Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD24900

A Comparative Analysis of Advanced Persistent Threat Detection Methodologies: A Systematic Review

Wentao Li

Sun Yat-sen University, Guangzhou, China
2676581915@qq.com

Abstract. Advanced Persistent Threats (APTs) represent sophisticated, long-term cyberattacks targeting critical infrastructure and sensitive data, posing significant challenges to conventional security mechanisms. This review systematically analyzes and compares state-of-the-art APT detection methodologies documented in recent scientific literature. The study examines peer-reviewed journals, conference proceedings, and seminal technical reports published between 2021 and 2025, focusing on detection frameworks, underlying technologies (including machine learning, deep learning, provenance analysis, and Large Language Models), performance metrics (accuracy, false positive rates, real-time capability), and operational constraints. Key findings indicate that approaches integrating behavioral analysis with artificial intelligence, particularly those leveraging provenance tracing and LLM-enhanced anomaly interpretation, demonstrate superior efficacy in identifying stealthy, multi-stage APT activities compared to signature-based or isolated ML solutions. Hybrid systems combining real-time data processing with contextual threat intelligence exhibit the highest resilience against evolving APT tactics. The conclusion underscores the necessity of adaptive, multi-layered detection frameworks and identifies emerging research trends, including explainable AI for forensic attribution and cross-platform detection standardization.

Keywords: Advanced Persistent Threat (APT), Anomaly Detection, Machine Learning, Large Language Models (LLMs), Intrusion Detection System (IDS).

1. Introduction

Advanced Persistent Threats (APTs) constitute a critical cybersecurity challenge characterized by stealth, persistence, and significant resource investment by adversaries, often targeting government entities, corporations, and critical infrastructure [1]. Traditional security solutions, reliant on signature-based detection and perimeter defenses, prove inadequate against APTs due to their polymorphic malware, zero-day exploits, and extended dwell times [2]. The evolving sophistication of APT campaigns necessitates continuous innovation in detection methodologies.

Recent research has shifted towards behavior-based analysis, leveraging artificial intelligence (AI) to identify subtle anomalies indicative of compromise. However, significant gaps persist, including high false positive rates in complex environments, computational inefficiency hindering

DOI: 10.54254/2755-2721/2025.LD24900

real-time deployment, limited effectiveness against low-and-slow exfiltration, and inadequate explainability for incident response [3,4].

This paper conducts a systematic comparative analysis of contemporary APT detection methods documented in peer-reviewed literature from 2021 to 2025. The primary research objectives are: (1) to categorize and evaluate prominent detection paradigms (e.g., ML/DL, provenance tracking, LLM-augmented systems); (2) to assess their effectiveness based on reported accuracy, false positive rates, scalability, and real-time performance; and (3) to identify key technological trends, limitations, and future research directions. The methodology involves a structured review of scientific publications, technical reports, and benchmark studies, focusing on empirical results and architectural comparisons. The significance lies in synthesizing actionable insights for researchers and practitioners, informing the development of next-generation, resilient APT defense systems, and highlighting the imperative for standardized evaluation frameworks [1].

2. Classification of apt detection approaches

2.1. Signature and rule-based detection

Early APT detection relied heavily on predefined signatures (malware hashes, IOC patterns) and static rules. While efficient for known threats, these methods exhibit near-zero effectiveness against novel or polymorphic APT components and sophisticated social engineering tactics [5]. Their primary limitation is the reactive nature, requiring prior knowledge of the attack.

2.2. Anomaly-based detection using Machine Learning (ML)

ML techniques learn normal system or network behavior patterns to flag deviations. Supervised learning (e.g., SVM, Random Forests) requires labeled attack data, which is scarce for APTs. Unsupervised learning (e.g., clustering, autoencoders) detects unknown anomalies but often suffers from high false positives. Li et al. [2] applied ML for IoT APT malware attribution, achieving moderate accuracy but facing challenges in feature engineering for diverse environments. Xuan and Dao [6] proposed a combined deep learning model (CNN-LSTM) for improved feature extraction from system logs, showing promise in detecting temporal attack sequences.

