分布式带有溯源的时序图学习在供应链中用于 APT 检测

Zhuoran Tan
University of Glasgow, UK
z.tan.1@research.gla.ac.uk

Christos Anagnostopoulos
University of Glasgow, UK
Christos.Anagnostopoulos@glasgow.ac.uk

Jeremy Singer
University of Glasgow, UK
jeremy.singer@glasgow.ac.uk

摘要—网络供应链，包括数字资产、软件和硬件，已成为现代信息技术（ICT）提供的重要组成部分。然而，日益增长的相互依赖性引入了许多攻击途径，使供应链成为主要的利用目标。特别是，高级持续威胁（APTs）经常利用供应链漏洞（SCVs）作为入口点，得益于其固有的隐蔽性。当前的防御策略主要侧重于通过区块链保证完整性或使用开源软件（OSS）中的明文源代码分析进行检测来预防。然而，这些方法忽略了在无法获取源代码的情况下，以及未能解决运行时的检测和防御问题。为弥补这一差距，我们提出了一种新方法，该方法整合多源数据，构建全面动态的来源图，并使用时间图学习实时检测 APT 行为。鉴于行业和学术界缺乏量身定制的数据集，我们也旨在通过重现实世界的供应链漏洞利用来模拟一个自定义数据集，同时进行多源监控。

I. 介绍与动机

自从 2020 年的 Solarwind 攻击报告 [1] 发布以来，该攻击影响了多家大型企业和数百万最终用户，利用第三方供应链软件发起攻击的趋势日益增长，包括勒索软件 [2] 和 APT 攻击 [3]。这种利用也扩展到了人工智能供应链，涵盖了预训练模型、数据集和依赖库 [4]。为了应对这些威胁，大多数现有的解决方案集中在源代码层面的检测或采用预防性方法。常见的方法包括通过程序分析 [5]、图挖掘 [6] 以及机器学习或深度学习技术 [7] 来识别恶意代码或功能。预防策略通常利用区块链框架 [8] 引入额外的身份验证层以检测潜在的篡改。然而，这些方法假设第三方源代码易于获取用于评估和基准测试。实际上，由于知识产权保护的原因，这种情况很少发生，这限制了现有防御机制的应用范围。

当前关于 APT 检测的主要工作依赖于基于来源的框架，其中系统行为被表示为从源数据（如审计日志）中提取实体组成的动态图 [9]。这些方法通常设计用于单一数据源，并使用反映一般 APT 活动的标准基准数据集 [10] 进行评估。然而，通过 SCVs 的利用遵循一条独特的攻击链，这可能无法被针对通用 APT 设计的检测方法有效捕捉到。

为了解决当前的挑战，我们提出以下研究问题 (RQs)：

a) **研究问题 *1*:** 恶意利用 SCVs 的 APT 有哪些独特特征？

b) **研究问题 *2*:** 如何在大规模源数据中准确检测此类 APT？

c) **研究问题 *3*:** 如何实现实时高效检测并持续适应新兴威胁？

II. 背景及相关工作

利用 SCV 的 APT 攻击将供应链作为入口点，并采用多阶段攻击策略，使其在早期更难以被发现和识别，从而最大化其影响。SCV 已被观察到针对广泛使用的开源软件仓库，如 PyPi、NPM 和 Ruby。此外，对 Hugging Face 等平台上托管的预训练模型日益依赖已经引入了 AI 供应链中的新攻击向量。一旦这些库和模型遭到破坏，它们可以将恶意负载传播到下游，可能影响数百万最终用户和设备。

SCV 的检测工作主要集中在识别源代码层面的漏洞。Tran 等人 [11] 提出了一个基于 BERT 的架构，该架构利用自注意力机制来检测 Python 代码中的漏洞。Lu 等人 [7] 结合了图形结构信息和上下文学习以增强基于大型语言模型（LLM）的漏洞检测，在此框架中根据语义、词汇和句法相似性识别相关的代码示例。然而，这些方法有一个显著的局限性：它们假设可以访问源代码，但在现实世界的情景中这并不总是可行的。

