



RED TEAMING AUTOMATION SIMULATING MULTI-STAGE CYBER ATTACKS

Dr. M. Kishore Kumar *Dept. of
CSE (Data Science) CMR
Technical Campus Hyderabad,
India
Hanmanthu Anji
Dept. of CSE (Cyber Security)
CMR Technical Campus
Hyderabad, India*

Domakonda Bhargavi
*Dept. of CSE (Cyber Security)
CMR Technical Campus
Hyderabad, India*

Kunta Hemanth
*Dept. of CSE (Cyber
Security) CMR
Technical Campus
Hyderabad, India*

Abstract— Modern cyber-attacks are highly organized, phased events that exploit weaknesses in networks, applications, users, and endpoints. Red team assessments (traditional security assessments) typically require a great deal of effort and time, cannot be conducted frequently, and lack clear and consistent documentation for compliance purposes. This research presents an automated red team framework to simulate the entire attack lifecycle, including reconnaissance, initial access, privilege escalation, lateral movement, persistence, and impact of an attack. This framework utilizes a series of cooperative attack modules and can base the next actions taken on the attack conditions due to the use of adaptive logic. All simulations will be executed in an isolated location to ensure safety during their execution and to eliminate the possibility of data leakages. Automated testing of security technology produces uniform, scalable, and repeatable security assessments and delivers intelligence to allow for improved detection engineering and improved efficiency within SOCs. Moreover, the automated red team assessment framework will give organizations the capability to constantly assess their security posture and detect gaps in their defensive countermeasures much faster than traditional methods.

I. INTRODUCTION

Cyberattacks are constantly growing in scale, becoming increasingly sophisticated and automated. Many traditional defensive tactics are falling by the wayside and are no longer capable of providing security for organizations. Nowadays, attackers are using coordinated attacks, which typically comprise multi-stage intrusions. They begin with reconnaissance, followed by exploitation, followed by privilege escalation, persistence, lateral movement, and finally, impactful actions. Organizations are looking to better assess and strengthen the security of their defenses, so they are turning to red teaming as a way of simulating how a real adversary would act to identify and fix vulnerabilities before they can be exploited. However, traditional red teaming has significant limitations - it requires a large amount of manual effort, and takes up a lot of resources and time, making it difficult to do on a large scale. Recent advancements in automated adversary emulation frameworks have demonstrated that there is a significant need for automating the red team. With automated red teaming, scenarios are run in a systematic manner, using predefined Tactics,

Techniques, and Procedures, while allowing for the possibility of adaptation. The MITRE Adversary Emulation guidelines are a key component of automated red teaming. According to MITRE, it is important to mimic the behavior of an attacker using a structured methodology while aligning the methodology to the ATT&CK framework. This structure allows for the accurate replication of the behavior of an actual adversary.

The main contributions of this work are:

- Design and implementation of a multi-stage automated cyber-attack simulation covering all phases of the cyber kill chain using real adversary TTPs.
- Execution of automated attacks in a sandbox environment using controlled tools with safety mechanisms such as kill switch and snapshot to avoid production impact.
- Creation of a central telemetry repository to develop ATT&CK-mapped modules, exploit frameworks, and vulnerability scanners.
- Implementation of continuous validation processes to detect security gaps, improve detection rules, strengthen configurations, and enhance incident response workflows.

The subsequent structure of this document is laid out in what follows. Section II deals with previously reported research regarding scanning websites, URLs or manual scanning for websites and how AI can be used to create simulated multi-stage cyber-attacks as a component of the red teaming process. Section III contains a problem statement and objectives of the research. Section IV provides a description of the architecture and flow of data through the system; Section V explains how the system was implemented; Section VI outlines key findings; and Section VII includes conclusions and descriptions of system functionality.

