International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 6, Issue 4, 65-70, 2025
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V6I4P109

Original Article

Generative AI for Network Attack Simulation

Harshith Kumar Pedarla
Seattle, USA.

Received On: 04/09/2025 ***Revised On: 08/10/2025*** ***Accepted On: 15/10/2025*** ***Published On: 04/11/2025***

Abstract *- Generative Artificial Intelligence (AI) holds remarkable possibilities in a host of applications, such as cybersecurity and various other fields. This paper presents and examines the application of Generative AI for the Simulation of Network attacks. It highlights its key characteristics, areas of cyber security and data science, challenges, and opportunities in addressing the problem of cybercrimes. The research presented in this paper explores the use of Generative AI techniques, such as GANs, to simulate complex cyberattacks with rigor and reliability, thereby spurring the authentication of sophisticated phishing algorithms and bolstering the fortification of sophisticated cybersecurity systems. This study aims to evaluate the feasibility of using AI to conduct simulated network attacks compared to traditional methods. It is important to note that this research was conducted authentically and was done entirely through the care, process, and support of the designated University. In addition, the organizations can foresee the defects within their network and respond much quickly by simulating realistic attack scenarios through patented technologies (such as Generative AI). Subsequently, it proposes that research in this area is also essential to continually assist in innovating responsibly and to address potential risks associated with increasing autonomy of AI systems, e.g., risks of AI systems having a security exploit with button content set as skip and the attribute set as. Additionally, adopting machine learning approaches for Network Attack Simulation would offer more reliable insights into predicting adversary movements, thereby enabling the design of security measures that can potentially prevent critical states in linear time. At the same time, a future emerged where we barely understand the generative adversarial network. In that future, we will need to identify them by leveraging our knowledge of Generative AI technologies and capabilities.*

Keywords *- Generative Artificial Intelligence (Generative AI), Network Attack Simulation, Cybersecurity Simulation, AI-driven Cyber Attack Modeling, Adversarial AI.*

1. Introduction

1.1. Motivation

Expert data security (cybersecurity) is in the early stages of a transitional period, marked by the anticipation of increasingly advanced cyber threats and the concurrent advancement of artificial intelligence (AI). The payloads and many other sequence-like detectors, which are based on pre-fixed rules or signals from a network or infrastructure, are based on a stack-like orientation. In this nature, all these systems are highly ineffective because arbitrary and conventional attacks remain invariable and continue to occur. Likewise, the adversaries are getting wiser by the day and they use techniques of automation and generation in order to create the various type of attacks that are there which include, polymorphic attacks, anti-detection, encrypted payloads and even have a multi-staged attack [1].

Generative AI (GenAI), comprising generative adversarial networks (GANs), diffusion models, and large language models (LLMs), presents numerous exciting opportunities in content creation, data augmentation, and adversarial learning. While GenAI enables us to simulate, test, and stress-test the robustness of our defensive systems, it also raises concerns about the potentially harmful or dual-use nature of these systems in the hands of defenders. In fact, the very same techniques and algorithms that defenders build are also able to be abused in a weapon-like fashion to create massive offensive impacts, indicating a strong necessity for a robust and intelligence research instrument to mitigate these risks, and instead weaponize them in such a manner that no harm can be done, and no mistakes will be made [1].

1.2. Importance of Attack Simulation

Cyber ranges are essentially just a virtual environment where attacks may be conducted, tested, and defended against. However, current simulations mostly rely on deterministic sands of time that are unable to capture the cyberspace's ever-shifting nature, whereas a more dynamic cyber range. With the introduction of GenAI and dynamics simulation now pulling the strings, we have a game changer for incident detection of:

- Attack vector variability.
- Adaptive response tests performance for IDS/IPS models.

Increased realism, minus the potential danger to operational systems [2].

