NETWORK-BASED HTTPS CLIENT IDENTIFICATION USING SSL/TLS FINGERPRINTING

Monday 24th August, 2015

Martin Husák
Milan Čermák
Tomáš Jirsík
Pavel Čeleda

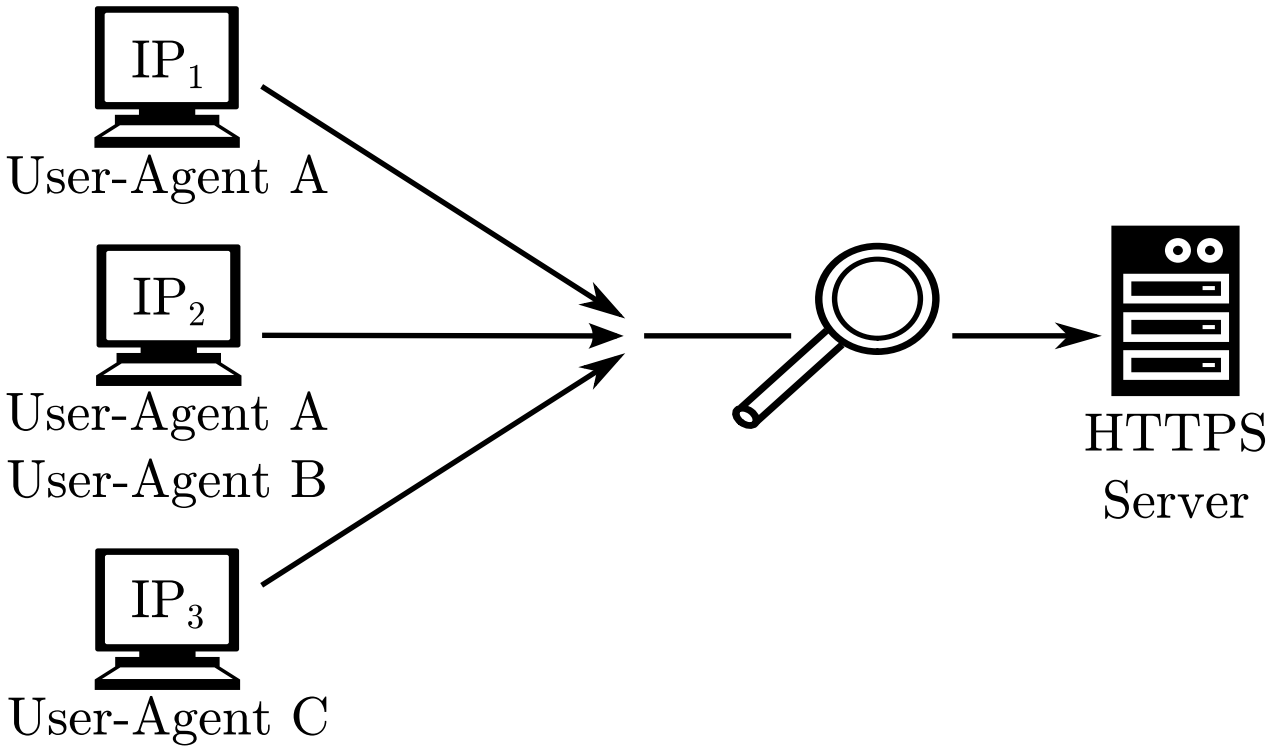
CSIRT-MU

Introduction

- Rising popularity of encrypted traffic secures the transmission, but also prevents legitimate monitoring and classification.
- Lot of work has been done on HTTP traffic identification and classification, but it is useless when dealing with HTTPS.
- The adversaries may evade disclosure by hiding malicious behavior in encrypted connections.
- Is there anything we can do to analyse encrypted traffic while preserving privacy of communication?
- For example, User-Agent is used often for analyses. Do we have anything similar in HTTPS?

