Technische Universität München

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

INTERNATIONAL COMPUTER SCIENCE INSTITUTE

THE UNIVERSITY OF SYDNEY

Freie Universität Berlin

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

IMC'18, Boston, Nov. 1st, 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)*

Certificate Transparency (CT)

- Goal: Provide transparency into issued certificates to detect certificate mis-issuances

Certificate Transparency (CT)

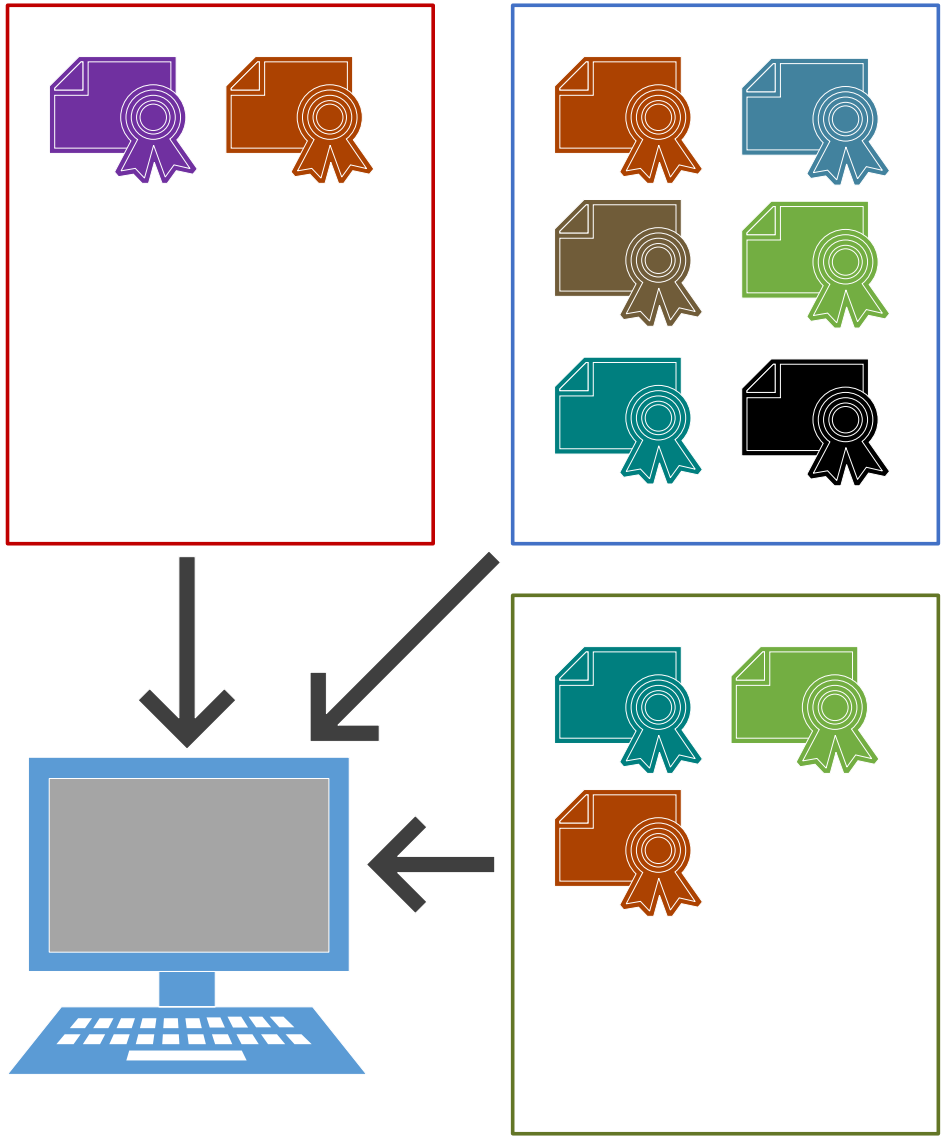
- Goal: Provide transparency into issued certificates to detect certificate mis-issuances

- Method: Uses public, append-only logs to record certificates

Certificate Transparency (CT)

