TLS BEYOND THE BROWSER

COMBINING END HOST AND NETWORK DATA TO UNDERSTAND APPLICATION BEHAVIOR

BLAKE ANDERSON, PhD, and DAVID McGREW, PhD
blake.anderson@cisco.com

November 18, 2019

TLS FINGERPRINTING

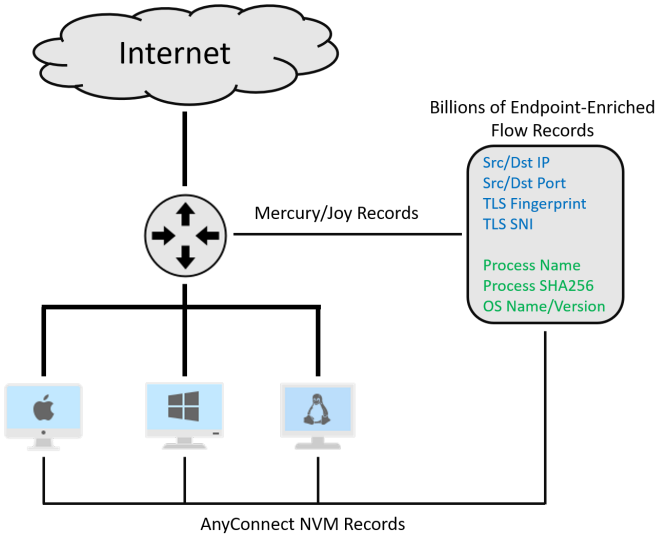
$$a : \begin{bmatrix} \texttt{ProtocolVersion:} & \texttt{0303} \\ \texttt{CipherSuites:} & \texttt{0a0a...000a} \\ \texttt{Extensions:} & \texttt{0000...} \end{bmatrix} \rightarrow \begin{bmatrix} \texttt{Process:} & \texttt{chrome.exe} \\ \texttt{Version:} & \texttt{76.0.3809.132} \\ \texttt{SHA-256:} & \texttt{5616...9acc} \\ \texttt{Category:} & \texttt{browser} \\ \texttt{OS:} & \texttt{WinNT} \\ \texttt{OSversion:} & \texttt{10.0.17134} \\ \texttt{OSedition:} & \texttt{Enterprise} \end{bmatrix}$$

TOOLS

- Mercury
  - https://github.com/cisco/mercury
  - 20+ Gbps protocol fingerprinting and analysis
  - minimal dependencies: Linux AF_PACKET/TPACKETv3
- pmercury
  - https://github.com/cisco/mercury → /python
  - Python 3-based implementation (`pip3 install pmercury`)

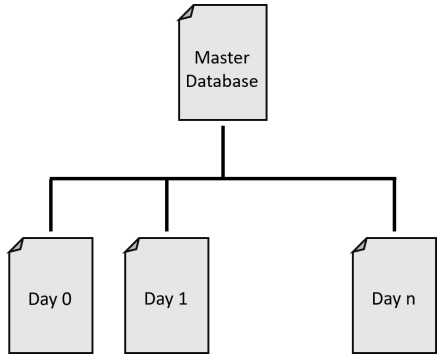
NETWORK / ENDPOINT FUSION PIPELINE

- 24,000 users
- 5 Distinct Geographies
- Primarily OSX/WinNT
- Identifying information is stripped
- Applicable to **all** protocols



BUILDING THE MASTER DATABASE

- Efficiently perform longitudinal studies
- Age out older data
- https://github.com/cisco/mercury → /resources/fingerprint_db.json.gz



DATABASE FORMAT


```
{
  "str_repr": "(0303)(0a0a...)((0000)...)",
  "first_seen": "2019-05-20",
  "last_seen": "2019-09-20",
  "max_implementation_date": "2018-10",
  "min_implementation_date": "1999-01",
  "total_count": 82701077,
  "process_info": [{
     "process": "chrome.exe",
     "sha256": "064F...638",
     "application_category": "browser",
     "count": 16139014,
     "classes_ip_as": {...},
     "classes_hostname_tlds": {...},
     "classes_hostname_domains": {...},
     "classes_port_applications": {...},
     "os_info": {...}},
    ...
```

DATABASE STATISTICS

Data Source	# Fingerprints	# Connections
Passive	64,214	4.10e10
Endpoint	7,909	5.43e9
Malware	5,633	3.61e7
Total	69,310	4.65e10

- Passive: maps to destinations
- Endpoint: maps to processes/destinations
- Malware: maps to malicious process/destinations
- $\sim$2 billion new sessions per day
- $\sim$200 million new Endpoint-labeled sessions per day

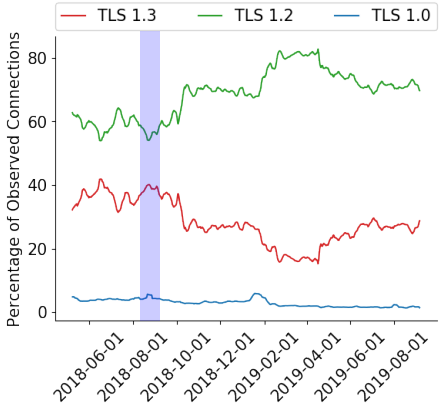
WHAT DOES A FINGERPRINT GIVE YOU?

