A Series of Fortunate Events:

Attacker behavior analysis from observable sequences

Azqa Nadeem

PhD candidate

Cyber Analytics Lab

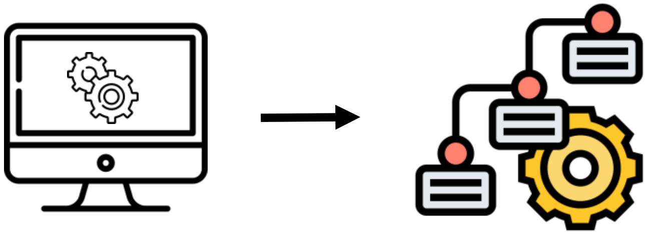
TU Delft

22 Dec 2020

Cyber Analytics Lab

Dynamic observables

- Program execution → observable data
- Network traffic, software logs, intrusion alerts



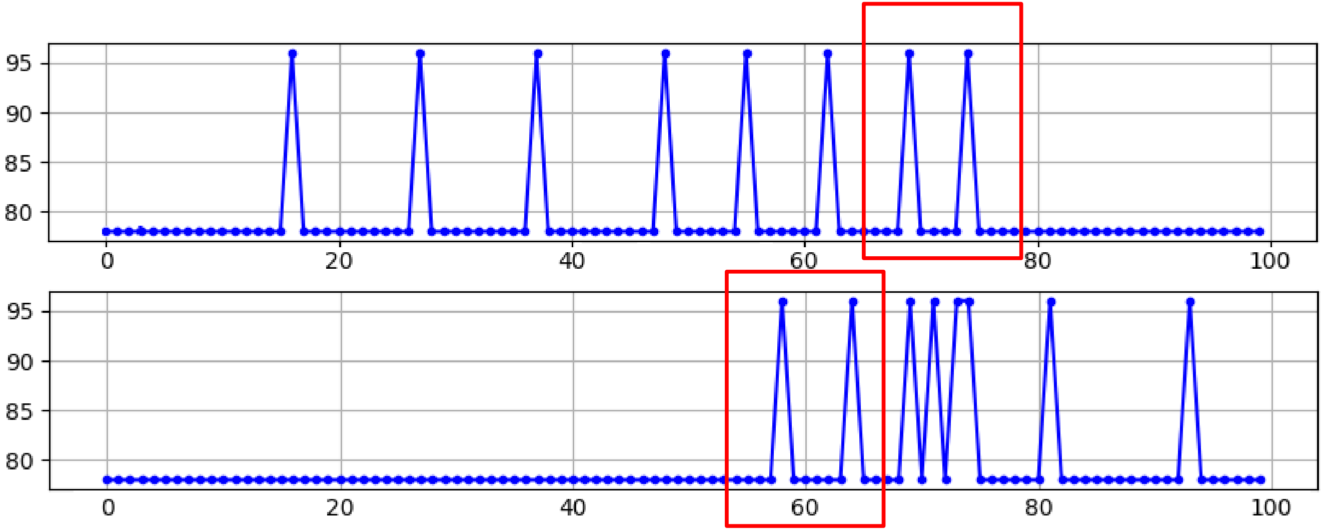
Dynamic observables

- Program execution → observable data
- Network traffic, software logs, intrusion alerts

- Proxy to attacker intent

Series of fortunate events → Sequences

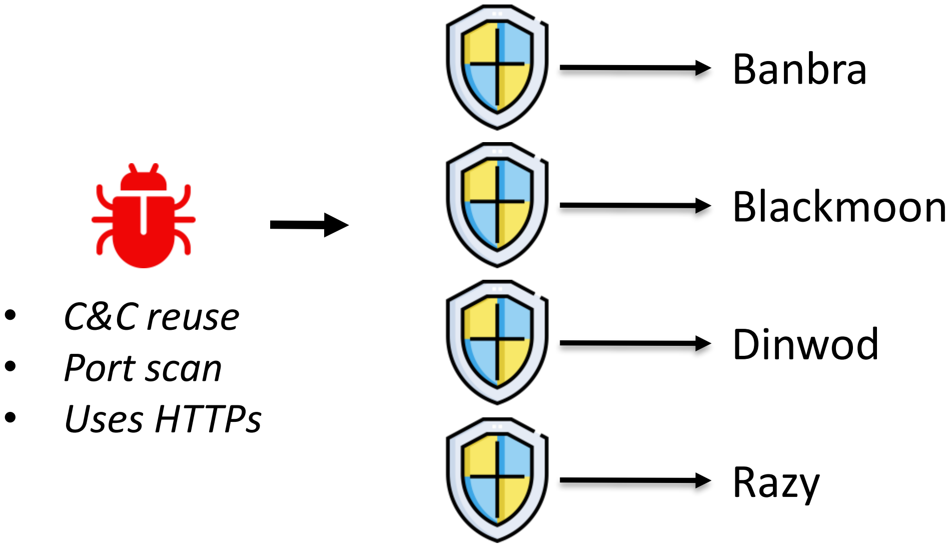
- Patterns in temporal data
- Limited data → insightful patterns
- Privacy non-invasive



Use cases

Case 1: Malware Behavior Profiles

- Malware labels are inconsistent and black-box



- *C&C reuse*
- *Port scan*
- *Uses HTTPs*

Banbra

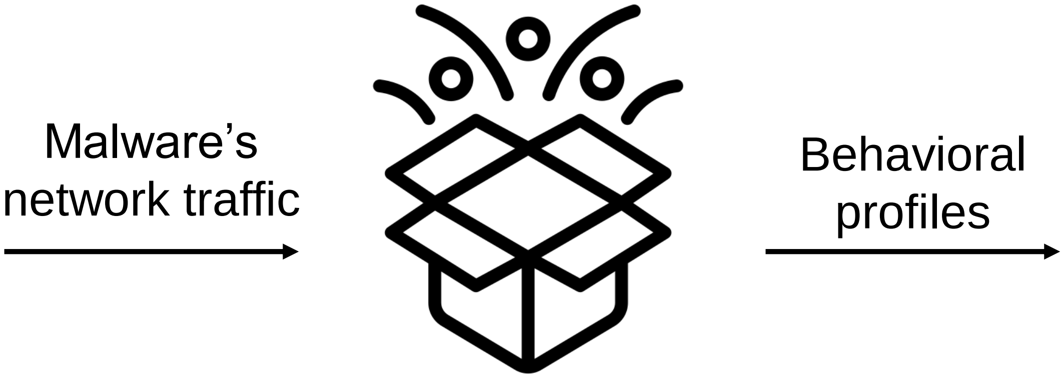
Blackmoon

Dinwod

Razy

Case 1: Malware Behavior Profiles

- Malware labels are inconsistent and black-box
- How to discover behaviors?