Motivation I

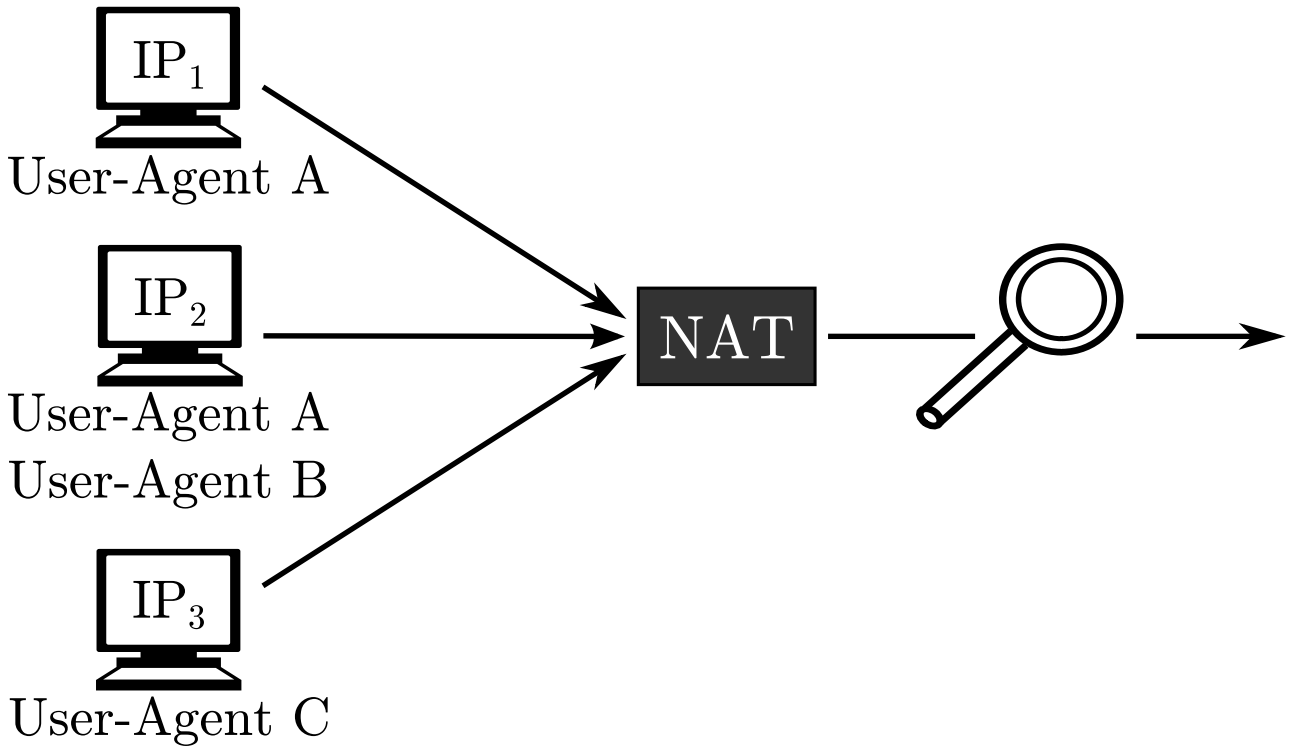
What can we tell about clients accessing an HTTPS server without access to system logs on the machine?



CSIRT-MU

Motivation II

What about clients behind NAT?
Can we enumerate them and estimate their types?

