AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

INTRODUCING A NEW ALERT DATA SET FOR MULTI-STEP ATTACK ANALYSIS

Max Landauer (AIT Austrian Institute of Technology), **Florian Skopik** (AIT Austrian Institute of Technology), **Markus Wurzenberger** (AIT Austrian Institute of Technology)

Workshop on Cyber Security Experimentation and Test (CSET 2024), August 13, 2024, Philadelphia, PA, USA

The work in this paper has received funding from the European Union - European Defence Fund under GA no. 101103385 (AInception) and GA no. 101121403 (NEWSROOM), and from the Austrian Research Promotion Agency (FFG) under GA no. FO999899544 (PRESENT). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them.

INTRUSION DETECTION

- Key element of cyber defense
- Autonomous monitoring of systems and networks for suspicious activities
- Types of IDS
  - **Data sources** → network packets (NIDS) or system/application log files (HIDS)
  - **Mode of operation** → expert rules or machine learning
  - **Triggers** → Simple string matching or statistical analysis
- Output: Low-level alerts
  - Attacks can cause **multiple alerts**
    - Many low-priority alerts from scanning activities
  - **False positives** are frequent
  - → **overwhelming** amount of alerts for analysts, causing fatigue
  - → relevant alerts are **concealed** in flood of alerts

BEYOND INTRUSION DETECTION

- Alert prioritization
- Enrichment of alerts with contextual information
  - Alert is part of an attack step, or part of a complex **attack chain**
- Multi-step attack analysis
  - Aggregation and correlation of single alerts into higher-level abstractions
  - Common issues
    - Multiple alerts per attack step
    - Alerts are dispersed across **several data sources** on the same machine
    - Alerts are dispersed across **several machines**
    - **False positives** occur at the same time as relevant alerts
    - Attack steps are **overlapping**
    - Difficult to map alerts to kill-chains

PROBLEM STATEMENT

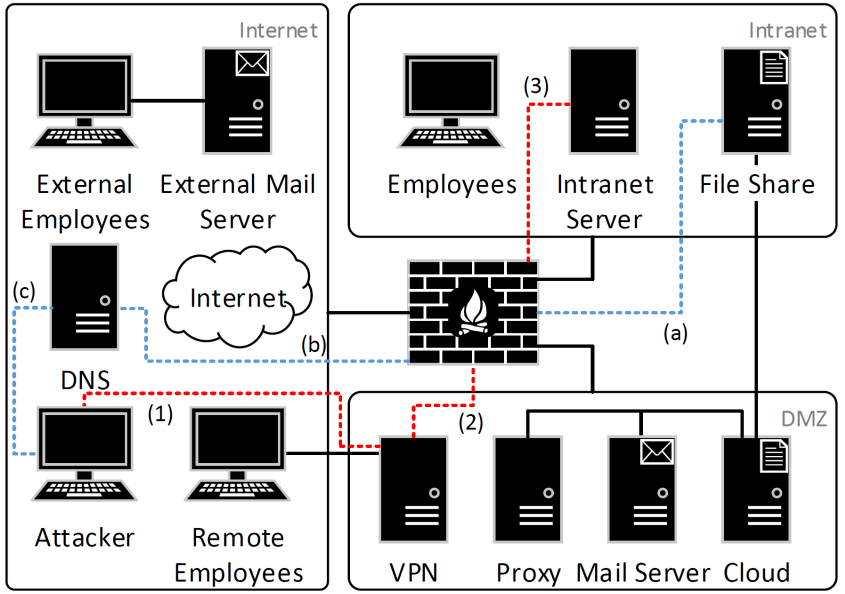
- New and innovative approaches are needed
- One of the main issues is the **lack of publicly available data sets**
- Problems with existing data sets:
  - Outdated and oversimplified
  - Single source of data
  - Designed for intrusion detection rather than multi-step attack analysis
  - Lack of ground truth
  - → researchers resort to private data sets that prevent reproducibility

PROPOSED SOLUTION

- AIT Alert Data Set (AIT-ADS)
  - High volumes of alerts
  - Many false positives
  - Heterogeneous IDS
    - Diverse detection techniques
    - Diverse alert formats
  - Alerts from multiple network components and data sources
  - Anomaly-based alerts that lack contextual information
  - Changes of attack step order and attack parameters
  - Repeatable attack plan
  - Repeated attack execution

GENERATION: LOG DATA SET

- Only few public log data sets
- AIT-LDSv2
  - Virtual test environment for data collection
  - Small enterprise network
  - State machines for normal behavior
    - Multi-step attack
    - Scans (Nmap, Dirb, WPScan)
    - Exploits (WordPress vulnerability)
    - Password cracking
    - Reverse shell + privilege escalation
    - Data exfiltration
  - Executed eight times with variations



GENERATION: SIGNATURE-BASED IDS

- Wazuh
  - Host-based
  - Comes with set of expert rules for various log sources
  - Some advanced rules (dependencies, event counts)
- Suricata
  - Network-based
  - Network packet inspection
  - Pattern matching
  - Already available in AIT-LDSv2

SAMPLE ALERTS - WAZUH


```
{
  "data": {
    "srcuser": "www",
    "dstuser": "data:jhall"
  },
  "rule": {
    "description": "User successfully changed UID.",
    "firedtimes": 1,
    "id": "5304",
  },
  "full_log": "Jan 24 04:37:40 intranet-server su[27950]: + /dev/pts/1 www-data:jhall",
  "@timestamp": "2022-01-24T04:37:40.000000Z",
  "location": "/var/log/auth.log",
}
```

SAMPLE ALERTS - SURICATA

{ "data": {
"tx_id": "0",
"app_proto": "http",
"in_iface": "ens3",
"src_ip": "192.168.230.122",
"src_port": "34642", "dest_ip": "172.19.130.68",
"proto": "TCP",
"dest_port": "80", },
"rule": {
"firedtimes": 15,
"mail": false,
"level": 3,
"description": "Suricata: Alert - ET SCAN Possible Nmap User-Agent Observed„ },
"@timestamp": "2022-01-24T03:57:01.687867Z", }