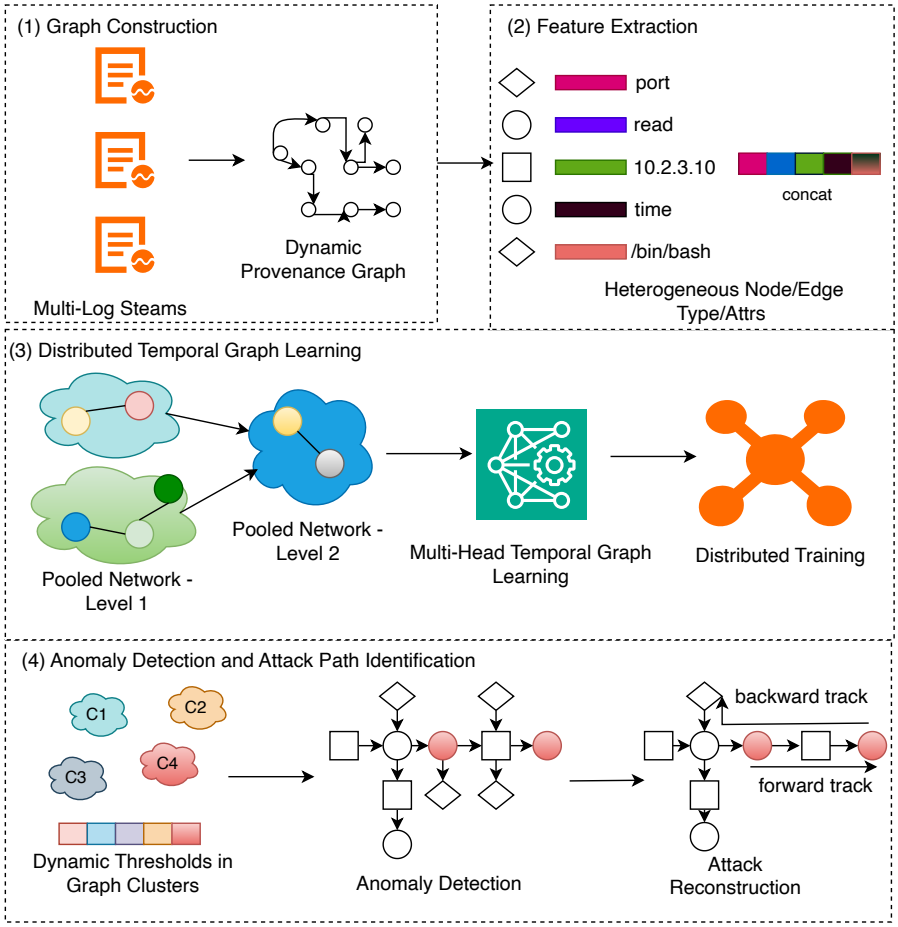

图 1. 利用供应链漏洞的 APT 检测框架

APT 检测侧重于通过关联来自多个数据源的指标来识别利用链，旨在重建攻击路径并追溯其根本原因 [12]。在现有的方法中，数据血缘关系已被广泛采用以关联基于时间线的数据源并捕捉系统活动 [13, 14]，特别是在分布式设置中 [15]。然而，这些研究中常用的数据库并不能有效表示供应链漏洞的利用。此外，大多数检测方法依赖单一类型的数据源，这通常不足以识别复杂和高级的 APT。

III. 方法论与实验

为了解决**研究问题 1**，我们进行了一项全面的调查 [16]，其中包括对比通用 APT 检测技术和专门针对利用 SCV 的 APT 技术的统计分析。研究结果揭示了 APT 在 SCV 中特有的攻击链。

对于**研究问题 2**，我们首先模拟了一个数据集，该数据集捕获了运行时的通用 OSS 利用行为 [17]。此数据集包括来自 npm、PyPI、crates.io、NuGet 和 Packagist 等生态系统中的 9,461 份报告，通过分析动态行为模式，提供了对运行时利用情况的洞察。此外，我们还模拟了第二个数据集，该数据集模拟了 Azure 云上的现实世界高级供应链攻击。此数据集重点关注初始触发后的多阶段利用，填补了用于研究 SCVs 中 APT 的数据集空白。

为了将多源数据处理成动态的综合来源图，我们开发了可扩展工具 UTLParser[18]，该工具旨在将多种结构化数据解析为时间来源图。这些图作为后续检测方法的基础。

我们的当前检测方法遵循两个方向。第一个利用基于注意力的图神经网络从数据集 [17] 中提取关键特征。第二个探索使用时间图模型进行精确攻击检测和重建，利用第二个数据集。为了提高检测准确性，我们实现动态阈值评分以识别并重建最关键的攻击路径，有效减少

误报。

对于**研究问题 3**，我们计划实施分布式学习技术以加速图数据集中的采样，并有效地将训练任务分配到多个集群或机器上。此外，在构建图时，我们的目标是修剪不必要的结构以优化消息传递并减少计算开销。

关于持续模型更新，我们提出了将诸如弹性权重固化（EWC）等技术整合到持续学习过程中 [19]。该方法在模型更新期间缓解了灾难性遗忘问题，使模型能够适应新识别的威胁。

如图 1所示，我们的方法包含四个关键组件：图形构建、特征提取、分布式时间图学习和带有攻击重构的异常检测。此框架有效解决了识别出的挑战并克服了现有解决方案的局限性。

IV. 评估和讨论

为了评估我们的检测框架的性能，我们将使用两个专注于 SCVs 和基于 SCVs 的 APT 的模拟数据集。此外，我们还将使用一个通用数据集 [10] 进行基准测试。为进一步评估实际性能，我们计划实施红队行动，采用高级利用技术，评估现实场景中的检测效率和准确性。

我们预计会遇到与实时图构建和分布式训练相关的挑战，原因是数据量庞大。为减轻存储和计算负担，我们计划应用聚类和剪枝技术来过滤冗余信息并减小图的规模，同时保留有效的检测所需的基本结构。

参考文献

[1] S. Ozarslan, "Tactics, techniques, and procedures (ttps) used in the solarwinds breach," 2020. [Online]. Available: https://www.picussecurity.com/resource/blog/ttps-used-in-the-solarwinds-breach

[2] Mandiant, "Special report: Mandiant m-trends 2024," 2024. [Online]. Available: https://services.google.com/fh/files/misc/m-trends-2024.pdf

[3] European Union Agency for Cybersecurity., *ENISA Threat Landscape for Supply Chain Attacks.* LU: Publications Office, 2021.