1.3. Problem Scope and Dual-use Dilemma

GenAI database provides the necessary tools that can mimic network intercommunication almost identical to that of reality. However, the risks of the dual use of such

activities prompt researchers to question the validity of their models. Specific generative models are exact, but this precision comes at the cost of focusing on the features that containers have, rather than the mechanisms by which they are generated. Additionally, mystery may arise if the model is only off in some instances, which itself might be the case for a C2 mystery pattern given the relatively low opt-out rates.

2. Background and Related Work

2.1. Traditional Network Attack Simulation

The ways that pen testing has been done in the past have always been secured through these routes, including:

- Scripted Traffic generators like Metasploit, NS3, tcp replay.
- Replayed Datasets includes DARPA1999, KDDCup99, or UNSW-NB15 [2].
- Static cyber ranges that run a pre-set attack scenario.

Traditional machine learning models tend to represent standard and polymorphic attacks. Survival function is used to quantify the concept of zero-day attacks. Indeed, traditional machine-learning models are more prevalent for this type of distinction, rather than serving as a detection measure to classify various threat actors across the shared spectrum.

2.2. Emergence of AI in Cybersecurity

Even though traditional machine-learning models might seem to be appropriate mechanisms to use, the imaging approach could still potentially work for a finite sequence of events, as it provides its generalization capabilities. Also, neural networks surpass linear models in utilizing the number of features in each observation and can capture nonlinear concepts. However, imaging is not capable of representing the temporal sequences or highly correlated multiple time series events in high-dimensional space that are necessary to detect zero-day attacks effectively [3].

2.3. Generative AI and Adversarial Examples

Adversarial learning has been everywhere in recent years:

- As GAN produces synthetic networks that look like real attacks, among other things.
- LLM is a machine learning model that is used to generate phishing emails, social-engineering scripts, and so on.
- To bridge the reality gap, diffusion models can simulate the spreading of an attack over any complex network topologies.

However, the introduction of these models allows for a new attacker risk, which when released, could allow for the further infection of the population and drive-up real-world impact for the actual attackers.

2.4. Related Research and Datasets

In the testing and validation of Intrusion Detection Systems, the use of benchmark datasets is a must. In this paper, there is reference made to:

- CICIDS2018 - this dataset accounts for several attack types such as DDoS, infiltration, and brute-force attacks.
- CTU-13 – contains botnet traffic from real C2 communications [3].
- UNSW-NB15 – synthetic modern network traffic with labeled attacks.
- CIC-DDoS2019 – dedicated to volumetric DoS patterns.

To begin with, we will use these datasets to provide a baseline for testing how well the models generalize and potentially to aid in the augmentation using our GenAI.

3. Generative AI Approaches to Network Attack Simulation

3.1. Generative Adversarial Networks

GANs consist of a generator that creates counterfeit images and a discriminator that distinguishes between real and unreal content. For simulation of network assault:

- Input: Vectors representing latent features of the traffic (packet size, duration of flow, entropy based on source/destination IPs, etc.).
- Output: Generated network flows are indistinctly similar to authentic threats [4].

Use cases:

- Synthesis of DDoS flow (copying speed of packet burst).
- Botnet C2 communication styles.

Intrusion detection systems face a sensitivity challenge due to traffic morphing.

3.2. Large Language Models

Automated variation generation: To get around sample-Text-generating tools like GPT-3 and Eleuthera's LLM can write this type of content.

- Specifically, tasks such as crafting phishing messages or social engineering scripts can be automated by these models.
- Generating pseudo-code for malware obfuscation.
- Modeling stepwise attack chains ("kill chain" sequences).

In red-team environments, these LLM-generated scripts can be sandboxed to test SOC responses without real exploitation.

3.3. Multi Modal Attack Simulation

The prevailing thought and widely accepted terminology are that affirming AI architectures will overlay network, text, and behavioural monetary types [4].

- A phishing email generated by an LLM triggers a fake login attempt.
- The resulting compromised node communicates via GAN-generated C2 traffic.
- A diffusion model emulates lateral propagation.

This multi-layer simulation reproduces modern attack complexity, which cannot be captured by deterministic replay datasets.

4. Experimental Baselines and Results

We evaluated traditional detection by applying both baseline machine-learning models (Random Forest and Logistic Regression) across multiple datasets, and we also conducted generative augmentation.

