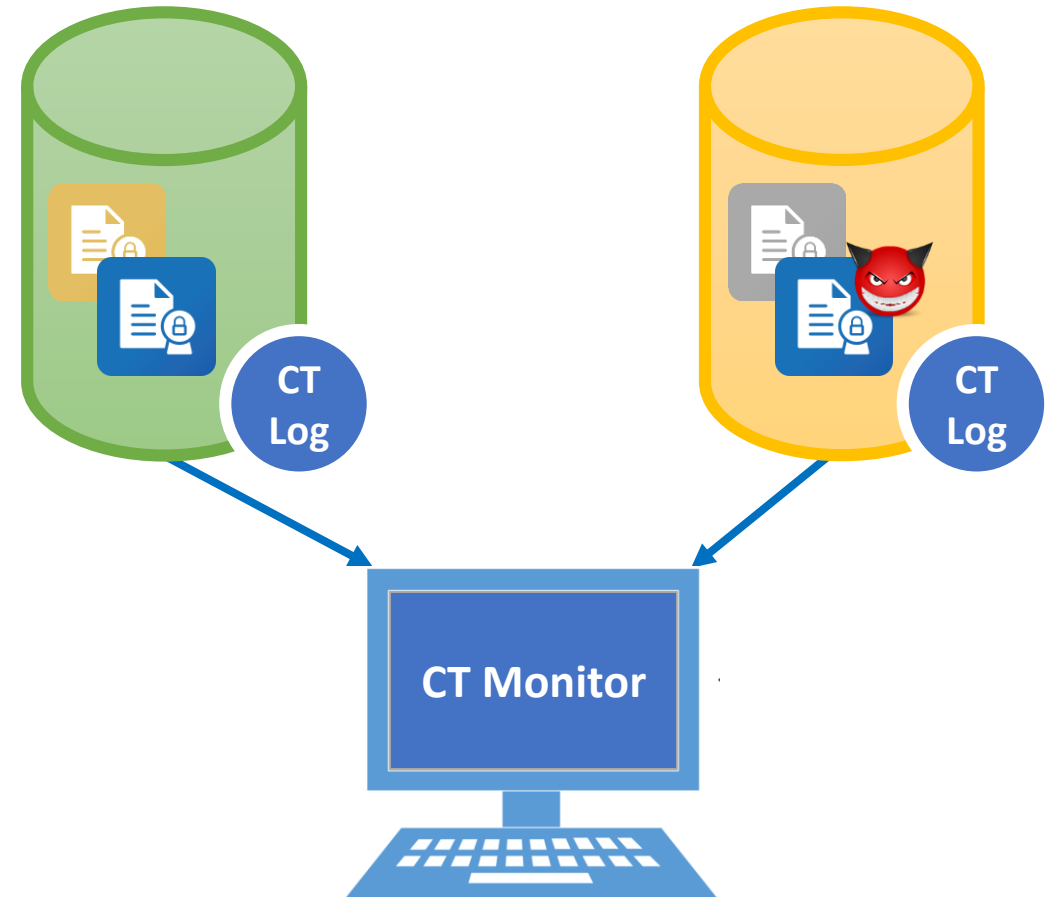www.ietf.org

# Certificate Transparency (CT) in a nutshell

# Certificate Transparency (CT) in a nutshell

**Goal**

Provide transparency into issued certificates to detect certificate mis-issuances