GENERATION: ANOMALY-BASED IDS

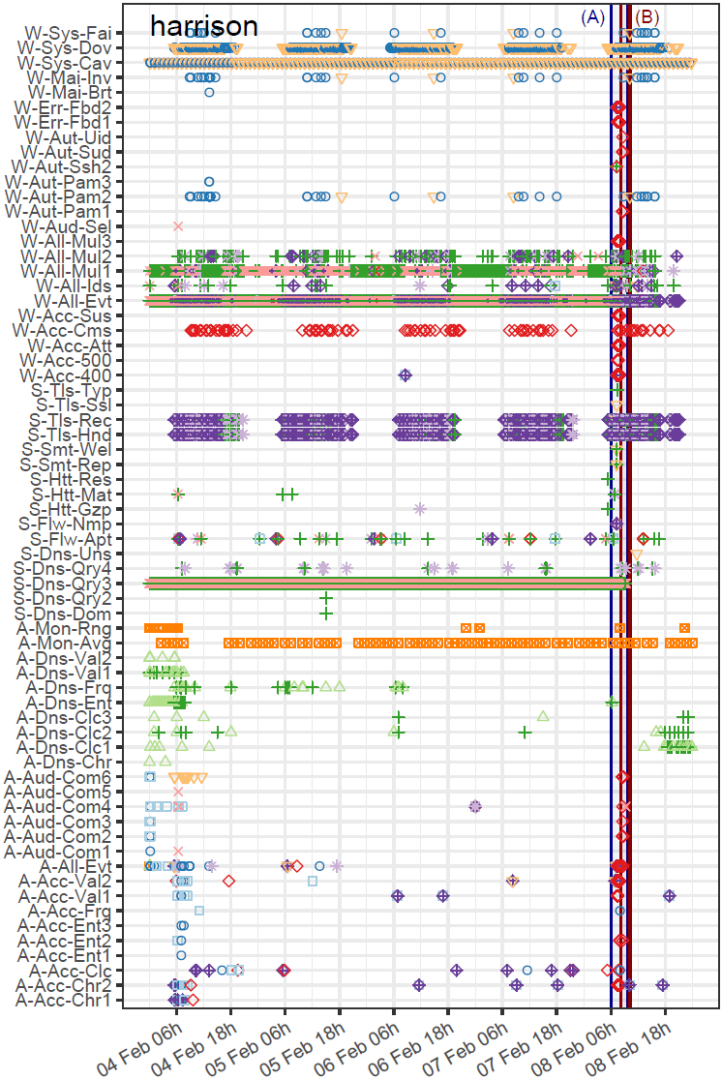
- AIT‘s AMiner
  - Host-based
  - Learn model of normal behavior, detect deviations as anomalies
  - Semi-supervised - requires training (first two days of AIT-LDSv2)
  - Detectors specifically configured for AIT-LDSv2
    - New events
    - New event parameters (e.g., Apache access status code)
    - New parameter combinations (e.g., audit syscall + uid + exe)
    - Unusual entropy/characters in event parameters (e.g., Apache access request)
    - Unusual event frequencies
    - Unusual numeric parameters (e.g., sudden spikes in CPU utilization)

SAMPLE ALERTS - AMINER

{"AnalysisComponent": {
"AnalysisComponentType": "EntropyDetector",
"AnalysisComponentName": "AMiner: High entropy in Apache Access request.",
"Message": "Value entropy anomaly detected",
"CriticalValue": 0.04173736650922487,
"ProbabilityThreshold": 0.05 },
"LogData": {
"RawLogData": [
"172.19.131.174 - - [24/Jan/2022:03:59:22 +0000] \"GET /wp-content/uploads/2022/01/ekmkimzkps-1642996700.9285.php?wp_meta=WyJ3Z2V0IiwgImh0dHBzOi8vZ2l0aHViLmNvbS9haXQtYWVjaWQvd3BoYXNoY3JhY2svYXJjaGl2ZS9yZWZzL3RhZ3MvdjAuMS50YXIuZ3oiXQ%3D%3D HTTP/1.1\" 200 506741 \"-\" \"python-requests/2.27.1\""
], "LogResources": [
"/var/log/apache2/intranet-access.log"
]}}

SCENARIO TIMELINE

- Alerts with 93 different signatures
  - 34 AMiner, 29 Suricata, 30 Wazuh
- 10 log sources
- 5 days (daily patterns)
- Many false positives outside of attack phases
  - Software updates, account login, training phase
- Attacks trigger some new alert types
- Data exfiltration already active from start of simulation

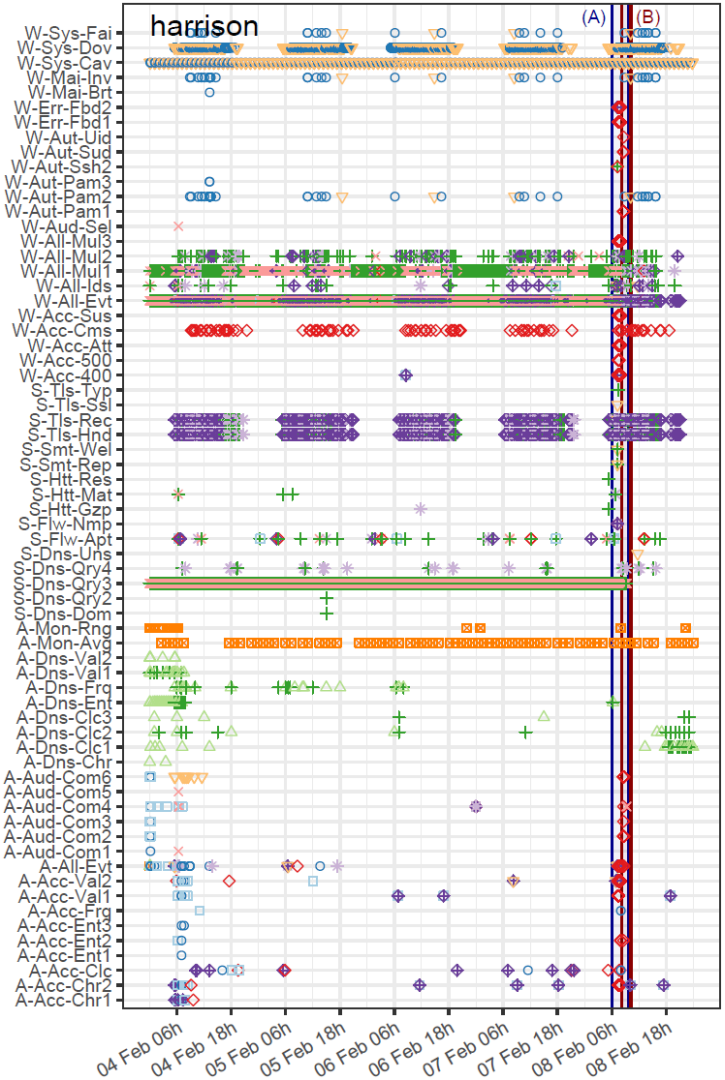


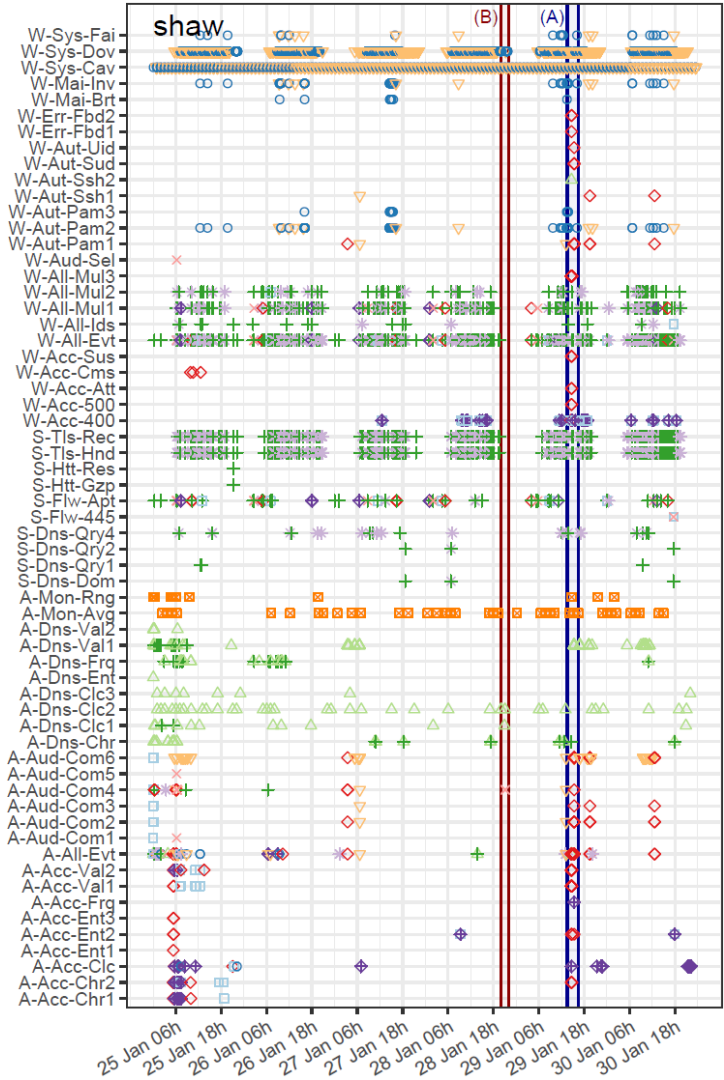
Location

Cloud share | Ext. mail server | Firewall | Mail server | VPN server

DNS server | File share | Intranet server | Monitoring | Web server

TWO DATA SETS

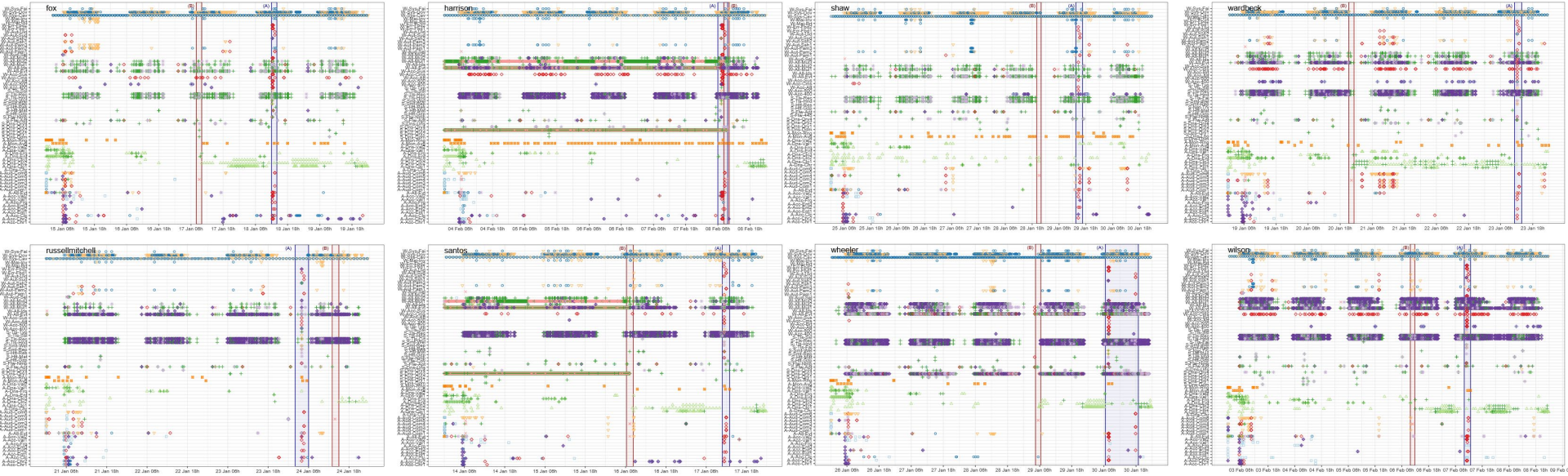




09/08/2024

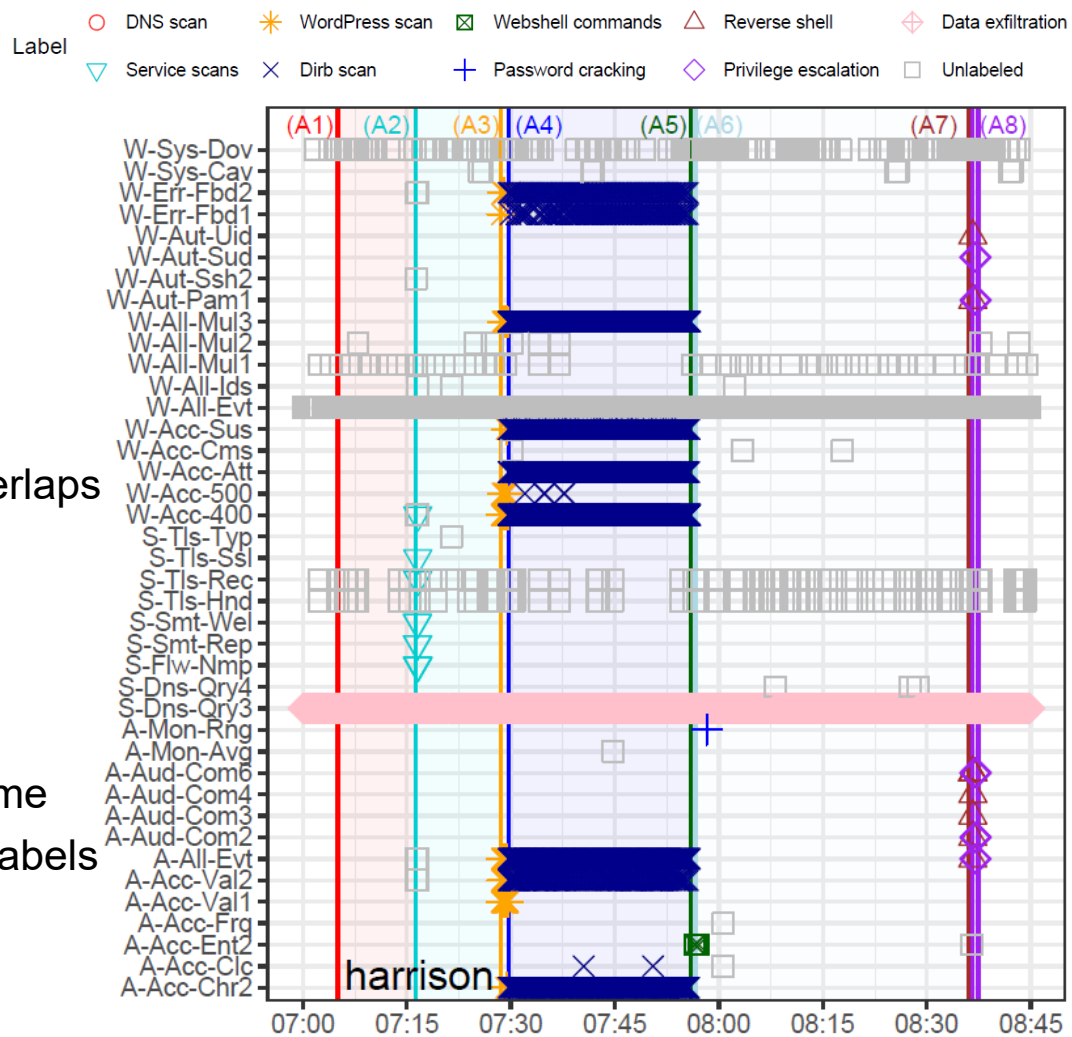
EIGHT DATA SETS

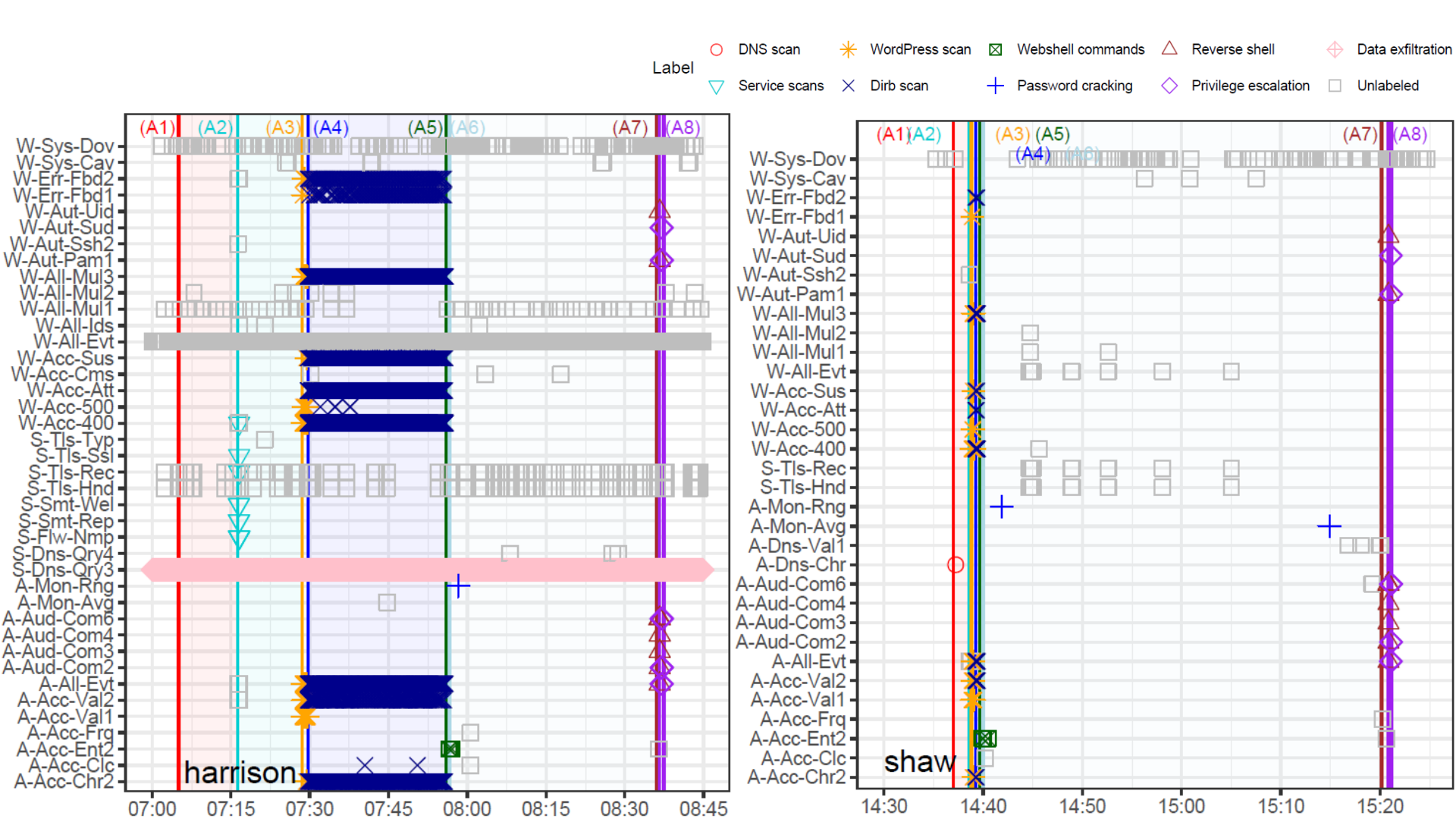
- Different simulation lengths, duration of attack phases, order of attack steps, number of users causing false alerts, etc.



LABELING

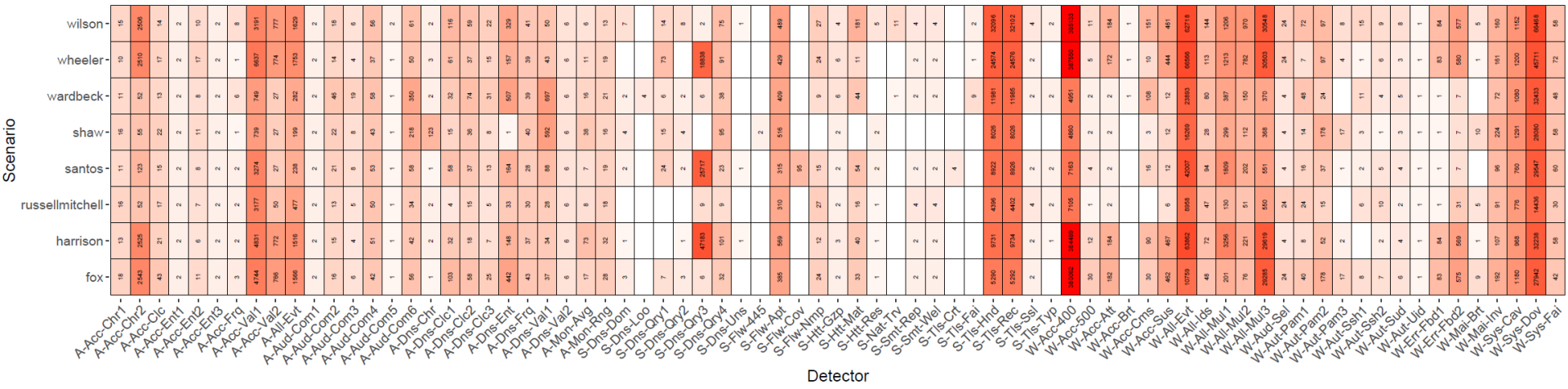
- AIT-LDSv2 is labeled
- Time-based labeling
  - Attack schedule is known
  - Start and stop time of attacks
  - Problems: Delays, false pos., overlaps
  - Shaded intervals
- Event-based labeling
  - Expert rules
  - HIDS: Match log line from alerts
  - NIDS: Match protocol, IP, port, time
  - Problem: Accuracy relies on log labels
    - Some alerts remain unlabeled




Label
DNS scan
WordPress scan
Webshell commands
Reverse shell
Data exfiltration
Service scans
Dirb scan
Password cracking
Privilege escalation
Unlabeled
(A1)
(A2)
(A3)
(A4)
(A5)
(A6)
(A7)
(A8)
W-Sys-Dov
W-Sys-Cav
W-Err-Fbd2
W-Err-Fbd1
W-Aut-Uid
W-Aut-Sud
W-Aut-Ssh2
W-Aut-Pam1
W-All-Mul3
W-All-Mul2
W-All-Mul1
W-All-Ids
W-All-Evt
W-Acc-Sus
W-Acc-Cms
W-Acc-Att
W-Acc-500
W-Acc-400
S-Tls-Typ
S-Tls-Ssl
S-Tls-Rec
S-Tls-Hnd
S-Smt-Wel
S-Smt-Rep
S-Flw-Nmp
S-Dns-Qry4
S-Dns-Qry3
A-Mon-Rng
A-Mon-Avg
A-Aud-Com6
A-Aud-Com4
A-Aud-Com3
A-Aud-Com2
A-All-Evt
A-Acc-Val2
A-Acc-Val1
A-Acc-Frg
A-Acc-Ent2
A-Acc-Clc
A-Acc-Chr2
harrison
07:00
07:15
07:30
07:45
08:00
08:15
08:30
08:45
shaw
14:30
14:40
14:50
15:00
15:10
15:20

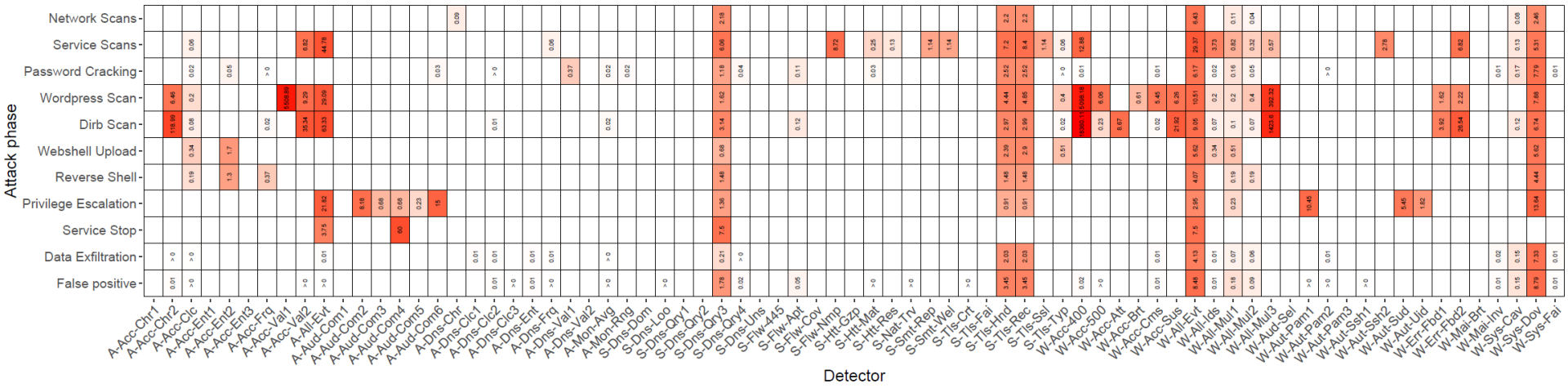
ALERT COUNTS

- 2,655,821 alerts across all scenarios
  - 86% Wazuh, 12% Suricata, 2% AMiner
  - Variations across scenarios
    - Depends on number of users, simulation length, attack parameters (scan mode)



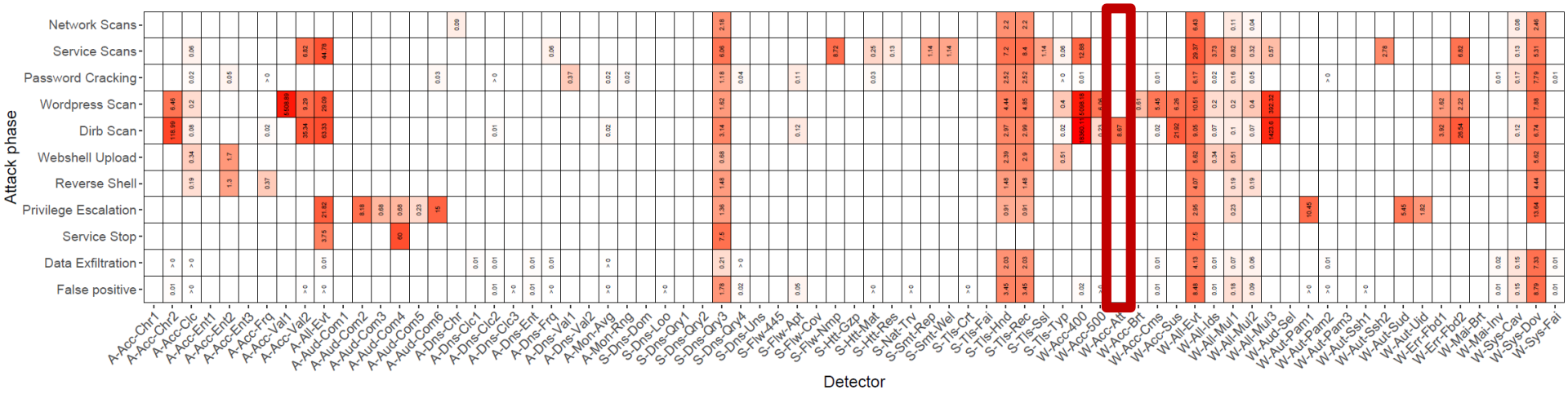
ALERT RATES

- Count alerts in attack time windows and during normal operation
- Indicates „useful“ detectors
  - Many detections during one or more attack phase
  - No or few detections during normal operation



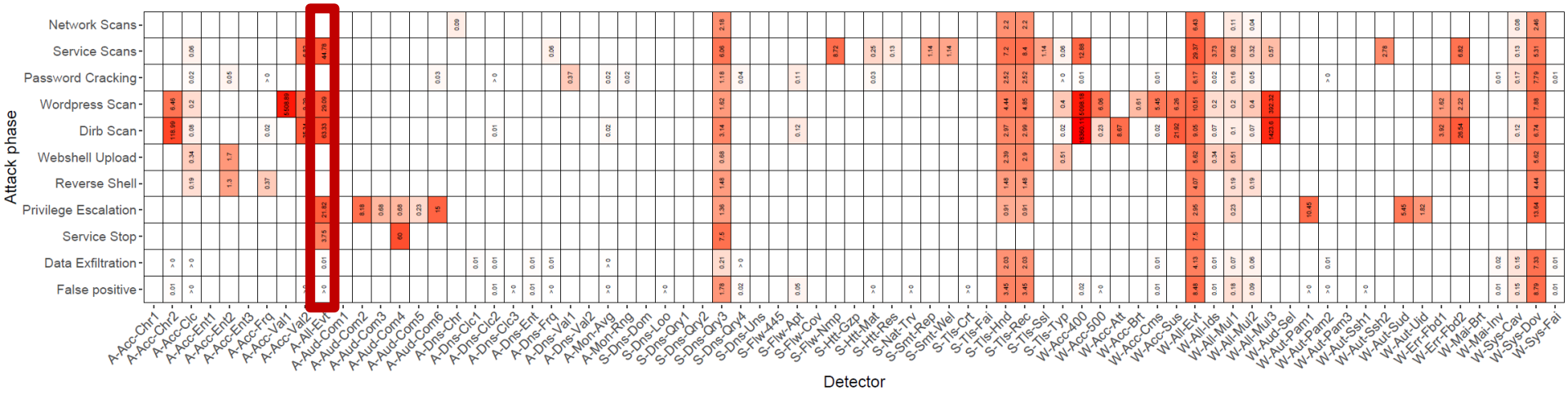
ALERT RATES

- Count alerts in attack time windows and during normal operation
- Indicates „useful“ detectors
  - Many detections during one or more attack phase
  - No or few detections during normal operation
- More than 8 alerts/minute during dirb scan, no false positives