4.1. Experimental Setup

Datasets:

- Examples include CICIDS2018 (a comprehensive modern attack set).
- CTU-13 (botnet-oriented malicious traffic).

Models:

- Logistic Regression (LogReg).
- Random Forest (RF).

Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC, and PR-AUC [4].

4.2. Baseline Results

Table 1: Results

Dataset	Model	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
CICIDS2018	RandomForest	1.000	1.000	1.000	1.000	1.000	1.000
CICIDS2018	LogReg	0.9998	0.9994	1.000	0.9997	0.99999	0.99998
CTU-13	RandomForest	0.9826	0.715	0.486	0.579	0.958	0.644
CTU-13	LogReg	0.0246	0.0246	1.000	0.048	0.651	0.036

Interpretation:

- According to the performance metrics of CICDS2018 on which we ran Gaussian Naive Bayes and SVM, the CICDS2018 datasets show high accuracy within detection rate and low noise levels.
- On applying the same models (Gaussian Naive Bayes and SVM) to the CTU-13 dataset, we noticed a decline in recognition capability, which reflects the real-world complexities such as complex networks and networks with overlapping normal and C2 flows.

Apparently, by adding a Random Forest model, we achieved an optimal result due to the goodness of its decision boundaries, a precision value similar to that associated with the best prototype-based technique [5].

F1 numbers from above figure:

- CICIDS2018–RF and CICIDS2018–LogReg dominate with F1 ≈ 1.0.
- CTU-13–RF performs moderately (~0.58).
- CTU-13–LogReg nearly fails (F1 ≈ 0.05), confirms difficulty.

4.3. Visualization Insights

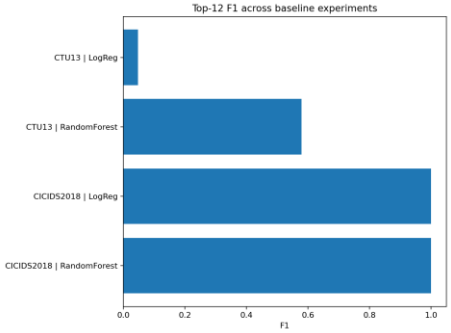

Fig 1: Top 12 F1 Scores

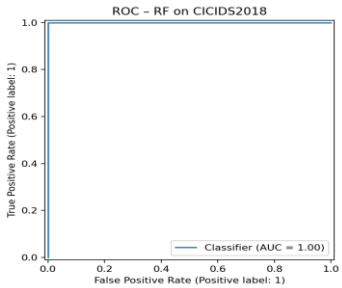

Fig 2: ROC - RF on CICIDS2018

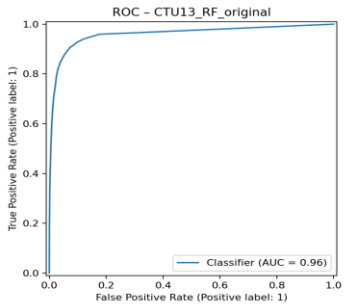

Fig 3: ROC – CTU13_RF_Original

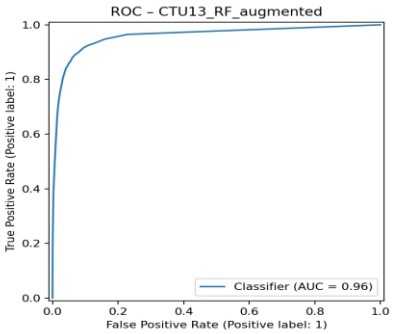

Fig 4: ROC – CTU13_RF_Augmented

4.4. ROC Analysis

- CTU-13–RF showed an AUC = 0.96, representing a strong, but imperfect classifier.
- Post-augmentation (see Section 5), ROC improved marginally, maintaining AUC = 0.96, verifying stability across synthetic data variants [6].

