

Generative AI-Powered Cybersecurity Digital Twin for Proactive Defence

Shrusti Panigrahi, Department of CSE, GIET University, 24cseaiml333.srustipanigrahi@giet.edu
Vamsi Krishna Patnaik, Department of CSE, GIET University, 24cseaiml137.vamsikrishnapatnaik@giet.edu
Smruti Suman Badjena, Lecturer, Department of CSE, GIET University, smrutisumanbadjena@giet.edu
Sanket Nayak, Department of CSE, Lecturer, GIET University, sanketnayak@giet.edu

Abstract: Currently, enterprise cybersecurity remains more reactive than proactive, where security testing and red team exercises take place following an attack or any other security breach. In this paper, we suggest the Generative AI-Powered Cybersecurity Digital Twin (GenCDT), which could assist enterprises in identifying possible risks and dealing with them even before their impact on real systems becomes evident. In contrast to other digital twins, which are predominantly based on analysis of previously happened events, our proposal will generate realistic cyberattack scenarios in real time. Following each iteration of such a scenario, the AI engine will analyse the outcome and will create new scenarios. In this way, the suggested tool will be able to learn over time, allowing for discovering vulnerabilities, analysing the efficiency of existing controls, and increasing the security of the enterprise in general.

Keywords: Enterprise Cybersecurity, Generative AI-Powered Cybersecurity Digital Twin (GenCDT), Generative AI, Digital Twin, Cyber Threat Prediction, Attack Simulation, Vulnerability Assessment.

1. Introduction

Modern enterprises' IT environment has grown into a tightly dynamic ecosystem. They have failed to recognise the networks, multi-cloud infrastructure, and hybrid identity systems because of the interruption of third parties. This complexity has grown over the decades and expanded the attacks. Despite significant investments in firewalls, endpoint detection, SIEM platforms, and IAM frameworks, breaches continue to occur at scale [1], [2]. They only work on known threats, which will definitely affect the ecosystem of enterprises. Security teams generally wait for threats to happen. System checks occur on a set schedule, and advanced threat testing is expensive, infrequent, and limited by human availability [3], [4]. Security defenses usually focus on past attacks, but hackers use AI to invent new ones faster than companies can adapt [5], [6]. This delay makes it much easier for breaches to happen.

This paper presents a fundamental argument for the traditional digital twin, which will not only predict minor realities but also predict future attacks. We propose the Generative AI-Powered Cybersecurity Digital Twin (GenCDT), a virtual replica of an enterprise's network topology, cloud infrastructure, users, IAM policies, security controls, and applications, including a generative AI engine mechanism that continuously generates and launches new attack scenarios and ransomware campaigns directly at the twin. Instead of waiting for hackers or human testers to find flaws, the GenCDT automatically simulates attacks against your evolving system, closing security gaps before they are exploited [7], [8], [9].

2. Literature Review

In cybersecurity, digital twins create an environment in which security analysts can practice enterprise system hacking without interfering with the live production environment [4], [5]. Digital twins allow analysts to build replicas of an organisation's infrastructure that are then used to simulate cyberattacks, test security mechanisms, and test incident response plans, even before attacks occur [5]. By using this technique, organisations can be able to find any potential problems at an early stage without risking the security of their operating systems.

Currently, digital twins are widely adopted across enterprise IT environments to model network topology, cloud infrastructure, user identities, applications, and security controls [1], [2], [3]. Some of the activities performed in this regard include vulnerability assessment, attack simulations, intrusion detection, and cybersecurity training. Currently, most digital twin platforms use SIEMs and other monitoring tools to gather information to recreate past security attacks and analyse attack patterns [1], [5]. Although this enables security analysts to understand how previous attacks occurred, these types of systems rely heavily on historical data and offer little help in predicting future attacks. In addition to manual security testing methods, many companies are using automated adversary emulation frameworks like MITRE Caldera or commercial Breach and Attack Simulation (BAS) systems [8], [9]. The main goal is to emulate TTPs of the known threat actors so that the security team can assess the ability of its security controls to detect and stop these attacks. However, most BAS platforms use predefined attack patterns and rules that are used during attack execution [8], [9]. With the evolution of cyber threats and the increase in identity-based attacks in cloud environments, predefined attacks may become irrelevant due to a lack of real-world behaviour

New advances in the area of Generative AI have made it possible to use large language models and AI agents to generate attack simulations that are not defined by predefined patterns [10], [11]. In particular, AI can generate various types of attacks depending on the specificities of the target environment. It includes phishing attacks, ransomware, privilege escalation, malware execution, and others [10], [11]. This approach allows assessing the efficacy of security controls against attacks that are not seen before. Even though Generative AI can automate the process of attack simulations, security professionals are required to review and verify attacks created by Generative AI.