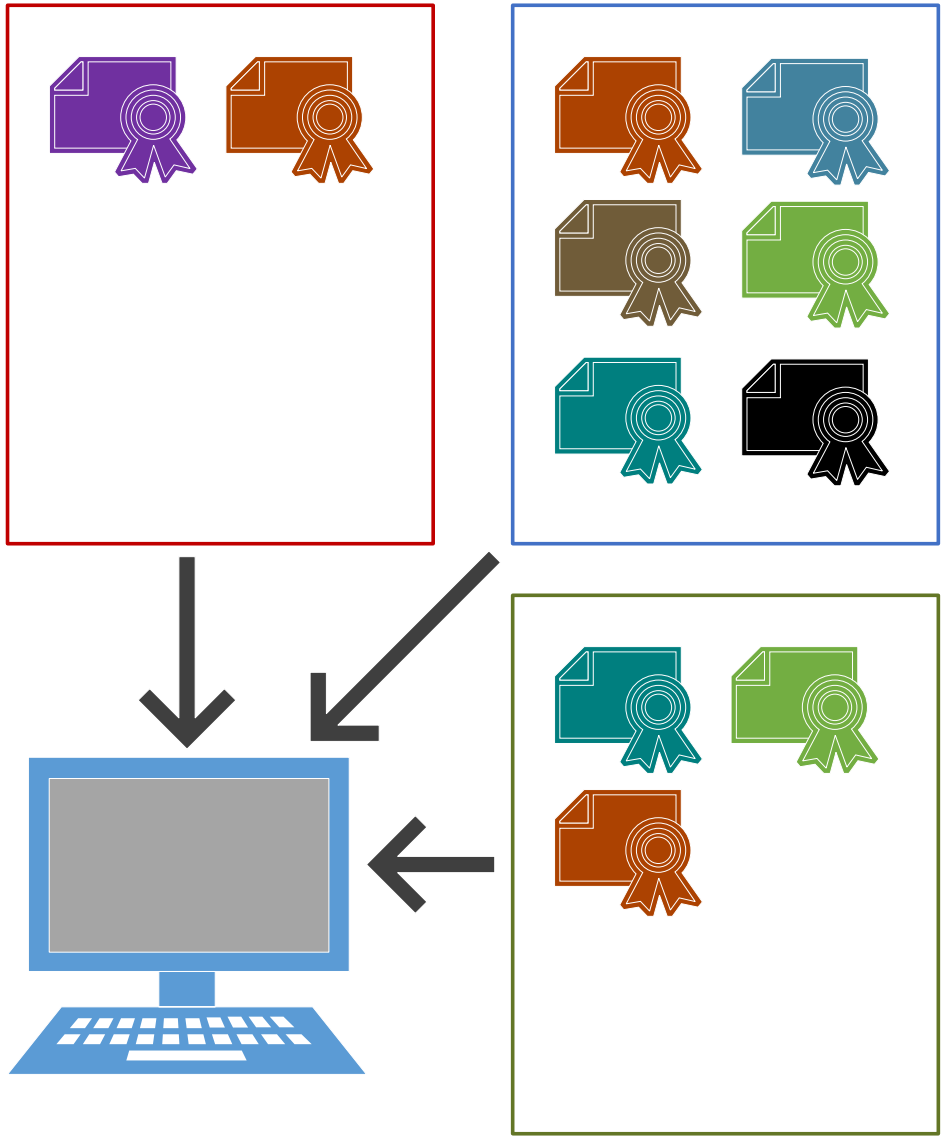
- Goal: Provide transparency into issued certificates to detect certificate mis-issuances