II. RELATED WORK

Automated red-teaming research relies on standardized models of attackers, like the MITRE ATT&CK framework. This framework outlines real-world tactics and techniques used by attackers to create complex simulations. Previous

studies have defined how adversaries behave by using TTP (Tactics, Techniques, and Procedures) modeling and decision-making algorithms. These tools help guide automated attack sequences, allowing for realistic simulations of actions like reconnaissance, exploitation, and post-exploitation in controlled settings. Modern tools, such as CALDERA, enhance this process by providing automated adversary-emulation features. These tools use scripted behaviors tied to ATT&CK techniques, making red-team operations repeatable and scalable. Metasploit, an open-source and modular framework, is crucial for safely testing exploits, checking for vulnerabilities, and simulating payloads. It greatly influences automated exploitation practices in both research and industry. These automated testing methods align with the NIST Cybersecurity Framework, which stresses the importance of validating detection capabilities, ensuring configuration robustness, and preparing Security Operations Centers (SOC). Network detection systems like Snort contribute to this field by demonstrating how attack simulations can relate to real-time intrusion-detection signatures, helping assess defensive visibility and response effectiveness during automated red-team tests. Improvements in the field highlight the need for automated planning and actions for emulating adversaries. This focus is on creating realistic attacker sequences and making decisions based on the environment to enhance the accuracy of simulations. The MITRE guidance on adversary emulation and red-teaming practices provides a framework for structuring safe and traceable simulations that mirror real attacker behavior while lowering operational risks. Recent research has introduced frameworks for simulating cyber-attacks that include reconnaissance scanning, vulnerability mapping, exploitation simulations, lateral movement modeling, and telemetry validation. This supports ongoing security assessments in business environments. New studies are exploring how AI and automation can boost adversary planning, dynamic TTP selection, and adaptable attack-path generation. This trend shows a movement towards smarter red teaming, with automation systems that can adjust to evolving threats. In our project, the attack simulation starts with automated setup and telemetry integration. The orchestrator then conducts reconnaissance using tools like Nmap, Nikto, and Wappalyzer CLI or WhatWeb for identifying ports, services, and technology. The scan results are compared to known adversary techniques from the ATT&CK framework and vulnerability profiles generated through TTP modeling. Next, the system carries out controlled exploitation simulations using safe Metasploit modules and adversary actions inspired by CALDERA, all within a sandbox environment. Results are analyzed alongside IDS/IPS detections from tools like Snort and SIEM telemetry, ensuring they meet NIST visibility standards. The framework also checks for privilege escalation and lateral movement by comparing observed system behavior to expected responses mapped in ATT&CK. Telemetry data is examined to find detection gaps, misconfigurations, and response delays, which support continuous security improvements using the automated adversary-emulation research. Compared to existing systems, Mind Mend aims to integrate these advancements.

AI-driven attack orchestration allows for flexible and smart choices in planning and executing steps for gathering information, exploiting weaknesses, and actions after

exploitation. These steps are based on the real tactics, techniques, and procedures used by actual threats.

- A thorough simulation process includes steps like gathering information, checking for weaknesses, carefully testing security flaws, increasing access levels, and moving laterally within a controlled space.
- Centralized monitoring and analysis of data helps connect events from security systems, intrusion detection systems, and endpoint monitoring to evaluate the readiness of the Security Operations Center and identify any visibility gaps.
- The management of the environment is done safely and automatically, using features like snapshots, emergency shutdowns, and separate network areas to avoid affecting production systems.

This approach really puts the project in a comprehensive spot. We're looking at a red-team automation ecosystem here, which goes beyond just being a single exploitation framework or a vulnerability scanner. Instead, it creates a repeatable, secure, and smart way to simulate multi-stage cyber-attacks.

III. PROBLEM STATEMENT AND OBJECTIVES

These days, many security assessment tools tend to operate in isolation. They might offer features like vulnerability scanning, exploit testing, or manual adversary emulations, but they often miss the mark when it comes to creating a comprehensive, automated, multi-stage red-teaming pipeline. Some of the common hurdles we face include:

- A lack of integration across the different phase reconnaissance, exploitation, and post-exploitation.
- Automation is quite limited, which means we still rely heavily on manual decisions from operators.
- There's also a significant gap in isolation and safety controls, which can elevate operational risks during testing.
- Telemetry often ends up being fragmented, making it tough to assess the overall detection quality of the Security Operations Center (SOC).
- Lastly, there's a noticeable absence of adaptive feedback mechanisms that could help refine attack paths or detection strategies over time.