Another topic that has attracted considerable interest is cloud security and identity security. In the case of increasing cloud migration of organisations, there is an increased threat from cyberattacks targeting Identity and Access Management systems rather than relying on malicious software attacks alone. Cloud misconfiguration, over-provisioning of user permissions, and compromised identity have emerged as the main attack vectors in enterprise networks. For research in such attack vectors, cloud digital twins of enterprise networks have been created to recreate the attack pathways, which include identity compromise, privilege escalation, and lateral movement. It becomes clear that modern enterprises need cloud security and identity security equally to protect themselves from threats.

3. Comparison between Traditional Digital Twin and Proposed GenAI-Powered Cybersecurity Digital Twin

Dimension	Traditional Digital Twin	Proposed GenAI-Powered Cybersecurity Digital Twin
Primary Objective	Replicates the current state of enterprise infrastructure for monitoring and analysis.	Identifies previously unexplored attack paths and supports proactive cyber defense.
Threat Modelling	Evaluates predefined or previously observed attack scenarios.	Continuously anticipates potential attack paths and emerging security risks.
Attack Generation	Based on historical attack datasets and manually developed rules.	Dynamically generates adaptive attack scenarios using Generative AI.
Update Mechanism	Periodically updated using operational telemetry.	Continuously evolves as enterprise conditions and threat landscapes change.

Adversary Representation	Models known attacker techniques and predefined TTPs.	Simulates adaptive external attackers, insider threats, and cloud-based attack campaigns.
Simulation Coverage	Focuses on specific components such as networks or operational technology.	Models the complete enterprise ecosystem, including networks, cloud infrastructure, IAM, endpoints, applications, and security controls.
Expected Outcome	Validates existing security mechanisms against known threats.	Identifies previously unexplored attack paths and supports proactive cyber defense.

4. Methodology

This research uses the Design Science Research (DSR) approach to design and validate the proposed Generative AI-Powered Cybersecurity Digital Twin (GenCDT). The entire design framework is structured in five major phases: digital twin design, threat creation, attack emulation, evaluation, and feedback. In the first phase, a digital twin of a medium-sized enterprise is designed by taking into account network configuration, cloud resource details, IAM policies, software applications and security measures that currently exist. A graph database, for example, Neo4J, can be used to model relationships among network devices, users and permissions, allowing for analysis of attack paths and privilege escalation vectors.

Once the digital twin has been created, a well-tuned LLM combined with the Retrieval-Augmented Generation (RAG) framework is used to generate realistic scenarios of cyberattacks.

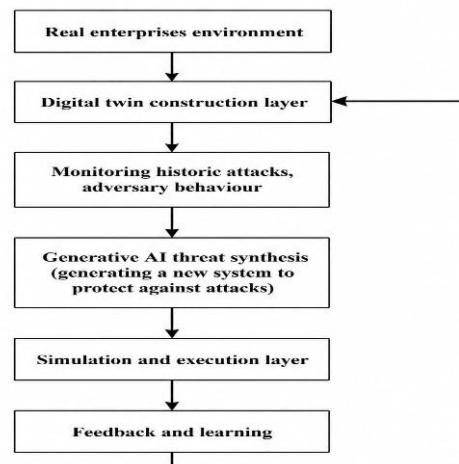


Fig. 1 (Workflow of GenCDT)

This LLM leverages the MITRE ATT&CK framework and the CVE database for generating scenarios of attacks, attacker models, insider threats, and ransomware attacks that would be applicable to the enterprise's environment. Each of the generated scenarios undergoes validation to make sure it is feasible in the digital twin.

Subsequently, the confirmed attack scenarios are executed within a controlled sandboxed environment where the corporate network is simulated through virtualisation, containerization, and adversarial emulation. The orchestration framework executes the multi-stage attack simulations and enables multiple attack scenarios to be executed in parallel, thus allowing various attack paths to be assessed. Performance metrics recorded during the experimentation phase include the detection ratio, Mean Time to Detection (MTTD), false negatives ratio,

vulnerabilities coverage, and realism of the generated attack scenarios. The results obtained from each round of experimentation are used for enhancing the Generative AI model, updating the digital twin with the new security controls, and executing the simulation process iteratively.

5. Results and Discussion

The presented GenCDT framework is an example of how the integration of a cybersecurity digital twin and Generative AI technologies will positively impact organisations' approaches to security testing. While traditional digital twins depend mainly on pre-defined attack scenarios, [1], [5] the suggested framework will provide the capability of generating new attack scenarios according to the current state of the enterprise infrastructure [10], [11]. Security teams will be able to test security controls against different attacks, instead of checking how the infrastructure is protected from known threats only. As a consequence, security vulnerabilities that could be overlooked by traditional testing will be revealed sooner. The suggested solution has the capability of modelling both cloud infrastructure and IAM systems. Nowadays, there are several digital twins designed mainly for network infrastructure. Simulations of cloud resources and identity-based attacks are under-represented in these frameworks. In modern enterprise infrastructure, identity-based attacks and privilege escalation become more frequent and cannot be neglected. Another aspect of the framework is the process of continuous feedback [5], [11]. After the simulation, the obtained information is used to refine the process of attack generation and update the digital twin whenever changes occur in security measures. As a result, the system will be able to run security assessments periodically and not only during testing. With the passage of time, the framework will be able to try various attack vectors and identify the areas of the enterprise environment that need to be protected further. However, there are also certain difficulties with the suggested approach. The accuracy of the simulations depends on how accurate the representation of the enterprise environment in the form of a digital twin is. If some assets, configurations or users' activity were not included in the model, the attack scenarios will not represent real-life situations. Besides, the AI-generated attack scenarios have to be evaluated by cybersecurity specialists since some of them might be impractical.

6. Conclusion

A framework named the “Generative AI-Powered Cybersecurity Digital Twin (GenCDT)” was suggested in this paper for proactive enterprise security. The key feature of this framework is that unlike the existing frameworks which only react to known attacks, this framework can simulate and detect future attacks on real systems through digital twin and Generative AI. Though the proposed framework is conceptual in nature, it opens up an exciting direction for research in predictive cybersecurity. The next step of research would be implementation of the framework in an actual enterprise setting.

References

1. C. Alcaraz and J. Lopez, "Digital Twin: A Comprehensive Survey of Security Threats," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1475–1503, 2022.
2. R. Minerva, G. M. Lee, and N. Crespi, "Digital Twin in the IoT Context: A Survey on Technical Features, Scenarios, and Architectural Models," *Proceedings of the IEEE*, vol. 108, no. 10, pp. 1785–1824, 2020.
3. S. Suhail, R. Jurdak, and R. Hussain, "Security Attacks and Solutions for Digital Twins," *arXiv:2202.12501*, 2022.
4. A. Saad et al., "On the Implementation of IoT-Based Digital Twin for Networked Microgrids Resiliency Against Cyber Attacks," *IEEE Trans. Smart Grid*, vol.11, no.6, pp.5138–5150, 2020.



5. M. Vielberth et al., "A Digital Twin-Based Cyber Range for SOC Analysts," Data and Applications Security and Privacy XXXV, Springer, pp.293–311, 2021.
6. A. Alzubaidi and H. K. Kalutarage, "Policy-Based Access Control for Zero Trust Architecture in Smart Industrial Systems," Computers & Security, vol.106, 102286, 2021.
7. E. C. Balta et al., "Digital Twin-Based Cyber-Attack Detection Framework for Cyber-Physical Manufacturing Systems," IEEE TASE, 2023.
8. A. Applebaum and D. Miller, "Analysis of Automated Adversary Emulation Techniques," SummerSim, 2017.
9. H. Kim et al., "Dynamic Cyberattack Simulation: Integrating Improved Deep Reinforcement Learning with the MITRE ATT&CK Framework," Electronics, vol.13, no.14, 2024.
10. Md. Raz et al., "Ransomware 3.0: Self-Composing and LLM-Orchestrated," arXiv:2508.20444, 2025.
11. "Forewarned Is Forearmed: A Survey on Large Language Model-Based Agents in Autonomous Cyberattacks," arXiv:2505.12786, 2025.
12. A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys & Tutorials, vol.19, no.2, pp.828–861, 2017.