LogCraft: Crafting CVE-Aware Synthetic Worlds (Logs)

Kai-Xian Wong[1]*, Chan-Jien Tan[1]*, Yi-Ting Huang[2], Ying-Ren Guo[3],
Yu-Zih Jheng[4], Guo-Wei Wong[4], and Meng Chang Chen[3]

[1]Universiti Malaya, [2]National Taiwan University of Science and Technology,
[3]Research Center for Information Technology Innovation, Academia Sinica, [4]National Taiwan University,

ACM CCS 2025

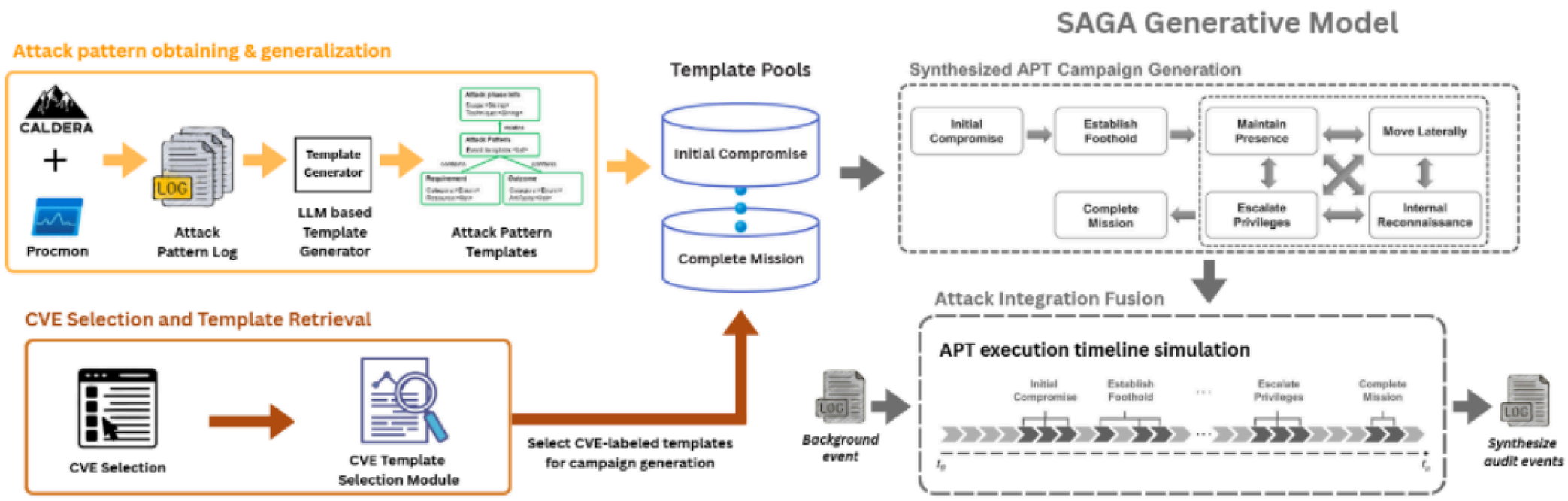
Overview

This work presents *SAGA [1]*, a framework that generates synthetic, labeled audit logs by (1) aligning with MITRE ATT&CK, (2) capturing realistic APT campaign behaviors, and (3) supporting both rule-based and ML-based detection research. This overcomes the limitations of existing datasets that lack diversity, scalability, and rich labeling.

We further extend SAGA by (1) simulating additional lateral movement techniques in realistic multi-host Active Directory settings and (2) introducing CVE-driven log synthesis using LLMs to link vulnerabilities with attack traces. These enhancements improve the fidelity, automation, and scalability of synthetic log generation for training and evaluating detection systems.