So, the core issue we're tackling here is straightforward: we need to design and develop a fully automated red teaming framework. This framework should be capable of simulating multi-stage cyber-attacks that align with real adversary tactics, techniques, and procedures (TTPs). It's important that this all happens within a secure, sandboxed environment, enabling centralized detection analysis, continuous validation, and adaptive improvements to our security posture.

Based on this statement, the specific objectives of the systems are:

Following this statement, the systems have clear goals:

1. Each one simulation a multi-stage attack automatically: The simulations have all steps of cyber-attacks filled out and include recon, pinpointing weak points, breaching weak points, taking more rights, and showcase movement of weak points. We do so while sticking to techniques of MITRE ATT&CK.
2. Each one drives the attack by AI: Decision is involved in the attack system--or by rules--whereby it picks paths based on the services it finds can be breached, what it leaks, and what is around in the setting.
3. Each one evaluates detection in one place: i.e. Collecting and making sense of logs coming from SIEM, IDS/IPS, the end point telemetry to tell how good SOC visibility is, how right alerts are, and how well defenses work.
4. each one is safe to run and can be switched via sandboxing: Network Go back, snapshots, and kill switches made it possible to keep action simulation inside a set amount of space and not change the places where work happens.
5. Toolchain works well together: It transversely paints the reconnaissance tools like Nmap, Nikto, weak point scanners, Exploitation like Metasploit, and adversary emulators like CALDERA from a unified timeline.
6. Learning is flexible: Learn from the results of the validation, missed detections, and response analysis and change what simulations to do, attack ways, and detection rules for future buildup.
7. Unknown security activity shows: Provide dashboards to understand detection span, attack course, used vectors, and repeatable weak points from test cycles.
8. All should respect safety and compliance: The framework should follow safety testing, safe test tips and follow step procedures by the company's cybersecurity norms.

IV. SYSTEM ARCHITECTURE AND DESIGN

Red Teaming Automation Architecture is made to be a Fully-stack system for Security, Scale, and Modularities where Continuous simulation and the study of cyber-attacks Automated.

A. Overall Architecture

The Red Teaming Automation Architecture creating and changing The Operator Console (Player UI/API) is also a type of simulation. provides snappy controls for setting up scenarios. and watching the actual attack happen and what it will be like. All the computer's instructions are done in the Back end, and it exposes APIs. liable for the scenario definition and communication with the AI decision unit, control of deployed agents, and storage of Forensic data A centrally located, cloud-based relational database, called the Reporting Database - POSTSGGUL, maintains many logical tables for audit and analysis on the front end: The last one offers fast controls to define scenarios, and to see the attack steps today in real-time, and bring back the bots to rear-ender. The back end oversees all automation logic, exposing APIs that are used for scenario definition, communication with the AI decision unit, control of the deployed agents, and storage of forensic data.

A centrally located cloud-based relational database, called the Reporting Database (POSTSGGUL), maintains logs for

multiple logical tables for audit and analysis about the previous:

- Simulation Runs: Metadata for each automated attack cycle.
- Attack Paths: Step-by-step logs of sequential execution of successful and failed TTPs.
- Detection Events: Records of security alerts triggered in the target environment (e.g., EDR/SIEM).
- Vulnerability Findings: Gaps in security automatically identifiedeldequately representing existing timings.
- Remediation/Remediation Status: View of the team.