- Method: Uses public, append-only logs to record certificates



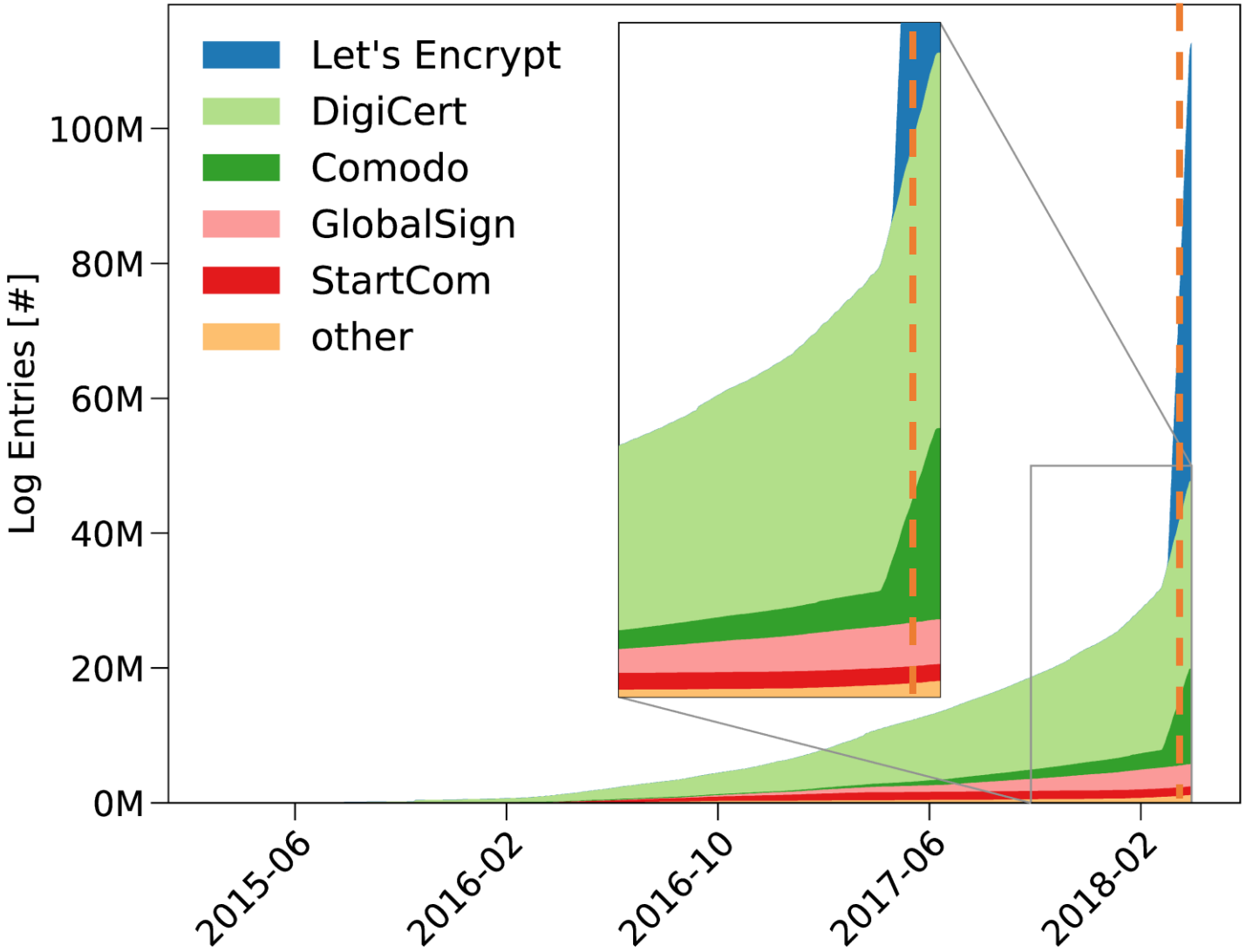
Certificate Transparency (CT)

- Goal: Provide transparency into issued certificates to detect certificate mis-issuances
- Method: Uses public, append-only logs to record certificates
- **Rise and Implications?**

