Aalborg Universitet

ISP-level identification of malicious traffic: Challenges and opportunities

Andersen, Martin Fejrskov; Vasilomanolakis, Emmanouil; Pedersen, Jens Myrup

Creative Commons License
CC BY 4.0

Publication date:
2019

Document Version
Other version

Link to publication from Aalborg University

Citation for published version (APA):
Andersen, M. F., Vasilomanolakis, E., & Pedersen, J. M. (2019). *ISP-level identification of malicious traffic: Challenges and opportunities*. Poster presented at The 24th Nordic Conference on Secure IT Systems, Aalborg , Denmark.

ISP-level identification of malicious traffic: Challenges and opportunities

Martin Fejrskov Andersen, Emmanouil Vasilomanolakis, Jens Myrup Pedersen

Introduction

Introduction

- The number and diversity of IoT devices is rapidly increasing
- Traditional host-based anti-malware products cannot be installed on the typical IoT device
- Network-based anti-malware solutions hosted by an Internet Service Provider (ISP) can be an alternative

Problem

- Which technically and legally available methods and data sources does an ISP have that can provide malware detection on an individual and network-wide level?

Challenges

- European legislation is designed to protect the privacy of the subscribers, restricting which data can be used.
- Can existing methods be adapted or improved for this use case, or are new methods needed?

Legislation

ePrivacy Directive

- Regulates how ISPs are allowed to handle data related to the subscribers data traffic and location

> - Data *already* being processed for the purpose of transmission must be **anonymized** before additional processing
> - Data *not* being processed for the purpose of transmission or as part of a value added service *cannot* be processed
> - Data can be processed for a specific value added service but only if a relevant **contract or consent** is available

- Anonymization techniques and requirements are not specified any further
- A notable conclusion is that no personal data can be used by ISPs without explicit consent or anonymization
- Malware detection and prevention is no exception to this

General Data Protection Regulation (GDPR)

- Is Lex Generalis to the ePrivacy Directive
- The GDPR only applies to matters not covered by the ePrivacy Directive