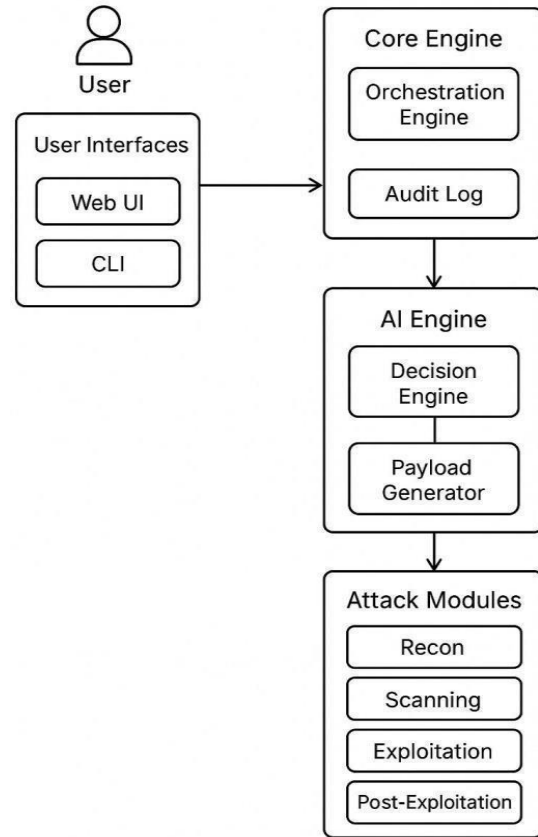


Fig. 1. Overall Architecture for red teaming automation.

B. Data Flow

The data flow can be described in six main processes:

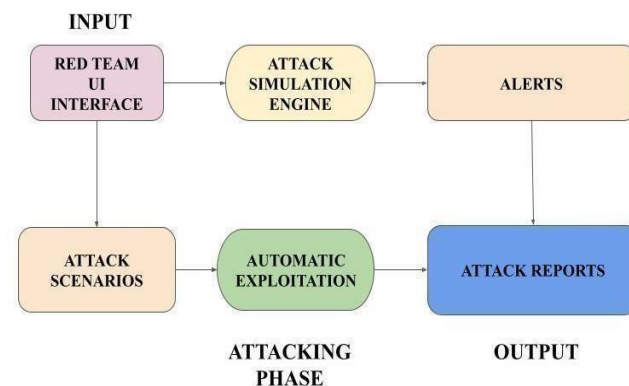


Fig. 2. Data flow diagram for red teaming automation.

1. **Red Team UI Interface:** This is the operator's control panel that is used to define the scope of the simulation, set

goals, and set Rules of Engagement (ROE). It is the first point that allows one to pick the target environment and choose the TTP based scenarios.

2. **Attack Scenarios:** This applies to the logic and data that are structured and use pre-set attack graphs and TTP sequences that are aligned with MITRE ATT&CK. It supplies the automation engine with the multi-stage plan that drives the path from initial access to the completion of the goal.
3. **Attack Simulation Engine:** This is the core that compiles the attack logic into sequential steps for the inaction agents to follow. It manages the attack state and has decision flow algorithms that chooses the next TTP based on feedback from the target during the attack process.
4. **Automatic Exploitation:** This is where the offensive actions happen in a controlled Environment. It uses tools like safe Metasploit Modules and CALDERA action to mimicurfexploiting a weakness, privilege raise, and lateral movement.
5. **Alerts:** This provides the observability of the objective environment's defense in real time. It receives logs and detections and links each step choice to the response of the security system (e.g., Snort and SIEM telemetry) as a measure of resistance.
6. **Attack Reports:** This condenses all the data for records from the simulation and the defense monitoring system. It provides a BCP biased report that documents the attack path, maps of how the initial attack is defeated, and analysis for gaps in detection, improper configuration, and response time that could lead to increased security.

V. IMPLEMENTATION

Red Teaming Automation was able to transform the blueprints of the system into a working, modular prompting system to map out cyber-attack kill chains. Important implementation points were provided in the Frontend, Backend/Orchestration and Data/Deployment levels.

A. Frontend

The user interface has been made to make it possible for users to control complicated ways of attacking a system and to make it easy to show us how well (or how bad) our security measures are working.

