The Generation and Use of TLS Fingerprints

Blake Anderson, PhD; David McGrew, PhD; Keith Schomburg
Cisco
Reducing the Visibility Gap
TLS Fingerprinting Overview

- TLS parameters offered in the ClientHello can provide library/process attribution [1-6]

- Applications
  - Network forensics
  - Malware detection [2]
  - Identifying obsolete/vulnerable software
  - OS fingerprinting [3]

- Advantages
  - No endpoint agent required
  - Completely passive
Fingerprinting Goals

- **Efficacy**: Maximize discerning power by including all informative data features
- **Flexibility**: Enable approximate matching where needed
- **Compatibility**: Accommodate missing data and new protocol features
- **Reversibility**: Fingerprint format is interpretable and forensically sound
- **Performance**: Fast and compact extraction and matching
Network and Endpoint Data Fusion

• Problem: Current fingerprint databases are slow to update and lack real-world, contextual data.

• Solution: Continuously and automatically fuse network and endpoint data.
TLS Feature Extraction and Pre-Processing

- Cipher Suites
  - Generalize GREASE cipher suites: 0x0a0a,...,0xfafa -> GREASE

- Extensions
  - Generalize GREASE extension types/data
    - 0x0a0a,...,0xfafa -> GREASE
  - Remove session specific extension data
    - server_name, padding, session_ticket
## Comparison with Previous Work

<table>
<thead>
<tr>
<th></th>
<th>Database Size</th>
<th>Automatically Updated</th>
<th>GREASE Support</th>
<th>Static Extension Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Work</td>
<td>~1,500</td>
<td>Yes</td>
<td>Yes</td>
<td>supported_groups, ec_point_formats, status_request, signature_algorithms, application_layer_protocol_negotiation, supported_versions, psk_key_exchange_modes</td>
</tr>
<tr>
<td>Kotzias et al. [4]</td>
<td>~1,684</td>
<td>No</td>
<td>Discards Locality</td>
<td>supported_groups, ec_point_formats</td>
</tr>
<tr>
<td>JA3 [5]</td>
<td>158</td>
<td>No</td>
<td>Discards All Data</td>
<td>supported_groups, ec_point Formats</td>
</tr>
<tr>
<td>FingerprinTLS [6]</td>
<td>409</td>
<td>No</td>
<td>No</td>
<td>supported_groups, ec_point_formats, signature_algorithms</td>
</tr>
</tbody>
</table>
TLS Fingerprint Database Schema

Metadata

TLS Information

Attribution
TLS Fingerprint Database Schema

Metadata

TLS Information

Attribution
TLS Fingerprint Database Schema

Metadata

```
"process_info": [
  {
    "process": "chrome.exe",
    "application_category": "browser",
    "prevalence": 0.72,
    "sha256": "C0EDC58622B6FA296A439DA2328C8BF74D7BF5F8E83446441048687EB69A472"
  },
  {
    "process": "Google Chrome",
    "application_category": "browser",
    "prevalence": 0.18,
    "sha256": "E42240A8038B687AEE9D999DB5F7215509A9FDF0A84BC307688E1788F4494790E"
  },
  {
    "process": "chrome.exe",
    "application_category": "browser",
    "prevalence": 0.82,
    "sha256": "EB23FF00CC2C6B1D4C5FC9454CACF07C8BA9F94695621AFC0702422C5E9FD082"
  }
],
```

TLS Information

```
"os_info": [
  {
    "os": "WinNT",
    "os_version": "10.0.15063",
    "os_edition": "Windows 10 Enterprise",
    "prevalence": 0.27
  },
  {
    "os": "WinNT",
    "os_version": "10.0.17134",
    "os_edition": "Windows 10 Enterprise",
    "prevalence": 0.25
  },
  {
    "os": "WinNT",
    "os_version": "6.1.7601",
    "os_edition": "Windows 7 Enterprise",
    "prevalence": 0.24
  }
]
```

