© APR 2026 | IRE Journals | Volume 9 Issue 10 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I10-1717535

Evolutionary Computation-Enhanced White-Box Penetration Testing for Lateral-Movement Prevention in Zero-Trust Architectures

MARCELO ARAUJO

Abstract- ***Preventing lateral movement remains a central cybersecurity challenge even in environments designed according to Zero Trust principles. Although this paradigm reduces implicit trust and enforces continuous verification, its effectiveness ultimately depends on access-policy quality, identity-to-resource segmentation, and the ability to detect abusive chains built from seemingly legitimate permissions. In parallel, recent automated penetration-testing research has advanced through reinforcement learning, graph-based modeling, and simulation frameworks for exploring complex attack surfaces [1-4]. Building on this state of the art, this article proposes a conceptual white-box penetration-testing framework for Zero-Trust architectures in which evolutionary algorithms perform global search over the internal blueprint of the environment, while reinforcement learning adaptively refines promising action sequences. The model assumes authorized defensive access to ZTNA policies, identity and privilege graphs, workload dependencies, and continuous-verification logs. Its fitness function is multi-objective and jointly considers success probability, stealth, and evasion rate. We argue that this combination may improve the identification of plausible lateral-movement routes and generate more useful remediation outputs, provided that it is applied in controlled environments with telemetry sufficiently faithful to the real system.***

Keywords: Automated Penetration Testing, Evolutionary Computation, Reinforcement Learning, Lateral Movement, Zero Trust.

I. INTRODUCTION

Zero Trust architecture emerged as a response to the limits of perimeter-centric security. Instead of assuming that users, devices, or internal flows are trustworthy simply because they are “inside” the network, the model relies on continuous authentication, least privilege, contextual evaluation, and fine-grained segmentation [12-14]. Still, adopting Zero Trust principles does not automatically eliminate lateral movement. In practice, identity-modeling flaws, inherited permissions, operational exceptions, and workload dependencies can allow an attacker to move internally without breaching a traditional perimeter boundary. For that reason, the relevant question is not merely whether an organization “uses Zero Trust,” but whether its policies actually prevent the sequential composition of individually acceptable access decisions that, when chained together, create an unsafe adversarial route [10-14].

This problem becomes clearer when lateral movement is represented as a graph. Prior work shows that authentication events, host relations, service-account links, and temporal transitions can be modeled as graph structures able to reveal unlikely or dangerous trajectories [9-11]. Bowman et al. showed that unsupervised graph learning can detect lateral movement in enterprise environments using standard authentication logs [9]. King and Huang reframed the problem as anomalous-edge detection in scalable temporal graphs, reinforcing the usefulness of dynamic structures for modeling intranetwork progression [10]. Khoury et al. extended this line of work with temporal inductive learning, bringing the problem closer to realistic enterprise settings in which topology, accounts, and access relations continuously evolve [11]. In parallel, Zero-Trust-oriented studies such as GAZETA and the edge-computing attack-detection framework indicate that adaptive authentication, trust evaluation, and conditional policies can constrain lateral spread, but only when they are aligned with the actual behavior and dependencies of the environment [13,14].

Within this context, automated penetration testing offers a promising form of offensive validation for defensive purposes. Rather than relying exclusively on manual campaigns, recent literature explores intelligent agents able to learn exploration,

exploitation, pivoting, and goal-reaching strategies in simulated or emulated environments [1-8]. Moreno et al. critically review this trend and show that reinforcement learning and recommender systems are already being used to support autonomy in penetration testing, although important limitations remain regarding generalization, environment realism, and reward design [1]. Wang et al. argue for a unified modeling framework for automated penetration testing, noting that the current diversity of simulators and abstractions still hinders reproducibility and meaningful comparison [3]. Chen et al., in their survey on penetration-path planning, show that the problem is inherently combinatorial and therefore compatible with heuristics, metaheuristics, and hybrid search strategies [4]. Together, these studies suggest that complex security environments require mechanisms able to combine global exploration with local adaptation [1-4].

The framework proposed here is positioned exactly at that intersection. It is white-box because it assumes authorized defensive access to the environment's internal blueprint: ZTNA policies, identity and privilege graphs, trust relations, historical authentication telemetry, continuous-verification logs, and workload dependencies. This assumption should be stated explicitly to avoid ambiguity. The goal is not to emulate a fully blind attacker, but to evaluate how a sufficiently informed adversary could exploit permission combinations, authentication pathways, and operational exceptions in order to move laterally. In Zero-Trust environments, this stance is methodologically coherent because risk often arises not from one catastrophic vulnerability, but from the aggregation of minor policy deviations, excessive privileges, or residual connectivity across segments [12-14].

In the framework, the evolutionary-computation layer is responsible for global attack-path search. Genetic algorithms and differential evolution are appropriate because the problem involves a high-dimensional state space, multiple valid combinations, and a significant risk of local optima [4,15].