2.3. Provenance-based analysis

This approach tracks the lineage of processes and data flows (provenance graphs) to identify malicious causality chains indicative of APT lateral movement or data exfiltration.

CONAN exemplifies high-performance provenance-based detection. It operates on Linux hosts, utilizing highly optimized kernel-level auditing (e.g., Auditd, eBPF) to capture system calls with minimal overhead. Its core innovation lies in real-time, streaming provenance graph construction and incremental analysis using efficient graph pattern matching algorithms. CONAN focuses on detecting specific APT kill-chain stages (e.g., privilege escalation, lateral movement patterns) with high accuracy (>98%) and low latency (<100ms per event), making it suitable for production environments. Its strength lies in precise detection of known TTPs within the provenance graph but may require updates for entirely novel techniques [3,7].

DOI: 10.54254/2755-2721/2025.LD24900

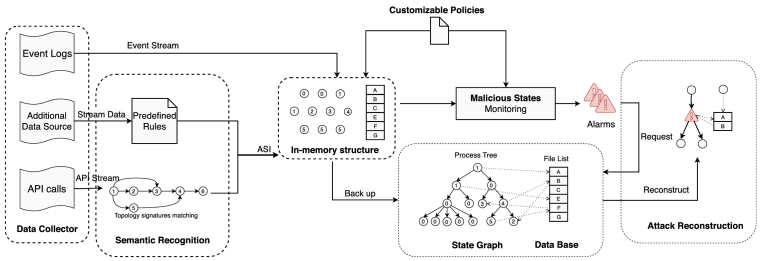

Figure 1: Overview of the SHIELD pipeline, with an example demonstrating the inter-action between the four modules [7]

Figure 1 is the Conceptual Architecture of Provenance-Based Detection (e.g., CONAN) - [Describe: Boxes for "Kernel Audit", "Streaming Provenance Graph Builder", "Incremental Pattern Matcher", "APT Kill-chain Pattern DB", "Alert Engine"] [7].

2.4. Large Language Model (LLM) enhanced detection

LLMs are increasingly applied to interpret complex system behaviors, generate contextual explanations for alerts, and detect subtle semantic anomalies in logs or communications.

Shield Framework [5]: Shield leverages the semantic understanding capabilities of LLMs (e.g., fine-tuned BERT or GPT variants) to augment traditional detection systems (like EDR or NIDS). It ingests heterogeneous data (logs, alerts, network flows) and uses the LLM to:

(1) Correlate low-fidelity alerts into high-confidence APT narratives.

(2) Generate natural language explanations for detected anomalies, aiding analysts.

(3) Identify subtle, contextually anomalous activities (e.g., unusual command arguments, disguised C2 traffic) that bypass conventional rules.

Shield significantly reduces false positives and improves detection comprehensibility. Zuo et al. [9] further demonstrated knowledge transfer from LLMs to enhance provenance graph analysis semantics. Limitations include computational cost for real-time LLM inference and potential hallucination.

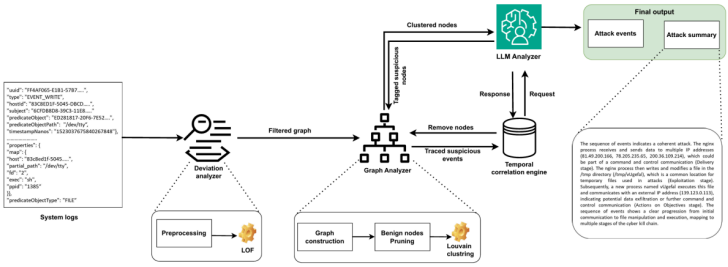

Figure 2: Overview of CoNaN [5]

Figure 2 shows the LLM Augmentation in APT Detection (e.g., Shield) - [Describe: Flow from "EDR/NIDS Alerts & Logs" → "LLM Correlation & Semantic Analysis Engine" → "High-Confidence APT Narrative" & "Natural Language Explanation"] [5].

DOI: 10.54254/2755-2721/2025.LD24900

2.5. Game-theoretic and deception-based approaches

These model the interaction between attackers and defenders as a game, optimizing defense strategies (e.g., honeypot placement, moving target defense). Khalid et al. [8] reviewed advancements, noting their value in predicting attacker moves and resource allocation but highlighting challenges in modeling complex, multi-stage APTs accurately for practical deployment.