Attribution
General Stats

• Generated from 30M+ real-world TLS sessions
• 1,567 fingerprints
  • 454 unique cipher suite vectors
  • 1,092 unique cipher suite + extension type vectors
• 12,644 unique process hashes
  • 2,411 unique process names
Operating System Representation

Total: 24,586,176

Mac: 8,998,267
- High Sierra
- Mojave
- El Capitan
- Sierra
- Yosemite

WinNT: 15,587,909
- 10 Enterprise
- 7 Enterprise
- Other
- Other

Other:
- Other
- Other
- Other
Application Representation

Total: 27,326,885

Browser: 14,529,118
Firefox: 4,700,830
Chrome: 8,508,891

Communication

System

Enterprise
MongoDB Compass Helper

Sysadmin

Email

Other

Storage

Productivity

Other browser
CiscoCollabHost.exe
CiscoSparkHelper.exe
Webex Teams
Other communication
com.apple.WebKit.Networking
svchost.exe
SearchProtocolHost.exe
Other system

Other enterprise

osqueryd

Other sysadmin

Microsoft Outlook

OUTLOOK.EXE

Mail

Other email

nsrlsessiond

Box Sync.exe

Box Sync

Other storage

EXCEL.EXE

POWERNP.EXE

Evernote.exe

Other productivity

Other Chrome

67.0.3396.99
69.0.3497.100
70.0.3538.77
68.0.3440.106
70.0.3538.102

Other Firefox

62.0.3
61.0.1
63.0.1
63.0.0
63.0.3
Similarity Matrix
Approximate TLS Fingerprinting

- String alignment over TLS features

<table>
<thead>
<tr>
<th>True Label</th>
<th>Inferred Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filename:</td>
<td>firefox.exe</td>
</tr>
<tr>
<td>File Version:</td>
<td>59.0.2.6656</td>
</tr>
<tr>
<td>Process Name:</td>
<td>Firefox</td>
</tr>
<tr>
<td>Process Version:</td>
<td>59.0.2.0</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Alignment</td>
<td></td>
</tr>
<tr>
<td>1301 1303 1302 c02b c02f cca9 cca8 c02c c030 - - c013 c014 - - 002f 0035 000a</td>
<td></td>
</tr>
<tr>
<td>- - - c02b c02f cca9 cca8 c02c c030 c00a c009 c013 c014 0033 0039 002f 0035 000a</td>
<td></td>
</tr>
</tbody>
</table>
Fingerprint Matching Overview

1. Identify TLS
2. Extract FP Data
3. Find Match
   - True: Report Match
   - False: Find Approximate Match
4. Update Database with Approximate Match

Data Plane
- Identify TLS
- Extract FP Data
- Find Match
- Report Match

Control Plane
- Find Approximate Match
- Update Database with Approximate Match
Performance (Unoptimized Python)

Average Time to Fingerprint a TLS Flow

- Preprocessing Time
- Known Lookup Time
- Previously Seen Lookup Time
- Approximate Match Time

Time (ms)

Observed TLS Flows (in Millions)
Fingerprint Prevalence
TLS Fingerprint Visibility

The diagram shows the percentage of unique fingerprints over time from 2018-07-01 to 2018-11-01.

Key:
- **Ground Truth Fingerprints**
- **Similar Fingerprints**
- **Incomplete Ground Truth Fingerprints**
Implementation

- Fingerprint database and relevant code has been open-sourced:
  - https://github.com/cisco/joy

- Joy
  - Packet parsing and fingerprint extraction

- Python Scripts
  - Exact and approximate matching
  - Generation of custom fingerprint database from Joy output
Next Steps

- More data!
  - iOS, Android, and Linux
- Incorporate other fingerprint databases
- Time window analysis
References

[1] https://github.com/cisco/joy


[4] Platon Kotzias, Abbas Razaghpanah, Johanna Amann, Kenneth G. Paterson, Narseo Vallina-Rodriguez, Juan Caballero; Coming of Age: A Longitudinal Study of TLS Deployment; IMC, 2018


[6] Lee Brotherston; FingerprintTLS
Thank You