Consider $G = (V, E)$ to be the white-box Zero-Trust graph built from ZTNA policies, identity-privilege relations, workload dependencies, and telemetry [9–11,12–14]. Each evolutionary individual π encodes a candidate lateral-movement path

$$\pi = (v_0 \xrightarrow{a_1} v_1 \xrightarrow{a_2} \cdots \xrightarrow{a_k} v_k),$$

where v_i are assets or identities and a_i are transitions via credentials or services [4,15]. The multi-objective fitness is defined as

$$f(\pi) = w_1 P_{\text{success}}(\pi) + w_2 S_{\text{stealth}}(\pi) + w_3 E_{\text{evasion}}(\pi),$$

with components estimated from logs and policy preconditions [9–11,16]. Reinforcement learning then locally refines promising paths [5–8].

Operationally, each individual in the population may encode a candidate chain of transitions involving assets, identities, credentials, tokens, policies, and services. Selection preserves better-performing sequences according to the fitness function; mutation introduces non-trivial alternatives; and recombination enables new compositions between promising subpaths.

This formulation is more precise than the generic claim that the algorithm simply "finds the best attack," because it acknowledges that the result depends on the model, the available telemetry, and the objectives defined by the defender [4,15]. Figure 1 summarizes the conceptual workflow of the proposed white-box framework, showing how authorized Zero-Trust internal data support evolutionary attack-path search, reinforcement learning refinement, and remediation-oriented defensive output.

© APR 2026 | IRE Journals | Volume 9 Issue 10 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I10-1717535

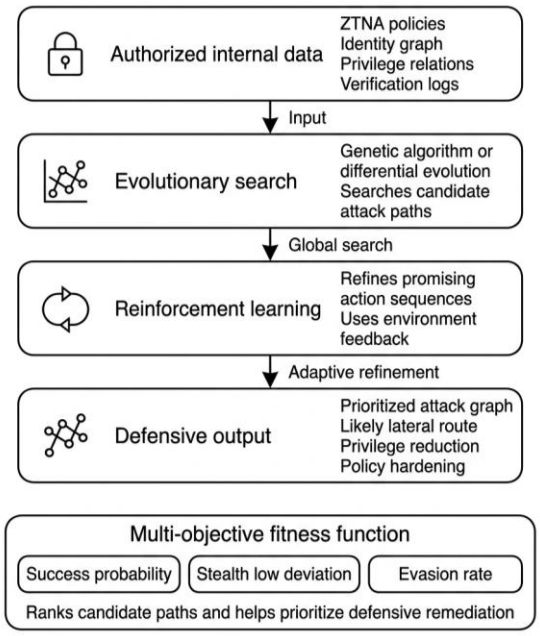

Figure 1. White-Box Penetration Testing Workflow for Lateral-Movement Prevention in Zero-Trust Architectures

Source: Created by author.

The fitness function should be explicitly multi-objective. A model optimized only for shortest path length would tend to favor compact routes, but not necessarily realistic ones. In real offensive operations, the shortest path may be excessively noisy or trivially detected. For that reason, the proposed fitness formulation combines three dimensions. The first is success probability, estimated from access preconditions, the value of obtained privilege, and the sequential feasibility of the chain. The second is stealth, measured through low deviation from behavioral patterns observed in logs, such as low temporal anomaly or limited entropy change in authentication sequences. The third is evasion rate, understood as the estimated ability to pass through controls such as adaptive MFA, UEBA, segmentation, and graph-based detection [9-11,16]. This design is consistent with multi-objective cyber-security optimization studies and avoids the oversimplified assumption that a single criterion is sufficient to capture realistic offensive risk [16].

After population-based exploration, reinforcement learning acts as a local refiner. This division of roles improves methodological coherence. Population-based metaheuristics are effective for broad search, whereas RL is especially useful for adjusting sequences of actions as the environment reacts [1,5-8]. Instead of claiming that the agent "learns everything automatically," it is more accurate to say that it optimizes decision policies within a modeled environment, conditioned by the defined rewards and observed transitions [1,5]. Studies by Ghanem and Chen, Janisch et al., and Becker et al. show that RL agents can learn relevant penetration strategies in simulated environments, although performance strongly depends on environment design, abstraction level, and transferability to unseen scenarios [5-8]. Li et al. further suggests that embedding structural knowledge into reward design can improve learning efficiency, which supports the use of identity, policy, and lateral-movement information during path refinement [7].

An important contribution of this framework, if the article is to remain faithful to the theme and technically credible, is to treat the output as dual: a prioritized attack graph and a set of remediation rules. The first output identifies the most plausible lateral-progression chains according to the defined objectives. The second maps critical edges to concrete policy adjustments such as privilege reduction, removal of transitive trust relations, tighter segmentation between workloads, stronger contextual requirements, and revision of persistent exceptions [12-14]. This formulation is more defensible than asserting that the system "delivers the exact fixes to block every path," because remediation depends on operational trade-offs, business constraints, and governance decisions. The framework should therefore be presented as a technical prioritization instrument rather than as an infallible auto-remediation mechanism.