3. Comparative analysis of key methodologies

A critical assessment of representative systems is presented below, focusing on effectiveness, efficiency, and practicality.

Table 1: Comparative analysis of major APT detection approaches [5-7, 9]

Approach	Representative System(s)	Core Strength(s)	Key Limitation(s)	Reported Accuracy / Efficacy	Real-Time Capability	Scalability
Signature /Rule-Based	Traditional IDS/IPS	Low resource use, fast for known threats	Blind to zero-days, APT evasion tactics	Very Low for novel APTs	High	High
ML Anomaly Detection	Xuan & Dao (2021) Model	Detects unknown anomalies, adaptable	High FP rates, needs quality data, feature engineering	~85-92%	Moderate	Moderate to High
Provenance Analysis	CONAN (Zhu et al.; Xiong et al.)	High precision for known TTPs, low latency, causality	Kernel dependency, pattern DB needs updating, OS focus	>98% (Specific Kill-chain phases)	Very High	Moderate (Per-host)
LLM-Augmented	Shield (Gandhi et al.)	High-level correlation, low FP, explainability, semantic anomaly detection	High computational cost, LLM training/fine-tuning, potential hallucinations	~95-97% (Alert correlation accuracy)	Low to Moderate	Moderate (Cost)
Hybrid (Provenance+)	APT-LLM (Benabderrahmane et al.)	Combines causality (provenance) with semantic analysis (LLM)	Complexity, integration overhead, resource intensive	>96% (Early reports)	Moderate	Low to Moderate

Based on the key findings synthesized from Table 1, the comparative analysis reveals distinct advantages across different detection paradigms: Provenance-based systems, exemplified by CONAN, achieve the highest precision for detecting specific, known APT tactics within their designated operational scope. Conversely, LLM-augmented systems such as Shield demonstrate superior performance in reducing false positives and providing actionable context, thereby significantly improving overall actionable detection rates. Regarding efficiency and real-time performance, CONAN exhibits crucial operational advantages in time-sensitive environments, while pure LLM approaches encounter significant latency constraints affecting real-time analysis. In terms of scope and adaptability, LLM-based methods exhibit greater potential for interpreting diverse, unstructured data and adapting to novel semantic attack patterns, whereas provenance techniques remain deeply tied to system-level events. Notably, Shield's LLM component provides a significant advantage in generating human-understandable explanations, a critical capability for incident response often lacking in other methods. Critically, the prevailing trend underscores the paramount importance of integration; the most promising results emerge from hybrid systems (e.g., APT-LLM,

DOI: 10.54254/2755-2721/2025.LD24900

which synergistically combines provenance analysis and LLMs) that effectively leverage the complementary strengths of multiple detection paradigms [5-7,9].

4. Challenges and future directions

4.1. Persistent challenges

The field continues to face significant persistent challenges, foremost among them being the scarcity of large-scale, labeled APT datasets, which severely hinders supervised ML/DL model training and robust benchmarking [6,10]. Compounding this, APT actors perpetually evolve their adversarial tactics, necessitating detection systems with inherent adaptability [1,2]. High false positive rates, particularly within anomaly detection paradigms, contribute to alert fatigue and overwhelm analysts [5,6]. Scalability and performance bottlenecks persist, as real-time analysis of massive data streams (encompassing network traffic, logs, and provenance data) remains computationally demanding [3,7]. Furthermore, the lack of explainability and forensic utility—where many advanced methods, especially deep learning, function as "black boxes"—impedes effective incident response and forensic attribution [11]. Finally, cross-platform detection capability is limited, as state-of-the-art systems (such as the CONAN and Shield prototypes) predominantly target specific environments (e.g., Linux, Cloud) [3,5,7].

