Cyber security cOmpeteNCe fOr Research anD InnovAtion

Using TLS Fingerprints for OS Identification in Encrypted Traffic

Martin Laštovička, Stanislav Špaček, Petr Velan, Pavel Čeleda

{lastovicka|spaceks|velan|celeda}@ics.muni.cz

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 830927.

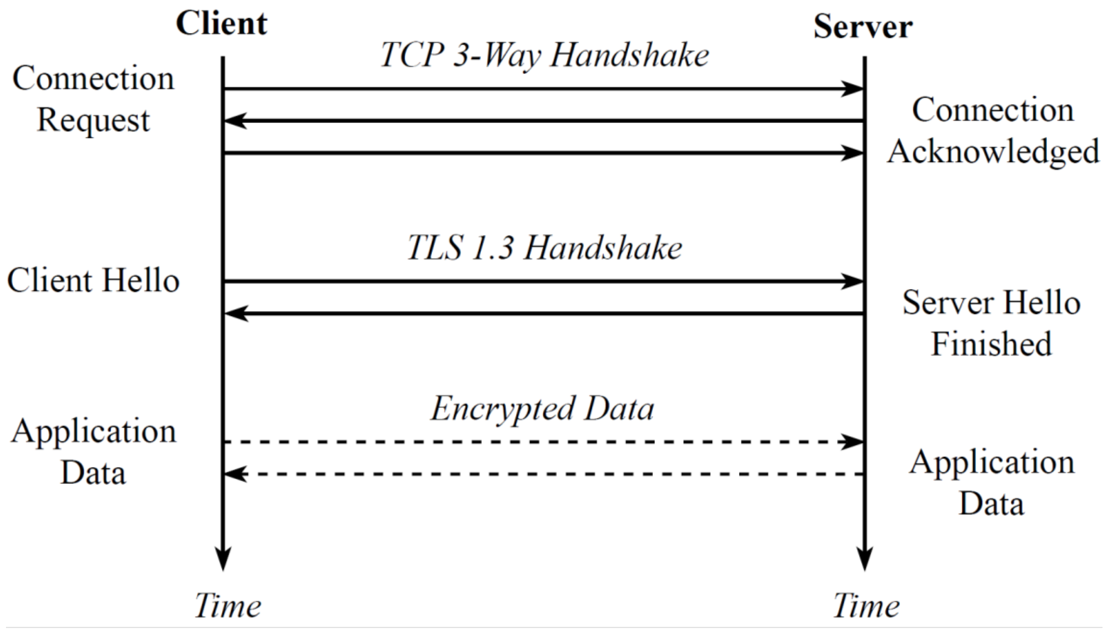
Motivation

- Asset identification in dynamic networks
  - Bring Your Own Device (BYOD) detection
  - Vulnerable devices detection
- Traffic encryption
  - Google reports up to 95% encrypted traffic on their servers
  - Focus on TLS 1.3

Methods

- Follow-up for NOMS 2018 paper – *Passive OS Fingerprinting Methods in the Jungle of Wireless Networks* – http://dx.doi.org/10.1109/NOMS.2018.8406262

- TCP/IP parameters
- TLS handshake parameters
- Specific domains
- HTTP User-Agent



TLS Handshake Parameters

- Client hello messages
- One-hot encoding
- Decision tree
    - Training on paired HTTP and HTTPS flows of the same device and Wi-Fi session

- TLS client version
- TLS cipher suites
- TLS extension types
- TLS extension length
- TLS supported groups
- TLS elliptic curves point formats

TCP/IP Parameters

- Every TCP packet
- No encoding used
- TTL rounded up to 2^n
- Decision tree
  - Training on HTTP of the same flow

- TCP SYN packet size
- TCP window size
- TTL of TCP SYN packet

SYN packet size	TCP window size	TTL	OS
52	8192	128	Windows 6.1
52	65535	128	Windows 10.0
60	65535	64	Android 6.0
60	29200	64	Ubuntu
64	65535	64	Mac OS X 10.12