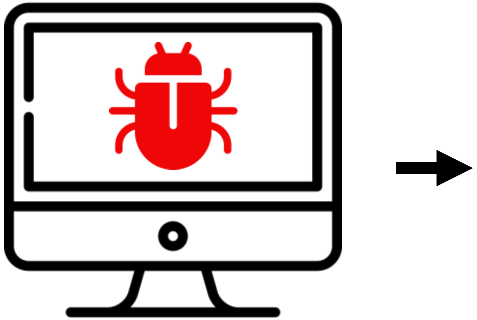
Malware's network traffic

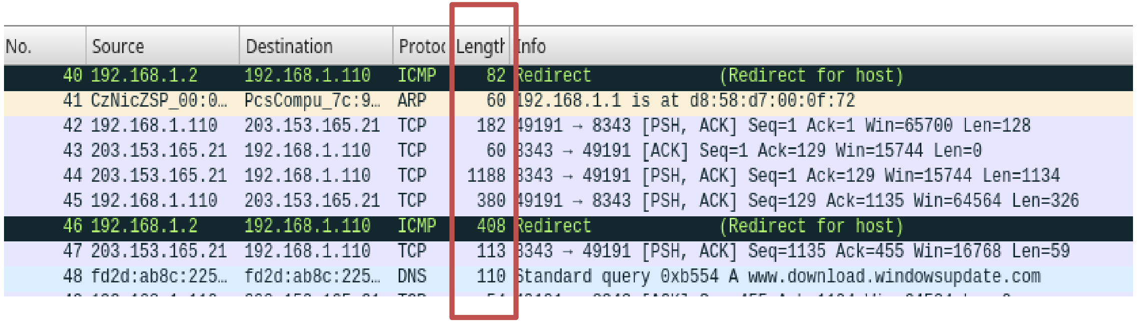
Behavioral profiles

MalPaCA

Case 1: Malware Behavior Profiles

- Malware infected machine generates network traffic



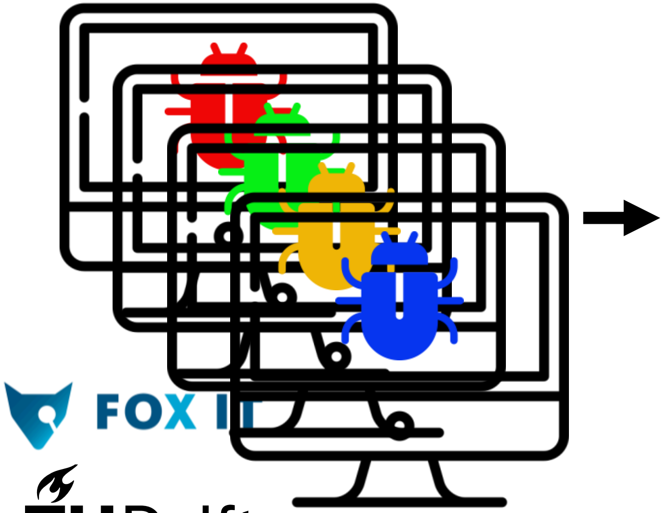


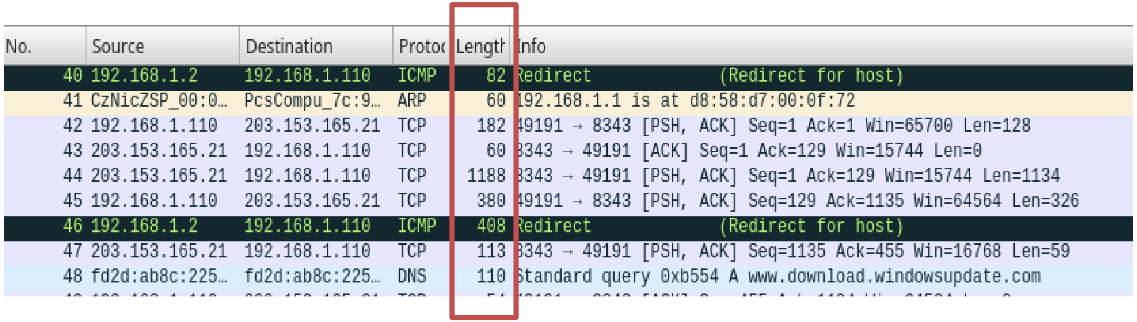
No.	Source	Destination	Protoc	Length	Info
40	192.168.1.2	192.168.1.110	ICMP	82	Redirect (Redirect for host)
41	CzNicZSP_00:0…	PcsCompu_7c:9…	ARP	60	192.168.1.1 is at d8:58:d7:00:0f:72
42	192.168.1.110	203.153.165.21	TCP	182	49191 → 8343 [PSH, ACK] Seq=1 Ack=1 Win=65700 Len=128
43	203.153.165.21	192.168.1.110	TCP	60	8343 → 49191 [ACK] Seq=1 Ack=129 Win=15744 Len=0
44	203.153.165.21	192.168.1.110	TCP	1188	8343 → 49191 [PSH, ACK] Seq=1 Ack=129 Win=15744 Len=1134
45	192.168.1.110	203.153.165.21	TCP	380	49191 → 8343 [PSH, ACK] Seq=129 Ack=1135 Win=64564 Len=326
46	192.168.1.2	192.168.1.110	ICMP	408	Redirect (Redirect for host)
47	203.153.165.21	192.168.1.110	TCP	113	8343 → 49191 [PSH, ACK] Seq=1135 Ack=455 Win=16768 Len=59
48	fd2d:ab8c:225…	fd2d:ab8c:225…	DNS	110	Standard query 0xb554 A www.download.windowsupdate.com