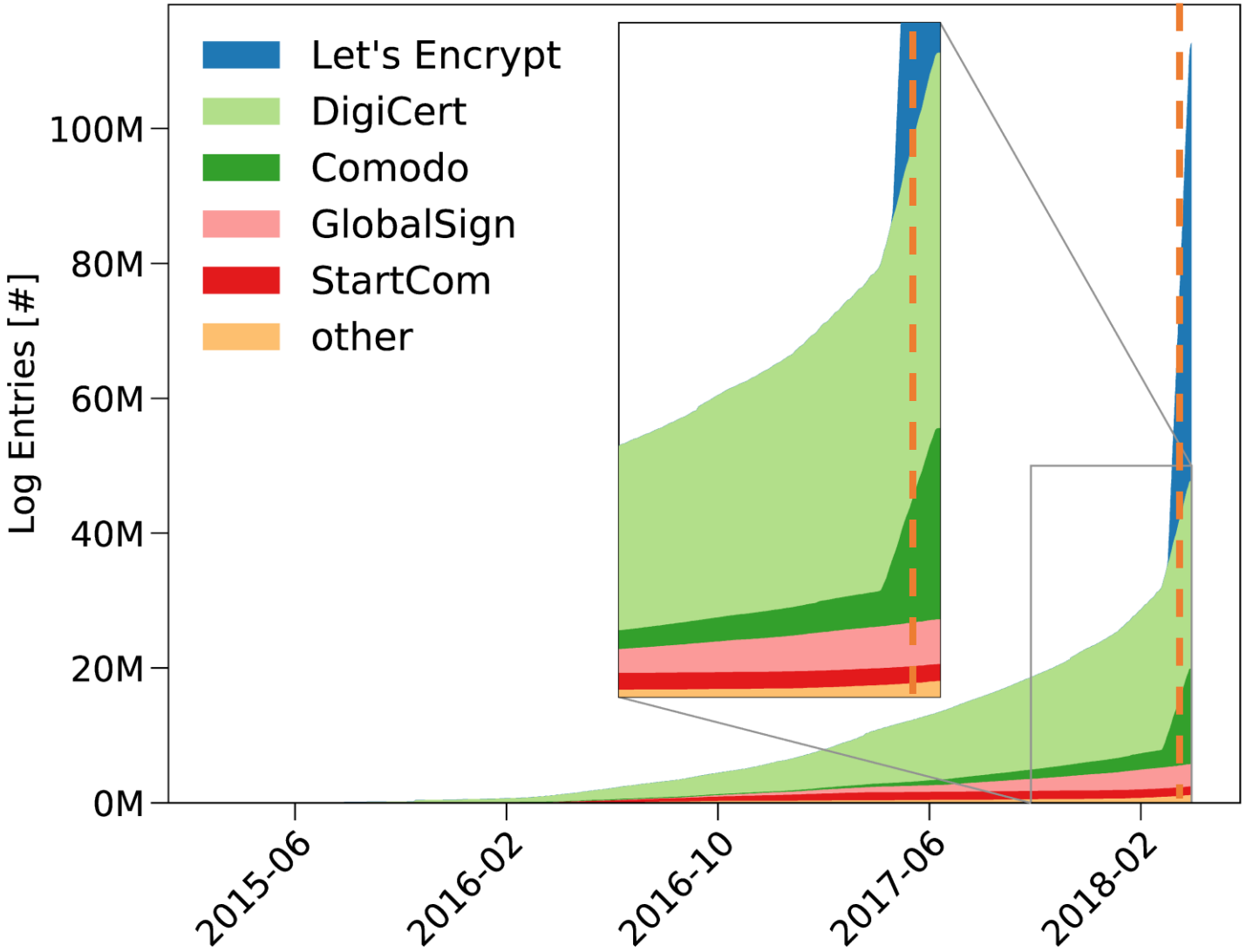


The Rise of Certificate Transparency and Its Implications on the Internet Ecosystem

How did the log volume change over time?