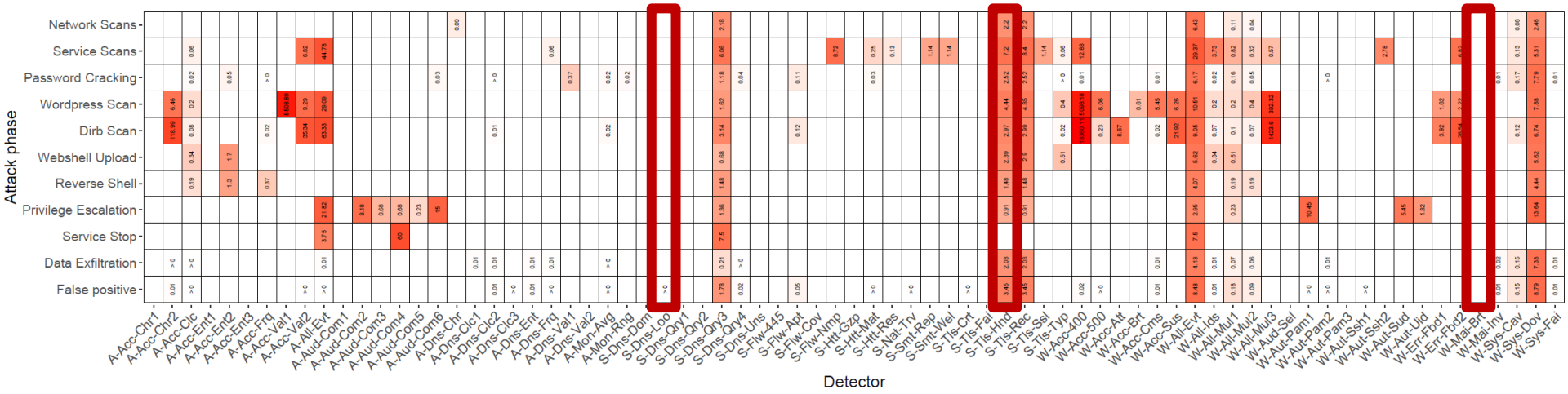
ALERT RATES

- Count alerts in attack time windows and during normal operation
- Indicates „useful“ detectors
  - Many detections during one or more attack phase
  - No or few detections during normal operation
- Many alerts during various attacks, few false positives



ALERT RATES

- Count alerts in attack time windows and during normal operation
- Indicates „useful“ detectors
  - Many detections during one or more attack phase
  - No or few detections during normal operation
- No attacks detected or too many false positives



DETECTOR SCORES

- Compute quantitative scores based on insights from alert rates
  - Compare number of alerts reported during attack interval with false positives
  - Weigh by duration to compensate uniformly occurring false positives
  - Average over all scenarios S
  - → Measures robustness against false positives

$$s_{rob}(A, D) = \frac{1}{\#S} \sum_{S} \left(1 - min\left(1, \frac{\#(\mathcal{A}_{D,S} \text{ in } \Delta_{T,S})}{\#(\mathcal{A}_{D,S} \text{ in } \Delta_{A,S})} \cdot \frac{\Delta_{A,S}}{\Delta_{T,S}}\right)\right)$$

  - Detection should work independent from attack parameters or system setup
  - → Measure whether attack is detected across all scenarios

$$s_{det}(D) = \max_{A} \left(s_{rob}(A, D) \cdot \frac{\#(S : A \in S \wedge \#(\mathcal{A}_{D,S} \text{ in } \Delta_{A,S}) > 0)}{\#(S : A \in S)}\right)$$

  - Detecting multiple attacks is nice, but not required → use maximum for any attack

Detector	Network Scans	Service Scans	WordPress Scan	Dirb Scan	Webshell Upload	Password Cracking	Reverse Shell	Privilege Escalation	Service Stop	Data Exfiltration	False positives	Robustness Score	Detection Score
W-All-Mul3		5	8	8								1.0	1.0
W-Acc-Sus			6	8								1.0	1.0
W-Acc-Att				8								1.0	1.0
W-Err-Fbd2		5	3	8								1.0	1.0
W-Aut-Ssh2		8										1.0	1.0
W-Aut-Uid								8				1.0	1.0
W-Aut-Sud								8				1.0	1.0
W-Err-Fbd1			8	4								1.0	1.0
A-Aud-Com4								3	8			1.0	1.0
A-Aud-Com2								8				1.0	1.0
A-Aud-Com6						1		8				1.0	1.0
A-Acc-Val1			8									1.0	1.0
A-Acc-Ent2					8	7	7					1.0	1.0
W-Acc-400		7	8	8		1					4	1.0	1.0
A-All-Evt		8	8	8				8	1	1	2	1.0	1.0
W-Acc-500			8	4							1	1.0	1.0
A-Acc-Val2		5	8	8							2	1.0	1.0
W-Aut-Pam1								8			1	1.0	1.0
A-Acc-Chr2			8	8						1	1	1.0	1.0
S-Smt-Val		7										1.0	0.88

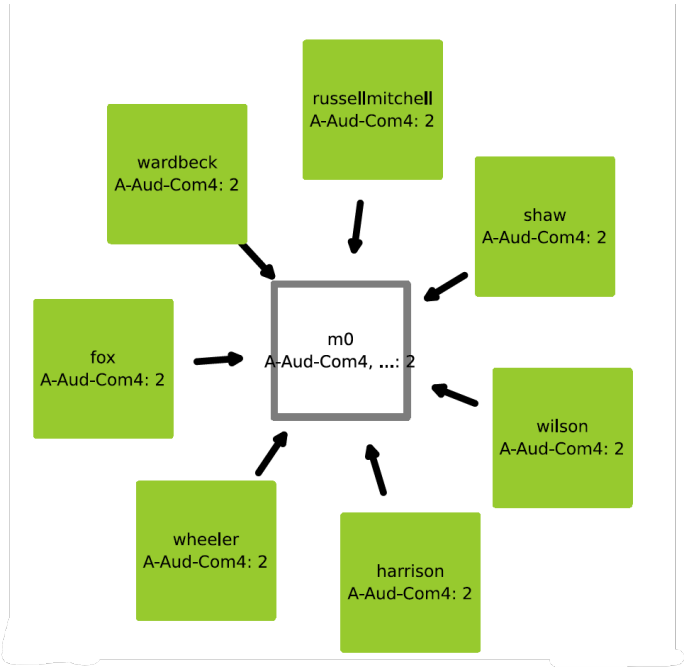
A-Mon-Rng						5						1.0	0.71
W-All-Evt	5	7	5	4	5	7	3	2	1	7	8	0.8	0.7
W-All-Mul1	5	6	1	3	3	6	1	1		5	8	0.81	0.61
S-Tls-Rec	5	7	5	4	6	6	3	1		7	8	0.57	0.5
A-Acc-Clc		1	1	4	2	3	1			1	1	0.99	0.49
W-All-Mul2	4	4	2	3		5	1			4	7	0.9	0.45
S-Htt-Mat		1				3					2	0.94	0.4
S-Tls-Typ		1	2	1	3	1						1.0	0.38
A-Aud-Com3								3				1.0	0.38
W-Acc-Brt			3									1.0	0.38
W-Acc-Cms			3	1		2				4	5	1.0	0.37
S-Flw-Apt				1		3					8	0.82	0.35
W-Mai-Inv						1				3	5	0.8	0.3
W-Sys-Fai						1				3	5	0.8	0.3
W-Aut-Pam2						1				3	5	0.8	0.3
W-Sys-Dov	7	3	5	4	3	6	5	5		7	8	0.46	0.29
S-Tls-Hnd	5	3	3	4	3	6	3	1		7	8	0.42	0.26
S-Htt-Res		2										1.0	0.25
A-Dns-Clc1										2		1.0	0.25
A-Dns-Frq		1								2	1	1.0	0.25
A-Acc-Frq				2		1	2					1.0	0.25
W-Sys-Cav	1	1		2		7				8	8	0.24	0.24
S-Dns-Qry4						2				2	6	0.85	0.24
A-Dns-Clc2				1		1				3	5	0.5	0.19
A-Dns-Val1						1						1.0	0.14
A-Dns-Chr	1											1.0	0.12
A-Aud-Com5								1				1.0	0.12
S-Dns-Qry3	2	1	1	2	1	1	2	1	1	1	2	0.88	0.11
A-Dns-Ent										1	2	0.63	0.08