Specific Domains and User-Agent

- Dictionary lookup
- Serve as a comparison to encrypted traffic identification

msftconnecttest.com
msftncsi.com
update.microsoft.com

clients3.google.com/generate_204
connectivitycheck.android.com

- HTTP Hostname
- TLS Server Name Indication
- HTTP User-Agent

Mozilla/5.0 (Linux; **Android 7.0**; SM-G930VC Build/NRD90M; wv) AppleWebKit/537.36 (KHTML, like Gecko) Version/4.0 Chrome/58.0.3029.83 Mobile Safari/537.36

Dataset

Campus Wireless Network

- 5 days of traffic
- 18 708 983 enriched flows
- 10 734 unique users
- 11 962 unique MAC addresses
- 45 602 unique Wi-Fi sessions
- 8 071 unique IPv4 addresses assigned
- Publicly available at https://zenodo.org/record/3461771
    - Dataset Using TLS Fingerprints for OS Identification in Encrypted Traffic

Ground Truth

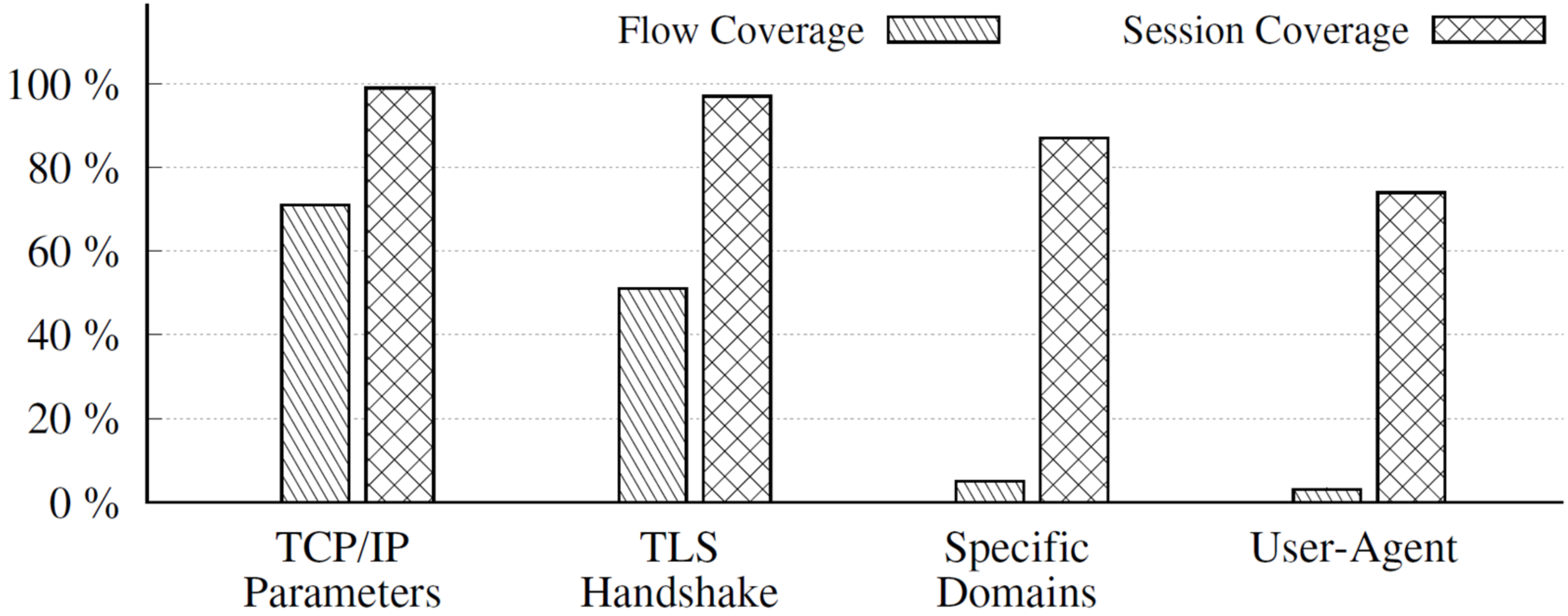
- Logs from DHCP (PPPoE concentrators)
  - MAC addresses
  - Device names
- Logs from RADIUS
  - Users
  - Sessions
- Manual mapping to OS
- Mapping of log records to flow data