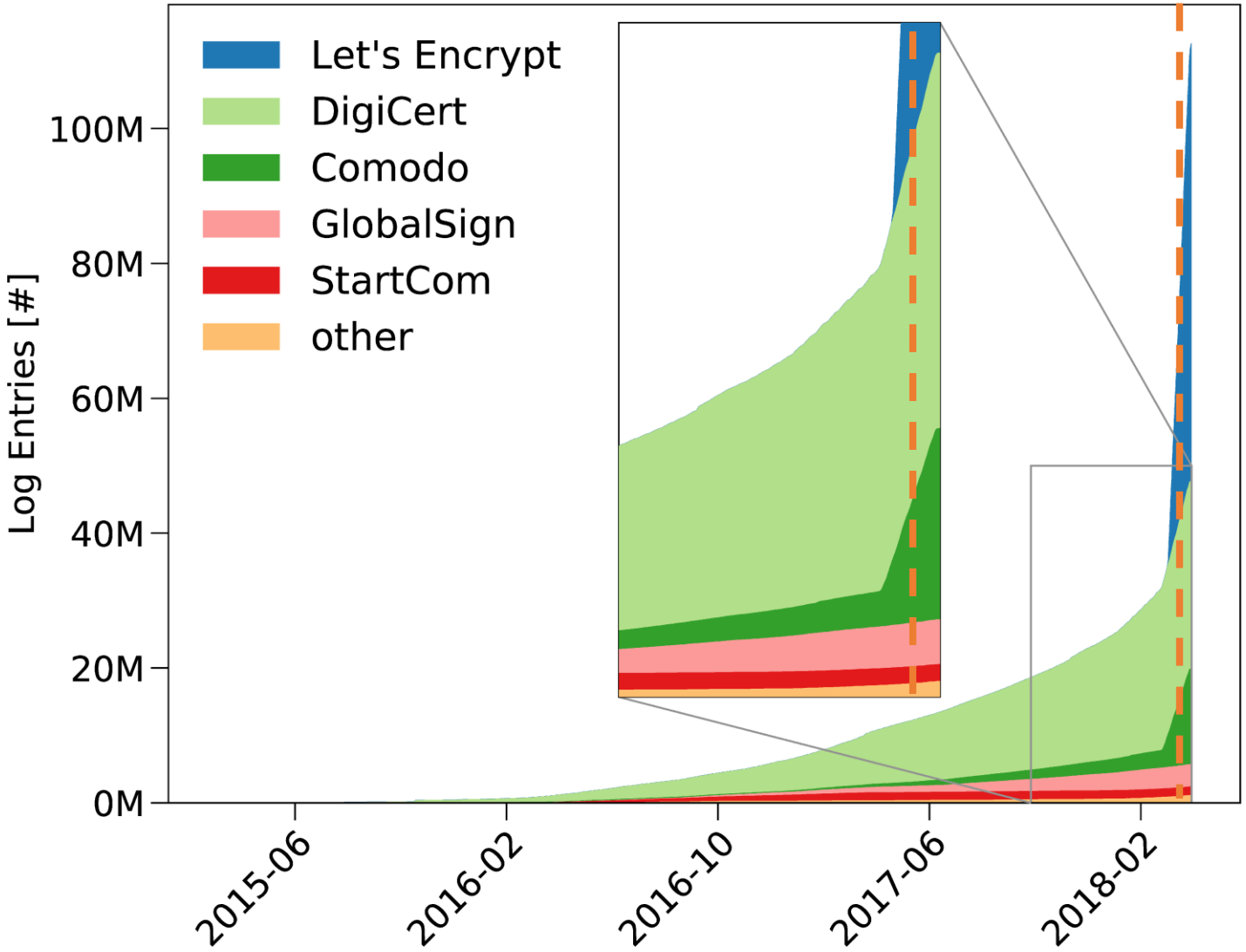


How did the log volume change over time?