Beyond Labeling: Using Clustering to Build Network Behavioral Profiles of Malware Families. *Azqa Nadeem, Christian Hammerschmidt, Carlos H. Ganan, Sicco Verwer. In Malware Analysis using Artificial Intelligence and Deep Learning, Springer, 2020. (Forthcoming)*

Case 1: Malware Behavior Profiles

- Malware infected machine generates network traffic



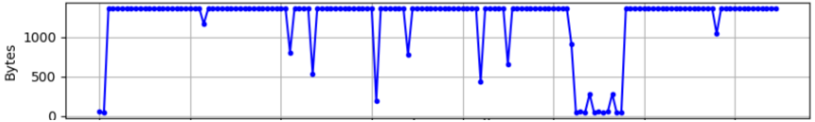


No.	Source	Destination	Protoc	Lengtl	Info
40	192.168.1.2	192.168.1.110	ICMP	82	Redirect (Redirect for host)
41	CzNicZSP_00:0…	PcsCompu_7c:9…	ARP	60	192.168.1.1 is at d8:58:d7:00:0f:72
42	192.168.1.110	203.153.165.21	TCP	182	49191 → 8343 [PSH, ACK] Seq=1 Ack=1 Win=65700 Len=128
43	203.153.165.21	192.168.1.110	TCP	60	8343 → 49191 [ACK] Seq=1 Ack=129 Win=15744 Len=0
44	203.153.165.21	192.168.1.110	TCP	1188	8343 → 49191 [PSH, ACK] Seq=1 Ack=129 Win=15744 Len=1134
45	192.168.1.110	203.153.165.21	TCP	380	49191 → 8343 [PSH, ACK] Seq=129 Ack=1135 Win=64564 Len=326
46	192.168.1.2	192.168.1.110	ICMP	408	Redirect (Redirect for host)
47	203.153.165.21	192.168.1.110	TCP	113	8343 → 49191 [PSH, ACK] Seq=1135 Ack=455 Win=16768 Len=59
48	fd2d:ab8c:225…	fd2d:ab8c:225…	DNS	110	Standard query 0xb554 A www.download.windowsupdate.com

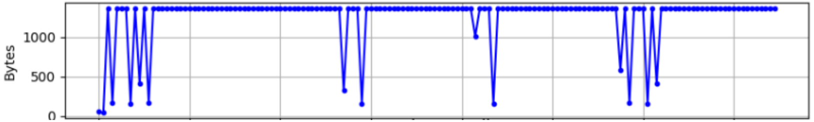
Beyond Labeling: Using Clustering to Build Network Behavioral Profiles of Malware Families. Azqa Nadeem, Christian Hammerschmidt, Carlos H. Ganan, Sicco Verwer. In Malware Analysis using Artificial Intelligence and Deep Learning, Springer, 2020. (Forthcoming)

Case 1: Malware Behavior Profiles

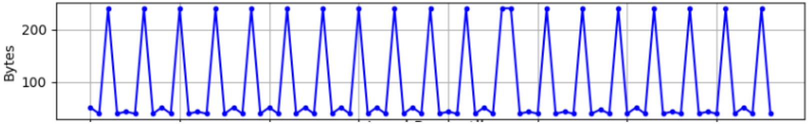
Zeus-738f →



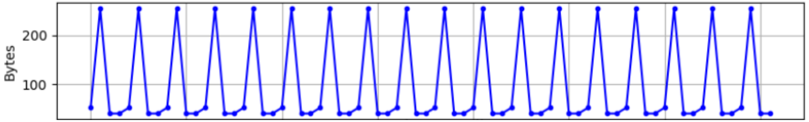
Gozi-4bd7 →



Zeus-78de →

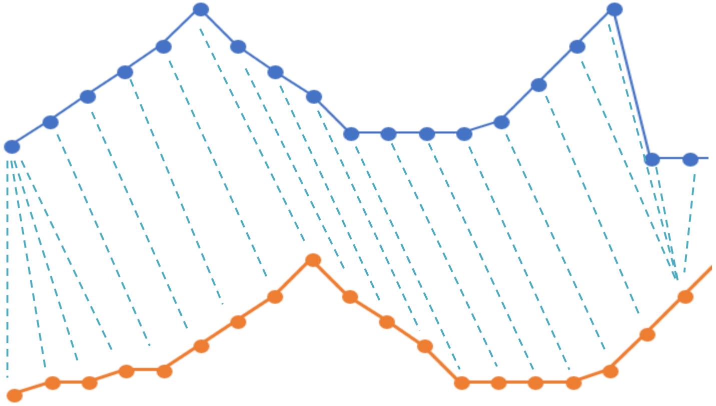


Zeus-6631 →



Beyond Labeling: Using Clustering to Build Network Behavioral Profiles of Malware Families. Azqa Nadeem, Christian Hammerschmidt, Carlos H. Ganan, Sicco Verwer. In Malware Analysis using Artificial Intelligence and Deep Learning, Springer, 2020. (Forthcoming)

Case 1: Malware Behavior Profiles

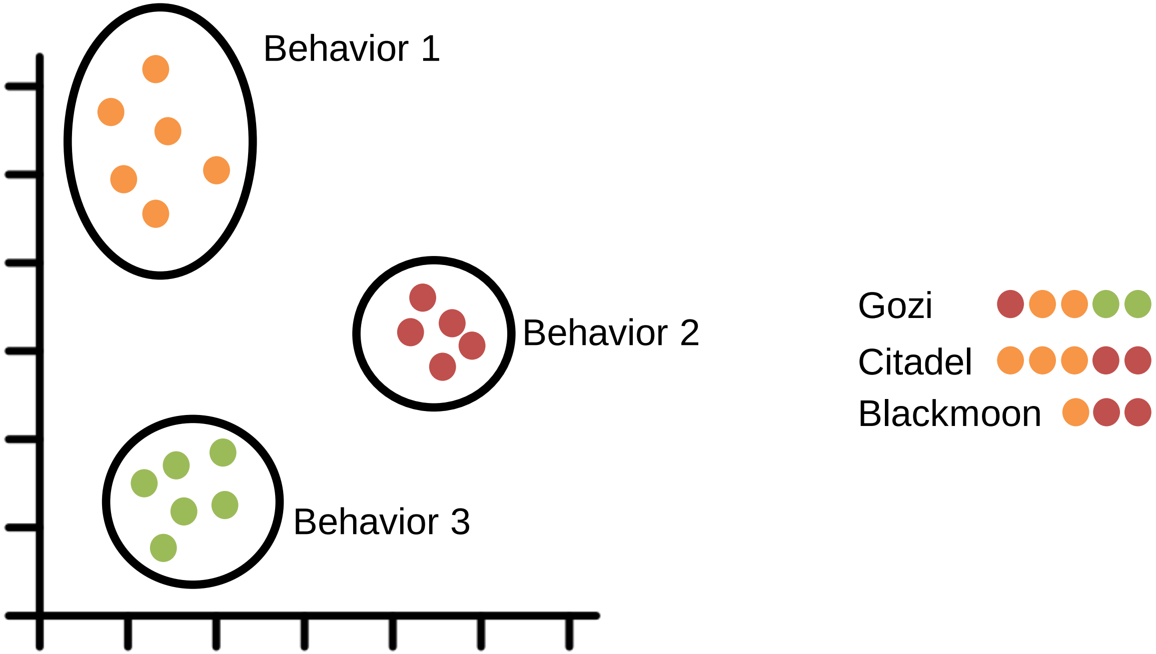


Dynamic Time Warping

$$D(i,j) = \left|A_i - B_j\right| + \min(D(i-1,j), D(i,j-1), D(i-1,j-1))$$

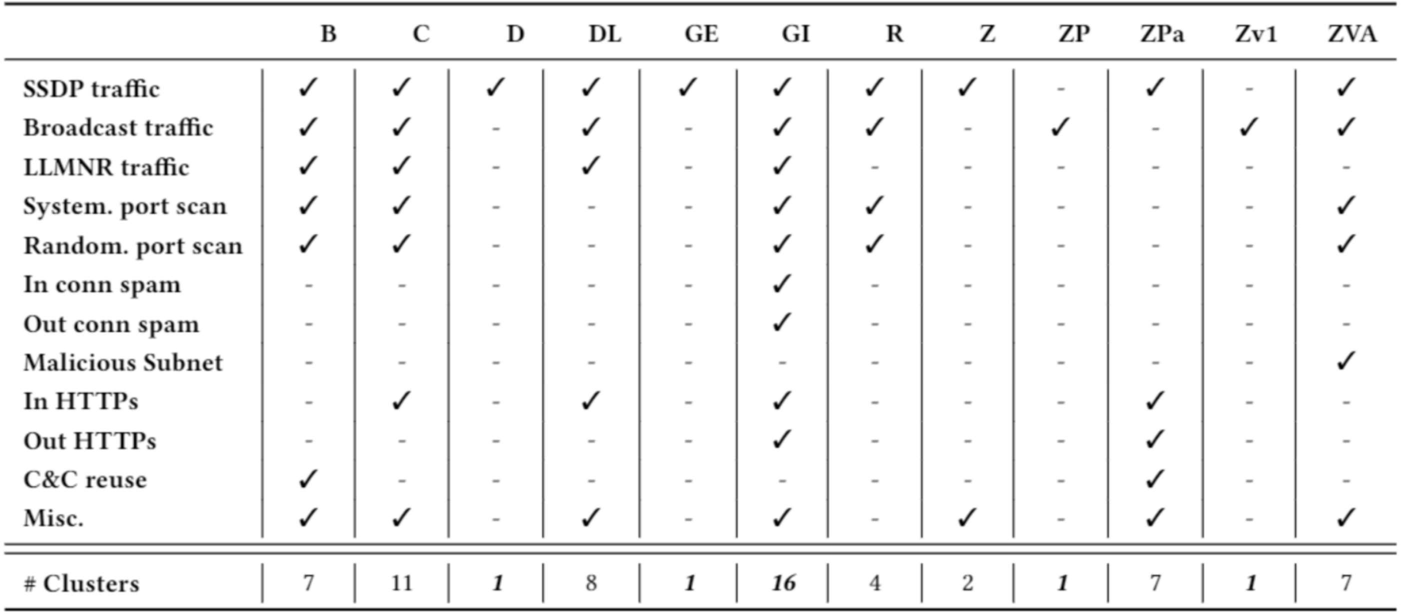
Beyond Labeling: Using Clustering to Build Network Behavioral Profiles of Malware Families. Azqa Nadeem, Christian Hammerschmidt, Carlos H. Ganan, Sicco Verwer. In Malware Analysis using Artificial Intelligence and Deep Learning, Springer, 2020. (Forthcoming)

Case 1: Malware Behavior Profiles



Beyond Labeling: Using Clustering to Build Network Behavioral Profiles of Malware Families. Azqa Nadeem, Christian Hammerschmidt, Carlos H. Ganan, Sicco Verwer. In Malware Analysis using Artificial Intelligence and Deep Learning, Springer, 2020. (Forthcoming)

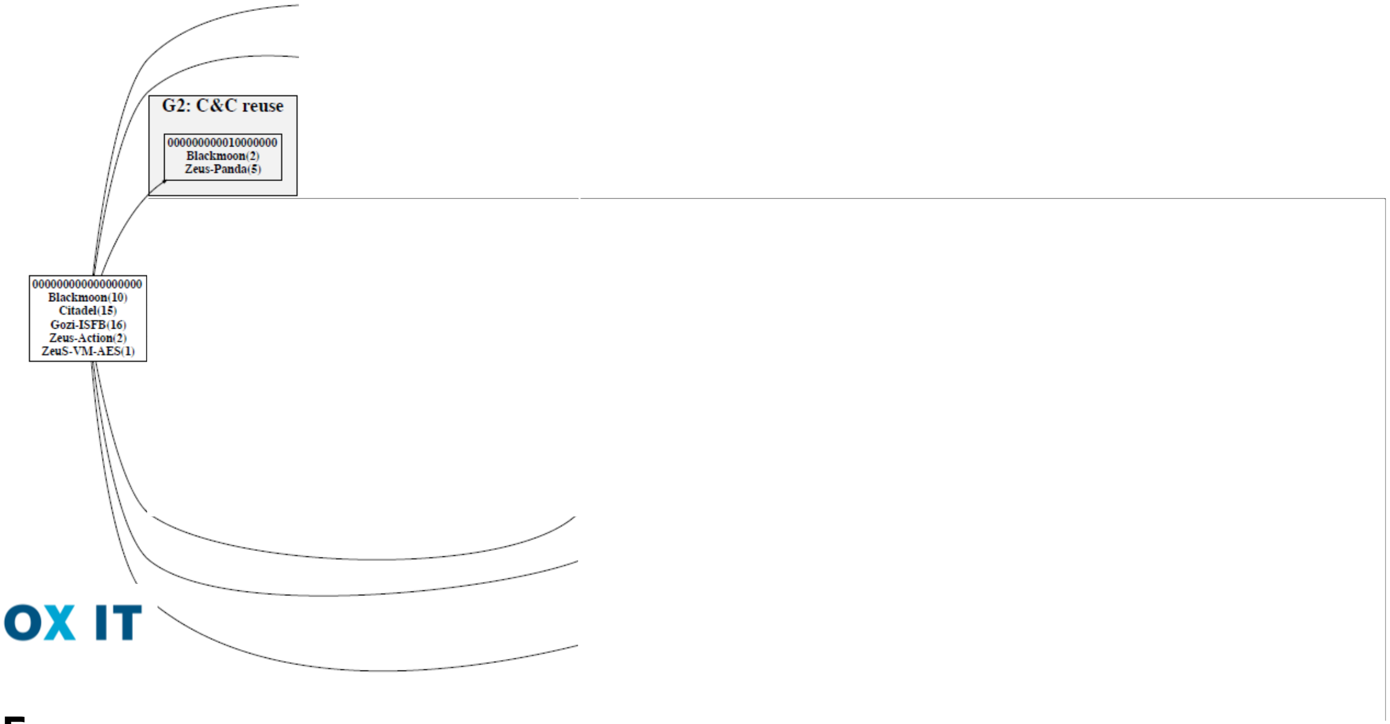
Case 1: Malware Behavior Profiles



	B	C	D	DL	GE	GI	R	Z	ZP	ZPa	Zv1	ZVA
SSDP traffic	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓
Broadcast traffic	✓	✓	-	✓	-	✓	✓	-	✓	-	✓	✓
LLMNR traffic	✓	✓	-	✓	-	✓	-	-	-	-	-	-
System. port scan	✓	✓	-	-	-	✓	✓	-	-	-	-	✓
Random. port scan	✓	✓	-	-	-	✓	✓	-	-	-	-	✓
In conn spam	-	-	-	-	-	✓	-	-	-	-	-	-
Out conn spam	-	-	-	-	-	✓	-	-	-	-	-	-
Malicious Subnet	-	-	-	-	-	-	-	-	-	-	-	✓
In HTTPs	-	✓	-	✓	-	✓	-	-	-	✓	-	-
Out HTTPs	-	-	-	-	-	✓	-	-	-	✓	-	-
C&C reuse	✓	-	-	-	-	-	-	-	-	✓	-	-
Misc.	✓	✓	-	✓	-	✓	-	✓	-	✓	-	✓
# Clusters	7	11	***1***	8	***1***	***16***	4	2	***1***	7	***1***	7

Beyond Labeling: Using Clustering to Build Network Behavioral Profiles of Malware Families. Azqa Nadeem, Christian Hammerschmidt, Carlos H. Ganan, Sicco Verwer. In Malware Analysis using Artificial Intelligence and Deep Learning, Springer, 2020. (Forthcoming)

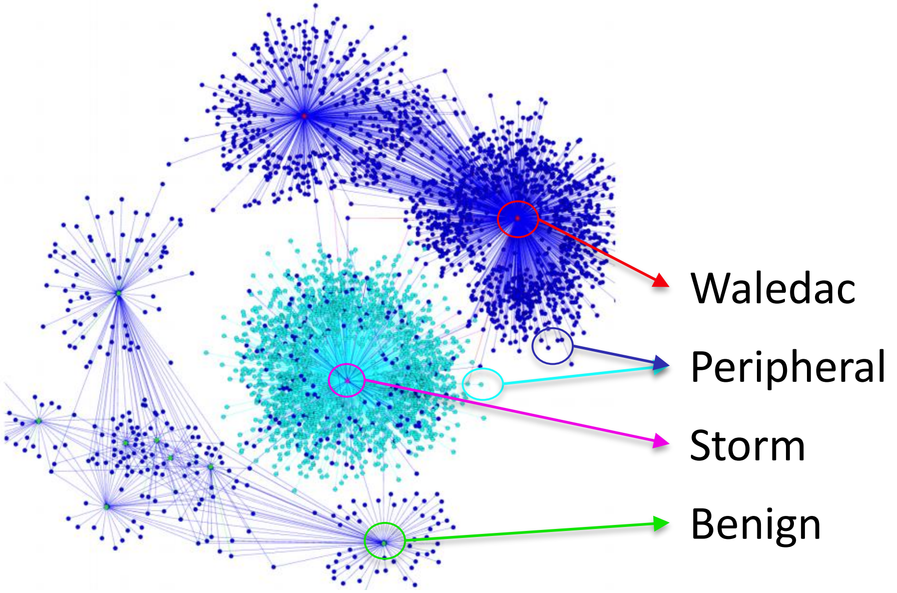
Case 1: Malware Behavior Profiles



Beyond Labeling: Using Clustering to Build Network Behavioral Profiles of Malware Families. Azqa Nadeem, Christian Hammerschmidt, Carlos H. Ganan, Sicco Verwer. In Malware Analysis using Artificial Intelligence and Deep Learning, Springer, 2020. (Forthcoming)

Case 1 (ext.): Detecting network communities

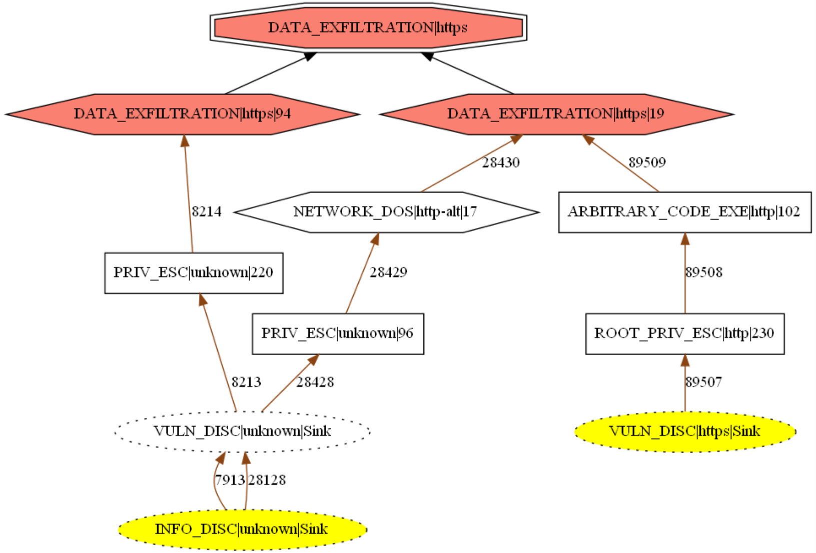
- What type of hosts are present in a network?
  - Connection + Host clustering



Hybrid Connection and Host Clustering for Community Detection in Spatial-temporal Network Data. Mark Patrick Roeling, Azqa Nadeem, Sicco Verwer. In Machine Learning for Cybersecurity (MLCS), 2020

Case 2: Attacker strategy analysis

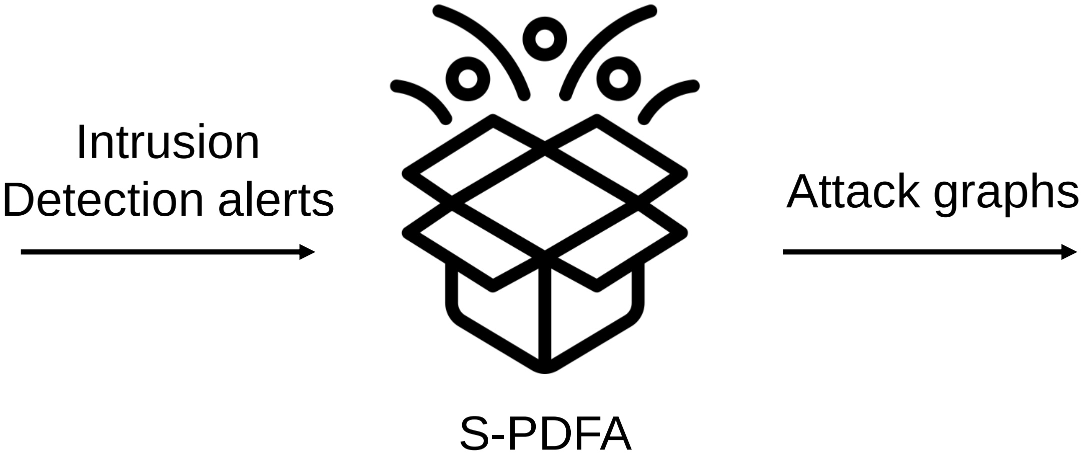
- Alert correlation groups related alerts
  - But how did the attack happen?
- How to get attacker strategies automatically?



Ideally…

- *From intrusion alerts*
- *Without expert knowledge*

Case 2: Attacker strategy analysis

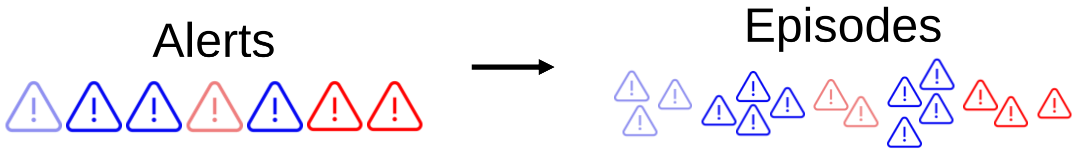


Intrusion Detection alerts

Attack graphs

S-PDFA

Case 2: Attacker strategy analysis

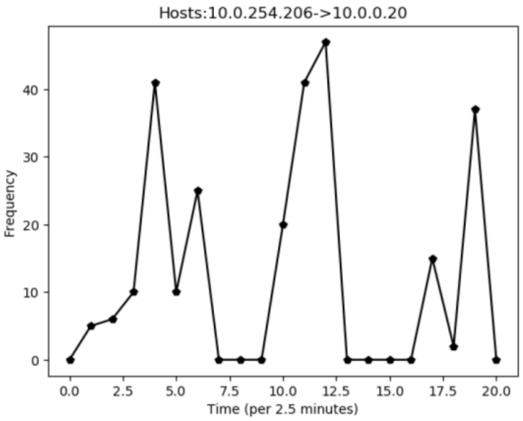


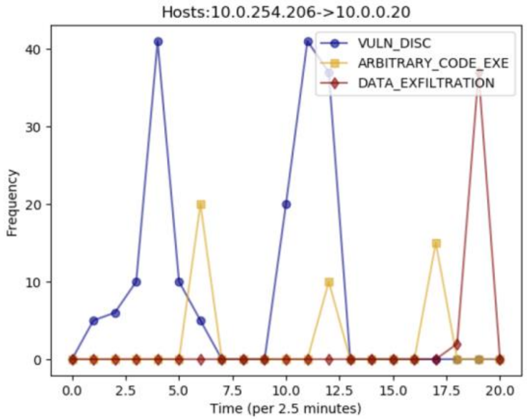
Alerts

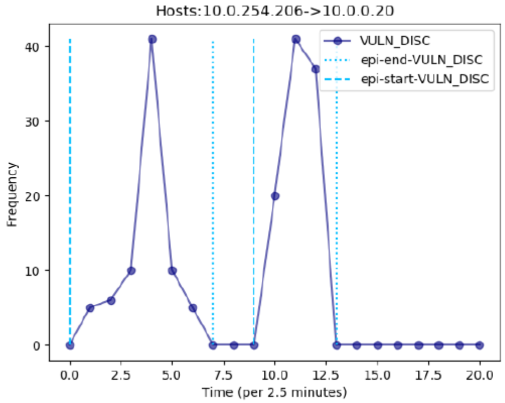
Episodes

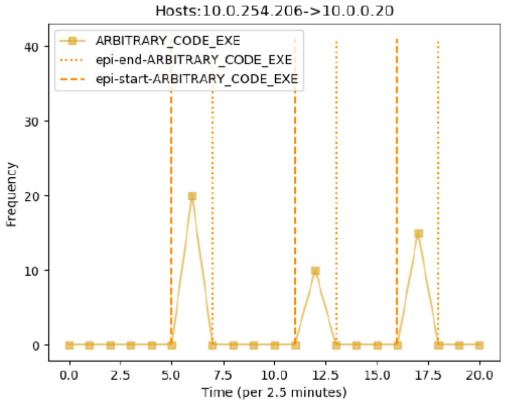
Alert-driven attack graphs using S-PDFA. Azqa Nadeem, Sicco Verwer, Stephen Moskal, Shanchieh Jay Yang. In IEEE Transactions on Dependable and Secure Computing, 2021. (Submitted)