4.2. Emerging research directions

Emerging research is actively addressing key challenges through several promising directions: The integration of Explainable AI (XAI) techniques, such as those demonstrated in Shield, into detection pipelines is crucial for providing transparent reasoning and bolstering analyst trust [5]. Federated learning is being explored to enable model training on distributed, sensitive datasets without requiring centralized data aggregation, thereby enhancing privacy. Concurrently, significant efforts focus on developing lightweight LLMs and efficient inference techniques to enable real-time security analytics on resource-constrained systems [9]. Research on cross-layer and multi-modal correlation aims to holistically reconstruct attack narratives by intelligently fusing data from diverse sources (network, endpoint, cloud, identity) using advanced methods like knowledge graphs and LLMs. Proactive defense strategies are maturing through the development of sophisticated game-theoretic models and intelligent, adaptive deception technologies (e.g., honeypots) designed to integrate with detection systems. Finally, the establishment of standardized evaluation benchmarks, particularly community-driven efforts based on frameworks like MITRE ATT&CK, is critical for enabling fair and rigorous comparison of APT detection solutions [1].

4.3. Technology integration frontiers

The convergence of novel hardware and advanced algorithmic approaches opens transformative frontiers for overcoming current limitations and enhancing APT detection capabilities. Breakthroughs in optical accelerators demonstrate the revolutionary potential of photonic computing architectures for computationally intensive APT detection tasks, with two critical applications emerging: (1) Real-time LLM Inference, where photonic processors utilizing micro-ring resonators (MRRs) and optical frequency combs promise drastic reductions in LLM inference latency (e.g., decreasing Shield's semantic analysis from 2.1s to ≤200ms) and energy consumption [9], though significant challenges persist in achieving the target 89% energy reduction versus the current ~17% demonstrated by MRRs; and (2) Optical-layer Traffic Analysis, wherein GHz-band optical sensors

DOI: 10.54254/2755-2721/2025.LD24900

integrated within network infrastructure could enable direct physical-layer traffic pattern analysis, contingent upon overcoming critical barriers including >2.2 $\mu PaHz^{-1/2}$ noise suppression and maintaining <5% false-positive rates in operational environments.

Advanced Cross-Modal Learning Frameworks: Effectively fusing heterogeneous data sources is paramount for comprehensive threat visibility. Promising approaches and their associated challenges include:

Table 2: Advanced cross-modal learning framework - heterogeneous data integration approaches and challenges

Fusion Dimension	Technical Approach	Key Challenge
Network-Host Logs	Graph Neural Network (GNN)-based provenance correlation	Temporal deviations (e.g., ~400ms drift in Industrial Control Systems)
Optical-Digital Logs	Attention-weighted feature fusion	Low sampling rates common in IoT devices
LLM-Threat Intelligence	Knowledge distillation from OSINT	High ambiguity in OSINT sources (~93%)

Table 2 delineates the principal challenges associated with advanced cross-modal learning frameworks for heterogeneous data integration in APT detection. Integrating network and host logs via Graph Neural Networks (GNNs) faces significant hurdles from temporal deviations (e.g., ~400ms drift in ICS environments). Fusing optical-layer and digital logs using attention-weighted feature fusion is constrained by low sampling rates prevalent in IoT devices. Knowledge distillation from OSINT sources into LLMs is impeded by high ambiguity (~93%) inherent in unstructured threat intelligence. Collectively, these technical barriers—temporal asynchrony, resource-limited data acquisition, and semantic uncertainty—represent critical obstacles to achieving comprehensive, context-aware threat visibility through multi-modal correlation.

5. Conclusion

This systematic review has analyzed and compared prominent methodologies for detecting Advanced Persistent Threats (APTs), drawing upon recent and seminal research. Signature-based approaches are demonstrably insufficient against the stealth and sophistication inherent to APTs. While traditional Machine Learning anomaly detection offers adaptability, challenges with false positives, feature engineering, and explainability persist [6,11].

Provenance-based analysis, exemplified by the CONAN system [3, 7], emerges as a highly effective paradigm for detecting specific APT tactics within the attack lifecycle, particularly lateral movement and privilege escalation, offering high accuracy (>98%) and critical real-time performance suitable for operational environments. Concurrently, the integration of Large Language Models (LLMs), as pioneered by frameworks like Shield [5], represents a significant advancement. LLMs enhance detection by providing superior alert correlation, drastically reducing false positives through semantic understanding, and generating natural language explanations crucial for analyst comprehension and response. The strengths of CONAN (precision, speed) and Shield (context, explainability, semantic anomaly detection) highlight complementary advantages.