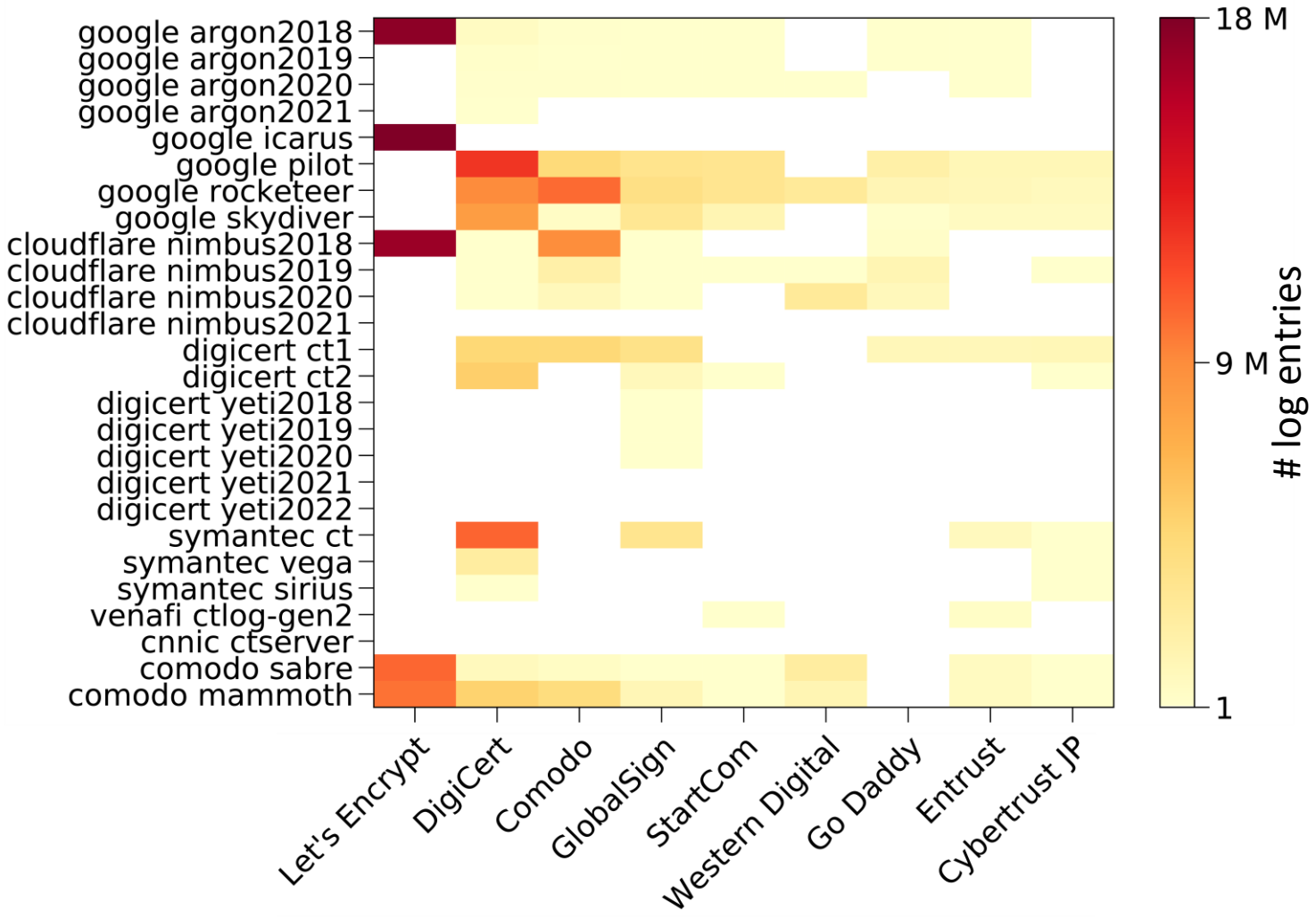
- Large increase of log entries before CT deadline
- Let’s Encrypt dominates

How did the log volume change over time?