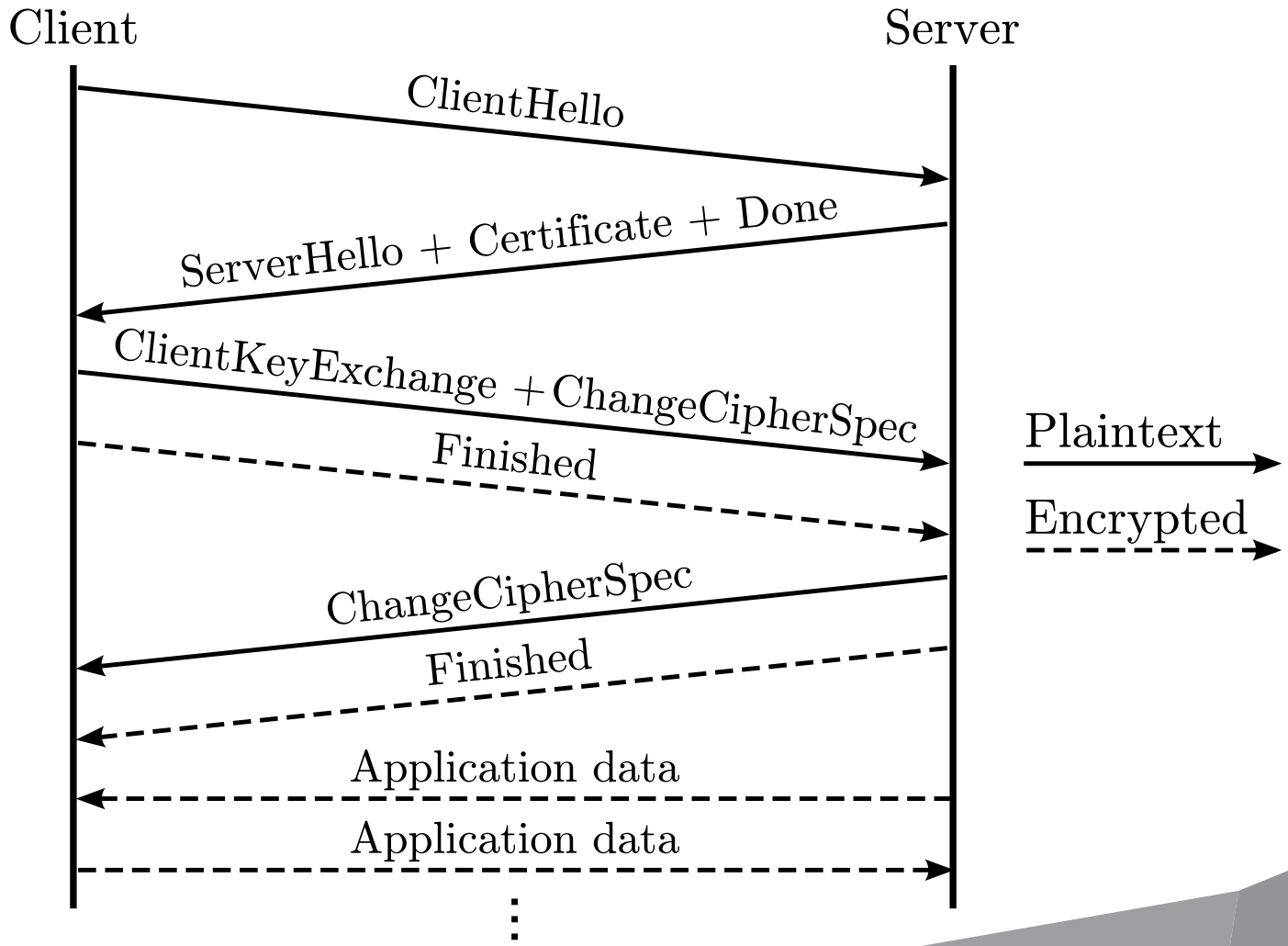


CSIRT-MU

Hypothesis

It is possible to estimate a User-Agent of a client in HTTPS communication **knowing only the parameters of SSL/TLS handshake**.

SSL/TLS Traffic Measurement



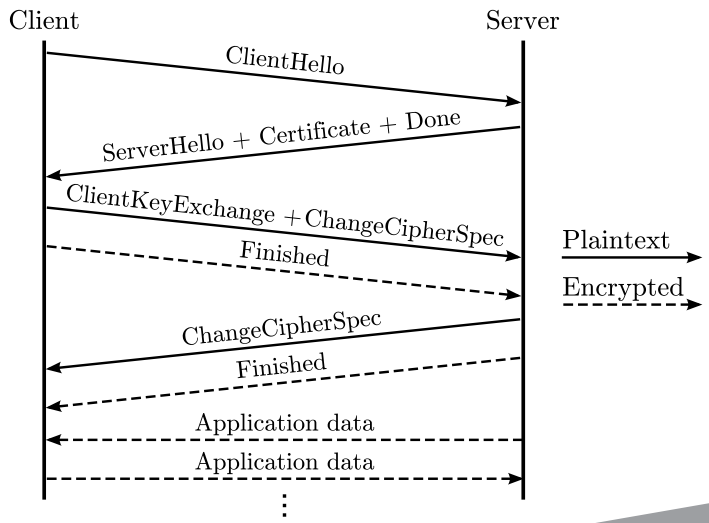
CSIRT-MU

SSL/TLS Traffic Measurement

ClientHello

- Protocol version,
- **cipher suite list**,
- extensions.