Ground Truth

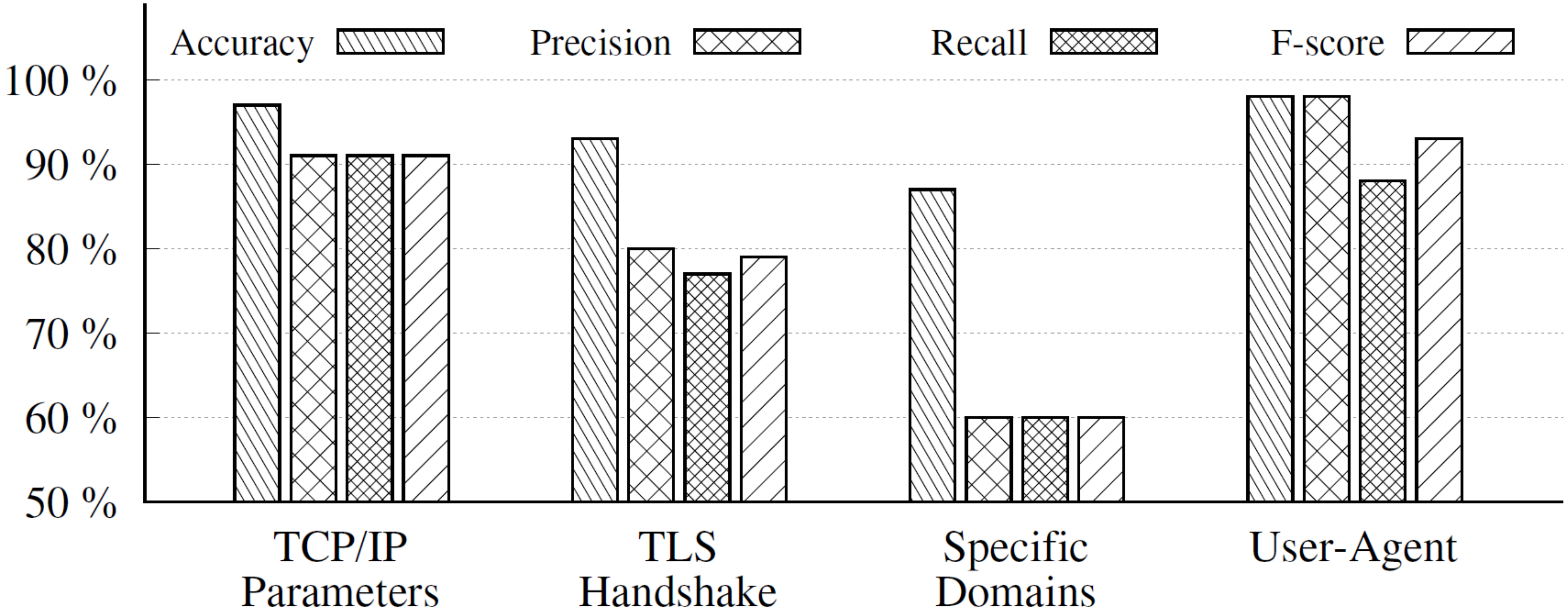
Method	Vendor	OS Name	Major Version	Minor Version
TCP/IP parameters	✓	✓	(✓)	(✓)
TLS handshake	✓	✓	(✓)	(✓)
User-Agent	✓	✓	✓	✓
Specific domains	✓	✓	✗	✗
Ground truth	✓	(✓)	✗	✗