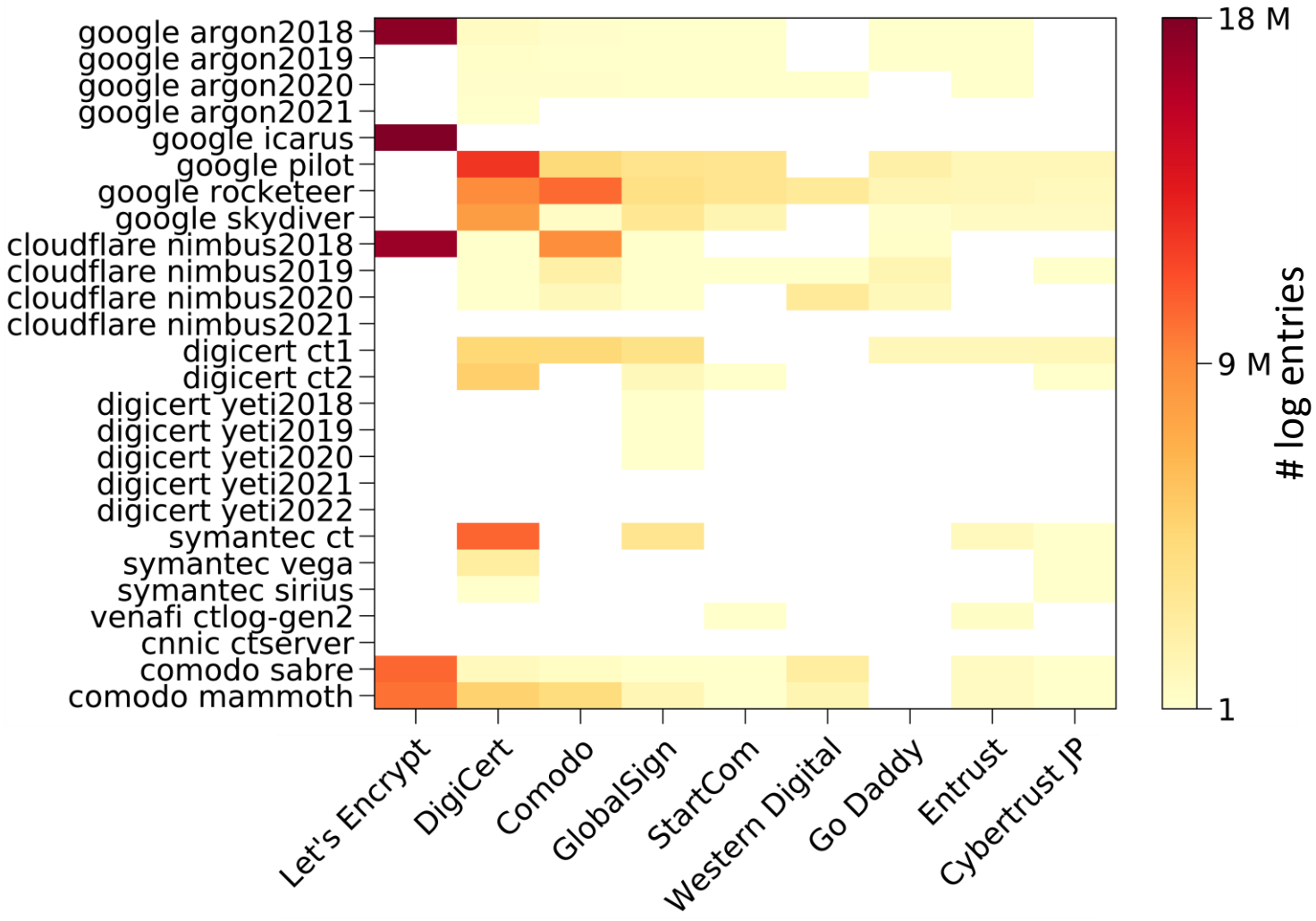


- Large increase of log entries before CT deadline
- Let’s Encrypt dominates
- **Strong rise**

Are CAs distributing certificates over many CT logs?