Category	Processes per Fingerprint		Fingerprints per Process	
browser	22.19	(50.79)	3.18	(3.12)
email	65.73	(118.84)	2.67	(2.76)
communication	64.00	(108.78)	1.87	(1.51)
system	41.07	(94.41)	1.63	(1.30)
productivity	56.42	(108.72)	1.86	(1.47)
security	76.77	(138.60)	2.05	(2.64)
storage	48.51	(110.38)	1.59	(1.40)
other	52.99	(113.56)	1.53	(0.99)

- Many processes map to the same TLS fingerprint!
  - Chrome/Firefox bundle their own TLS library and use more extensions
  - Destination information helps to disambiguate processes
- $\sim$70% of fingerprints were strong indicators of Major OS

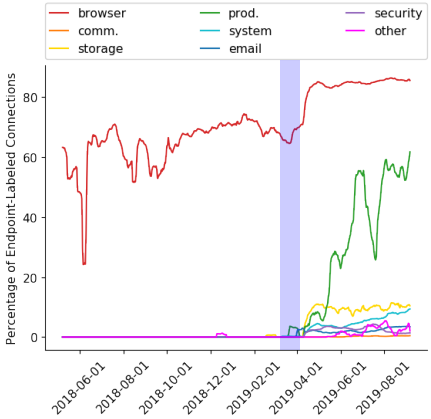
TLS 1.3 ADOPTION

- TLS 1.3 support peaked when the RFC was published (August 2018)
- Support then declined over the following 6 months
- Support seems to rebound in March 2019

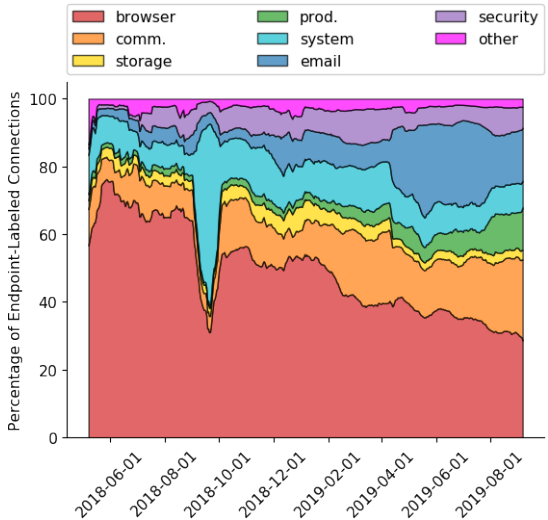


TLS 1.3 ADOPTION BY APPLICATION CATEGORIES

- Chrome/Firefox were initially the only applications supporting 1.3
- March 2019: MacOS 10.14.4/CoreTLS turn 1.3 on by default



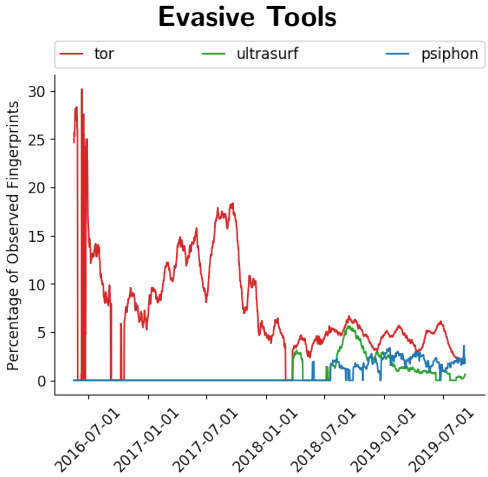
NON-BROWSER TLS MARKET SHARE

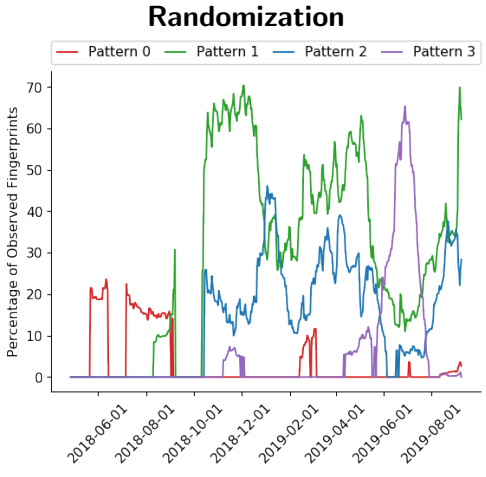


TLS FINGERPRINTING FOR MALWARE DETECTION

- Analyzed https://sslbl.abuse.ch/ja3-fingerprints/
  - successfully reverse engineered 64 of the 67 malware fingerprints
  - from abuse.ch: *These fingerprints have not been tested against known good traffic yet and may cause a significant amount of FPs!*
- 55 of the fingerprints were regularly used by benign software
- The remaining 9 fingerprints were associated with older versions of Windows or OpenSSL

MALWARE'S ABUSE OF CENSORSHIP CIRCUMVENTION TOOLS





CONCLUSIONS

- Endpoint/Network datasets provide valuable insights into the TLS/Network ecosystem
- TLS Fingerprinting must:
  - use an up-to-date database
  - leverage additional data outside the `client_hello`
- TLS Beyond the Browser (IMC'19)


THANK YOU!
cisco

APPLICATION CATEGORIES

Category	# Connections (millions)	% Connections
browser	173.3	37.1%
communication	90.0	19.3%
productivity	80.1	17.2%
email	42.2	9.0%
system	31.7	6.8%
security	28.9	6.2%
other	12.5	4.6%
storage	7.8	1.7%

FINGERPRINT LONGEVITY

- Purple: Applications using system libraries and some browsers
- Blue: Browsers
- Green: Unpopular applications
- Red: Scanners and evasive applications