**Approach**

Uses public, append-only logs to record certificates

# Additional advantages and new challenges

**Does CT introduce new dependencies?**

Log servers are operated by multiple companies

Concentration on few log operators should be prevented

**What do we lose (or gain) by exposing domain names?**

Logs provide data to easily search for names

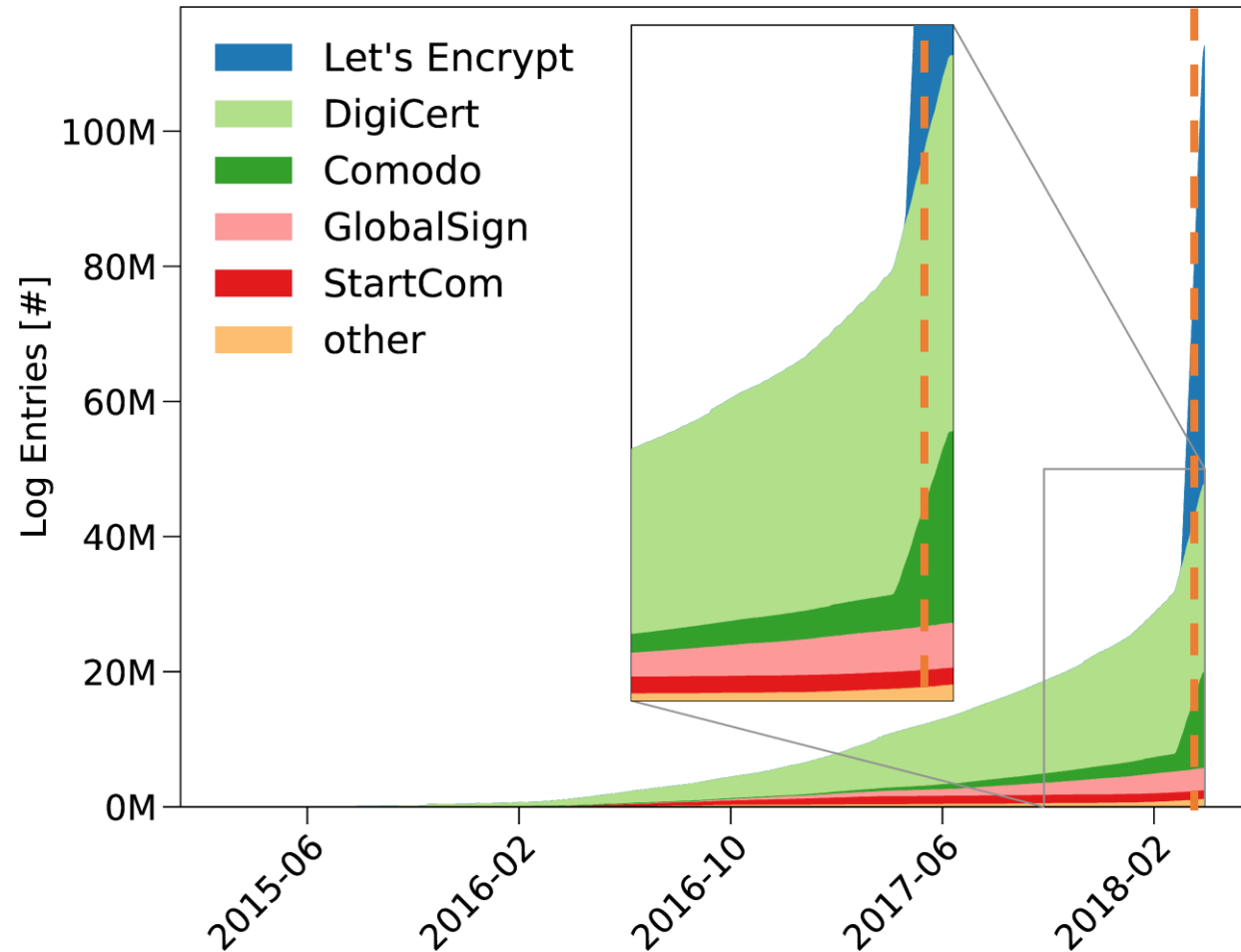Might help to identify malicious domain names but might also help attackers to find victims