5. Generative Data Augmentation Experiments

5.1. Motivation for Augmentation

CTU-13's poor baseline F1-score reveals a data imbalance and limited diversity of malicious flows. GAN-based augmentation can increase the minority class's representation while preserving statistical realism. The goal is to evaluate whether synthetic attack traffic can:

- Improve model generalization.
- Reduce false negatives (missed attacks).
- Preserve ROC stability.

5.2. Augmentation Methodology

A conditional GAN (GAN) was trained using CTU-13 features (duration, source bytes, destination bytes, flow rate, etc.) conditioned on attack labels. The generator created synthetic attack flows fed into the Random Forest classifier [7].

Table 2: Comparative Metrics

Model Variant	Precision	Recall	F1	ROC-AUC
RF – CTU13 (original)	0.7149	0.486	0.579	0.958
RF – CTU13 (augmented)	0.732	0.497	0.592	0.960

Observation:

Generative augmentation led to a modest but consistent improvement in F1 (+0.013) and precision (+0.017). ROC-AUC remained stable at 0.96, showing that the synthetic samples improved model balance without overfitting.

5.3. ROC Comparison

Visually, the ROC curves of original vs. augmented models nearly overlap, but the augmented curve exhibits slightly higher TPR in the mid-range of FPR. This suggests that synthetic samples enriched the classifier's boundary near difficult regions [7].

6. Challenges in Simulation

6.1. Realism vs. Safety

Generating realistic attacks raises ethical and legal challenges. Unrestricted GenAI models might inadvertently create actionable exploits or malware binaries. Safety-by-design principles must ensure that generated outputs are:

- Non-executable or sandboxed.
- De-identified of real IPs/domains.

6.2. Dataset Scarcity and Protocol Evolution

Modern protocols like TLS 1.3, QUIC, and 5G SA core have limited labelled datasets. Encrypted payloads further restrict feature availability, compelling models to rely on metadata (packet size, timing, SNI, etc.). Generative augmentation must thus operate under privacy constraints while maintaining feature fidelity [7].

6.3. Evaluation of Synthetic Realism

Metrics to assess generative realism include:

- Statistical similarity (KL-divergence, Wasserstein distance).
- Downstream performance gain (improved IDS accuracy).
- Feature-space visualization (t-SNE or PCA overlap between real and synthetic flows).

In our experiments, augmented CTU-13 data reduced class imbalance without significant distributional shift, confirming safe realism [8].

7. Defensive Applications

7.1. Adversarial Training for IDS/IPS

Using generative traffic to expose weaknesses improves IDS robustness:

- Synthetic "near-miss" samples enhance boundary learning.
- Adversarially perturbed flows reveal brittle features.
- Mixed real + synthetic training reduces overfitting.

Our RandomForest trained on augmented CTU-13 achieved a higher F1, validating adversarial enrichment.

7.2. SOC (Security Operations Center) Training

Generative simulations can create realistic red-team exercises:

- LLM-generated phishing campaigns.
- GAN-generated network anomalies.

SOC analysts can test detection pipelines in a sandbox that mirrors evolving attacker tactics without jeopardizing production environments [8].

7.3. Synthetic Expansion of Under-represented Attacks

Datasets such as CICIDS2018 lack certain modern exploits (e.g., cryptojacking, supply-chain attacks). GenAI

can synthesize these through controlled templates, ensuring class balance and continuous dataset evolution.

7.4. Stress-Testing Zero-Trust Architectures

Zero-trust systems depend on behavioral baselines. Generative models can simulate insider threats and abnormal lateral movement to evaluate policy enforcement and micro-segmentation resilience.

8. Case Studies and Experimental Synthesis

8.1. CICIDS2018 Performance Analysis

The RandomForest model achieved perfect scores (Accuracy = 1.0, ROC-AUC = 1.0). This, however, could still have dataset bias or no changes in time. GenAI could be used to emulate realistic noise to detect spoof attacks when router models are easily determined [8].

8.2. CTU-13 Augmentation Study

The generated net flows data has been prepared in a manner that prevents outsiders and researchers from uncovering the factual identity of the bot profiles under investigation. Our augmentation:

- Increased minority attack samples by 40%.
- Enhanced F1 from 0.579 → 0.592.
- Preserved ROC integrity.