Important limitations should remain visible in the text to avoid overstatement. First, blueprint fidelity is critical. If policies, identity relations, or service dependencies are incomplete or outdated, discovered paths may reflect the model rather than the real environment. Second, multi-objective optimization over large graphs can be computationally expensive. Third, external validity remains limited: results obtained in simulators or emulated settings should not be automatically generalized to heterogeneous enterprise networks [1,3,4]. The automated-

penetration-testing literature itself acknowledges reproducibility issues, lack of standardized scenarios, and difficulty in comparing approaches [1,3]. These caveats strengthen the paper's credibility because they replace absolute claims with a technically verifiable formulation.

This work contributes a hybrid evolutionary-computation and reinforcement-learning framework for white-box validation of lateral-movement prevention, strengthening existing Zero-Trust architectures. By combining global search over permission graphs with adaptive sequence refinement and a multi-objective fitness function focused on success, stealth, and evasion, the approach generates prioritized attack paths and actionable remediation recommendations. Future work will implement the framework in realistic simulators (e.g., NASimEmu), evaluate it against baselines on synthetic and emulated Zero-Trust environments, and assess transferability to production telemetry. Addressing challenges in blueprint fidelity and computational scalability will further strengthen its practical utility for proactive Zero-Trust policy hardening.

REFERENCES

[1] Moreno AC, Hernandez-Suarez A, Sanchez-Perez G, Toscano-Medina LK, Perez-Meana H, Portillo-Portillo J, Olivares-Mercado J, García Villalba LJ. Analysis of Autonomous Penetration Testing Through Reinforcement Learning and Recommender Systems. Sensors (Basel). 2025;25(1):211.

[2] Lei H, Ge Y, Zhu Q. ADAPT: A Game-Theoretic and Neuro-Symbolic Framework for Automated Distributed Adaptive Penetration Testing. arXiv [Preprint]. 2024:2411.00217.

[3] Wang Y, Liu S, Wang W, Zhou C, Zhang C, Jin J, Zhu C. A Unified Modeling Framework for Automated Penetration Testing. Comput Secur. 2026;162:104787.

[4] Chen Z, Kang F, Xiong X, Shu H. A Survey on Penetration Path Planning in Automated Penetration Testing. Appl Sci. 2024;14(18):8355.

[5] Ghanem MC, Chen TM. Reinforcement Learning for Intelligent Penetration Testing. In: 2018 World Conference on Smart Trends in Systems, Security and Sustainability. 2018. p. 185-192.

[6] Janisch J, Pevný T, Lisý V. NASimEmu: Network Attack Simulator & Emulator for Training Agents Generalizing to Novel Scenarios. In: Machine Learning and Knowledge Discovery in Databases. Cham: Springer; 2023. p. 565-582.

[7] Li Y, Dai H, Yan J. Knowledge-Informed Auto-Penetration Testing Based on Reinforcement Learning with Reward Machine. In: 2024 International Joint Conference on Neural Networks. 2024.

[8] Becker N, Reti D, Ntagiou EV, Wallum M, Schotten HD. Evaluation of Reinforcement Learning for Autonomous Penetration Testing using A3C, Q-learning and DQN. arXiv [Preprint]. 2024:2407.15656.

[9] Bowman B, Laprade C, Ji Y, Huang HH. Detecting Lateral Movement in Enterprise Computer Networks with Unsupervised Graph AI. In: Recent Advances in Intrusion Detection. Cham: Springer; 2020. p. 257-268.

[10] King IJ, Huang HH. Euler: Detecting Network Lateral Movement via Scalable Temporal Link Prediction. ACM Trans Priv Secur. 2023;26(3):1-30.

[11] Khoury J, Klisura D, Zand H, De La Torre Parra G, Najafirad P, Bou-Harb E. Jbeil: Temporal Graph-Based Inductive Learning to Infer Lateral Movement in Evolving Enterprise Networks. In: 2024 IEEE Symposium on Security and Privacy. 2024. p. 3644-3660.

[12] Gambo ML, Almulhem A. Zero Trust Architecture: A Systematic Literature Review. J Netw Syst Manage. 2026;34(1):25.

[13] Ge Y, Zhu Q. GAZETA: GAme-Theoretic ZEro-Trust Authentication for Defense Against Lateral Movement in 5G IoT Networks. IEEE Trans Inf Forensics Secur. 2024;19:540-554.

[14] Sedjelmaci H, Ansari N. Zero Trust Architecture Empowered Attack Detection Framework to Secure 6G Edge Computing. IEEE Netw. 2024;38(1):196-202.

[15] Alhomidi MA, Reed MJ. A Genetic Algorithm Approach for the Most Likely Attack Path Problem. In: 2013 International Conference on Availability, Reliability and Security. 2013.

[16] Khouzani MHR, Liu Z, Malacaria P. Scalable min-max multi-objective cyber-security optimisation over probabilistic attack graphs. Eur J Oper Res. 2019;278(3):894-903.