Results

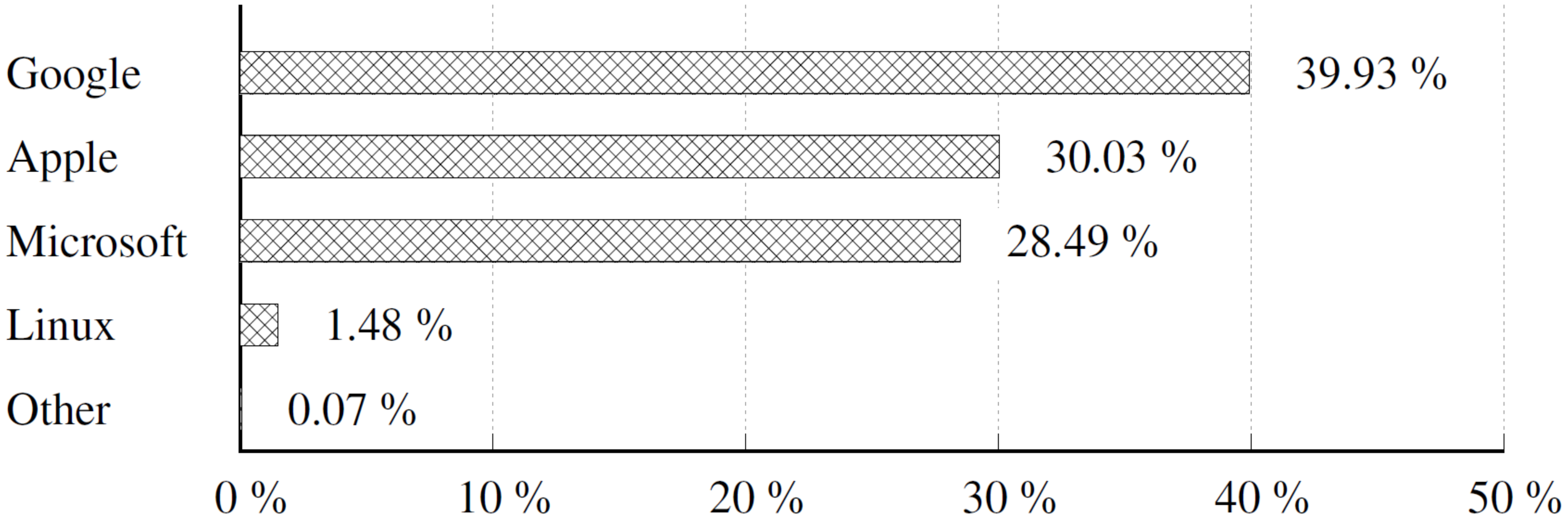
Methods’ Coverage



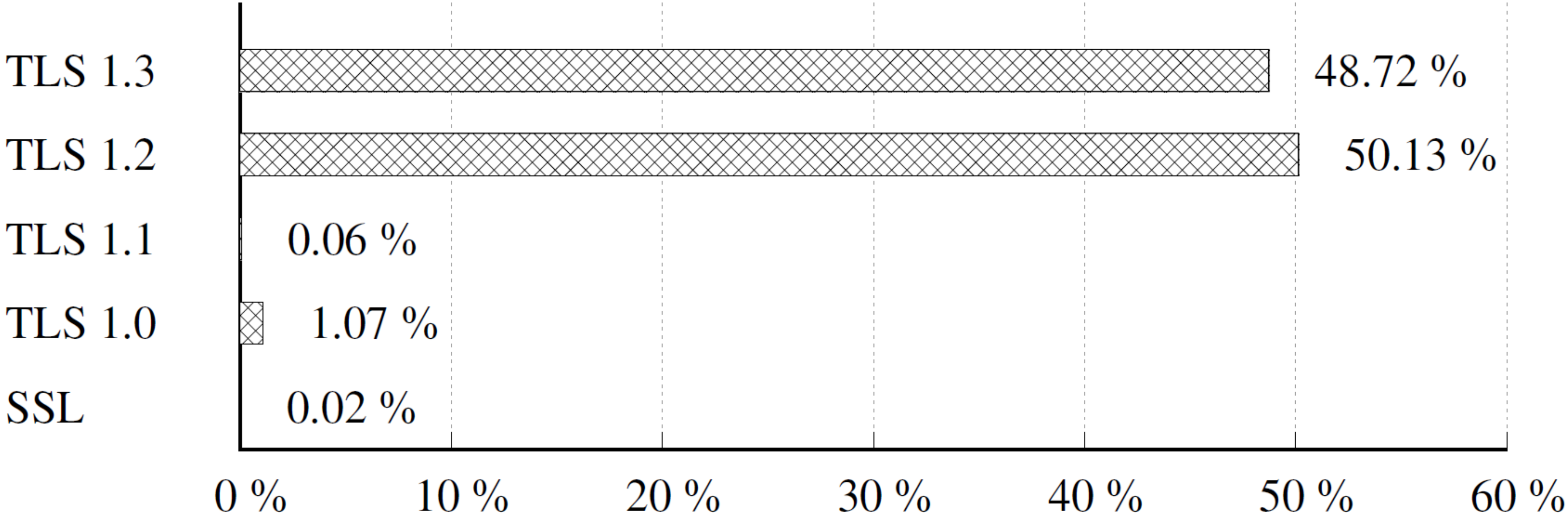
Performance Metrics



Vendors’ Share



TLS Usage – Clients



Lessons Learned

Which version of TLS is this?

Client request

TLS Record Layer

- Content Type: Handshake (22)
- Version: TLS 1.0 (0x0301)
- Handshake Protocol: Client Hello
  - Version: TLS 1.2 (0x0303)
  - Extension: supported_versions:
    - Supported version: Unknown (0xEAEA)
    - Supported version: TLS 1.3 (0x0304)
    - Supported version: TLS 1.2 (0x0303)
    - Supported version: TLS 1.1 (0x0302)
    - Supported version: TLS 1.0 (0x0301)

Server response

TLS Record Layer

- Content Type: Handshake (22)
- Version: TLS 1.2 (0x0303)
- Handshake Protocol: Server Hello
  - Version: TLS 1.2 (0x0303)
  - Extension: supported_versions:)
    - Supported version: TLS 1.3 (0x0304)

Grease Values (Generate Random Extensions And Sustain Extensibility)

TLS 1.3 RFC 8446

- Only 5 cipher suites to use:
  - TLS_AES_128_GCM_SHA256 (0x13,0x01)
  - TLS_AES_256_GCM_SHA384 (0x13,0x02)
  - TLS_CHACHA20_POLY1305_SHA256 (0x13,0x03)
  - TLS_AES_128_CCM_SHA256 (0x13,0x04)
  - TLS_AES_128_CCM_8_SHA256 (0x13,0x05)
- Named groups, 5x EC, 5x Finite:
  - secp256r1 (0x0017), secp384r1 (0x0018), secp521r1 (0x0019), x25519 (0x001D), x448 (0x001E)
  - ffdhe2048 (0x0100), ffdhe3072 (0x0101), ffdhe4096 (0x0102), ffdhe6144 (0x0103), ffdhe8192 (0x0104)