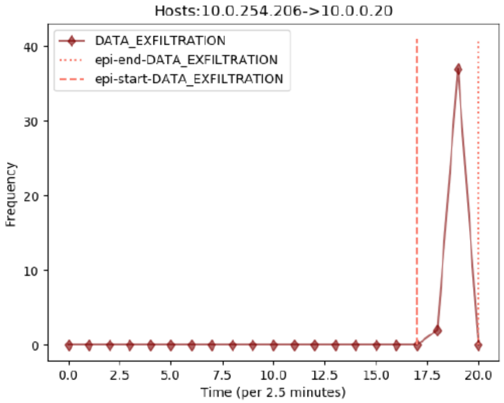
Case 2: Attacker strategy analysis





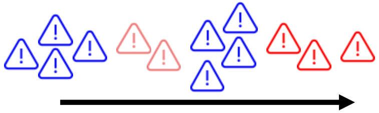


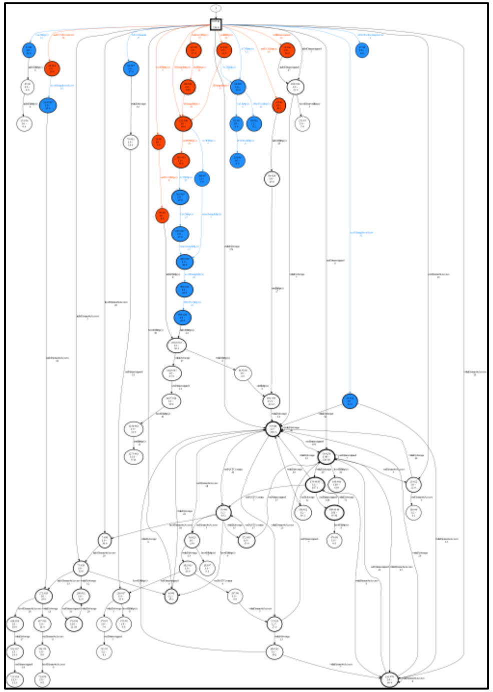




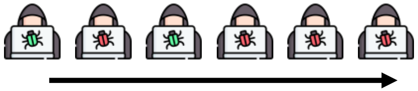
Case 2: Attacker strategy analysis

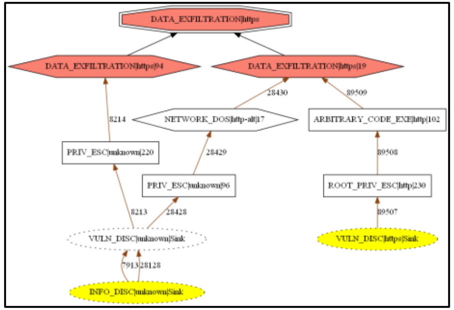
Episodes





Actions

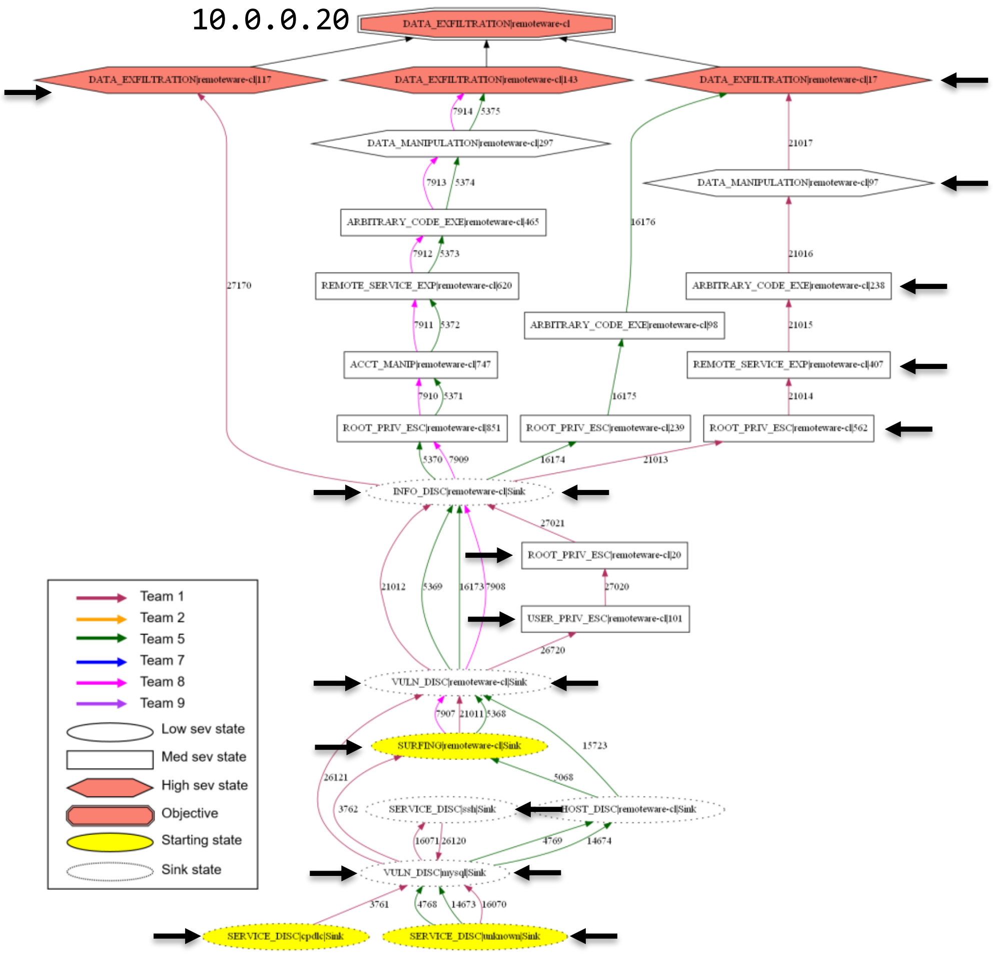




Alert-driven attack graphs using S-PDFA. Azqa Nadeem, Sicco Verwer, Stephen Moskal, Shanchieh Jay Yang. In IEEE Transactions on Dependable and Secure Computing, 2021. (Submitted)

Insights

- Attackers follow shorter paths after discovering longer ones


10.0.0.20
DATA_EXFILTRATION|remoteware-cl
DATA_EXFILTRATION|remoteware-cl|117
DATA_EXFILTRATION|remoteware-cl|143
DATA_EXFILTRATION|remoteware-cl|17
7914
5375
DATA_MANIPULATION|remoteware-cl|297
21017
DATA_MANIPULATION|remoteware-cl|97
7913
5374
ARBITRARY_CODE_EXE|remoteware-cl|465
16176
21016
7912
5373
27170
REMOTE_SERVICE_EXP|remoteware-cl|620
ARBITRARY_CODE_EXE|remoteware-cl|238
ARBITRARY_CODE_EXE|remoteware-cl|98
7911
5372
21015
ACCT_MANIP|remoteware-cl|747
REMOTE_SERVICE_EXP|remoteware-cl|407
7910
5371
16175
21014
ROOT_PRIV_ESC|remoteware-cl|851
ROOT_PRIV_ESC|remoteware-cl|239
ROOT_PRIV_ESC|remoteware-cl|562
5370
7909
16174
21013
INFO_DISC|remoteware-cl|Sink
27021
ROOT_PRIV_ESC|remoteware-cl|20
21012
5369
16173
7908
27020
USER_PRIV_ESC|remoteware-cl|101
26720
VULN_DISC|remoteware-cl|Sink
7907
21011
5368
SURFING|remoteware-cl|Sink
15723
26121
5068
3762
SERVICE_DISC|ssh|Sink
HOST_DISC|remoteware-cl|Sink
16071
26120
4769
14674
VULN_DISC|mysql|Sink
3761
4768
14673
16070
SERVICE_DISC|cpdlc|Sink
SERVICE_DISC|unknown|Sink
Team 1
Team 2
Team 5
Team 7
Team 8
Team 9
Low sev state
Med sev state
High sev state
Objective
Starting state
Sink state

CyberVSR RIT

TU Delft

Future research directions

- Additional use cases for sequential ML
- Defining explainability in security
- Running qualitative studies with analysts

Wrap-up

- Sequence of dynamic observables → attacker intent
- Input: observables | Output: Intelligence

- Unsupervised setting with limited expert knowledge

- 2 use-cases
  - Network traffic → Malware behavior profiles
  - Intrusion alerts → Attacker strategy attack graphs

Thank you!
Questions?

Sequence of dynamic observables → attacker intent

Input: observables | Output: Intelligence

Unsupervised setting with limited expert knowledge

2 use-cases

- Network traffic → Malware behavior profiles
- Intrusion alerts → Attacker strategy attack graphs

azqa.nadeem@tudelft.nl https://cyber-analytics.nl/

TU Delft

Icons courtesy Freepik, Those Icons, Dinosoft Labs, Pixel Perfect, Flat Icons, Kiranshastry from www.flaticon.com