- **Technology Stack:** Here is the rewritten version: Built on a well-known web framework for example a Web UI (Flask App) as shown in the diagram. It might have used a file of nice parts (like React/Next.js or something similar) for fast creating and making it look good. The layer above takes care of the part where a user uses words to get in.
- **Major UI Modules:**
 - **Web UI (Flask App User Authentication):** The primary interface for secure access. It handles user login and session management, providing validated access to the system.
 - **Dashboard:** The central control panel and visualization module. It highlights the core system status, active scenarios, and provides intuitive navigation to reports and configuration settings.
 - **Command-Line Interface (CLI):** Provides an

alternative, efficient interface for power users and for scripting interaction with the Core Orchestration Engine.

- **Report Viewer:** Renders detailed HTML and PDF Reports, allowing analysts to view and filter findings, attack chains, and remediation steps.
- **Configuration Pages:** Interfaces for defining new scan targets and customizing attack playbooks.

B. Backend and AI Integration

The backend acts as the Core Orchestration Engine.

Managing multi-stage attack simulation and integrating the AI decision-making capability.

- **Technology Stack:** The core logic of the system lies in Core orchestration Engine main.py - likely implemented in Python to leverage its strong libraries for security tools.
- **Core Backend Services:** These controllers manage the execution of the cyber-attack kill chain:
 - **Reconnaissance Phase:** Executes tools like whois and dnsenumeration to gather initial intelligence on the target.
 - **Scanning Phase:** Executes vulnerability and misconfiguration checks using tools like nuclei, arachni, and sqlmap-scan.
 - **Exploitation Phase:** Manages the execution of exploits using tools like sqlmap and metasploit, or running Custom Scripts for specific vulnerabilities like SQLi, XSS, RCE.
 - **Post-Exploitation Phase:** Orchestrates multi-stage attack completion, including establishing Persistence like corn jobs, Lateral Movement, and Privilege Escalation attempts.
- **AI Engine Integration:** The orchestration engine communicates with the AI Engine which acts as the Decision Engine and Payload Generator. This service uses AI/ML logic to:
 - **Decision Engine:** Analyze scan results and prior steps to intelligently select the next TTP (Tactics, Techniques, and Procedures) in the attack chain, simulating an adaptive attacker.
 - **Payload Generator:** Generate optimized, custom payloads based on identified target characteristics and vulnerabilities, enhancing the realism of the simulation.

C. Database and Deployment

The system's database strategy involves the usage of a robust platform. like PostgreSQL or MongoDB for Data Persistence, Organizing the data into three key logical data sets: the Scan History Database - forensic records of attack steps; the User Databases are authentication and RBAC, and the Audit Log like system activity records. For Deployment, the entire application is containerized example Docker to ensure tool stability and is hosted on a secure cloud platform including AWS/Azure/GCP for necessary scaling. Git is used for version control, enabling

The figure presents an outline of a dashboard analytics used in a red teaming security assessment platform. The information consists of stats relative to the process of scanning such of the total scanned (8) as well as the entire detached targets (1).

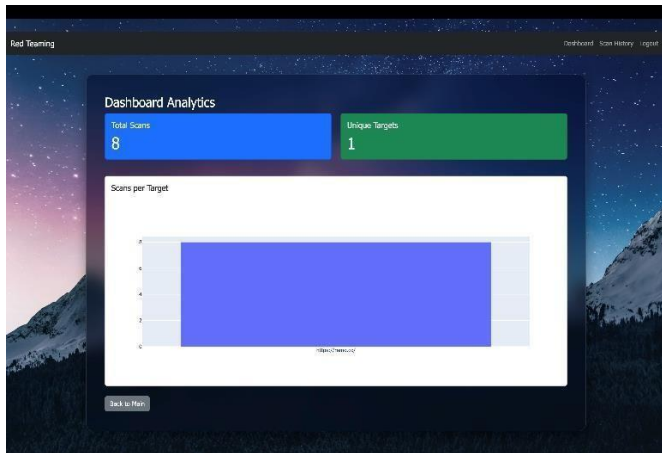


Fig. 11. Dashboard Showing Scan Statistics and Target Analysis.

There will be a bar graph titled “Scans per Target”, This graph shows the number of scans that are taken up on each target domain; the bar graph enables the user to find out about the speed of each scan on target system.