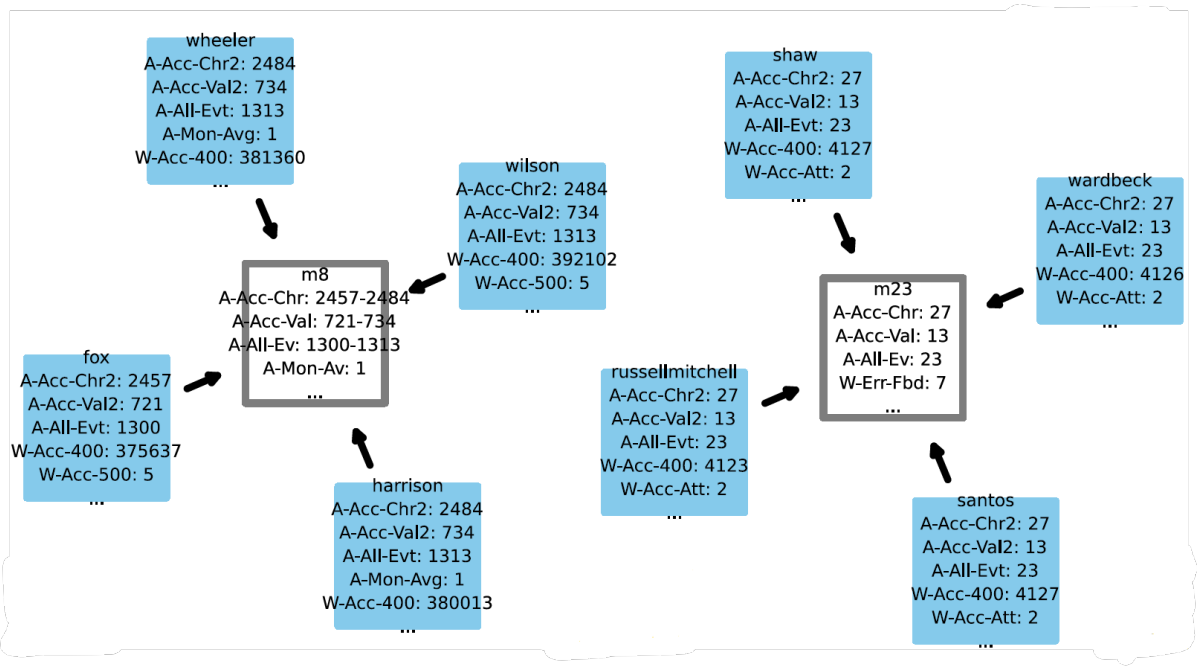
ALERT AGGREGATION

- Alerts from top 26 detectors evaluated in illustrative use-case
- Identify repeating patterns
- Generate abstract representations of activities and attacks steps
  - Merging two or more related alerts (e.g., based on similarity or co-occurrence)
- AECID-Alert-Aggregation
  - Groups alerts based on occurrence time
  - Incremental clustering of groups based on alert attributes, frequencies, and sequences
  - Merge highly similar groups (e.g., replace values with wildcards)

META-ALERTS

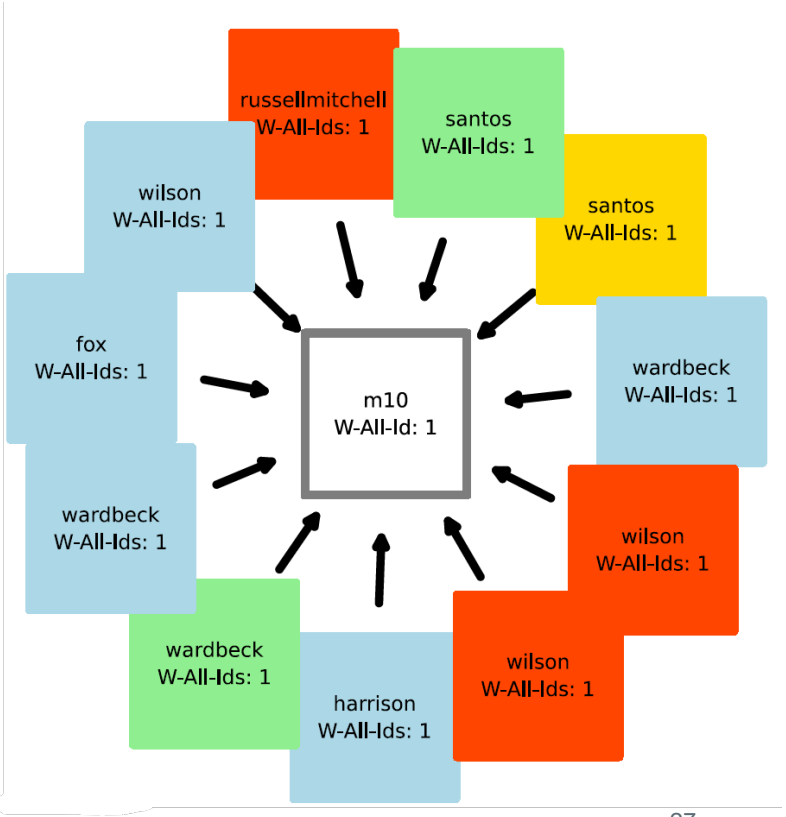
- Green: Service stopped from 7 out of 8 scenarios
- Blue: Dirb scan in basic and extensive mode (number of W-Acc-400)

