SAGE: Intrusion Alert-driven Attack Graph Extractor

Azqa Nadeem[*], Sicco Verwer[*], Stephen Moskal[^], Shanchieh Jay Yang[^]

[*]*Delft University of Technology, The Netherlands*
[^]*Rochester Institute of Technology, USA*

1st KDD Workshop on AI-enabled Cybersecurity Analytics (AI4Cyber)

Cyber Analytics Lab

TUDelft

RIT

15 August 2021

Background

- Security analysts receive > 1M intrusion alerts/day*


```
{   '_sourcetype': 'suricata:alert',
{   '_sourcetype': 'suricata:alert',
{   '_sourcetype': 'suricata:alert',
{   '_sourcetype': 'suricata:alert',
{   '_sourcetype': 'suricata:alert',
{   '_sourcetype': 'suricata:alert',
    'alert': {   'category': 'Attempted Information Leak',
                 'severity': 2,
                 'signature': 'ET POLICY Python-urllib\\/ '
                              'Suspicious User Agent'},
    'dest_ip': '169.254.169.254',
    'dest_port': 80,
    'src_ip': '10.0.0.20',
    'src_port': 56952,
    'timestamp': '2018-11-03T13:51:58.205548+0000'}}
```

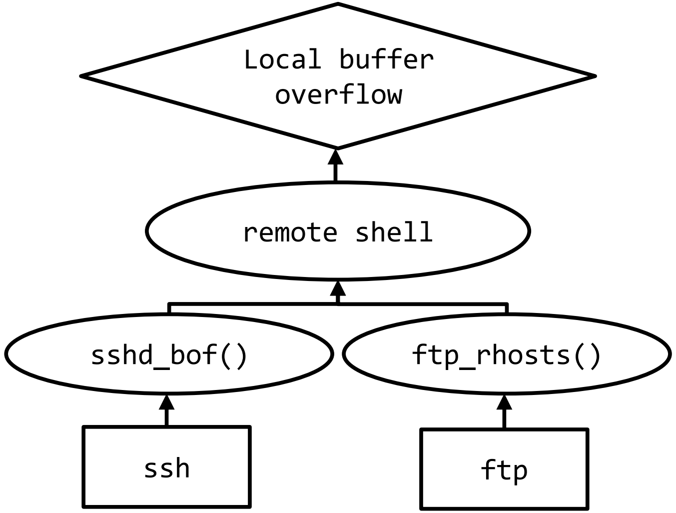

* *https://www.imperva.com/blog/27-percent-of-it-professionals-receive-more-than-1-million-security-alerts-daily/*

Background

- Security analysts receive > 1M intrusion alerts/day*

- Attacker strategy identification
  - How?
  - Multiple attackers?
  - Strategies similar?

- Often represented as Attack graphs

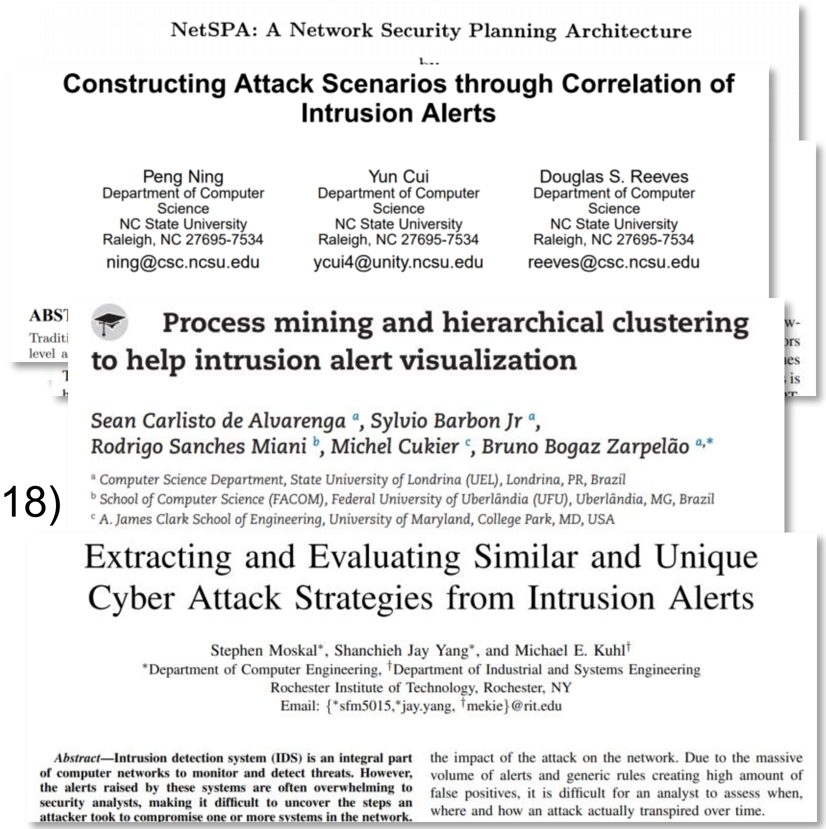


* https://www.imperva.com/blog/27-percent-of-it-professionals-receive-more-than-1-million-security-alerts-daily/
https://www.researchgate.net/publication/220202986_Measuring_Security_Risk_of_Networks_Using_Attack_Graphs

Existing approaches

- Expert-crafted attack graphs
  - `NetSPA` by ML Artz (MIT '02)
  - `MulVAL` by X Ou *et al.* (USENIX '05)

- Alert-driven attack scenario modelling
  - Ning *et al.* (Ning '02)
  - De Alvarenga *et al.* (Computers & Security '18)
  - Moskal *et al.* (ISI '18)


NetSPA: A Network Security Planning Architecture

Constructing Attack Scenarios through Correlation of Intrusion Alerts

Peng Ning, Department of Computer Science, NC State University, Raleigh, NC 27695-7534, ning@csc.ncsu.edu
Yun Cui, Department of Computer Science, NC State University, Raleigh, NC 27695-7534, ycui4@unity.ncsu.edu
Douglas S. Reeves, Department of Computer Science, NC State University, Raleigh, NC 27695-7534, reeves@csc.ncsu.edu

Process mining and hierarchical clustering to help intrusion alert visualization

Sean Carlisto de Alvarenga [a], Sylvio Barbon Jr [a], Rodrigo Sanches Miani [b], Michel Cukier [c], Bruno Bogaz Zarpelão [a,]*

[a] *Computer Science Department, State University of Londrina (UEL), Londrina, PR, Brazil*
[b] *School of Computer Science (FACOM), Federal University of Uberlândia (UFU), Uberlândia, MG, Brazil*
[c] *A. James Clark School of Engineering, University of Maryland, College Park, MD, USA*

Extracting and Evaluating Similar and Unique Cyber Attack Strategies from Intrusion Alerts

Stephen Moskal*, Shanchieh Jay Yang*, and Michael E. Kuhl†
*Department of Computer Engineering, †Department of Industrial and Systems Engineering
Rochester Institute of Technology, Rochester, NY
Email: {*sfm5015,*jay.yang, †mekie}@rit.edu

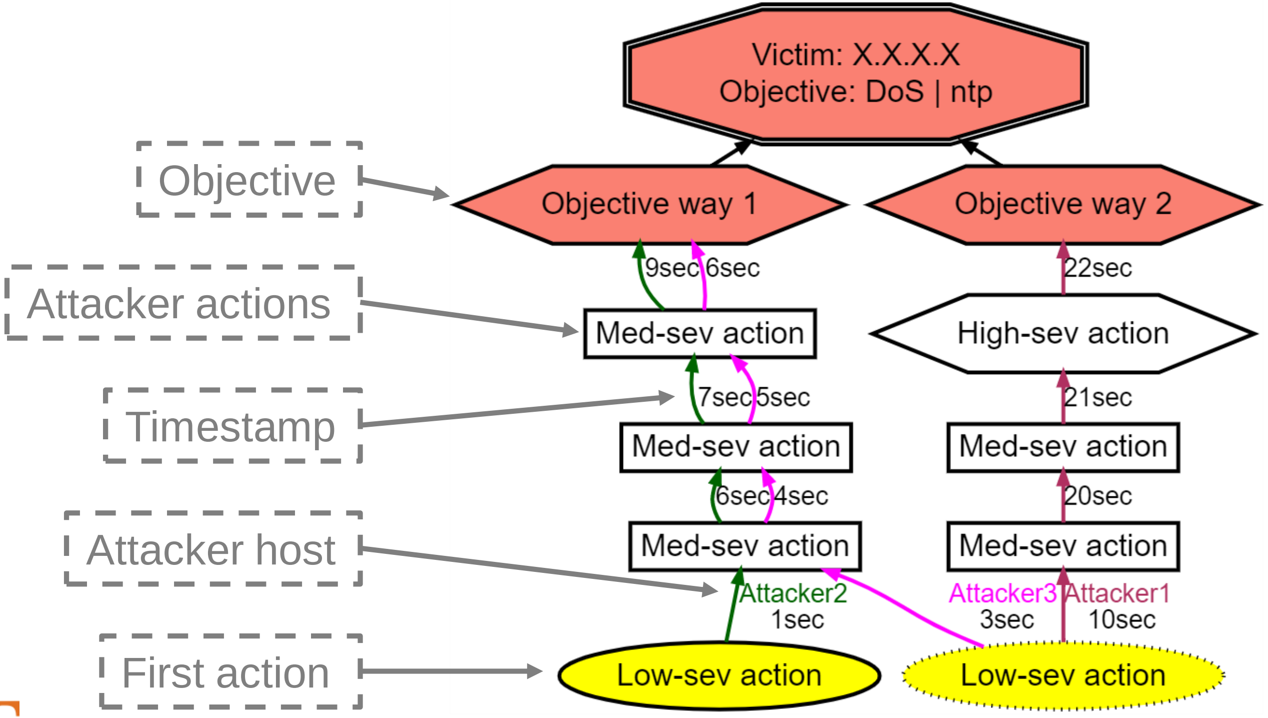
***Abstract*—Intrusion detection system (IDS) is an integral part of computer networks to monitor and detect threats. However, the alerts raised by these systems are often overwhelming to security analysts, making it difficult to uncover the steps an attacker took to compromise one or more systems in the network.** the impact of the attack on the network. Due to the massive volume of alerts and generic rules creating high amount of false positives, it is difficult for an analyst to assess when, where and how an attack actually transpired over time.

Research aim

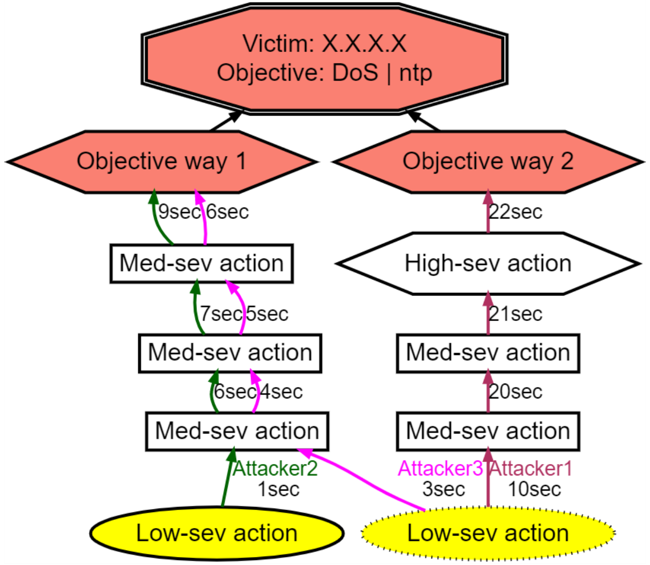
RQ: How to construct attack graphs directly from intrusion alerts?

- For extracting intelligence about attacker strategies
- Preferably without network dependence

Anatomy of an Alert-driven Attack Graph

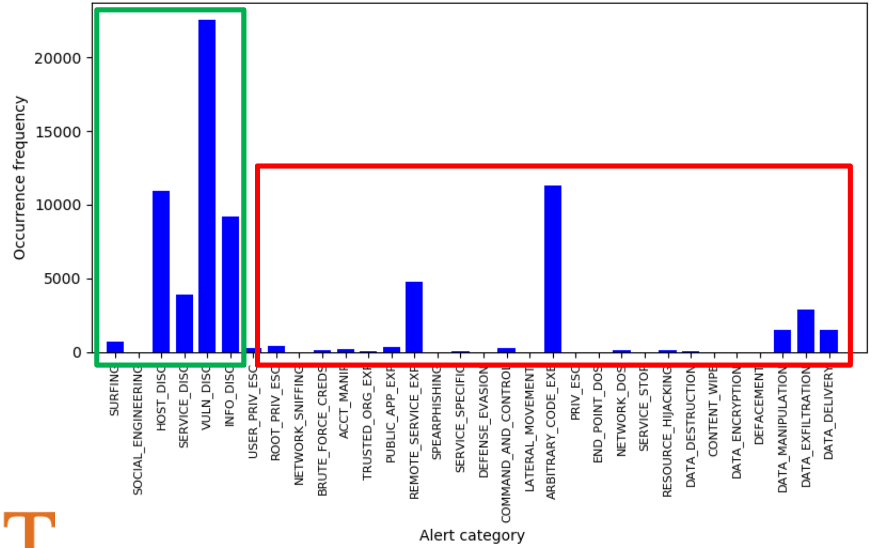


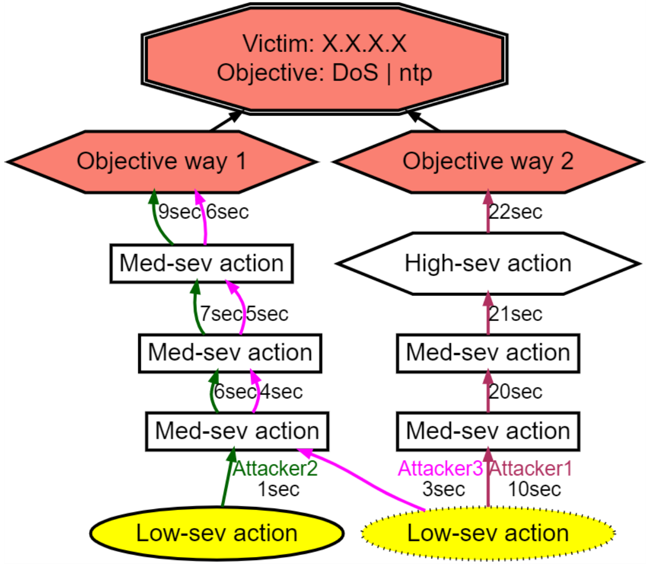
Key design challenges



Key design challenges

1. Alert-type imbalance

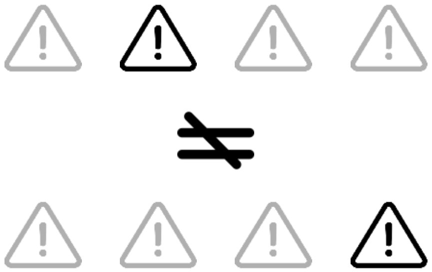


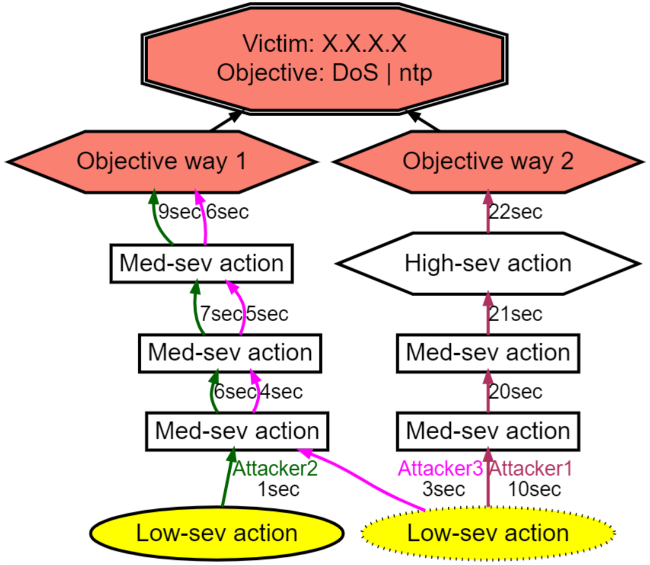


RIT
TU Delft

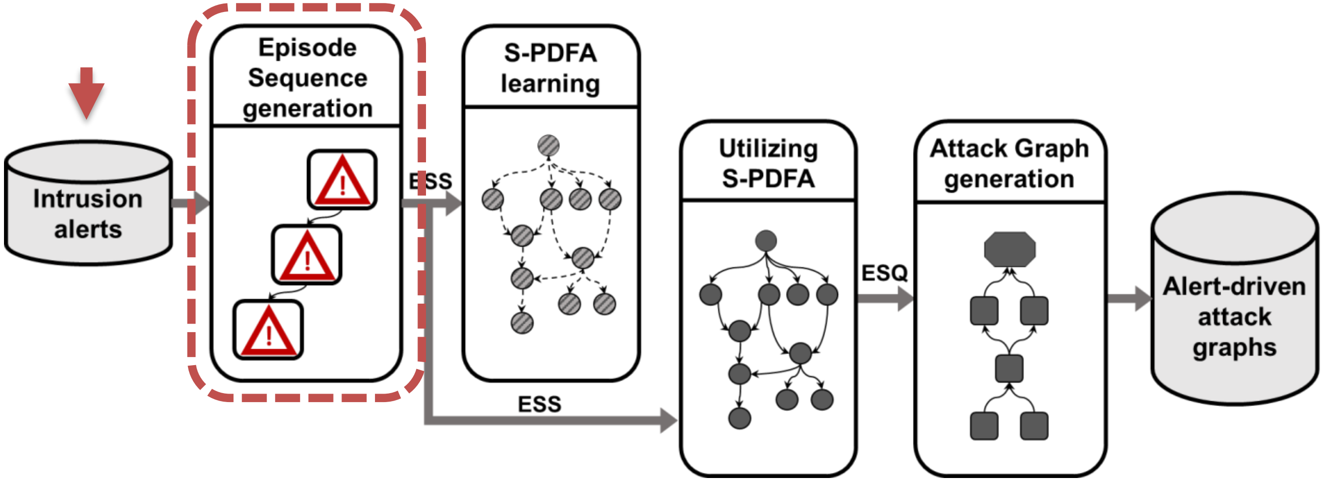
Key design challenges

1. Alert-type imbalance
2. Context modelling





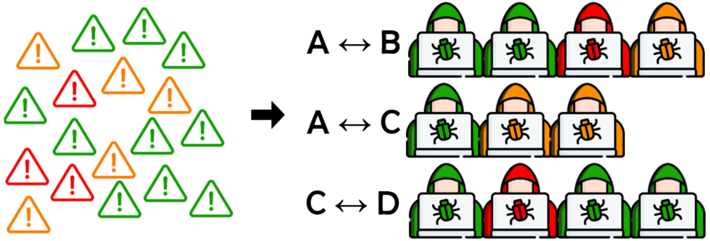
SAGE: IntruSion alert-driven Attack Graph Extractor

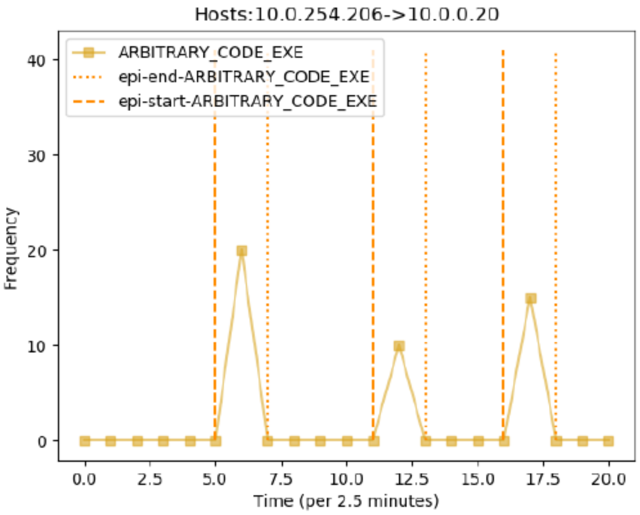


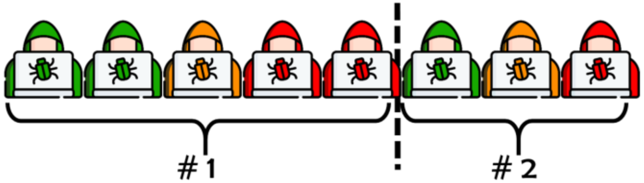
RIT
TU Delft

Alert → Episode sequences

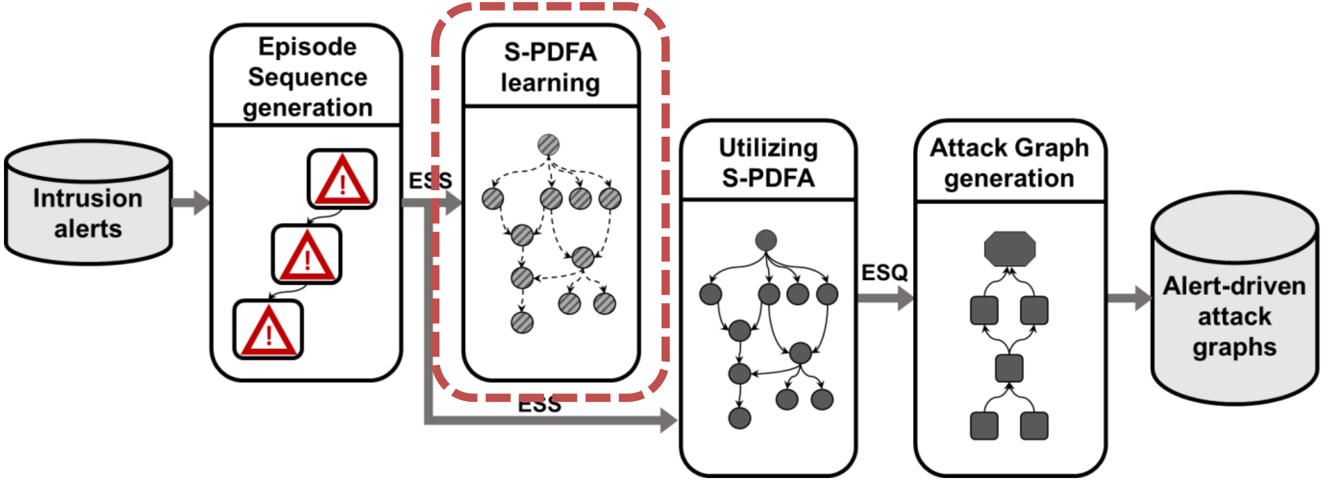

```
{   '_sourcetype': 'suricata:alert',
    'alert': {   'category': 'Attempted Information Leak',
                 'severity': 2,
                 'signature': 'ET POLICY Python-urllib\\/
                              'Suspicious User Agent'},
    'dest_ip': '169.254.169.254',
    'dest_port': 80,
    'src_ip': '10.0.0.20',
    'src_port': 56952,
    'timestamp': '2018-11-03T13:51:58.205548+0000'}}
```



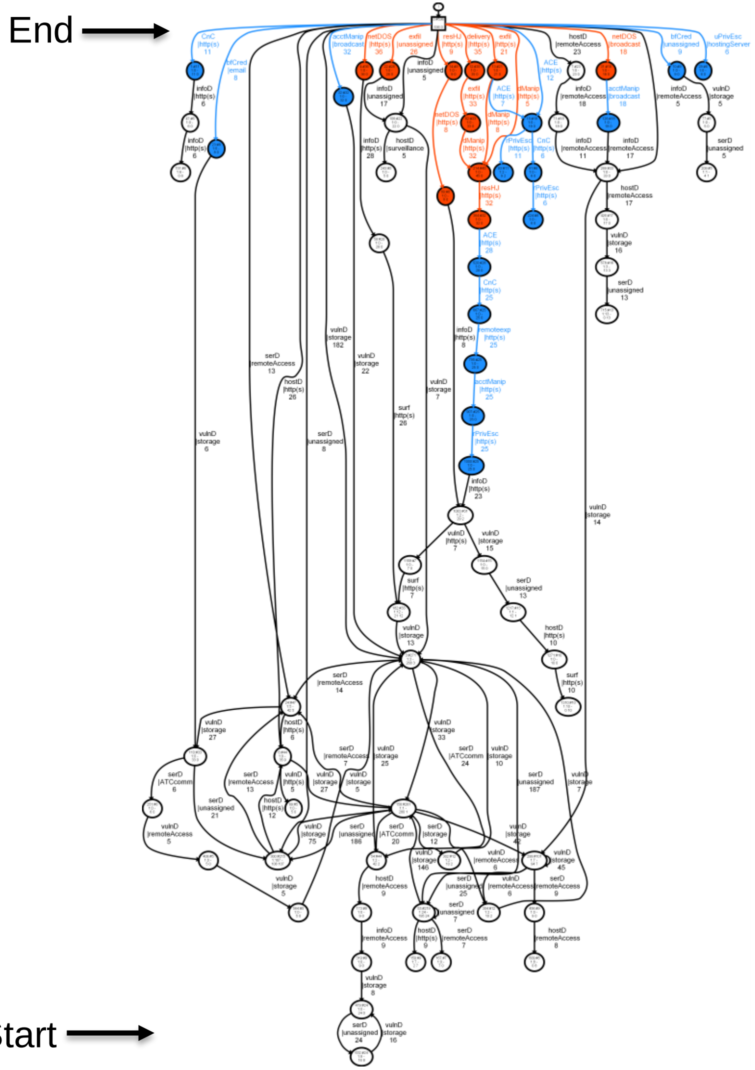


SAGE: IntruSion alert-driven Attack Graph Extractor



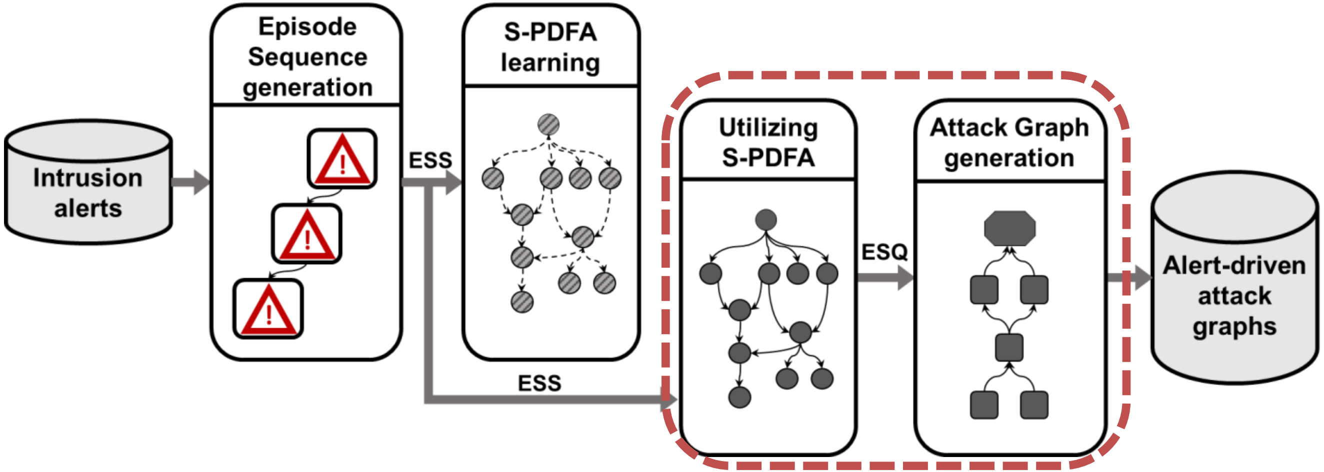
Suffix-based PDFA

- Summarizes attack paths
- Brings infrequent episodes to the top
  - Red → Severe | Blue → Medium severity
- States → milestones with context



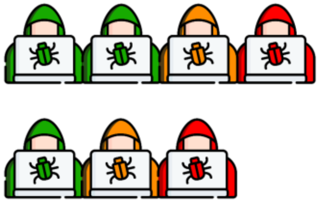
RIT
TUDelft

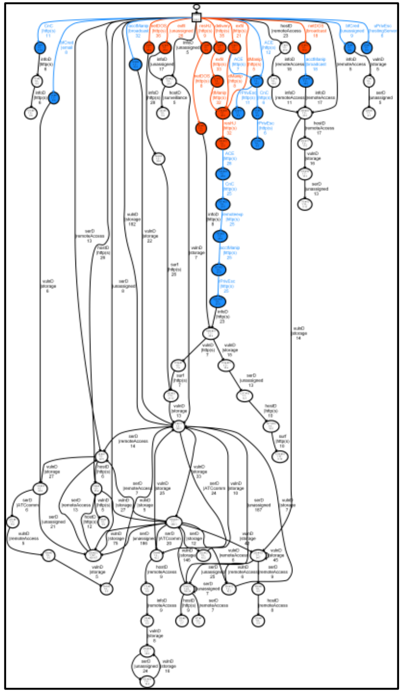
SAGE: IntruSion alert-driven Attack Graph Extractor



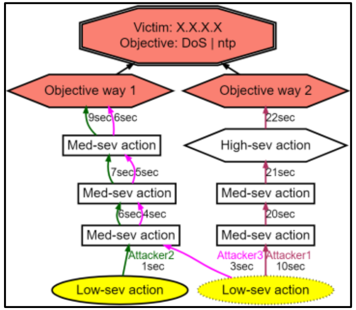
Adding context to sequences

Episode sequences





State sequences

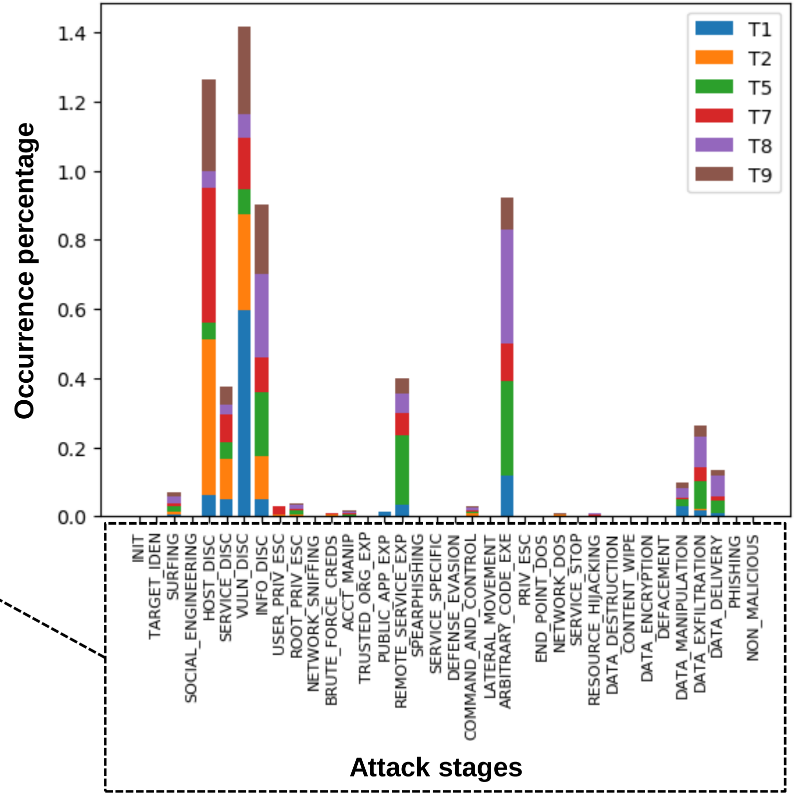


On a *per-victim*, *per-objective* basis

Experimental dataset

- Suricata alerts from Collegiate Penetration Testing Competition[1]
  - 6 multi-attacker teams
  - 1 fictitious network
  - 330,270 alerts

- Moskal's Action-Intent framework[2]
  - Alert signature → Attack stage
  - Based on MITRE ATT&CK



1. CPTC dataset: https://www.globalcptc.org/
2. S. Moskal and S. J. Yang, "Framework to describe intentions of a cyber attack action," arXiv preprint arXiv:2002.07838, 2020.

[1] Alert triaging

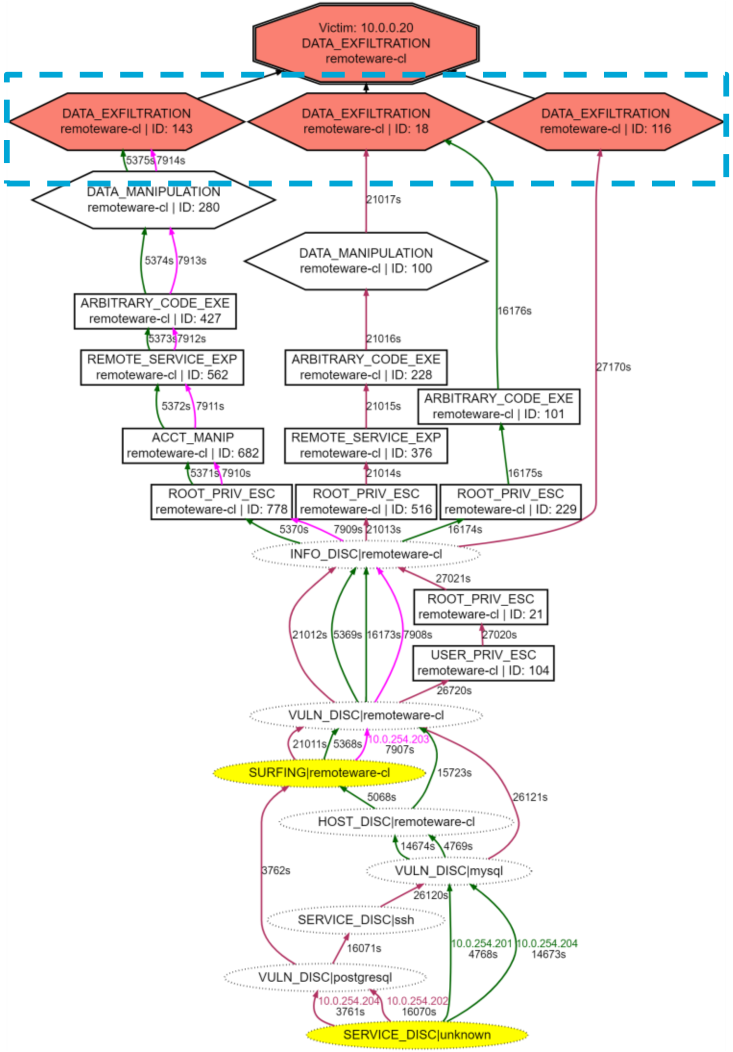
- 330,270 alerts → 93 alert-driven AGs
- ~500 alerts in < 25 vertices

	# alerts (raw)	# alerts (filtered)	#episodes	#ES/ #ESQ	#ESS	#AGs
T1	81373	26651	655	103	108	53
T2	42474	4922	609	86	92	7
T5	52550	11918	622	69	74	51
T7	47101	8517	576	63	73	23
T8	55170	9037	439	67	79	33
T9	51602	10081	1042	69	110	30

RIT
TU Delft

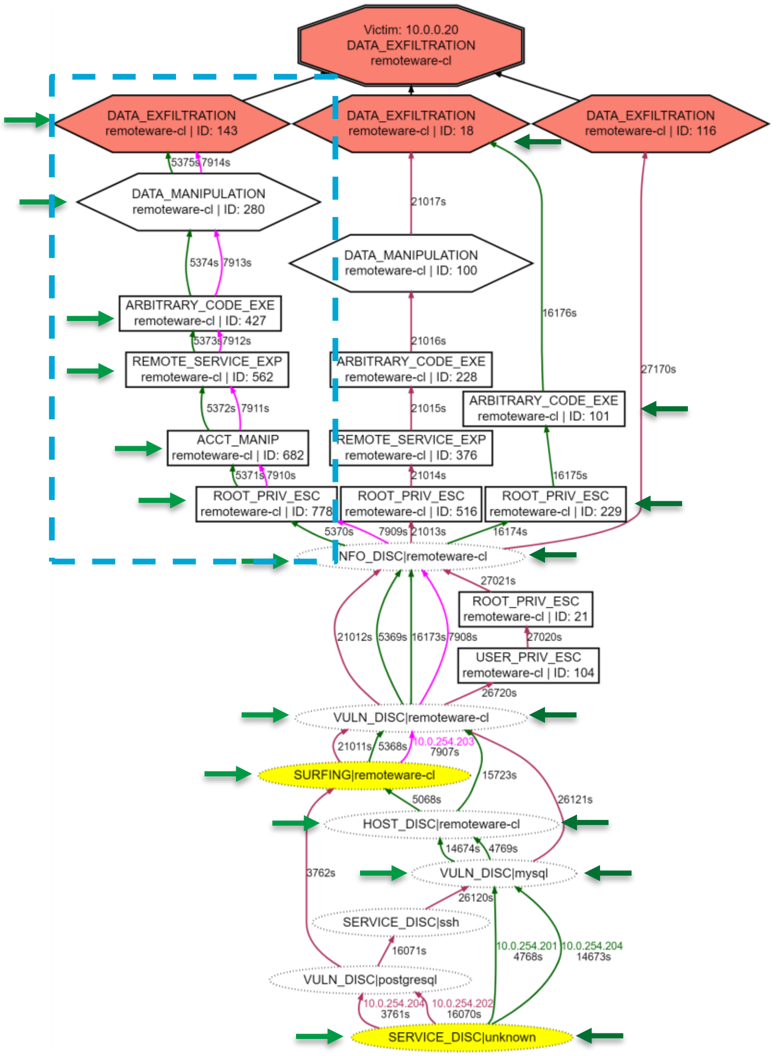
[2] Attacker strategy visualization

- Shows how the attack transpired
- 3 teams, 5 attempts
- 3 ways to reach objective



[3] Attacker strategy comparison

- T5 and T8 share a common strategy
- Some paths are shorter than others
- Attackers follow shorter paths to re-exploit an objective in 84.5% cases



RIT
TUDelft

[4] Ranking interesting attackers

- Rank on the uniqueness and severity of actions
- $Score = \frac{(2*sev)+(1*med)}{3}$

Team	Severe vertices (out of 70)	Medium vertices (out of 148)	Weighted average percentage
T5	28 (40%)	40 (27%)	35.67
T1	18 (26%)	62 (42%)	31.33
T9	23 (33%)	36 (24%)	30.0
T7	22 (31%)	26 (18%)	26.67
T8	15 (21%)	32 (22%)	21.33
T2	3 (4%)	8 (5%)	4.33

Take aways

- SAGE uses sequence learning to extract attacker strategies
  - Builds attack graphs from intrusion alerts without expert input

- The S-PDFA is critical for
  - Accentuating infrequent severe actions,
  - Identifying contextually different actions

- Alert-driven attack graphs
  - Compress millions of alerts in a few AGs
  - Provide insights into attacker strategies
  - Capture attackers' behavior dynamics

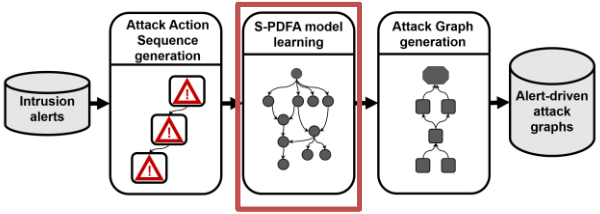
Thank you! Questions?

- SAGE uses sequence learning to extract attacker strategies
  - Builds attack graphs from intrusion alerts without expert input
- The S-PDFA is critical for
  - Accentuating infrequent severe actions,
  - Identifying contextually different actions
- Alert-driven attack graphs
  - Compress millions of alerts in a few AGs
  - Provide insights into attacker strategies
  - Capture attackers' behavior dynamics

SAGE is open-source!

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

RIT
TU Delft



Suffix Tree (Merged)

HostD	VulnD	ServD	Exfil
Surf	InfoD	ACExec	DoS
HostD	VulnD	ACExec	Exfil