[4] W. Jiang, N. Synovic, R. Sethi, A. Indarapu, M. Hyatt, T. R. Schorlemmer, G. K. Thiruvathukal, and J. C. Davis, "An Empirical Study of Artifacts and Security Risks in the Pre-trained Model Supply Chain," in *Proceedings of the 2022 ACM Workshop on Software Supply Chain Offensive Research and Ecosystem Defenses.* Los Angeles CA USA: ACM, 2022, pp. 105–114.

[5] P. Wang, S. Liu, L. Ai, and W. Jiang, "Detecting security vulnerabilities with vulnerability nets," *Journal of Systems and Software*, vol. 208, pp. 111 902–111 902, 2024.

[6] C. Liang, Q. Wei, Z. Jiang, Y. Wang, and J. Du, "A source code vulnerability detection method based on adaptive graph neural networks," in *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering Workshops*, ser. ASEW '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 187 – 196.

[7] G. Lu, X. Ju, X. Chen, W. Pei, and Z. Cai, "Grace: Empowering llm-based software vulnerability detection with graph structure and in-context learning," *Journal of Systems and Software*, vol. 212, p. 112031, 2024.

[8] M. D. Islam, H. Shen, and S. Badsha, "Integrating blockchain into supply chain safeguarded by puf-enabled rfid," *Internet of Things*, vol. 18, p. 100505, 2022.

[9] T. Poštuvan, C. Grohnfeldt, M. Russo, and G. Lovisotto, " Learning-Based Link Anomaly Detection in Continuous-Time Dynamic Graphs," May 2024, arXiv:2405.18050 [cs]. [Online]. Available: http://arxiv.org/abs/2405.18050

[10] R. Arantes, C. Weir, H. Hannon, and M. Kulseng, "Operationally transparent cyber (optc)," 2021. [Online]. Available: https://dx.doi.org/10.21227/edq8-nk52

[11] H.-C. Tran, A.-D. Tran, and K.-H. Le, "DetectVul: A statement-level code vulnerability detection for Python," *Future Generation Computer Systems*, p. 107504, Sep. 2024.

[12] B. Jiang, T. Bilot, N. El Madhoun, K. Al Agha, A. Zouaoui, S. Iqbal, X. Han, and T. Pasquier, "ORTHRUS: Achieving High Quality of Attribution in Provenance-based Intrusion Detection Systems," in *Security Symposium (USENIX Sec'25).* USENIX,

2025.

[13] C. Xiong, T. Zhu, W. Dong, L. Ruan, R. Yang, Y. Cheng, Y. Chen, S. Cheng, and X. Chen, "A Practical Real-Time APT Detection System With High Accuracy and Efficiency," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 1, pp. 551–565, 2022.

[14] A. A. Alsaheel, Y. Nan, S. Ma, L. Yu, G. Walkup, Z. B. Celik, X. Zhang, and D. Xu, "Atlas: A sequence-based learning approach for attack investigation," in *USENIX Security Symposium*, 2021.

[15] W. Ametepe, C. Wang, S. K. Ocansey, X. Li, and F. Hussain, "Data provenance collection and security in a distributed environment: A survey," *International Journal of Computers and Applications*, vol. 43, no. 1, pp. 11–25, 2021.

[16] Z. Tan, S. P. Parambath, C. Anagnostopoulos, J. Singer, and A. K. Marnerides, "Advanced persistent threats based on supply chain vulnerabilities: Challenges, solutions & future directions," *IEEE Internet of Things Journal*, pp. 1–1, 2025.

[17] Z. Tan, C. Anagnosstopoulos, and J. Singer, "Osptrack: A labeled dataset targeting simulated execution of open-source software," *CoRR*, vol. abs/2411.14829, 2024.

[18] Z. Tan, C. Anagnostopoulos, S. P. Parambath, and J. Singer, "Unified semantic log parsing and causal graph construction for attack attribution," 2024. [Online]. Available: https://arxiv.org/abs/2411.15354

[19] Z. Tan, Q. Wang, C. Anagnostopoulos, A. Marnerides, J. Singer, S. P. Paramba, and S. Temple, "Ledlog: Lightweight explainable real-time dual-model log anomaly detection system," *SSRN*, 2023. [Online]. Available: http://dx.doi.org/10.2139/ssrn.4743487