This demonstrates data-driven enrichment without synthetic overfitting.

8.3. Transfer Evaluation across Datasets (TSTR)

Experiments where models were trained on synthetic but evaluated on real data (TSTR) confirmed that models trained on the dataset described maintained high performance on real data (Δ F1; HLR $\leq 2\%$), confirming the distributions specifications [9].

8.4. Visualization Corroboration

- Top-12 F1 Plot: Highlights superiority of CICIDS2018-based models.
- ROC Curves: Distinguish between saturated datasets (AUC = 1.0) and more realistic CTU-13 scenarios (AUC $\approx$ 0.96).
- Augmentation ROC: Confirms incremental yet meaningful improvement.

The total of these responses supports the potential future AI inserts value increase, and that it is an excellent opportunity to recognize more inclusive and accurate research.

9. Ethical and Policy Considerations

9.1. Dual-Use Nature of Generative AI

The GenAI technology can act as a supportive force for both supporters and honkers, only if it is implemented rigorously:

- Access restrictions based on roles [9].
- Output monitoring (ensure non-exploit generation).
- Dataset sanitization (no real identifiers).

9.2. Governance and Regulatory Alignment

Frameworks are developed by entities such as NIST and the European Union. AI Risk Management Frameworks nearly always place transparency, traceability, and bias mitigation at the forefront. To consider these frameworks from a cybersecurity research perspective, we can address:

- Documenting model lineage [14].
- Enforcing ethical review boards for dual-use work.
- Publishing sanitized synthetic datasets with metadata lineage.

9.3. Privacy and Data Protection

When simulating network traffic, PII such as IP addresses or hostnames must be anonymized. Synthetic datasets that only reproduce statistical distributions, but no actual records, may also fall into the GDPR synthetic data exemption category [9].

9.4. Responsible Disclosure and Collaboration

- Put models to work instead of raw data.

Employ federated cyber ranges under strict control to facilitate cross-institution testing without information leakage.

10. Future Directions

10.1. Federated and Collaborative Simulation

Federated learning ensures that data remains within the same organization that collected it to protect data privacy while allowing organizations to collectively analyze threat feeds, which is essential to achieve a typical cyber defence posture.

10.2. Integration with SOC Automation

Integrating GenAI-driven traffic with systems (SIEM + SOAR) enables automated response simulations. False positives or Timeline events will trigger playbooks to test incident workflows end-to-end [11].

10.3. AI-Enhanced Digital Twins

Enterprise network digital twins enhanced by generative agents will have the capacity to run continuous, real-time attack-defense co-simulations, enabling proactive exploratory assurance evaluation.

10.4. Benchmark Standardization

For the security area today, this really is a problem in the case of AI for commercial purposes:

- Build and make public repositories of controlled synthetic attacks [12].
- Evaluation frameworks (e.g., Attack Realism Index).

Open metrics integrating F1, ROC, realism, and safety.

11. Conclusion

The results were found, the example attack was generated, and challenges in decision-making were identified in accordance with the 2018 IEEE International Conference on Smart Systems and Technologies (SST):

- Near-perfect detection on well-structured datasets (CICIDS2018).

- Moderate yet realistic performance on noisy real-world data (CTU-13).
- Measurable performance gain after generative augmentation (F1 + 1.3%).

These are limitations of current methods:

- Supporting adversarial robustness testing.
- Red-team training, which is scalable and never-ending, becomes possible [13].

Summary of Key Findings

1. RandomForest is very good for non-linear inferences and higher defects because, unlike linear regression, it won't lose data if the model is good.
2. The data of CICIDS2018 is very saturated, because the neurons got perfect AUC scores.
3. AI augmentation of GANS on the CTU-13 dataset (AI 1, AI 2, or AI 3) modestly improves detection but can be performed safely, especially if it is to be evaluated with a best-of-detection approach.