# The Rise of Certificate Transparency and Its Implications on the Internet Ecosystem

# How did the log volume change over time?
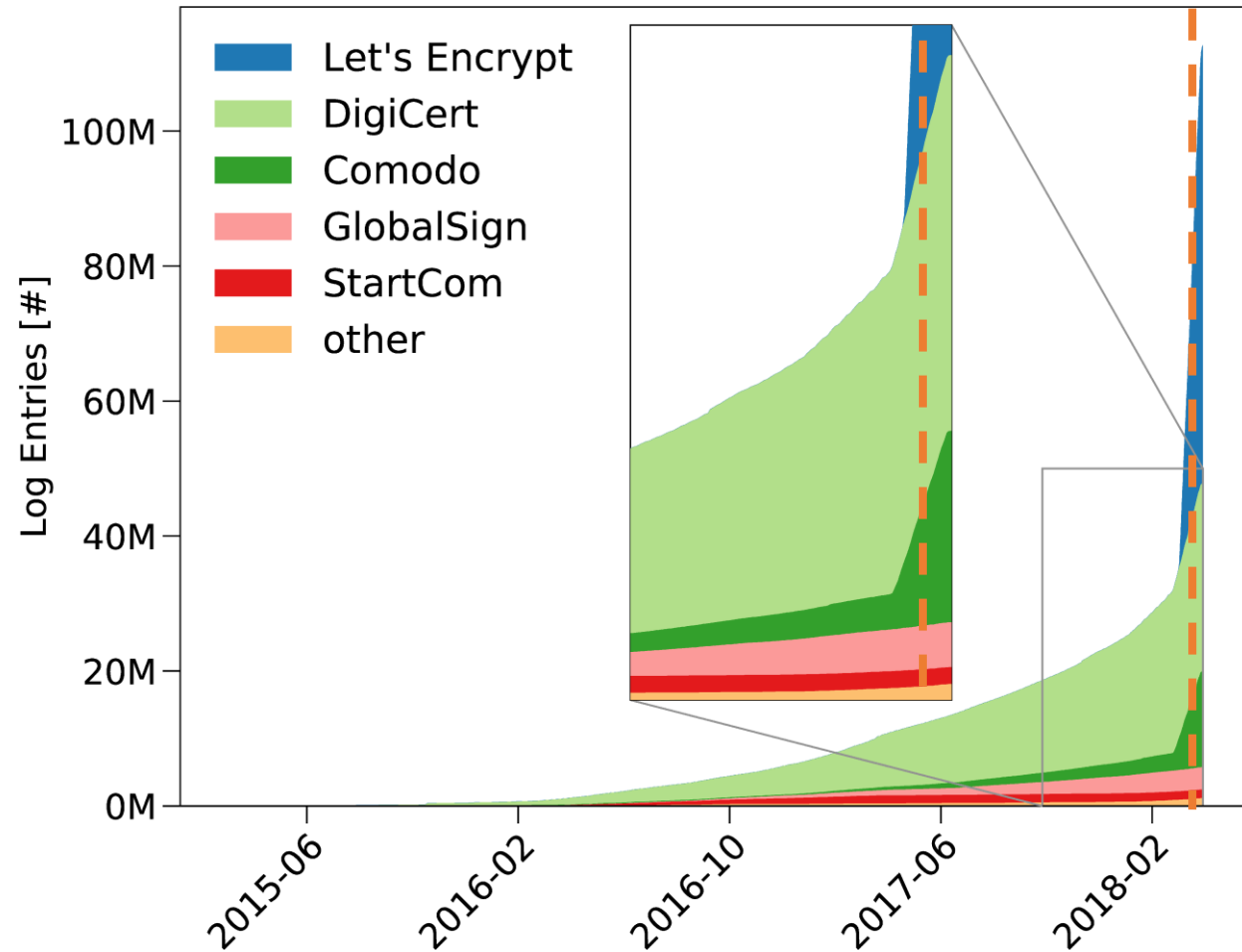
# How did the log volume change over time?