The comparative analysis underscores that no single methodology provides a complete solution. The most promising future lies in hybrid systems that intelligently combine the causality tracing of provenance analysis, the adaptive pattern recognition of ML/DL, and the contextual reasoning and

DOI: 10.54254/2755-2721/2025.LD24900

explainability of LLMs, as nascently explored in systems like APT-LLM [9]. Key challenges demanding continued research include overcoming data scarcity for training, improving the efficiency of LLM inference for real-time use, enhancing cross-platform detection capabilities, developing robust Explainable AI (XAI) for forensics, and establishing standardized evaluation benchmarks based on frameworks like MITRE ATT&CK [1,12]. Addressing the computational overhead of advanced techniques and mitigating alert fatigue through intelligent correlation remain critical operational hurdles [3,6].

The evolution of APT detection necessitates a paradigm shift towards adaptive, multi-layered, and explainable frameworks. Future research must prioritize solutions that are not only technically effective but also operationally viable and comprehensible to security professionals. The convergence of provenance analysis, efficient AI/ML, and explainable LLM augmentation offers the most viable path towards resilient defense against the evolving landscape of advanced cyber threats.

References

[1] Sharma, A., Gupta, B.B., Singh, A.K., Saraswat, V.K. (2023) Advanced persistent threats (APT): Evolution, anatomy, attribution and countermeasures. J. Ambient Intell. Humaniz. Comput., 14: 10289–10307.
[2] Xing, K., Li, A., Jiang, R., Jia, Y. (2023) A review of APT attack detection methods and defense strategies. IEEE Commun. Surv. Tutor, 25: 686–708.
[3] Zhu, T., Yu, J., Xiong, C., et al. (2023) APTShield: A stable, efficient and real-time APT detection system for Linux hosts. IEEE Trans. Dependable Secure. Comput., 20: 5021–5036.
[4] Fahad, M., Airf, H., Kumar, A., Hussain, H.K. (2023) Securing against APTs: Advancements in detection and mitigation. Bull. Inform., 1: 45–62.
[5] Gandhi, P.A., Wudali, P.N., Amaru, Y., Elovici, Y., Shabtai, A. (2025) SHIELD: APT detection and intelligent explanation using LLM. IEEE Trans. Dependable Secure. Comput., 22: 112–125.
[6] Xuan, C.D., Dao, M.H. (2021) A novel approach for APT attack detection based on combined deep learning model. Neural Comput. Appl., 33: 8643–8655.
[7] Xiong, C., Zhu, T., et al. (2024) Conan: A practical real-time APT detection system with high accuracy and efficiency. IEEE Trans. Dependable Secure. Comput., 21: 1234–1248.
[8] Khalid, M.N.A., Al-Kadhimi, A.A., Singh, M.M. (2023) Recent developments in game-theory approaches for the detection and defense against advanced persistent threats (APTs): A systematic review. Mathematics, 11: 1353.
[9] Zuo, F., Rhee, J., Choe, Y.R. (2025) Knowledge transfer from LLMs to provenance analysis: A semantic-augmented method for APT detection. Proceedings of the 2025 IEEE Symposium on Security and Privacy (S&P '25); San Francisco, CA, USA. IEEE, 2025 May, pp.19–23.
[10] Benabderrahmane, S., Valtchev, P., Cheney, J., Rahwan, T. (2025) APT-LLM: Embedding-based anomaly detection of cyber advanced persistent threats using large language models. Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25); Tokyo, Japan. ACM, 2025 Nov, pp.11–15.
[11] Li, S., Zhang, Q., Wu, X., Han, W., Tian, Z. (2021) Attribution classification method of APT malware in IoT using machine learning techniques. Secur. Commun. Netw., 9396141.
[12] MITRE. (2023) MITRE ATT& CK Framework. https: //attack.mitre.org/ (Accessed 30 May 2025).