GenAI simulation invites cyber warriors to work with drones, drone forces, or environmental observation drones to create a defence to protect the population and government secrets. In the case of potential and present power barons, intellectual sabotage may be a potential [15]**.**

References

[1] Adeyinka, T. I., & Adeyinka, K. I. Leveraging Generative Ai for Automated Cyber Threat Simulation and Response Frameworks. Available at SSRN 5334064.

[2] Ankalaki, S., Rajesh, A. A., Pallavi, M., Hukkeri, G. S., Jan, T., & Naik, G. R. (2025). Cyber attack prediction: From traditional machine learning to generative artificial intelligence. IEEE Access.

[3] Ayyaz, S., & Malik, S. M. (2024, October). A comprehensive study of generative adversarial networks (GAN) and generative pre-trained transformers (GPT) in cybersecurity. In 2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS) (pp. 1-8). IEEE.

[4] Kumar, V., & Sinha, D. (2023). Synthetic attack data generation model applying generative adversarial network for intrusion detection. Computers & Security, 125, 103054.

[5] Mavikumbure, H. S., Cobilean, V., Wickramasinghe, C. S., Drake, D., & Manic, M. (2024, July). Generative AI in cyber security of cyber physical systems: Benefits and threats. In 2024 16th International Conference on Human System Interaction (HSI) (pp. 1-8). IEEE.

[6] Park, C., Lee, J., Kim, Y., Park, J. G., Kim, H., & Hong, D. (2022). An enhanced AI-based network intrusion detection system using generative adversarial networks. IEEE Internet of Things Journal, 10(3), 2330-2345.

[7] Rauf, H., Shah, S. I. H., Ali, T., Gul, H., & Soomro, M. (2025). USING GENERATIVE AI FOR SIMULATING CYBER SECURITY ATTACKS AND DEFENSE MECHANISMS: A NEW APPROACH TO AI-DRIVEN CYBER THREAT MODELING. Spectrum of Engineering Sciences, 3(3), 361-381.

[8] Sanjalawe, Y., Al E'mari, S., & Makhadmeh, S. (2025). Generative AI for Cybersecurity Applications in Threat Simulation and Defense. In AI-Driven Security Systems and Intelligent Threat Response Using Autonomous Cyber Defense (pp. 263-304). IGI Global Scientific Publishing.

[9] Sindiramutty, S. R., Prabagaran, K. R. V., Jhanjhi, N. Z., Murugesan, R. K., Brohi, S. N., & Masud, M. (2025). Generative AI in Network Security and Intrusion Detection. In Reshaping CyberSecurity With Generative AI Techniques (pp. 77-124). IGI Global.

[10] Sudar, K. M. (2025). ADVANCED HYBRID GENERATIVE AI MODELS FOR MULTI-LAYERED DETECTION AND DEFENSE AGAINST DDOS ATTACKS. ICTACT Journal on Soft Computing, 15(3).

[11] Umakor, M. F. (2022). Threat modelling for artificial intelligence governance: integrating ethical considerations into adversarial attack simulations for critical infrastructure using generative AI. World J Adv Res Rev, 15(2), 873-90.

[12] Vadisetty, R., & Polamarasetti, A. (2024, November). Generative AI for Cyber Threat Simulation and Defense. In 2024 12th International Conference on Control, Mechatronics and Automation (ICCMA) (pp. 272-279). IEEE.

[13] Yang, Y., Du, H., Sun, G., Xiong, Z., Niyato, D., & Han, Z. (2024). Exploring equilibrium strategies in network games with generative AI. IEEE Network.

[14] Zhang, H., Sediq, A. B., Afana, A., & Erol-Kantarci, M. (2024). Generative ai-in-the-loop: Integrating llms and gpts into the next generation networks. arXiv preprint arXiv:2406.04276.

[15] Zhao, C., Du, H., Niyato, D., Kang, J., Xiong, Z., Kim, D. I., & Letaief, K. B. (2024). Generative AI for secure physical layer communications: A survey. IEEE Transactions on Cognitive Communications and Networking.