1. Scan History

The images show a dashboard for watching a robot’s security test work. The first big picture shows a grid of times when a test was run. The grid shows many times the test was run, and each test was run against a single target. The table has a number for each row, and a time so it’s clear when it was run then.

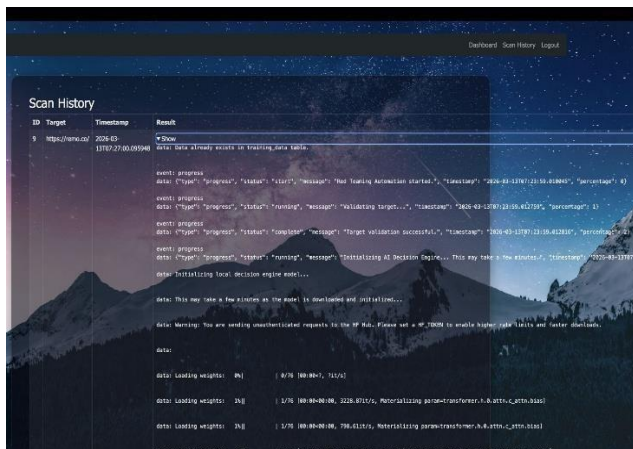


Fig. 12a. Automated Vulnerability Scan History Dashboard

Choosing a specific scan results in a detailed execution log showing real-time results of the underlying automated process. The technical output reveals specific critical stages of the process, including validating the target URL, checking for existing training data, and initializing a local AI Decision Engine.

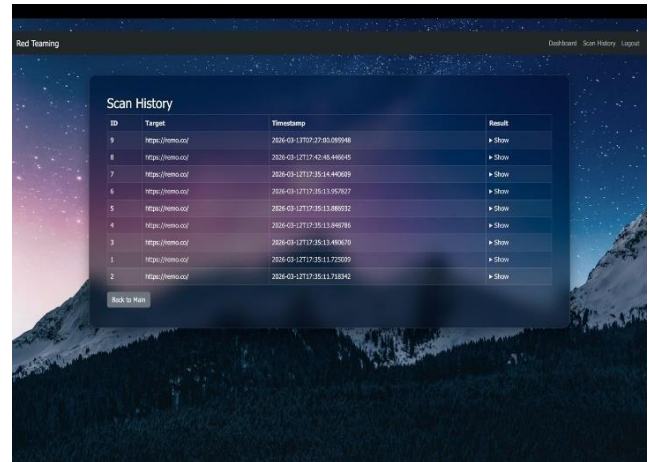


Fig. 12b. Automated Vulnerability Scan History Dashboard

The logs also reveal that the system utilizes transformer-based models, as seen by the loading of specific neural network weights and parameters to implement an advanced, AI-driven process for identifying security vulnerabilities.

J. Docker Logs.

The above image shows the interface of the Docker Desktop application monitoring a running container named “batch3project-red-teamer-1”. This container is currently executing several automated reconnaissance actions as part of a red team security assessment. The logs display DNS tracing, IP resolution, network mapping, etc., all of which are related to domains being proxied through Cloudflare. This is all part of the enumeration stage of the attack where the host makes use of tactics of the network such as domain records, IP ranges, etc. Further on in the logs, the container initiates a port scan using Nmap looking for open ports such as HTTP (80), HTTPS (443), HTTP proxy (8080) and HTTPS alt (8443).

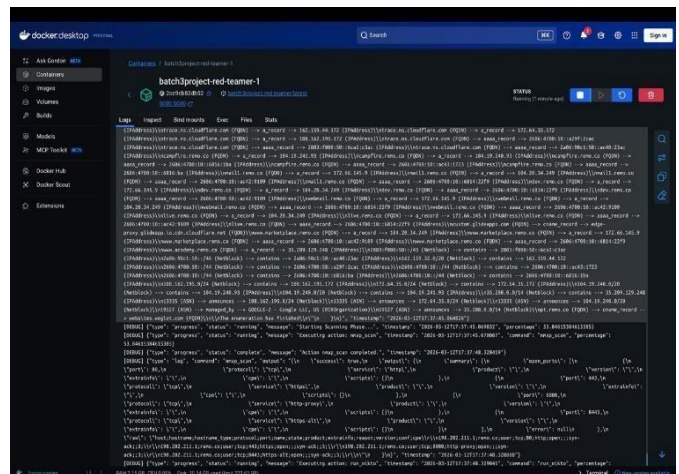


Fig. 13. Containerized Network Scanning Dashboard.