Google reality – draft-ietf-tls-grease-01

- Intentionally use random values
- Suggested non-existent Cipher suites
  - TBD (0x0A,0x0A) TBD (0x1A,0x1A)
  - TBD (0x2A,0x2A) TBD (0x3A,0x3A)
  - TBD (0x4A,0x4A) TBD (0x5A,0x5A)
  - TBD (0x6A,0x6A) TBD (0x7A,0x7A)
  - TBD (0x8A,0x8A) TBD (0x9A,0x9A)
  - TBD (0xAA,0xAA) TBD (0xBA,0xBA)
  - TBD (0xCA,0xCA) TBD (0xDA,0xDA)
  - TBD (0xEA,0xEA) TBD (0xFA,0xFA)
- Suggested named groups
  - 0x0A0A, 0x1A1A, 0x2A2A, 0x3A3A, 0x4A4A, 0x5A5A, 0x6A6A, 0x7A7A, 0x8A8A, 0x9A9A, 0xAAAA, 0xBABA, 0xCACA, 0xDADA, 0xEAEA, 0xFAFA

Summary

- OS identification methods are mature enough
  - work in large dynamic networks
  - can cope with majority of traffic encrypted
- Data acquisition is becoming more complex
  - flow data enhancement with application layer
  - protocols are continuously evolving and changing the specifications
- Correlation of data from multiple data sources
  - requires a lot of manual work
  - uses heuristics to correctly match log records to corresponding flows

Contact

Research Institute CODE
Carl-Wery-Straße 22
81739 Munich
Germany

contact@concordia-h2020.eu

Follow us

www.concordia-h2020.eu

www.twitter.com/concordiah2020

www.facebook.com/concordia.eu

www.linkedin.com/in/concordia-h2020

MUNI
C4E

https://c4e.cz

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS