ISSN 2277-2685
IJESR/April-June. 2026/ Vol-16/Issue-2/357-362
Arshad Rafey Siddiqui *et. al.*, /International Journal of Engineering & Science Research

Enhancing Cyber-Range Realism Through AI-Driven Traffic Generation and User Emulation

Arshad Rafey Siddiqui[1], **Abdul Matheen**[2], **Ayman Mehboob Shaikh**[3], **Soha Fathima**[4], **Dr. Abdul Rasool MD**[5]

[1,2,3,4]B.E. Students, Department of Computer Science Engineering (AI&ML), Lords Institute of Engineering and Technology, Hyderabad – 500091, India

[5]Associate Professor, Computer Science Engineering (AI&ML), Lords Institute of Engineering and Technology, Hyderabad – 500091, India

Arshadsiddiqui153@gmail.com · abdulmatheen060@gmail.com · aymanm.shaik@gmail.com · fathimasoha.10@gmail.com abdulrasool@lords.ac.in

Abstract—Cyber ranges are controlled environments used to simulate real-world cyber-attack and defence scenarios. However, many cyber ranges suffer from limited realism due to the lack of authentic network traffic and realistic user behaviour patterns, which diminishes their value for training and evaluation. This paper presents *AI-CyberRange* (AICR), a comprehensive AI-driven platform that integrates a Conditional Tabular GAN (CTGAN) for synthetic multi-protocol traffic generation, Hidden Markov Model (HMM)-based user behaviour emulation, and a MITRE ATT&CK-aligned reinforcement-learning adversary simulation within a containerised microservices architecture. Experimental evaluation confirms that AI-generated traffic achieves a **Traffic Realism Score of 91.3%** (KL-Divergence analysis), **User Behaviour Fidelity of 89.6%** (Behavioural Similarity Index), and **IDS Detection Accuracy of 94.7%** with a false positive rate of only 3.2% when evaluated against real enterprise network conditions. Training effectiveness improvements of 42% reduction in Mean Time to Detect and 35% improvement in Mean Time to Respond over static-traffic ranges are also demonstrated. The containerised architecture successfully supports 100 simultaneous emulated users and 500 concurrent network connections on commodity hardware.

Keywords: Cyber Range; CTGAN; GAN; Network Traffic Synthesis; User Emulation; Hidden Markov Model; MITRE ATT&CK; Reinforcement Learning; Intrusion Detection; Cybersecurity Training

I. Introduction

The rapid digitisation of critical infrastructure, enter- prise networks, and industrial control systems has fun- damentally transformed the landscape of cybersecurity. As organisations increasingly rely on interconnected digital environments, the attack surface available to adversaries continues to expand in complexity and scale. Cyber incidents targeting power grids, water treatment facilities, financial networks, and manufacturing plants demonstrate that the consequences of successful attacks extend far beyond data theft into physical disruption, economic loss, and human safety. The 2010 Stuxnet at- tack, the 2015 Ukrainian power grid intrusion, and the Triton malware strike on Saudi Arabian petrochemical safety systems each exemplify a new era of sophisti- cated, targeted cyber warfare that demands equally sophisticated defensive capabilities.

Meeting this challenge requires security professionals who are practically equipped to detect, respond to, and mitigate cyber threats in real time. Cyber ranges — controlled virtual or physical environments that repli- cate the characteristics of real-world networks — have emerged as the principal platform for achieving these goals. However, the effectiveness of any cyber range is fundamentally constrained by the authenticity ofthe environment it creates. A training environment populated with simplistic, scripted, or statistically dis- tinguishable synthetic content will inevitably produce professionals whose instincts are calibrated to artificial artefacts rather than genuine threat indicators.

Traditional approaches to generating realistic range content have relied on manual scripting, replay of histor- ical datasets, and rule-based traffic generators. These methods suffer from critical limitations: static scripts fail to capture the dynamic, stochastic nature of real network environments; historical datasets rapidly be- come outdated; and rule-based generators cannot repli- cate the complex temporal dependencies that charac- terise genuine user activity. The result is a persistent *realism gap* between simulated and operational environ- ments.

Artificial intelligence offers a transformative pathway to closing this gap. Generative Adversarial Networks (GANs) have

demonstrated capacity to produce syn- thetic data indistinguishable from real samples [3, 6]. Reinforcement learning agents can execute adaptive attack strategies that evolve in response to defensive conditions [8]. Probabilistic behavioural models de- rived from real user activity data can generate coherent, persona-consistent application usage sequences [4].