- Large increase of log entries before CT deadline
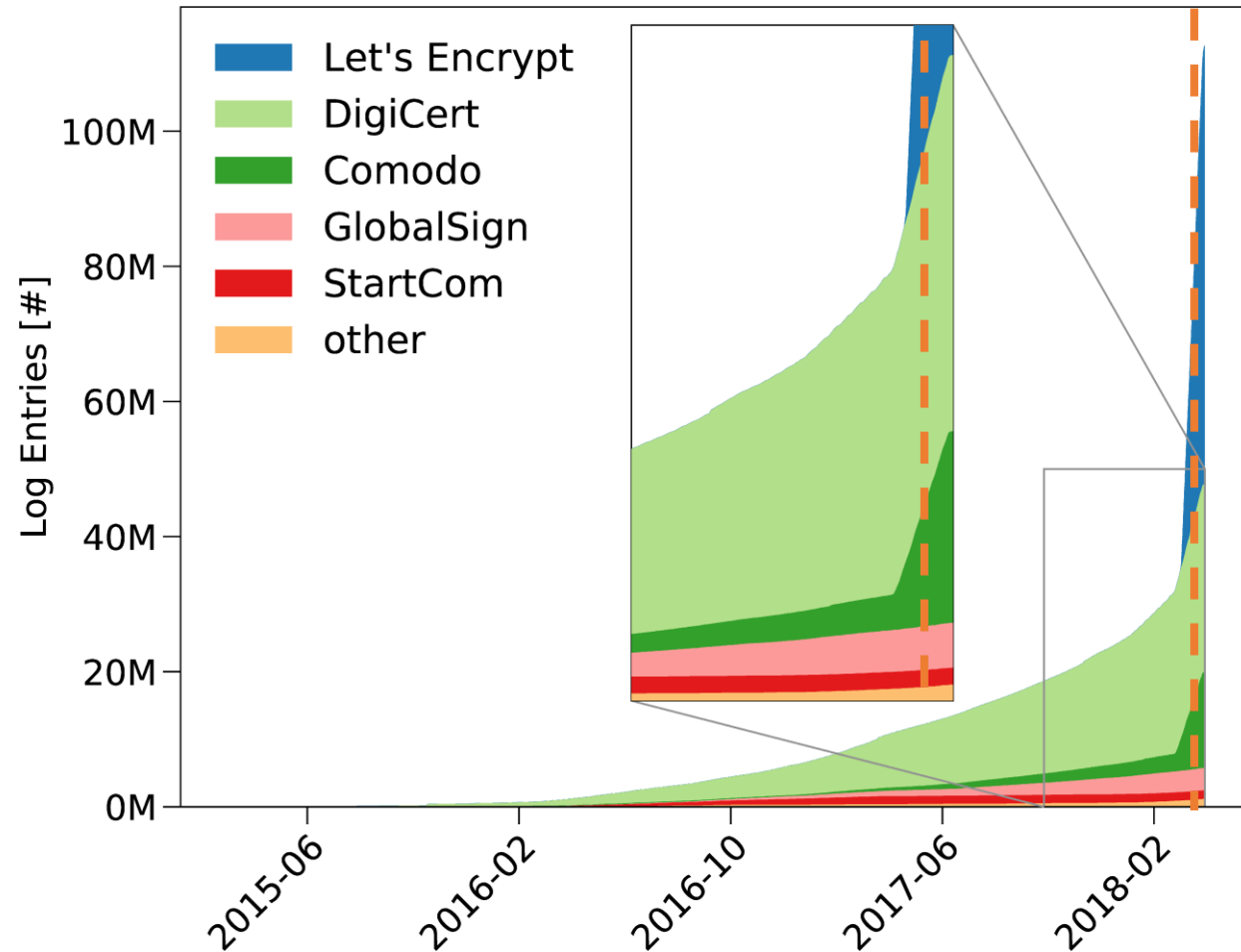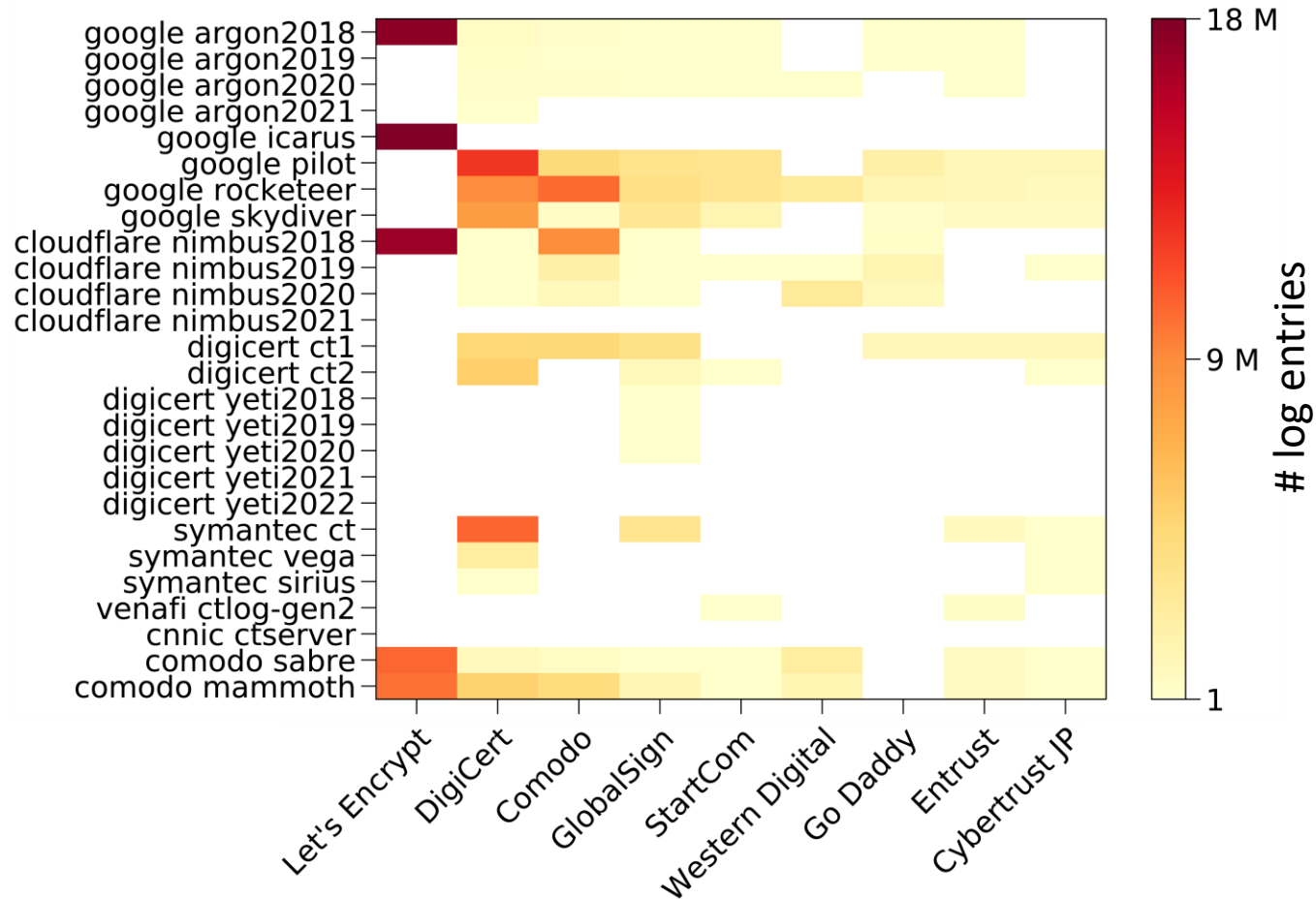
- Let's Encrypt dominates

# How did the log volume change over time?



- Large increase of log entries before CT deadline

- Let's Encrypt dominates

- **Strong rise**

# Are CAs distributing certificates over many CT logs?

# Are CAs distributing certificates over many CT logs?
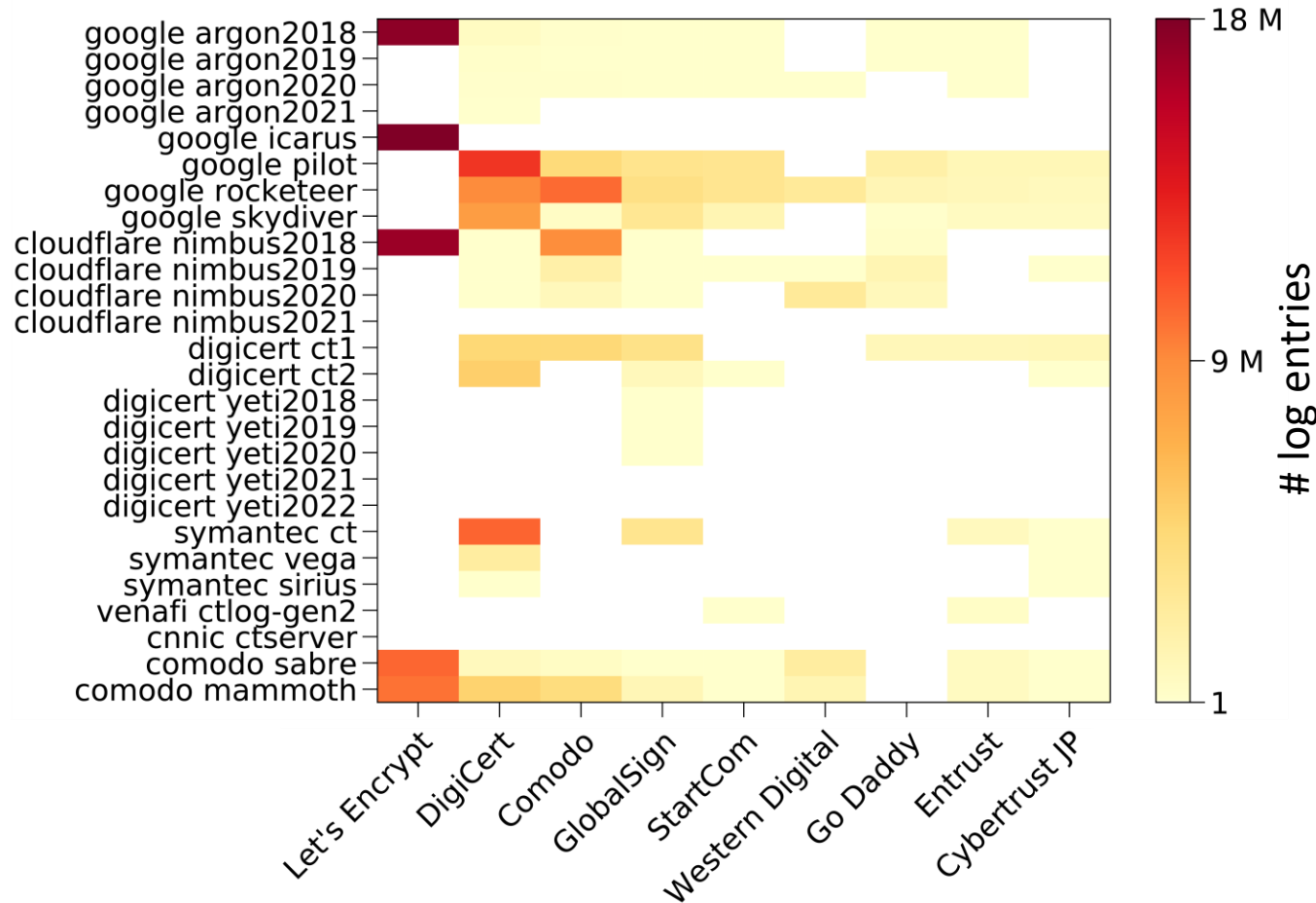


- System overly relies on few log servers

- Almost all CAs use few logs for their certificate

# Are CAs distributing certificates over many CT logs?
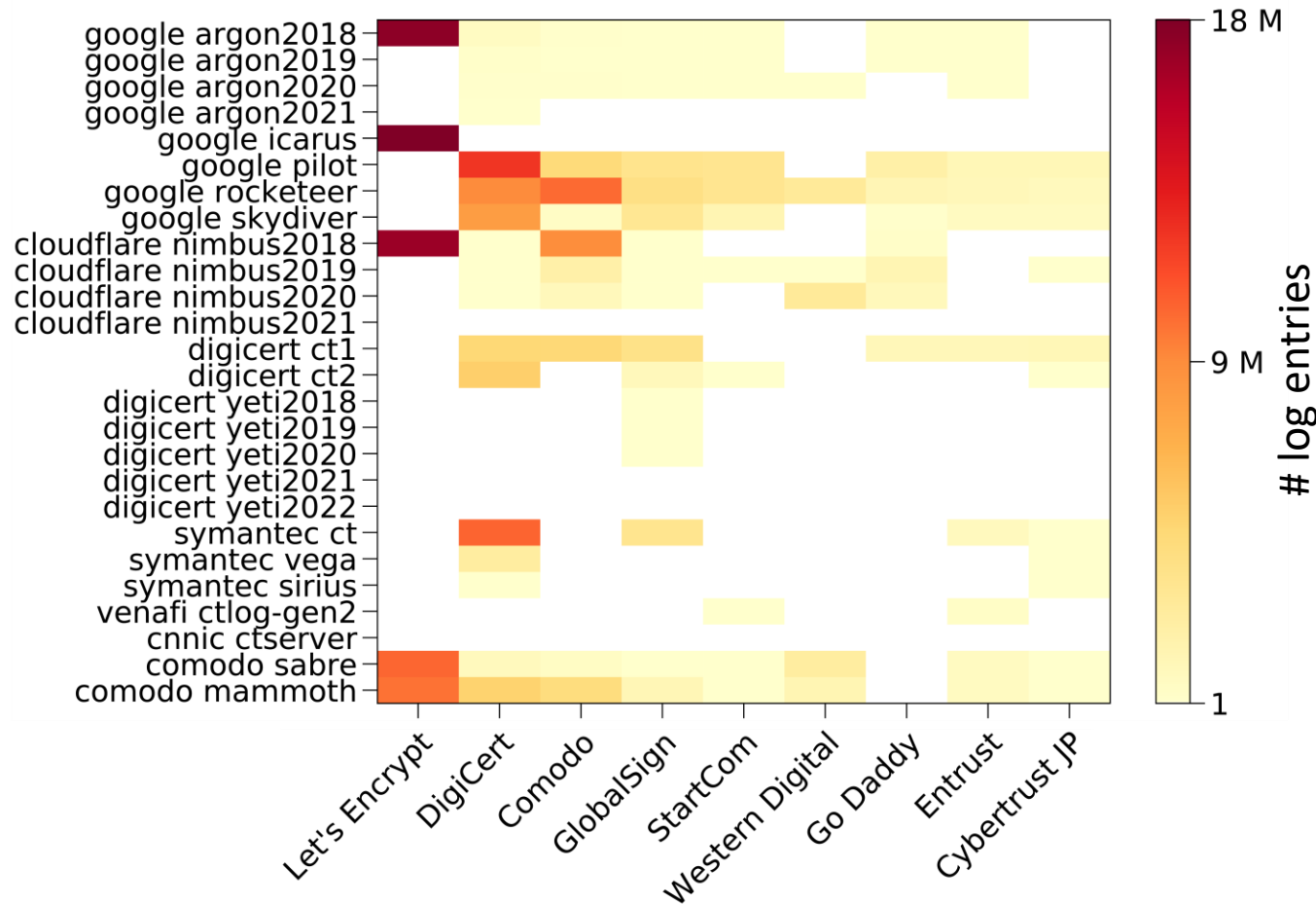


- System overly relies on few log servers

- Almost all CAs use few logs for their certificate

- **No**, CAs use few logs which limits reliability

# The Rise of Certificate Transparency and Its Implications on the Internet Ecosystem

# Can CT be used to find malicious domains?

# Can CT be used to find malicious domains?

Method

- Inspect domains with similarities
  to domains from
    - Apple
    - Paypal
    - Hotmail
    - Google
    - Ebay
- Example:
  appleid.apple.com-7etr6eti.gq

# Can CT be used to find malicious domains?

## Method

- Inspect domains with similarities to domains from
  - Apple
  - Paypal
  - Hotmail
  - Google
  - Ebay
- Example: appleid.apple.com-7etr6eti.gq

## Results 126k potential phishing domains

| Service | Count | Example |
|---|---|---|
| Apple | 63k | appleid.apple.com-7etr6eti.gq |
| PayPal | 58k | paypal.com-account-security.money |
| Microsoft | 4k | www-hotmail-login.live |
| Google | 1k | accounts.google.co.am |
| eBay | <1k | www.ebay.co.uk.dll7.bid |

# Can CT be used to find malicious domains?

## Method

- Inspect domains with similarities to domains from
  - Apple
  - Paypal
  - Hotmail
  - Google
  - Ebay

- Example: appleid.apple.com-7etr6eti.gq

## Results 126k potential phishing domains

| Service | Count | Example |
|---------|-------|---------|
| Apple | 63k | *appleid.apple.com-7etr6eti.gq* |
| PayPal | 58k | *paypal.com-account-security.money* |
| Microsoft | 4k | *www-hotmail-login.live* |
| Google | 1k | *accounts.google.co.am* |
| eBay | <1k | *www.ebay.co.uk.dll7.bid* |

- CERT confirmed that a subset was used to host malicious content

# Can CT be used to find malicious domains?

## Method

- Inspect domains with similarities to domains from
  - Apple
  - Paypal
  - Hotmail
  - Google
  - Ebay
- Example: appleid.apple.com-7etr6eti.gq

## Results 126k potential phishing domains

| Service | Count | Example |
|---------|-------|---------|
| Apple | 63k | appleid.apple.com-7etr6eti.gq |
| PayPal | 58k | paypal.com-account-security.money |
| Microsoft | 4k | www-hotmail-login.live |
| Google | 1k | accounts.google.co.am |
| eBay | <1k | www.ebay.co.uk.dll7.bid |

- CERT confirmed that a subset was used to host malicious content
- **Yes**, CT can be used to find malicious (i.e., phishing) domains

# Does CT help attackers to find new domains?

# Does CT help attackers to find new domains?

Method

- Extract subdomain labels from
  all CT logged certificates
  - `dev for .io`