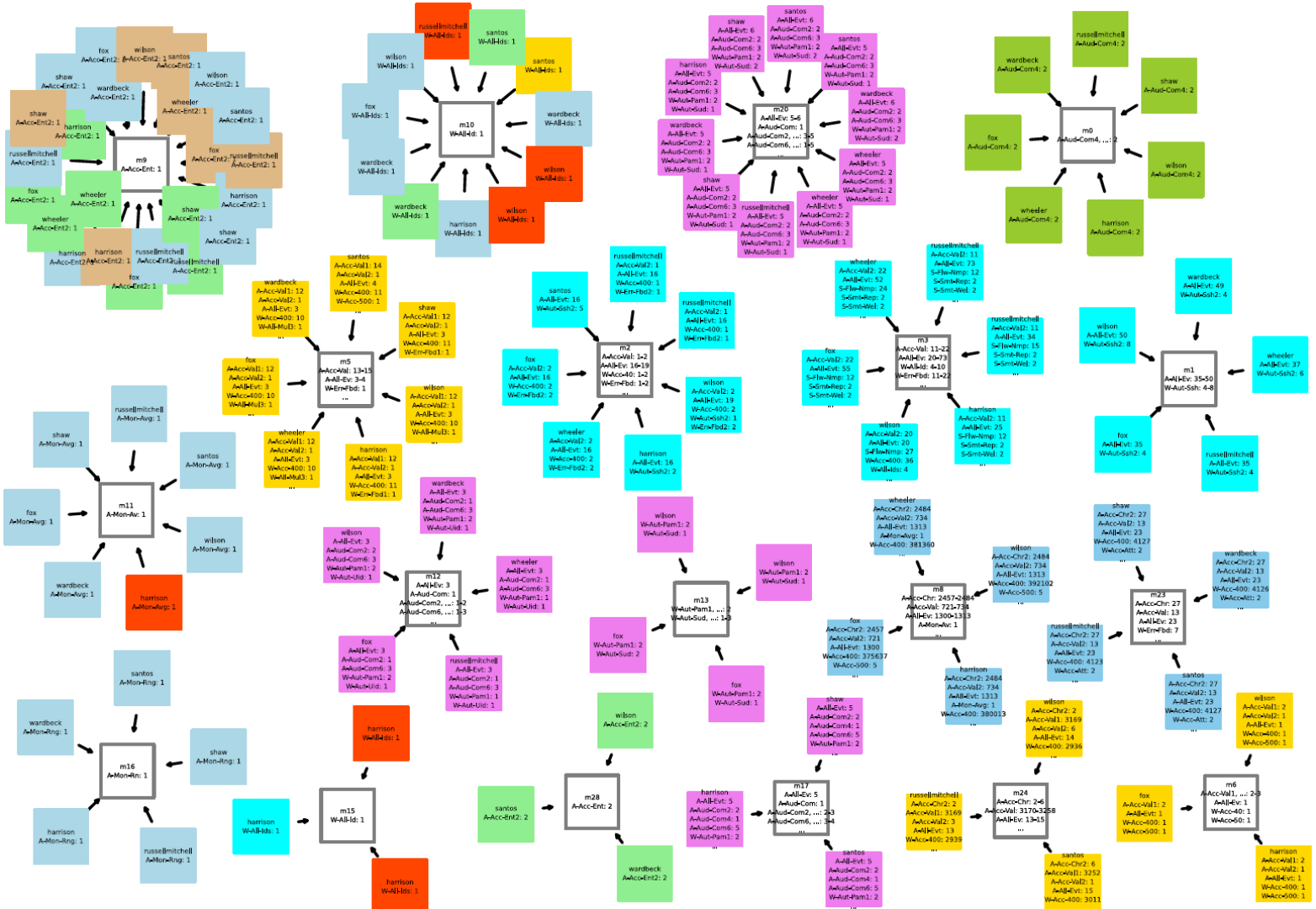




META-ALERTS

- Open issues
  - Works best for long alert patterns
  - Single alerts more difficult to group
  - Not robust to noise



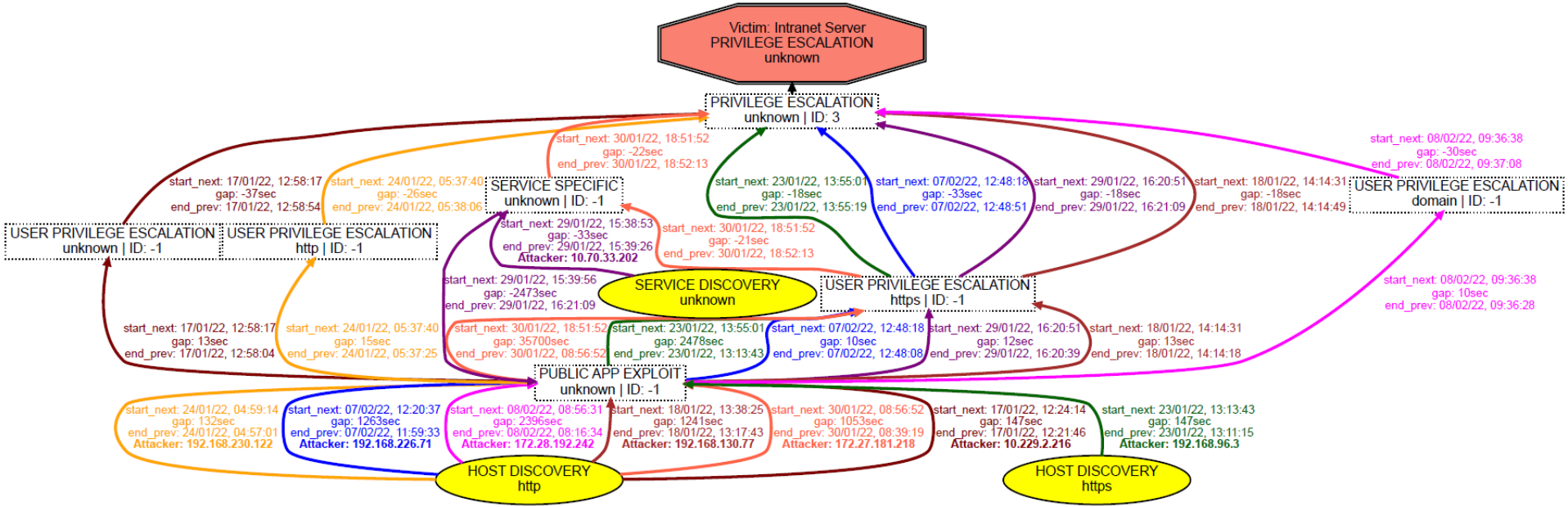


09/08/2024

ATTACK GRAPH

- Alert aggregation puts less focus on sequential execution stages of attacks
- Attack graphs visually summarize attack strategies
- SAGE is an open-source approach to automate attack graph extraction
  - Map alerts to attack steps
  - Filtering of irrelevant alerts
  - Grouping into episodes
  - Merge episodes into single graph
- Several attackers, same target: single end node, multiple start nodes

ATTACK GRAPH



ALERT REDUCTION RATES

- Key metric to compare alert filtering and aggregation approaches

Alerts	harrison	russellmitchell	santos	shaw	Avg. reduction rate
All	593,948	45,544	130,779	70,782	-
Filtered by prioritization	425,392 (28.38%)	11,705 (74.30%)	11,709 (91.05%)	6,667 (90.58%)	56.12%
In attack phases	431,492 (27.35%)	12,015 (73.62%)	13,004 (90.06%)	6,935 (90.20%)	55.6%
Filtered and in attack phases	424,974 (28.45%)	11,230 (75.34%)	11,217 (91.42%)	6,065 (91.43%)	56.57%
SAGE	6,515 (98.47%)	383 (96.59%)	238 (97.88%)	175 (97.11%)	97.73%
Alert aggregation	167 (99.96%)	167 (98.51%)	167 (98.51%)	167 (97.25%)	98.93%

DISCUSSION

- Prioritization
  - Our prioritization relies on labeled data, which is not available in practice
  - Semi- or unsupervised approaches required
- Meta-alert generation
  - Does not consider progression of attack
- Attack graph extraction
  - Depends on manual mapping of alerts to attack steps
  - Alerts are often generic and may fit into several steps of kill chain
- Future work
  - Combine meta-alert aggregation with attack graph extraction
  - Evaluations of federated and collaborative intrusion detection systems

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

THANK YOU!

Code to obtain and reproduce data sets available at

https://zenodo.org/records/8263181

https://github.com/ait-aecid/alert-data-set