Cipher suite list is the most variable SSL/TLS handshake parameter.



CSIRT-MU

Research Questions

Question I.

Which parameters of a SSL/TLS handshake can be used for client identification?

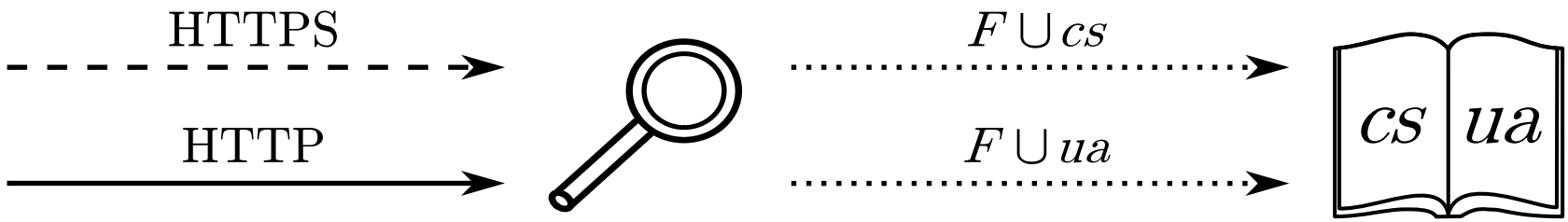
Question II.

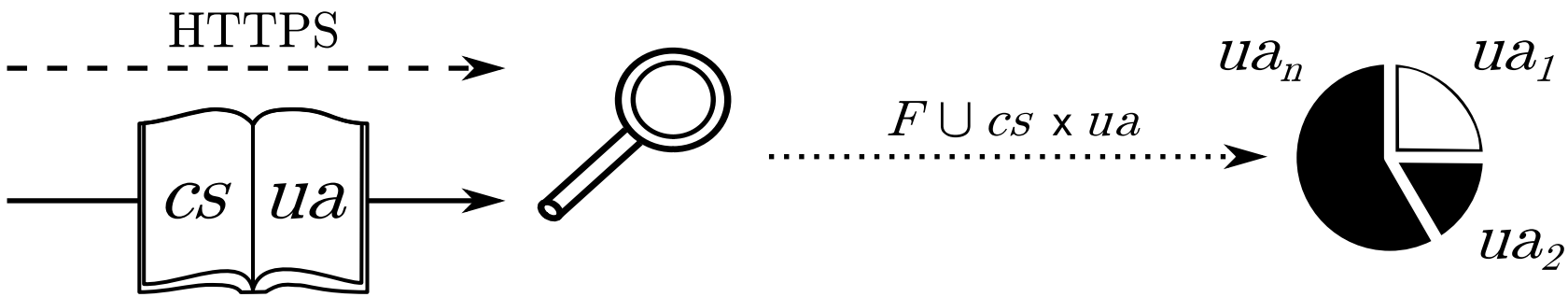
How can we build a dictionary of SSL/TLS handshakes and HTTP User-Agents?

Question III.

How large does the dictionary need to be to cover a significant portion of network traffic?

Experiment design





CSIRT-MU

Pairing Ciper Suite Lists and User-Agents

Host-based method

- Proposed earlier by Ristić et al.
- The results are exact, but it is difficult to obtain large dictionary.
- Limited to a single host (web server).
- Limited set of client types that can be observed.