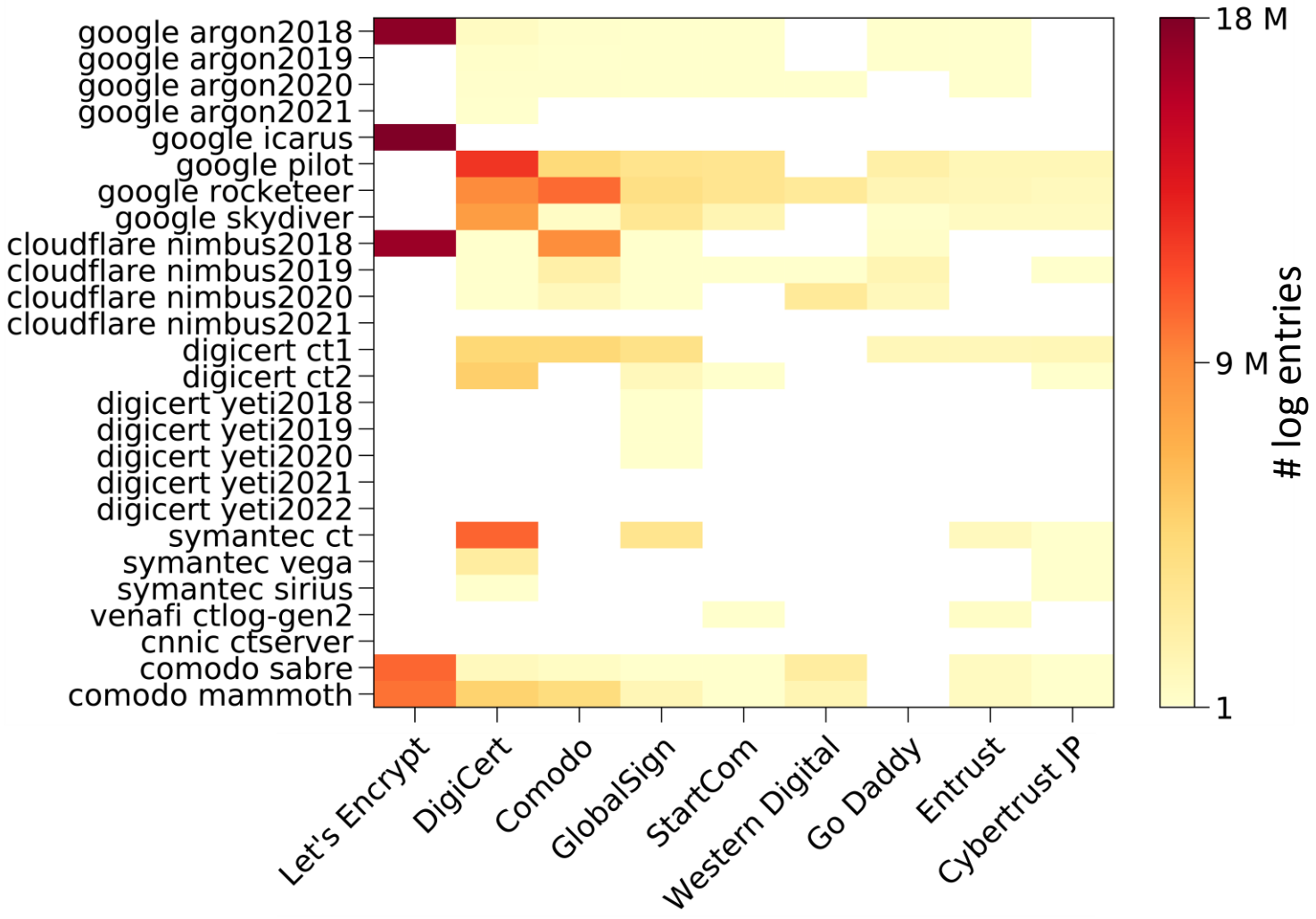


Are CAs distributing certificates over many CT logs?



- System overly reliant on few logs
- Almost all CAs use few logs for their certificate

Are CAs distributing certificates over many CT logs?



- System overly reliant on few logs
- 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
- `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
- `appleid.apple.com-7etr6eti.gq`

Results

- Identified 63k phishing domains for Apple alone
- 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
- `appleid.apple.com-7etr6eti.gq`

Results

- Identified 63k phishing domains for Apple alone
- CERT confirmed that a subset was used to host malicious content
- **Yes**, CT can be used to find malicious domains

Does CT leak private data to attackers?

Does CT leak private data to attackers?

Method

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

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

Does CT leak private data to attackers?

Method

- Extract subdomain labels from all CT logged certificates
  - `dev for .io`
- Generate new FQDNs with most common subdomain labels
  - `dev.foureyes.io`

Results

- 18.8M new FQDNs found

Does CT leak private data to attackers?

Method

- Extract subdomain labels from all CT logged certificates
  - `dev for .io`
- Generate new FQDNs with most common subdomain labels
  - `dev.foureyes.io`

Results

- 18.8M new FQDNs found
- **Yes**, CT helps attackers find previously unknown domains

Is CT actively being misused to find victims?

Is CT actively being misused to find victims?

Method

- Create CT honeypot for scanners
- Log certificate for pseudorandom subdomain
- Check DNS and webserver activity

Is CT actively being misused to find victims?

Method

- Create CT honeypot for scanners
- Log certificate for pseudorandom subdomain
- Check DNS and webserver activity

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

- Create CT honeypot for scanners
- Log certificate for pseudorandom subdomain
- Check DNS and webserver activity

Results

- First DNS lookups after 1 minute, HTTP(S) access after 1 hour
- Most scanners without info in rDNS, WHOIS, or on website
- **Yes**, CT is being misused by actors with undeclared intent

Data and code available ✓

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