VILNIUS TECH
Vilnius Gediminas Technical University

Correlation-Based Log Data Analysis for Cybersecurity Incident Detection

Dalius Mažeika, Dominykas Volodka

Vilnius Gediminas Technical University, Saulėtekio al. 11 Vilnius, Lithuania

Summary. Study investigates correlation-based techniques for analyzing distributed logs to improve incident detection. Two datasets were used: a public Windows authentication log dataset and a new dataset generated in a virtual infrastructure under controlled attack scenarios. Four aggregation methods—statistical, time-based, model-based, and behavior-based—were applied, followed by Min-Max or StandardScaler normalization. Correlation analysis employed Pearson and Kendall methods, complemented by Euclidean distance and cosine similarity measures. A modular testbed was developed to evaluate correlations across network traffic, DNS, and authentication logs. Simulated cyberattacks were executed to assess performance with particular focus on logs from the IIS server, firewall, VSFTPD, SSHD, and system auditing services. Results show that, with proper configuration, the proposed correlation-based approach can effectively detect anomalies and identify malicious activity.

DATASETS

Two datasets were used for cybersecurity incident detection:

1. **Comprehensive, Multi-Source Cyber-Security Events Dataset** from LANL
   - Authentication events, DNS lookups, network flow data, red teaming events
2. **Dataset generated by authors using GNS3** virtual environment
   - Windows logs: IIS server, firewall, VSFTPD, SSHD, system auditing services

Table 1. Dataset Features

Features	CMSCS-ED LANL	GNS3
Number of Features	28	30
Number of Records	1 648 275 307	357 673
Size, GB	12	0.05
Number of Computers	17 684	6