Pairing Ciper Suite Lists and User-Agents

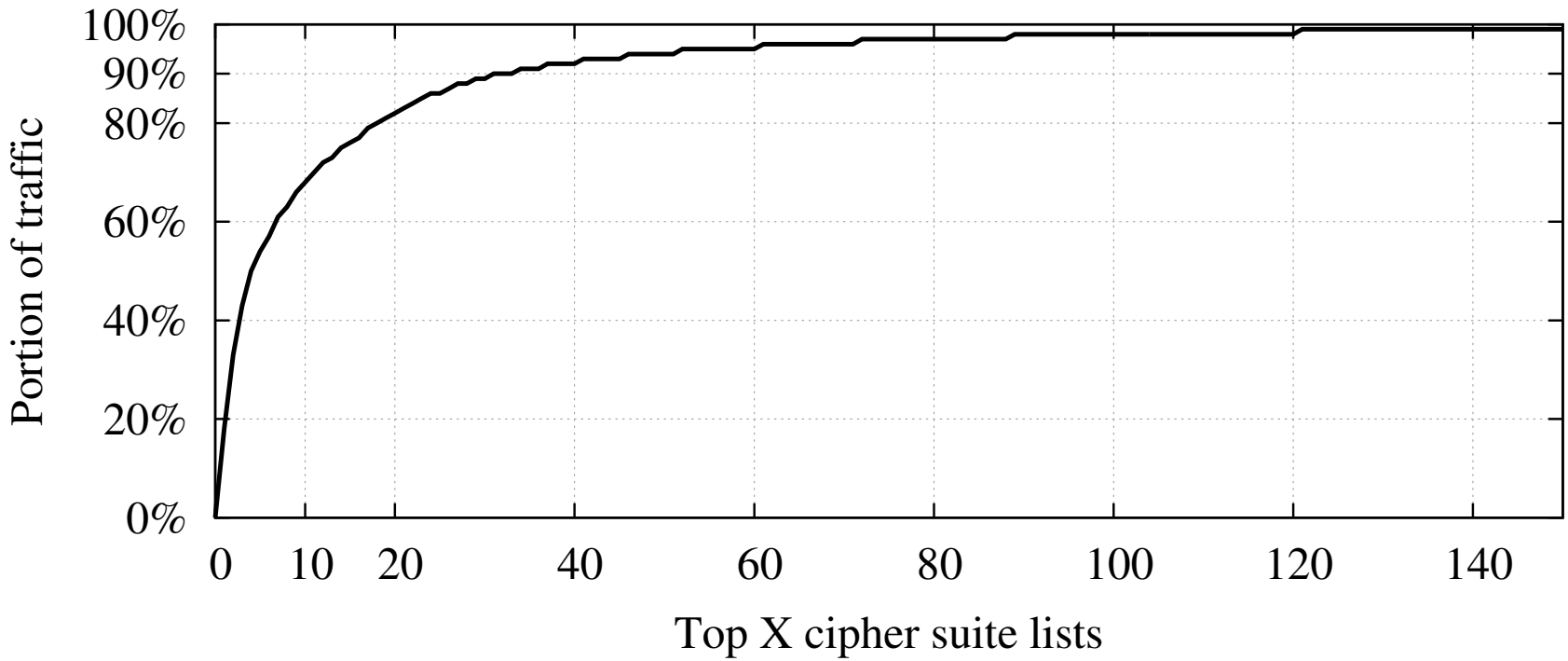
Network-based method

- Clients commonly communicate via both HTTP and HTTPS.
- HTTP and HTTPS connections with the same source IP address are selected.
- Cipher suite list from the HTTPS connection is paired to the User-Agent from the HTTP connection that is the closest in time.
- Not limited to a single host.
- Can detect any client type.
- Better reflects the structure of live network traffic.