The output is an outline of the services observed and the network reply for the services. This is only a sample of the use of containerization to automate pen test workflows while also watching the actions of scans of the reconnaissance and vulnerability in real time.

VII. CONCLUSION AND FUTURE WORK

The proposed Automation of the Red Team: Multistage Cyber-Attack Simulation with MITRE ATT&CK Framework offers an effective way to evaluate and improve organizational cybersecurity defences, through automated adversary emulation. The system creates a controlled and repeatable setting for simulating realistic multi-stage cyber-attacks based on MITRE ATT&CK techniques. It covers phases like reconnaissance, vulnerability scanning, exploitation, privilege escalation, and lateral movement. By integrating automated workflows, telemetry collection, and security monitoring tools like IDS/IPS and Snort, the framework allows for continuous checking of detection abilities and finding security gaps. The use of sandboxed environments makes sure that simulations are done safely, without impacting production systems. Overall, the automated red-teaming framework cuts down manual effort, boosts testing efficiency, and offers valuable insights that help organizations improve detection rules, enhance incident response strategies, and increase overall cyber resilience.

Future work will concentrate on:

- An AI-driven system selects attack paths that adapt based on current environments and changing threats.
- Enhanced systems now collect telemetry and perform standards, providing audit-ready reports for organizational policies.
- AI-powered simulations emulate adversaries using digital twins, synthetic data, and automation-as-code, allowing continuous evaluation of cyber-attacks in a safe manner.
- real-time analytics, boosting detection coverage and security monitoring details.
- The solution seamlessly integrates with both cloud-native and hybrid infrastructures, extending security validation across the enterprise.
- Dashboards offer advanced reporting and visualization, making it easy to pinpoint attack timelines, identify detection gaps, and set remediation priorities.
- Automated workflows recommend solutions and help teams quickly fix vulnerabilities that are detected.
- Compliance tracking ensures adherence to security standards providing audit-ready reports for organizational police

REFERENCES

1. Cisco, "Snort - Network Intrusion Detection & Prevention System," [online] Available: <https://www.snort.org>.
2. MITRE Corporation released a report on Automated Adversary Emulation, covering both planning and execution aspects in 2021. You can find it online at: <https://www.mitre.org/sites/default/files/2021-11/prs-18-0944-1-automated-adversary-emulation-planning-acting.pdf>.
3. To dive into adversary emulation and red teaming, check out the guide by MITRE ATT&CK, accessible here: <https://attack.mitre.org/resources/get-started/adversary-emulation-and-red-teaming>
4. SpringerOpen's Cybersecurity Journal provides insights into cyber attack simulations and security analysis. Access their research here: <https://cybersecurity.springeropen.com/articles/10.1186/s42400-025-00361-w>.
5. Cisco Systems, Snort – Network Intrusion Detection and Prevention System. [Online]. Available: <https://www.snort.org>.
6. Applied Generative AI in Security offers recent findings on generative AI security research, available for viewing at: <https://applied-gai-in-security.ghost.io>
7. The National Institute of Standards and Technology (NIST) offers the Cybersecurity Framework, which can be accessed online at <https://www.nist.gov/cyberframework>.
8. Also, the MITRE Corporation presents MITRE CALDERA, an automated platform for adversary emulation. You can find it at <https://caldera.mitre.org/>.
9. Rapid7, Metasploit Framework Documentation. [Online]. Available: <https://www.metasploit.com>.
10. MITRE Corporation, CALDERA: Automated Adversary Emulation System Documentation. [Online]. Available: <https://caldera.mitre.org>.