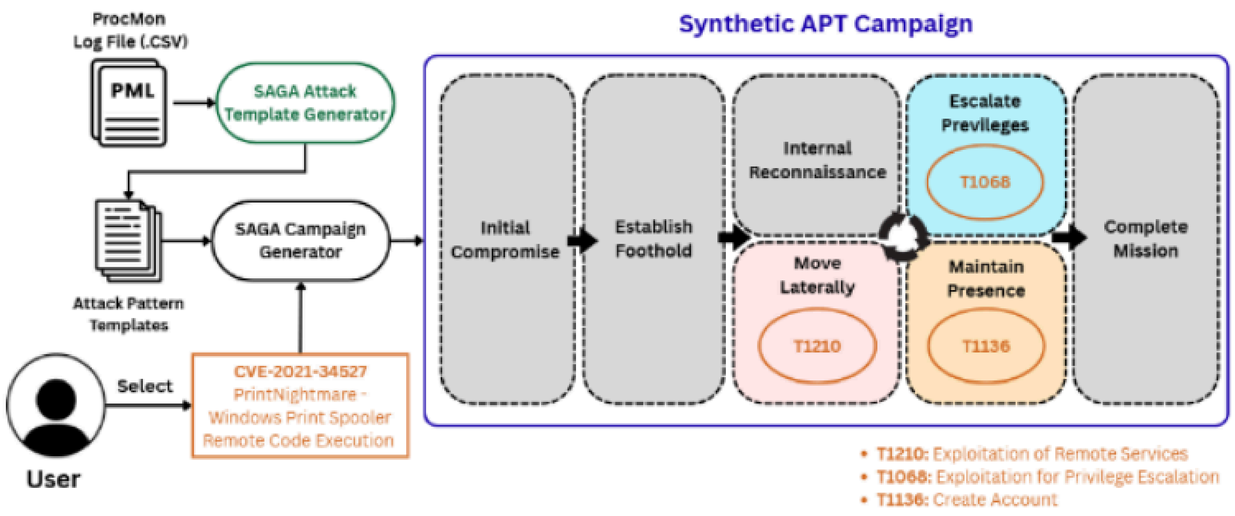
Methodology

Prior Work: SAGA generates synthetic, labeled audit logs aligned to MITRE ATT&CK for simulating APT campaigns.
Extensions: (1) Attack coverage expansion, (2) Template automation, (3) CVE-based log generation.



Case Study

CVE-2021-34527 : PrintNightmare



- CVE is mapped to MITRE ATT&CK techniques using the *Mapping Explorer [2]* methodology, supported by an LLM for consistent mappings.
- When a CVE is selected, SAGA inserts CVE-labeled templates into the matching lifecycle stages.
- Any stages not covered by the CVE are automatically populated with other SAGA templates to produce a coherent APT campaign.

Expected Deliverables

- **Security Research** – Produce realistic, CVE-driven audit logs for evaluating and comparing detection systems.
- **Machine Learning** – Provide structured, labeled datasets to train and benchmark models on adversarial behavior.
- **Defensive Training** – Support blue team exercises with reproducible, evolving attack campaigns.

Future Work

- Continuous CVE integration through automated CVE-to-ATT&CK mapping and simulation with LLM agents.
- Enhanced template generation by improving event filtering and exploring multi-agent LLM pipelines.
- Scalable adversary emulation extended to larger, heterogeneous environments.

Reference

[1] Yi-Ting Huang, Ying-Ren Guo, Yu-Sheng Yang, Guo-Wei Wong, Yu-Zih Jheng, Yeali Sun, Jessemyn Modini, Timothy Lynar, and Meng Chang Chen. 2024. SAGA: Synthetic Audit Log Generation for APT Campaigns. arXiv preprint arXiv:2411.13138 (2024).

[2] Center for Threat-Informed Defense. 2024. Mapping CVEs to ATT&CK Techniques.https://center-for-threat-informed-defense.github.io/mappings-explorer/about/.

[3] CVE Program. 2024. cvelist GitHub Repository. https://github.com/CVEProject/cvelistV5.