Experiment Results I

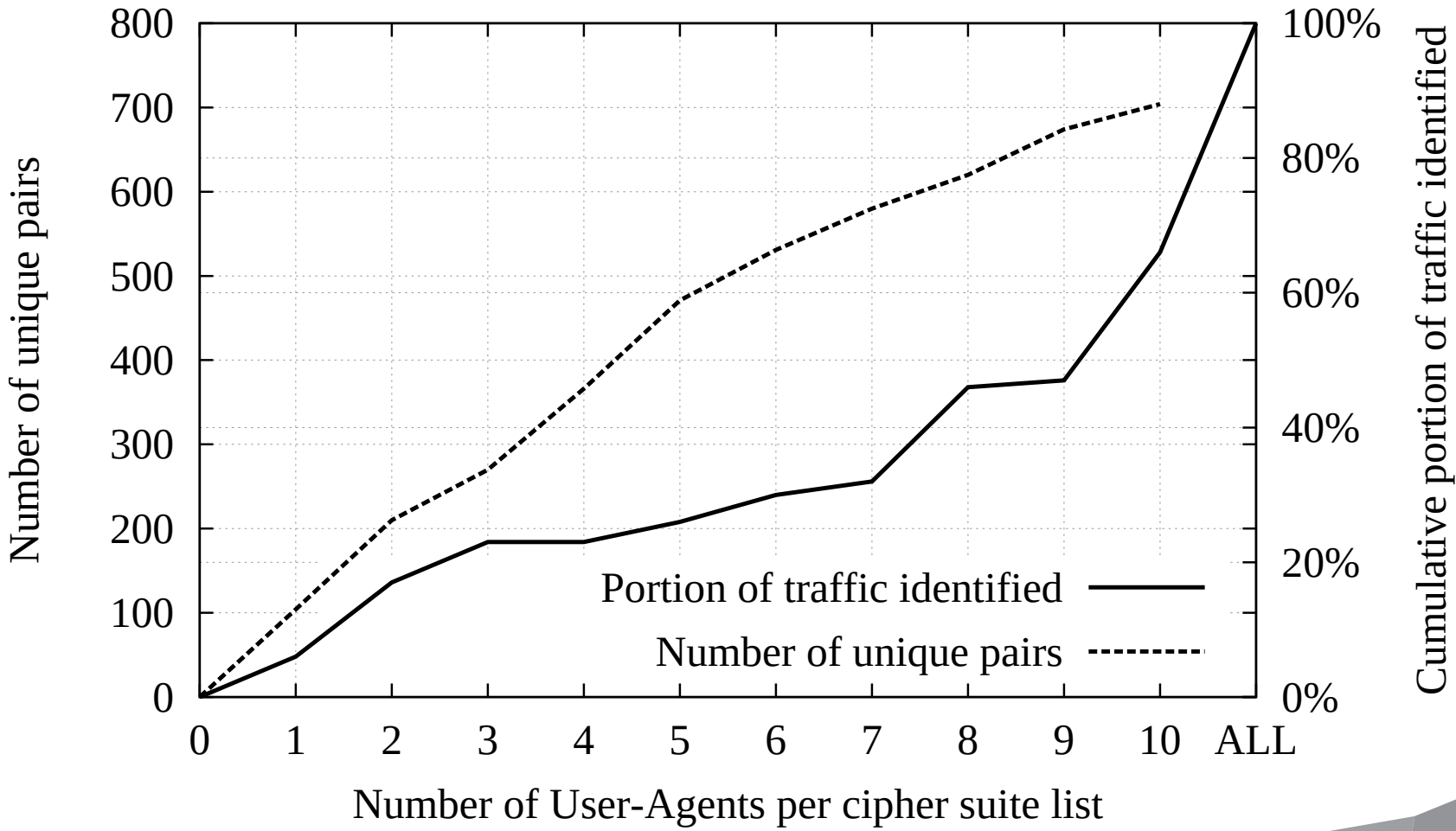
- Over 85 million HTTPS connection were processed during a week in our campus network.
- 307 pairs (72 unique cipher suite lists) were collected using host-based method on a single host.
- 12,832 pairs (305 unique cipher suite lists) were collected using network-based method in our campus network.
- The final dictionary is a union of the two (316 unique cipher suite lists).
- We were able to assign a User-Agent to 99.6 % of HTTPS connections.
- 57 % of connections used TLS 1.2, 40 % used TLS 1.0.