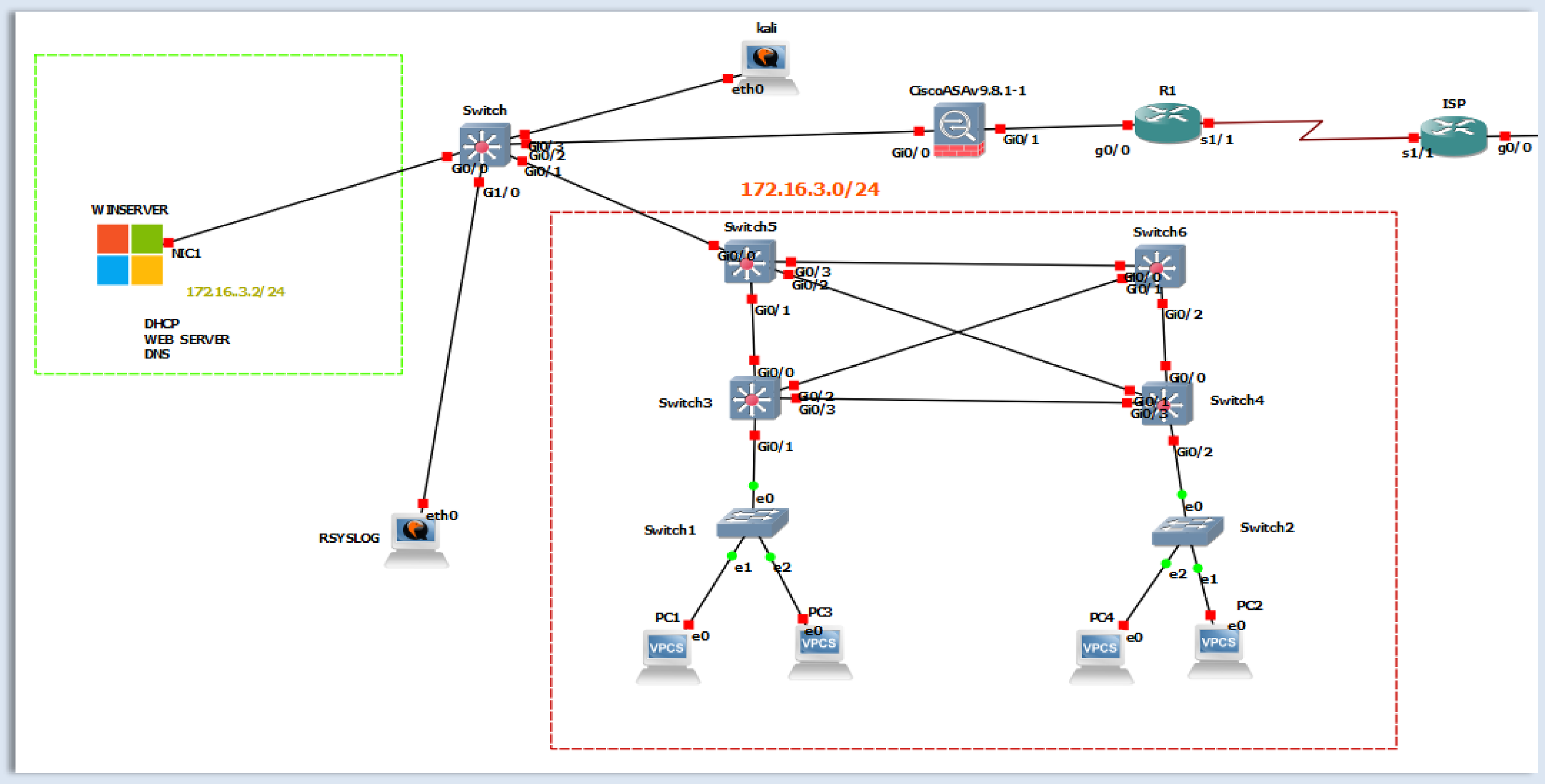

Fig 1. GNS3 Virtual Infrastructure

```
Mar 07 00:00:03 WINSRV01 DHCPServer[1022]: DHCPACK on
172.16.3.51 to 00:0c:29:11:22:02 via eth0
Mar 07 00:00:20 ASA-FIREWALL %ASA-6-302014: Teardown TCP
connection 1001 for outside:8.8.8.8/53 to
inside:172.16.3.2/54321 duration 0:00:30 bytes 256
Mar 07 00:10:10 PC3 ntpd[1345]: synchronized to 172.16.3.2,
stratum 3
Mar 07 00:15:02 kali systemd[1]: Started Network Manager.
Mar 07 00:15:05 kali NetworkManager[980]: <info>
[1741260905.1234] device (eth0): state change: config -> ip-
config (reason 'none')[501]
```