- Generate new FQDNs with most
  common subdomain labels
  - `dev.foureyes.io`

- Ignore `.com, .net, .org`

- Request A records

# Does CT help attackers to find new domains?

Method

- Extract subdomain labels from all CT logged certificates
  - `dev for .io`

- Generate new FQDNs with most common subdomain labels
  - `dev.foureyes.io`

- Ignore `.com, .net, .org`

- Request A records

Results

- 18.8M new FQDNs found
  17.7M more FQDNS compared to other public lists

# Does CT help attackers to find new domains?

Method

- Extract subdomain labels from all CT logged certificates
  - `dev for .io`

- Generate new FQDNs with most common subdomain labels
  - `dev.foureyes.io`

- Ignore `.com, .net, .org`

- Request A records

Results

- 18.8M new FQDNs found
  17.7M more FQDNS compared to other public lists

- **Yes**, CT helps attackers find previously unknown domains

# Does CT leak private data to attackers?

Top 20 subdomain labels in CT-logged certificates

| | SDL | Count | | SDL | Count | | SDL | Count |
|---|---|---|---|---|---|---|---|---|
| 1 | www | 61.1M | 8 | shop | 303k | 15 | secure | 176k |
| 2 | mail | 14.4M | 9 | **whm** | 280k | 16 | admin | 158k |
| 3 | **webdisk** | 8.7M | 10 | dev | 256k | 17 | mobile | 156k |
| 4 | webmail | 8.6M | 11 | remote | 253k | 18 | server | 146k |
| 5 | **cpanel** | 8.2M | 12 | test | 249k | 19 | cloud | 141k |
| 6 | autodiscover | 3.6M | 13 | api | 239k | 20 | smtp | 140k |
| 7 | m | 310k | 14 | blog | 235k | | | |