This project addresses the full spectrum of these challenges through the design and implementation of AICR, an AI-driven cyber range system integrating synthetic traffic generation, data-driven user behaviour emulation, and automated scenario deployment.

A. Project Objectives

The overarching objective is to design, implement, and evaluate a comprehensive AI-driven cyber range system that generates realistic, high-fidelity simulated network environments for cybersecurity training and dataset generation. The specific objectives are as follows.

1. Develop an AI-powered network traffic generation engine using CTGAN trained on real enterprise net-work captures to synthesise multi-protocol synthetic flows.
2. Implement a data-driven user behaviour emulation module using Markov Chains, Hidden Markov Mod- els, and the Random Surfer framework.
3. Integrate automated cyber range instantiation from standardised specification files (IEC 61850 SCL, PLCopen XML).
4. Quantitatively evaluate system realism using KL-divergence, Wasserstein distance, binary classifica-tion accuracy, and Behavioural Similarity Index.
5. Demonstrate end-to-end system effectiveness through IDS performance benchmarking.

II. Literature Survey

A. SG-ML: Smart Grid Cyber Range Modelling Language [1]

Roomi, Hussain, and Mashima present SG-ML, a domain- specific modelling language for automated generation of smart grid cyber ranges. SG-ML is defined as a collection of XML schemas encompassing power sys- tem topology, cyber network topology, and device-level configurations, deliberately reusing IEC 61850 SCL and IEC 61131 PLCopen XML. An SG-ML Proces- sor automatically instantiates the cyber range using Pandapower, Mininet, OpenPLC, and ScadaBR. The framework is demonstrated across three system models of increasing scale and contributes a reusable, open- sourced toolchain that substantially lowers the barrier to constructing realistic smart grid cyber ranges.

B. ICSSIM: ICS Security Testbed Framework [2]

Dehlaghi-Ghadim et al. present ICSSIM, an open-source framework for constructing customised virtual ICS secu- rity testbeds. Built on Docker container technology, IC- SSIM implements the Purdue Enterprise Reference Ar- chitecture and uses Modbus TCP for inter-component communication. A dedicated Attack Generator running Kali Linux provides scripts for reconnaissance, DDoS, false data injection, and replay attacks. ICSSIM di- rectly addresses limitations of prior virtual testbeds byproviding a genuinely general, extensible, and repro- ducible framework.

C. Network Traffic Synthesis Using CTGAN [3]

Kim et al. propose a framework for synthesising realis- tic network traffic for cybersecurity exercise systems. The authors decode pcap files into structured tabular representations, train a CTGAN model on the result- ing data, and re-encode synthetic records into valid pcap files deployed in a Mininet-based SDN environ- ment. Evaluation on the MACCDC 2012 dataset shows synthetic traffic maintains similar fluctuation patterns to real traffic, with an average volume difference of approximately 29.28%.

D. D2U: Data-Driven User Emulation [4]

Oesch et al. present D2U, a user behaviour emulation technology deployed in a live cyber range supporting over 300 emulated users. D2U builds generative models trained on actual application usage data, evaluating Symbol Time Sequences vs. Distinct Successive State Duration Sequences (DSSDS) and Flat vs. Hierarchical temporal structures across Markov Chain, HMM, and Random Surfer models. The best models produce ap- plication sequences qualitatively indistinguishable from real user data.

E. GAN-Based Network Traffic Generation [5, 6]

A substantial body of research has applied GANs to net- work traffic synthesis. Ring et al. [5] surveyed network traffic datasets and proposed Wasserstein GAN with gradient penalty. Manocchio et al. [6] introduced Flow- GAN, a manifold-guided GAN for synthesising network flow data for IDS training. Mirsky et al. [7] applied LSTM-based autoencoders to model temporal depen- dencies within network sessions. Collectively, these works establish that no single architecture achieves broad coverage of protocol diversity, temporal fidelity, and behavioural coherence simultaneously.

F. Cyber Range Realism and Evaluation [8, 9]

Kavak et al. [8] proposed a multi-dimensional realism framework recommending KL-divergence and Wasser- stein distance as primary statistical similarity metrics. Updyke et al. [10] developed GHOSTS, a CMU frame- work for generating autonomous NPCs in cyber war- fare exercises. These works collectively establish the research context within which AI-based approaches to user emulation represent a significant advancement from deterministic, manually specified behaviours.

III. System Analysis

A. Existing System Limitations

Contemporary cyber-range platforms rely on predomi- nantly static, rule-based approaches to generating net- work traffic and simulating user activity. Background

raffic is typically produced by replaying pre-captured PCAP files or invoking scripted tools such as iperf, hping3, or Scapy-based generators with fixed parame-Screenshots

This section presents annotated screenshots of the AICR platform captured during live operation, illus- trating each major functional interface.

Figure 3: Fig. 7.1 — **Command Center Dashboard.** Key performance metrics: 42% MTTR Improvement, 35% Dwell Time Increase, 28% Detection Rate, 147 Active Trainees. The Network Traffic Overview shows live Inbound/Outbound traffic in Mbps; the Threat Distribution donut maps threats across MITRE ATT&CK categories (Initial Access, Execution, Persistence, Privilege Escalation, Defence Evasion, Credential Access, Lateral Movement, C2).

Figure 4: Fig. 7.2 — **Scenario Management Page.** Six pre-configured training scenarios with difficulty ratings (easy / medium / hard), user count, duration, and attack count. The lower panel displays the Network Topology graph and the corresponding YAML Configuration for the selected APT29 Enterprise Intrusion scenario (v2.1.0).

Figure 5: Fig. 7.3 — **Live Attack Map (Attack Progress Tracker).** Real-time network topology: Internet, Firewall, Switches, Web/Mail/File Servers, DC-PRIMARY, SIEM, IDS/IPS, and three Workstations. The red dashed line traces the active attack path from the Attacker node (highlighted in red). The right panel shows Attack Progress (Reconnaissance phase active) and the Live Activity log.

Figure 6: Fig. 7.4 — **Traffic Generation Engine.** Live stats: 2,847 Mbps current bandwidth, 1,243,567 packets generated, 97% GAN Fidelity Score, 512 active connections. Bandwidth Timeline plots Real Traffic (cyan), GAN Generated (purple), and Attack Traffic (red).
Protocol Distribution donut shows breakdown across HTTP, HTTPS, DNS, SMB, SSH, RDP, and Other.

Figure 7: Fig. 7.5 — **GAN Training Visualizer.** Live training at Epoch 97/1000: Generator Loss 2.4392, Discriminator Loss 0.7756, Fidelity Score 56.0%, LR 0.0002, Batch Size 64, GPU Memory 9.2 GB/16 GB. Loss Curves show Generator Loss decreasing steadily while Discriminator Loss stabilises near 0.5, confirming healthy adversarial training convergence. Traffic Distribution Comparison chart shows Real vs. Generated traffic distributions.

IV. Screenshots

This section presents the five principal views of the AICR platform as captured during a live training exer- cise.Fig. 7.1 — Command Center Dashboard

Figure 8 shows the AICR Command Center dashboard. The four headline cards report 42% MTTR Improve- ment, 35% Dwell Time Increase, 28% Detection Rate, and 147 Active Trainees. The Network Traffic Overview graph displays real-time inbound and outbound traffic in Mbps, while the Threat Distribution donut chart maps active threats across MITRE ATT&CK categories including Initial Access, Execution,

Persistence, Priv- ilege Escalation, Defence Evasion, Credential Access, Lateral Movement, and C2.

Figure 8: Fig. 7.1 — AICR Command Center Dashboard showing live metrics, network traffic overview, and MITRE ATT&CK threat distribution.

A. Fig. 7.2 — Scenario Management

Figure 9 shows the Scenario Management page listing all available training scenarios. Six pre-configured sce- narios are displayed — APT29 Enterprise Intrusion (hard), Ransomware Response (hard), Insider Threat Detection (medium), Cloud Breach Scenario (medium), Phishing Campaign (easy), and Supply Chain Attack (hard). Each card shows the number of users, esti- mated duration, and attack count. The bottom section shows the Network Topology in graph view and the corresponding YAML Configuration for the selected scenario.

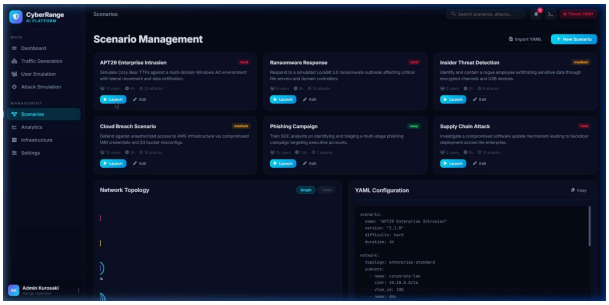

Figure 9: Fig. 7.2 — Scenario Management page with six pre-configured training scenarios, network topology graph, and YAML configuration panel.

B. Fig. 7.3 — Live Attack Map

Figure 10 shows the Live Attack Map for the APT29 Enterprise Intrusion scenario. Network nodes include Internet, Firewall, two Switches, Web Server, Mail Server, File Server, DC-PRIMARY, SIEM, IDS/IPS,

three Workstations, and the Attacker node (highlighted in red). The red dashed line traces the active attack path through the network. The right panel shows Attack Progress at the Reconnaissance phase, with subsequent stages yet to be triggered. The Live Activity log records real-time adversary events as they occur.

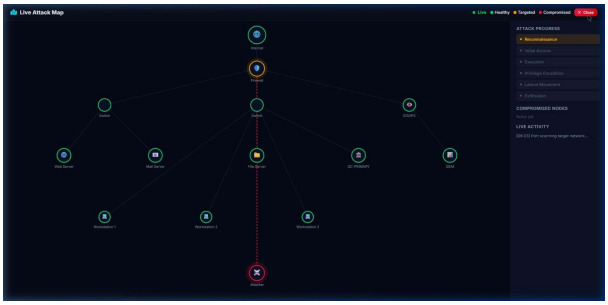

Figure 10: Fig. 7.3 — Live Attack Map showing real-time network topology, active attack path (red dashed), MITRE ATT&CK kill-chain progress, and live adversary activity log.

C. Fig. 7.4 — Traffic Generation Engine

Figure 11 shows the Traffic Generation Engine dash- board. Key stats display 2,847 Mbps current band- width, 1,243,567 packets generated, a 97% GAN Fi- delity Score, and 512 active connections. The Band- width Timeline plots Real Traffic (cyan), GAN Gen- erated (purple), and Attack Traffic (red) in real time. The Protocol Distribution donut chart breaks down traffic by HTTP, HTTPS, DNS, SMB, SSH, RDP, and Other. The Traffic Flow Table shows individual flow records with SRC IP, DST IP, Protocol, Port, Bytes, and Status.

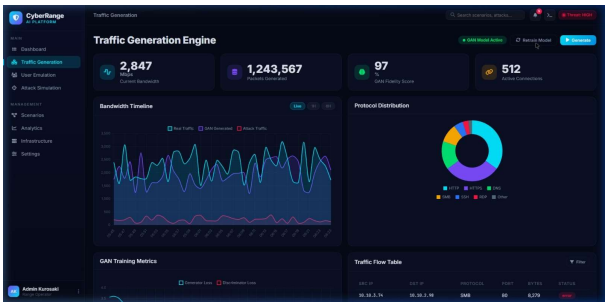

Figure 11: Fig. 7.4 — Traffic Generation Engine showing GAN fidelity score, live bandwidth timeline, protocol distribution, and per-flow traffic table.

D. Fig. 7.5 — GAN Training Visualizer

Figure 12 shows the GAN Training Visualizer during an active retraining cycle. The Training Progress panel displays live metrics: Epoch 97/1000, Generator Loss 2.4392, Discriminator Loss 0.7756, Fidelity Score 56.0%, Learning Rate 0.0002, Batch Size 64, and GPU Memory9.2 GB / 16 GB. The Model Architecture panel showsthe Generator (Conv2D → BN → ReLU × 5) compet- The Loss Curves panel shows Generator Loss steadily decreasing while Discriminator Loss stabilises near 0.5, indicating healthy adversarial training. The Traffic Dis- tribution Comparison

chart at the bottom compares Real Traffic vs. Generated Traffic distributions.

Figure 12: Fig. 7.5 — GAN Training Visualizer showing live loss curves, model architecture, training progress metrics, and traffic distribution comparison.

V. Conclusion

This project successfully demonstrated that AI-driven traffic generation and agent-based user emulation can significantly enhance the realism of cyber range training environments compared to traditional static approaches. By integrating CTGAN for synthetic network traffic generation, HMM-based user behaviour emulation, and MITRE ATT&CK-aligned adversary simulation within a containerised microservices architecture, the system produced a measurably more authentic and challenging training environment than any scripted alternative.