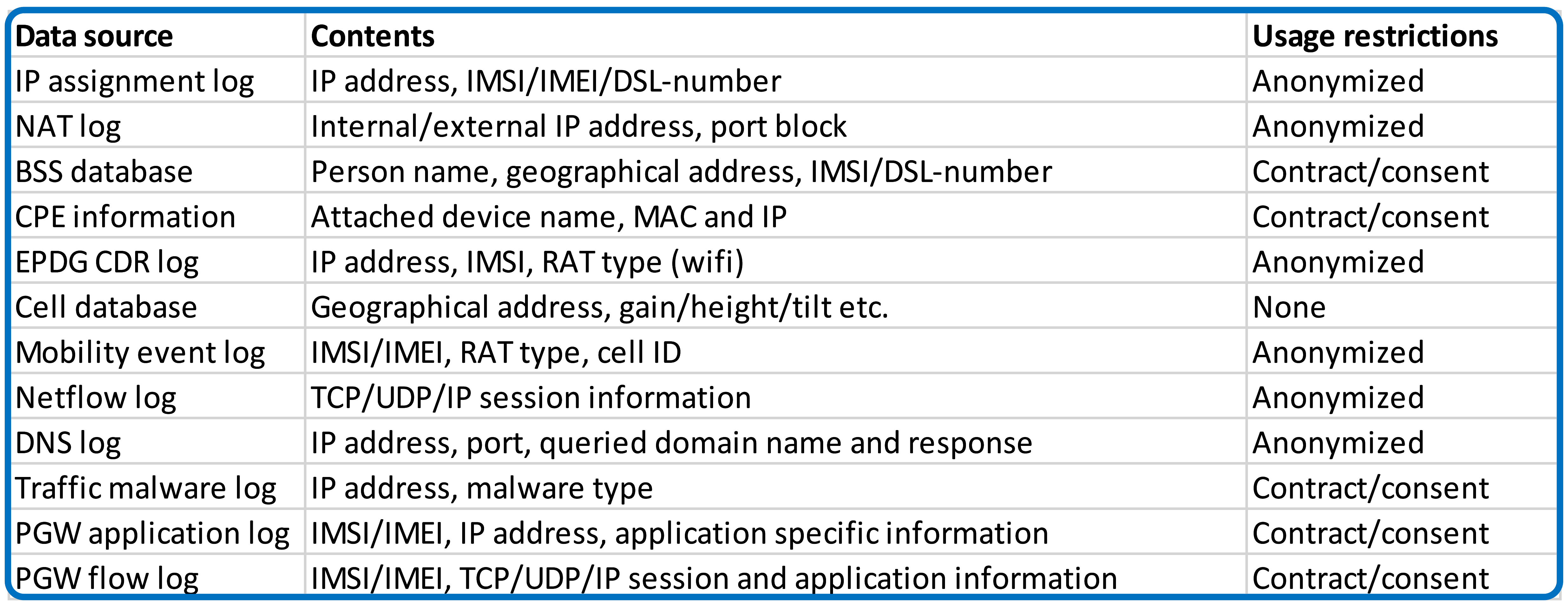
Data Sources

Data sources

- Data sources available to a typical ISP using Telenor Denmark as example
- Consent cannot be obtained from all subscribers.
- Mobility related logs are irrelevant for malware detection
- Logs like the IP assignment log are irrelevant if anonymized

Anonymization

- Hashing of subscriber IP addresses etc. to prevent direct identification
- Aggregation of subscribers to prevent event correlation based identification



Data source	Contents	Usage restrictions
IP assignment log	IP address, IMSI/IMEI/DSL-number	Anonymized
NAT log	Internal/external IP address, port block	Anonymized
BSS database	Person name, geographical address, IMSI/DSL-number	Contract/consent
CPE information	Attached device name, MAC and IP	Contract/consent
EPDG CDR log	IP address, IMSI, RAT type (wifi)	Anonymized
Cell database	Geographical address, gain/height/tilt etc.	None
Mobility event log	IMSI/IMEI, RAT type, cell ID	Anonymized
Netflow log	TCP/UDP/IP session information	Anonymized
DNS log	IP address, port, queried domain name and response	Anonymized
Traffic malware log	IP address, malware type	Contract/consent
PGW application log	IMSI/IMEI, IP address, application specific information	Contract/consent
PGW flow log	IMSI/IMEI, TCP/UDP/IP session and application information	Contract/consent

Related work

Commercial

- Approximately 60 commercial products were surveyed
- Some analyse only Netflow or DNS data
- Some analyse both Netflow and DNS data, but as independent analyses
- None perform a behavioural analysis on the combined feature set of DNS and NetFlow data

Academic

- Primary focus also seem to be on either DNS or NetFlow, but not both
- Some approaches use features from both DNS and NetFlow:
    - Rinkel Hananto, Charles Lim and Heru Purnomo Ipung: "Detecting Network Security Threats Using Domain Name System and NetFlow Traffic", 2018
    - Kuochen Wang, Chun-Ying Huang, Shang-Jyh Lin and Ying-Dar Lina: "A fuzzy pattern-based filtering algorithm for botnet detection", 2011

Opportunities

Identifying individual malware infections

- Improvement of existing detection methods using NetFlow and/or DNS logs, for example by using knowledge of the device type
- Analysis of the combined feature set for matching DNS and NetFlow logs
- Analysis of flows without matching DNS requests and vice versa
- Analysis of indirectly related DNS and Netflow data
- What is the impact of anonymization by aggregation?

Evaluating network-wide infection level

- An infected mobile phone can use a Wifi connection
- Is this counted as one or two infected subscribers?
- The EPDG CDR log may be used to identify WiFi attached mobile subscribers