Sometimes Simpler is Better: A Comprehensive Analysis of State-of-the-Art Provenance-based Intrusion Detection Systems

paper

framework

Tristan Bilot[123], Baoxiang Jiang[4], Zefeng Li[5], Nour El Madhoun[2], Khaldoun Al Agha[1], Anis Zouaoui[3], Thomas Pasquier[5]

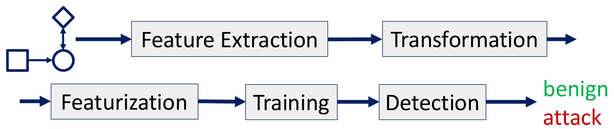
[1]Université Paris-Saclay, [2]LISITE, Isep, [3]Iriguard, [4]Xi'an Jiaotong University, [5]University of British Columbia

Overview

- We built an **efficient framework** integrating 8 SOTA anomaly-based Provenance-based Intrusion Detection Systems (PIDSs).
- We identified **9 key shortcomings** in their **evaluation** and real-world **applicability**.
- We demonstrate that a **much simpler** neural network outperforms all baselines while **addressing** all 9 shortcomings.

Framework

- A pipeline of 5 restart-able tasks.
- 8 SOTA systems / 9 DARPA datasets.
- 11 encoders / 6 decoders / 9 objectives.
- Integrated hyperparameter tuning.
- Open-sourced



Pipeline comprising 5 tasks, customizable via arguments in YML files

Studied Systems

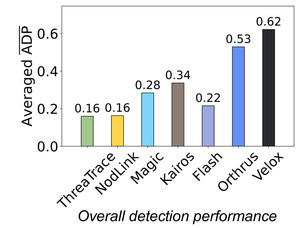
SIGL [1]	USENIX Sec'21
Kairos [2]	IEEE S&P'24
ThreaTrace [3]	IEEE TIFS'22
Flash [4]	IEEE S&P'24
NodLink [5]	NDSS'24
R-Caid [6]	IEEE S&P'24
MAGIC [7]	USENIX Sec'24
Orthrus [8]	USENIX Sec'25

Shortcomings

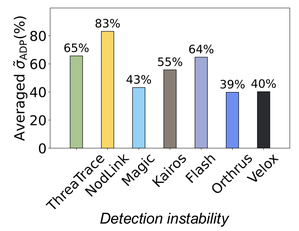
1. **Insufficient Detection Granularity**
   Systems detect at the graph or neighborhood levels, leading to thousands positives to analyze.
2. **Missing Metric to Measure Attack Detection**
   Traditional metrics don't account individual attacks and are biased toward thresholding.
3. **Impractical Thresholding Methods**
   Systems rely on fixed, arbitrary and manually set thresholds that fail to adapt dynamically.
4. **Unfair Comparison with Baselines**
   Evaluation baselines are left untuned, while proposed systems are typically extensively tuned.
5. **Not Measuring Instability**
   The detection performance of systems is extremely instable under identical configurations.
6. **Featurization Methods Trained on Test Data**
   Some systems rely on test data for training, leading to data snooping.
7. **Overly Complex Architectures**
   Systems are usually complex, yet they are rarely compared to much simpler models.
8. **Insufficient Scalability**
   Systems do not meet scalability and overhead requirements for a practical deployment.
9. **Lacking Real-Time Detection**
   Systems are poorly fitted for real-time setting due to their design and overhead.

Key Contributions

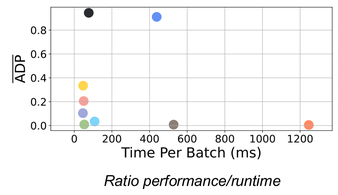
1. We introduce **Attack Detection Precision (ADP)** as a new metric to measure detection capability.
2. We show that all systems have high **detection instability** using relative ADP standard deviation.
3. Systems tend to become more complex, whereas **Velox**, a simple neural network on text features, surprisingly reaches SOTA on 8/9 datasets.



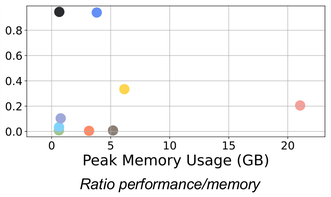
Overall detection performance



Detection instability



Ratio performance/runtime



Ratio performance/memory

Recommendations

- Use fine-grained evaluation methods such as node- or edge-level detection.
- Use cybersecurity-oriented metrics.
- Tune baselines fairly and consistently.
- Evaluate instability through repeated runs.
- Use ablations to find the simplest effective design.
- Ensure scalability and real-time capability.

References

[1] Han, Xueyuan, et al. "{SIGL}: Securing software installations through deep graph learning." 30th USENIX Security Symposium (USENIX Security 21). 2021.
[2] Wang, Su, et al. "Threatrace: Detecting and tracing host-based threats in node level through provenance graph learning." IEEE Transactions on Information Forensics and Security 17 (2022): 3972-3987.
[3] Li, Shaofei, et al. "Nodlink: An online system for fine-grained apt attack detection and investigation." arXiv preprint arXiv:2311.02331 (2023).
[4] Jia, Zian, et al. "{MAGIC}: Detecting advanced persistent threats via masked graph representation learning." 33rd USENIX Security Symposium (USENIX Security 24). 2024.
[5] Cheng, Zijun, et al. "Kairos: Practical intrusion detection and investigation using whole-system provenance." 2024 IEEE Symposium on Security and Privacy (SP). IEEE, 2024.
[6] Rehman, Mati Ur, Hadi Ahmadi, and Wajih Ul Hassan. "Flash: A comprehensive approach to intrusion detection via provenance graph representation learning." 2024 IEEE Symposium on Security and Privacy (SP). IEEE, 2024.
[7] Goyal, Akul, Gang Wang, and Adam Bates. "R-caid: Embedding root cause analysis within provenance-based intrusion detection." 2024 IEEE Symposium on Security and Privacy (SP). IEEE, 2024.
[8] Baoxiang Jiang, Tristan Bilot, et al. ORTHRUS: Achieving High Quality of Attribution in Provenance-based Intrusion Detection Systems. In *Security Symposium (USENIX Sec'25)*. USENIX, 2025.