Experiment Results II



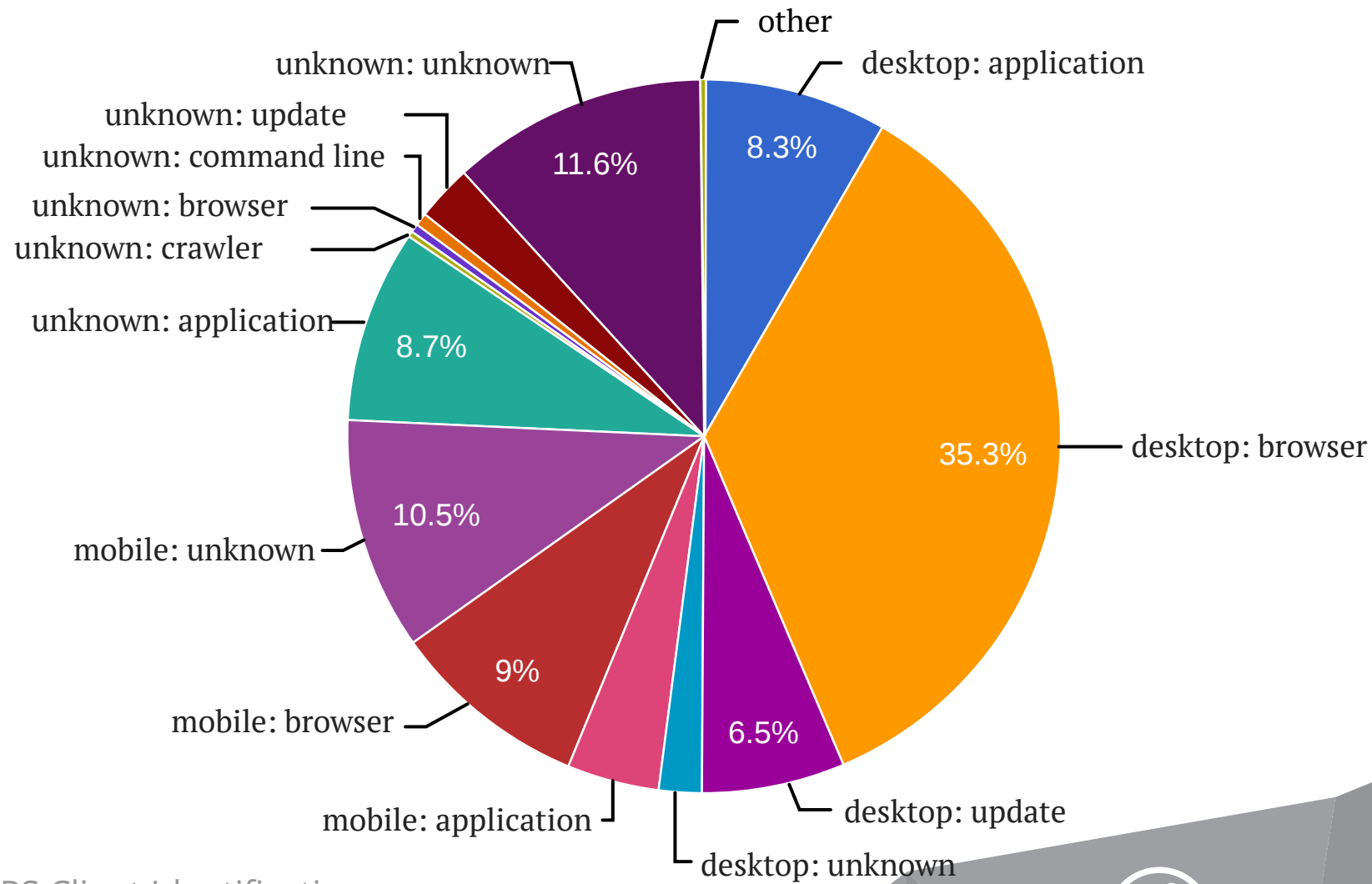
CSIRT-MU

Experiment Results III



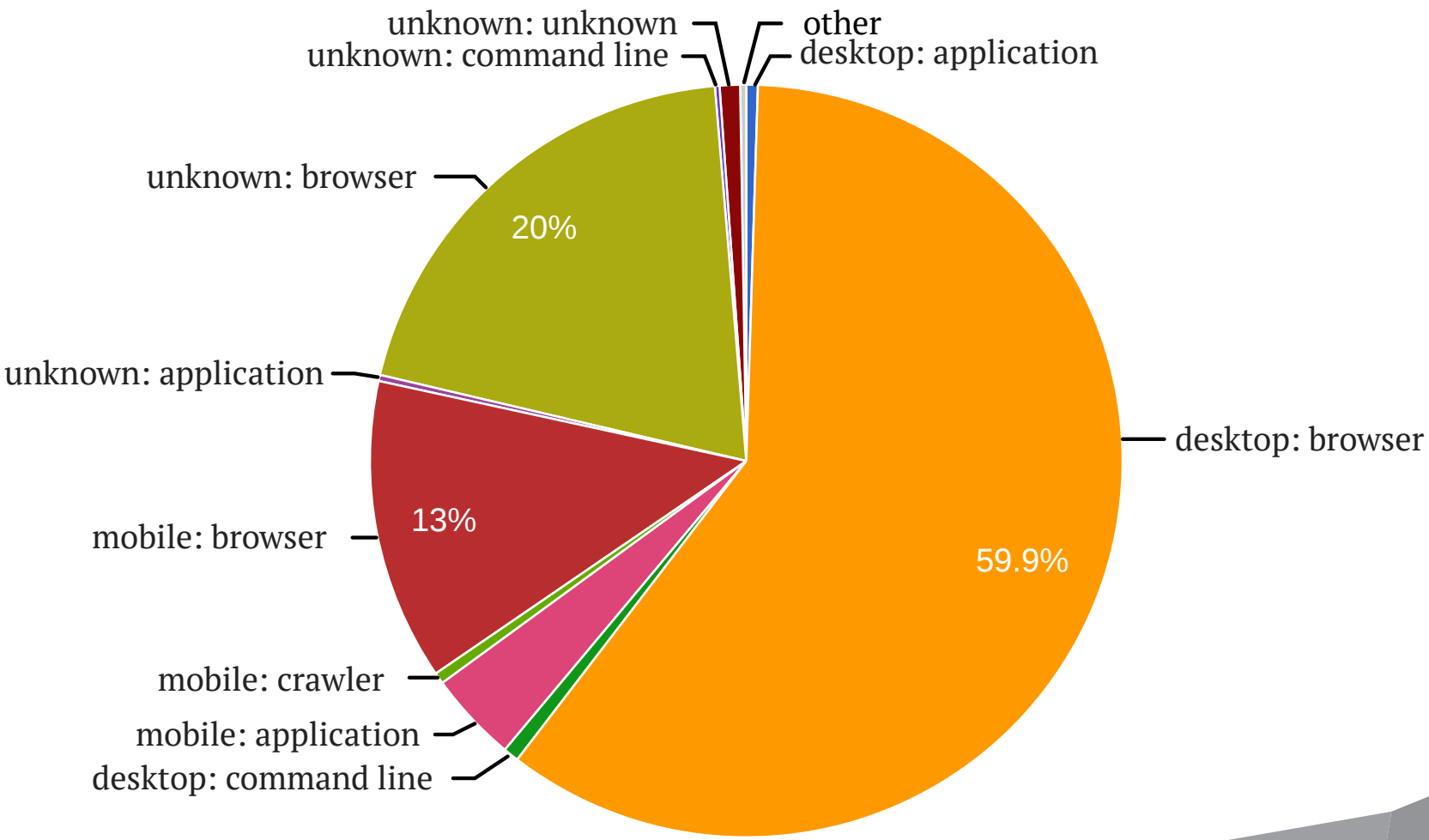
CSIRT-MU

Client Types in Dictionary



CSIRT-MU

Client Types in Network Traffic



CSIRT-MU

Conclusion

- Parameters of SSL/TLS handshake can be used for identification of clients in HTTPS communication.
- Cipher suite lists in SSL/TLS corresponds to HTTP User-Agents.
- Novel network-based of pairing cipher suite lists and User-Agents was proposed.
- The approach was tested in live network environment.
- Type of client can be estimated, while the privacy of communication is preserved.

THANK YOU FOR YOUR ATTENTION!

muni.cz/csirt

@csirtmu

Martin Husák

husakm@ics.muni.cz

UNIVERSITAS MASARYKIANA BRUNENSIS

ÚVT

CSIRT-MU