RESEARCH WORKFLOW

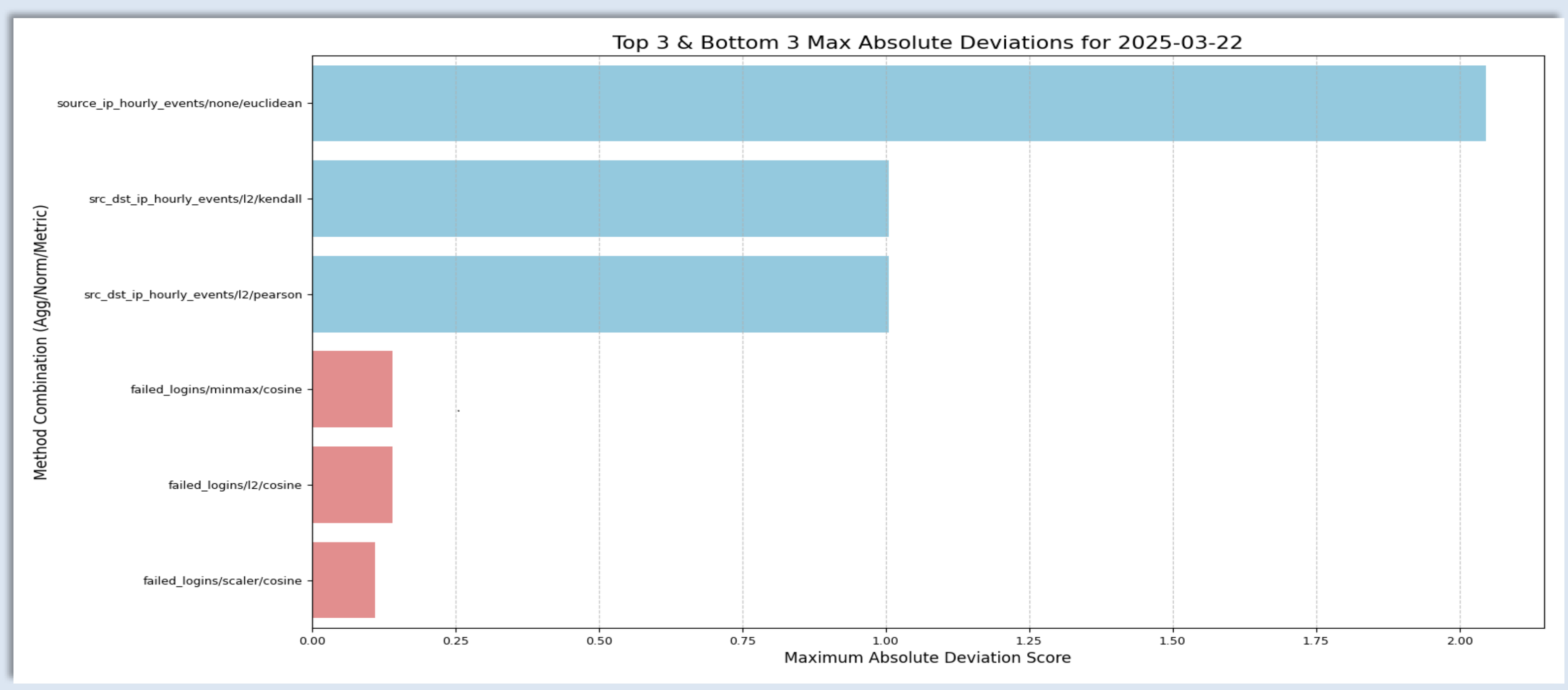

Fig 2. Summary of the methods

RESULTS OF THE EXPERIMENTS

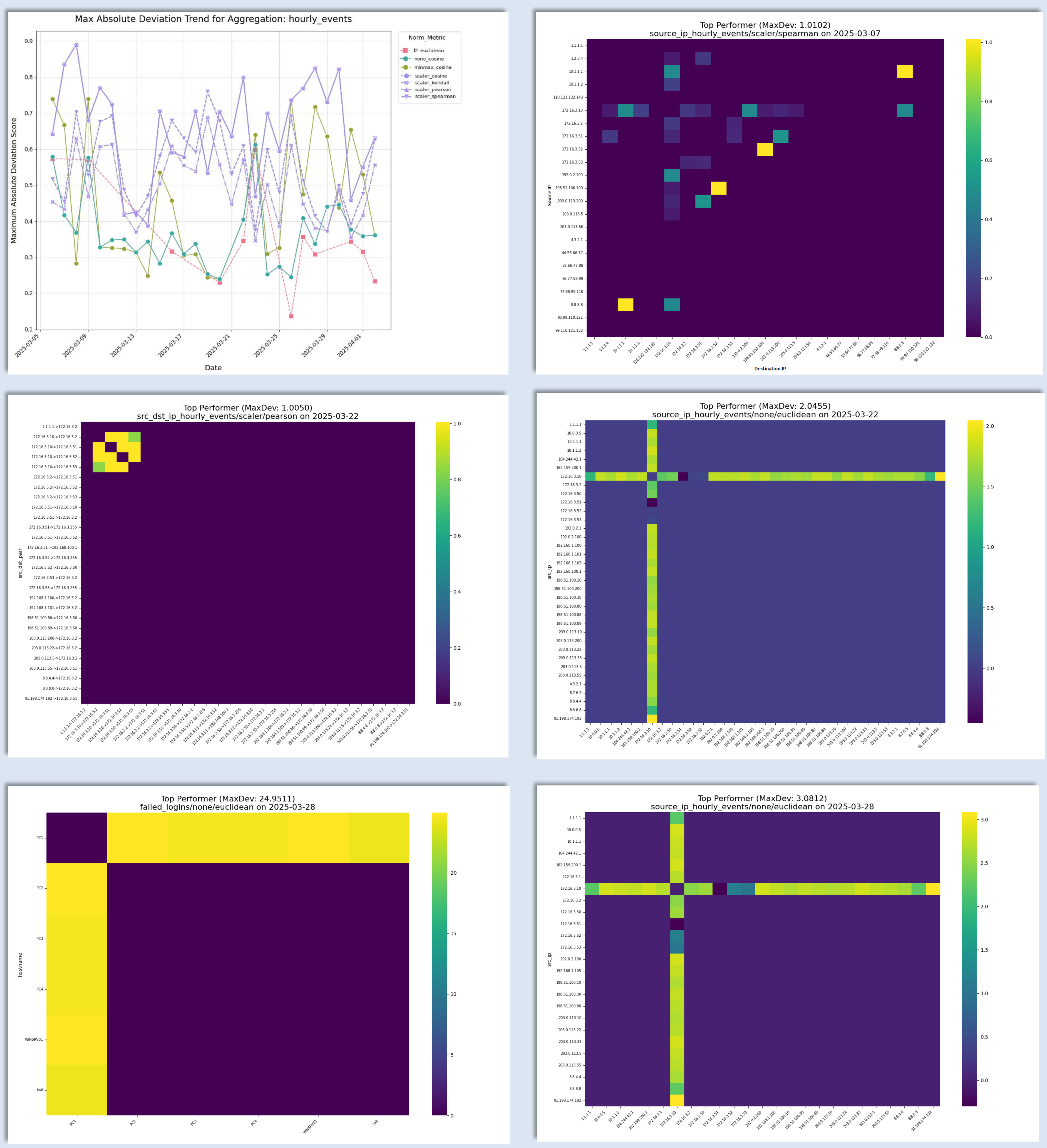

Fig 3. Correlation Diagrams

CONCLUSIONS

The experimental results indicate that, with appropriate configuration, the proposed method can effectively detect anomalies and identify malicious activities.