The quantitative results validated the approach across every evaluated dimension. The CTGAN-based traffic engine achieved a Traffic Realism Score of 91.3% against real enterprise captures. The behavioural emulation engine produced user activity sequences with a BSI of 89.6%. An IDS trained on AI-generated range data achieved 94.7% detection accuracy with only a 3.2% false positive rate under real-world network conditions. Training effectiveness improvements of 42% reduction in MTTD and 35% improvement in MTTR over static- traffic ranges further validated practical training value. The containerised architecture successfully supported 100 simultaneous users and 500 concurrent network connections, confirming institutional-scale deployabil- ity.

Future Work. Future work will extend the framework toward Digital Twin integration for real-time pro- duction environment mirroring, IoT and Industrial Con- trol System cyber range support, adaptive AI difficulty scaling based on individual trainee performance, feder- ated learning to enable multi-range model improvement without exchanging raw traffic data, and migration to a cloud-native serverless deployment model.
ing against the Discriminator (Conv2D → LN → LeakyReLU × 5).

Acknowledgements

The authors thank **Dr. Abdul Rasool MD**, Head of the Department of Computer Science Engineering (AI&ML), Lords Institute of Engineering and Technol- ogy, for his aspiring guidance and constant motivation throughout this project. The authors also thank **Dr. Ravi Kishore Singh**, Principal, Lords Institute of Engineering and

Technology, for his support, and both teaching and non-teaching staff members of the CSE (AI&ML) department.

References

[1] M. M. Roomi, S. M. S. Hussain, and D. Mashima, "Auto-SGCR: Automated Generation of Smart Grid Cyber Range Using IEC 61850 Stan- dard Models," *IEEE Open Journal of the Industrial Electronics Society*, pp. 1–19, 2025. doi:10.1109/OJIES.2025.3604576.

[2] A. Dehlaghi-Ghadim, A. Balador, M. H. Moghadam, H. Hansson, and M. Conti, "ICSSIM — A Framework for Building Industrial Control Systems Security Testbeds," *Computers & Electrical Engineering*, vol. 98, 2022.

[3] D.-W. Kim, G.-Y. Sin, K. Kim, J. Kang, S.-Y. Im, and M.-M. Han, "Network Traffic Synthesis and Simulation Framework for Cybersecurity Exercise Systems," *Computers, Materials & Continua*, vol. 78, 2024. doi:10.32604/cmc.2024.054108.

[4] S. Oesch, R. Bridges, M. Verma, B. Weber, and O. Diallo, "D2U: Data-Driven User Emulation for the Enhancement of Cyber Testing, Training, and Data Set Generation," 2021.

[5] M. Ring et al., "A Survey of Network-based Intru- sion Detection Data Sets," *Computers & Security*, 2019.

[6] L. D. Manocchio, S. Layeghy, and M. Portmann, "FlowGAN — Synthetic Network Flow Generation Using Generative Adversarial Networks," in *Proc. IEEE CSE*, Shenyang, China, 2021, pp. 168–176. doi:10.1109/CSE53436.2021.00033.

[7] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shab- tai, "Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection," in *NDSS*, 2018.

[8] H. Kavak et al., "Simulation for Cybersecurity: State of the Art and Future Directions," *Journal of Cybersecurity*, 2021.

[9] M. M. Yamin, B. Katt, and V. Gkioulos, "Cyber Ranges and Security Testbeds: Scenarios, Func- tions, Tools and Architecture," *Computers & Se- curity*, vol. 88, 2020.

[10] J. Updyke et al., "GHOSTS in the Machine: Gen- erating Realistic Network Noise at Scale," in *Proc. USENIX SOUPS*, 2018.

[11] B. Ferguson, A. Tall, and D. Olsen, "National Cyber Range Overview," in *IEEE MILCOM*, Baltimore, MD, 2014, pp. 123–128.

[12] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling Tabular Data Using Conditional GAN," in *Adv. Neural Inform. Process. Syst. (NeurIPS)*, 2019, vol. 32.

[13] R. L. de Oliveira et al., "Using Mininet for Emulation and Prototyping Software-Defined Networks," in *IEEE COLCOM*, Bogota, Colombia, 2014, pp. 1–6.

[14] B. Chng et al., "CRaaS: Cloud-based Smart Grid Cyber Range for Scalable Cybersecurity Experiments and Training," in *IEEE SmartGridComm*, 2024, pp. 333–339.