# Is CT actively being misused to find victims?

# Is CT actively being misused to find victims?

Method

- Deploy CT honeypot for scanners

- Leak existence of pseudorandom subdomains only via CT logs

- Check DNS logs and check requests on IP addresses belonging to A/AAAA records

- Use EDNS field to reveal location of stub resolvers

# Is CT actively being misused to find victims?

**Method**

- Deploy CT honeypot for scanners

- Leak existence of pseudorandom subdomains only via CT logs

- Check DNS logs and check requests on IP addresses belonging to A/AAAA records

- Use EDNS field to reveal location of stub resolvers

**Results**

- First DNS lookups after 1 minute, HTTP(S) access after 1 hour

- Most scanners without info in rDNS, WHOIS, or on website

# Is CT actively being misused to find victims?

Method

- Deploy CT honeypot for scanners

- Leak existence of pseudorandom subdomains only via CT logs

- Check DNS logs and check requests on IP addresses belonging to A/AAAA records

- Use EDNS field to reveal location of stub resolvers

Results

- First DNS lookups after 1 minute, HTTP(S) access after 1 hour

- Most scanners without info in rDNS, WHOIS, or on website

- One scanner requested A/AAAA records fast and scanned 30 ports

# Is CT actively being misused to find victims?

**Method**

- Deploy CT honeypot for scanners

- Leak existence of pseudorandom subdomains only via CT logs

- Check DNS logs and check requests on IP addresses belonging to A/AAAA records

- Use EDNS field to reveal locations of stub resolvers

**Results**

- First DNS lookups after 1 minute, HTTP(S) access after 1 hour

- Most scanners without info in rDNS, WHOIS, or on website

- One scanner requested A/AAAA records fast and scanned 30 ports

- **Yes**, CT is being misused by actors with undeclared intent

# Take-Aways

CT ecosystem dominated by few stakeholders

Majority of logging volume from few CAs to few logs

CT helps in finding phishing domains

Enables near-time detection and reaction

CT helps attackers

Find previously unknown domains
Scans from dubious actors within minutes

# More details? See ACM IMC'18 paper.

Q. Scheitle, O. Gasser, T. Nolte, J. Amann, L. Brent, G. Carle, R. Holz, T. C. Schmidt, M. Wählisch, **The Rise of Certificate Transparency and Its Implications on the Internet Ecosystem, In:** *Proc. of ACM Internet Measurement Conference (IMC) 2018.*

## The Rise of Certificate Transparency and Its Implications on the Internet Ecosystem

Quirin Scheitle[1], Oliver Gasser[1], Theodor Nolte[2], Johanna Amann[3], Lexi Brent[4], Georg Carle[1], Ralph Holz[4], Thomas C. Schmidt[2], Matthias Wählisch[5]

[1]TUM, [2]HAW Hamburg, [3]ICSI/Corelight/LBNL, [4]The University of Sydney, [5]FU Berlin

**ABSTRACT**

In this paper, we analyze the evolution of Certificate Transparency (CT) over time and explore the implications of exposing certificate DNS names from the perspective of security and privacy. We find that certificates in CT logs have seen exponential growth. Website support for CT has also constantly increased, with now 33% of established connections supporting CT. With the increasing deployment of CT, there are also concerns of information leakage due to all certificates being visible in CT logs. To understand this threat, we introduce a CT honeypot and show that data from CT logs is being used to identify targets for scanning campaigns only minutes after certificate issuance. We present and evaluate a methodology to learn and validate new subdomains from the vast number of domains extracted from CT logged certificates.

In this paper, we contribute to a better understanding of CT rollout and related security and privacy implications:

**CA and CT Log Evolution (§ 2):** Using data of all CT log servers deployed, we investigate the evolution of CT logs over time and the dependency of Certificate Authorities (CAs) on CT log operators.

**Server CT Deployment (§ 3):** Using passive and active measurements, we quantify the evolution of CT adoption among server operators and show positive effects.

**DNS Information Leakage (§ 4):** We investigate the mass leakage of Fully Qualified Domain Names (FQDNs), and use subdomain data to construct and query new FQDNs.

**Detecting Phishing Domains (§ 5):** We show that CT logs can be used to detect and study phishing domains.

**CT Honeypot (§ 6):** We introduce a CT honeypot to show that third parties monitor CT logs to initiate likely malicious scans.