Cyber Defense with Secure LLM

Sabitha K[1], Kadhiravan EG[2], Hanushree L[3], Dharshini L[4], Dharsni S [5]
[1]AP, [2-5]Student, Dept of Computer Science Engineering with cyber security
Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India

***Abstract*—The rapid growth of digital technologies and online systems has led to an increase in cybersecurity threats, making it difficult for individuals to understand and analyze attacks using traditional learning methods. This project proposes a cybersecurity simulation system called "Cyber Defense with Secure LLM" that automates the execution and analysis of common cyber attacks using Artificial Intelligence. The system performs attacks such as reconnaissance, brute-force login, SQL injection, and data theft in a controlled environment, and processes the results to extract meaningful insights. A locally deployed Large Language Model (LLM) is used to analyze the attack behavior and generate clear explanations for better understanding.**

The proposed system identifies attack patterns, detects vulnerabilities, and determines root causes such as weak authentication, insecure database queries, misconfigurations, and data exposure risks. It also presents the analysis from both red team (attack perspective) and blue team (defense perspective), providing practical knowledge of how attacks occur and how they can be prevented. In addition, the system generates simple human-readable explanations and suggests effective countermeasures, reducing the need for expert guidance and improving learning efficiency.

A user-friendly web interface is developed to display logs, attack summaries, analysis results, and recommended solutions in an easy-to-understand format. The system also includes features such as a chat assistant for clearing doubts, a history module to review past activities, and an option to download a complete report in PDF format. By integrating cybersecurity concepts with Generative AI, the project enhances practical understanding, strengthens defensive awareness, and supports modern approaches to intelligent cyber defense systems.

***Keywords*— Cybersecurity Simulation, Large Language Models (LLMs), Attack Analysis, Red Teaming, Blue Team Defense, Automation.**

I. INTRODUCTION

The rapid advancement of digital systems and internet-based applications has led to a significant rise in cybersecurity threats across various platforms. These threats include unauthorized access, data breaches, and application-level attacks that can compromise system integrity and user data. Understanding such attacks using traditional theoretical methods has become increasingly difficult, as they do not provide practical exposure or real-time analysis. Moreover, manual analysis of cyber attacks requires expert knowledge and is often time-consuming, making it challenging for beginners to clearly understand the nature and impact of different security threats.

To address these challenges, this project introduces a cybersecurity simulation system titled "Cyber Defense with Secure LLM", which leverages Large Language Models (LLMs) to automate the analysis and explanation of cyber attacks. The system simulates common attacks such as reconnaissance, brute-force login, SQL injection, and data theft within a controlled environment. It then processes the results to identify vulnerabilities, analyze attack behavior, and determine the root causes using AI-based reasoning. This approach reduces the need for manual intervention, improves analysis accuracy, and provides faster understanding of cybersecurity concepts.

In addition, the proposed system enhances user experience by generating simple human-readable explanations and suggesting effective defense strategies through an interactive interface. It also includes features such as activity logs, a chat-based assistant for queries, and a history module to review past simulations. A report generation option allows users to download detailed analysis in PDF format. By integrating cybersecurity practices with Generative AI, the system aims to improve practical learning, strengthen defensive awareness, and support intelligent approaches to modern cyber defense.

II. LITERATURE SURVEY

The increasing number of cybersecurity threats has made it essential to develop effective methods for understanding and analyzing attacks in modern

systems. In earlier approaches, cybersecurity learning and analysis were mainly based on theoretical study or static tools that provided limited practical exposure. These traditional methods lack real-time interaction, scalability, and the ability to clearly demonstrate how attacks occur and how defenses can be applied. As a result, users often find it difficult to gain a deep understanding of cyber threats and their impact.

To overcome these limitations, researchers have introduced Artificial Intelligence-based solutions in the field of cybersecurity. These approaches use machine learning techniques to detect anomalies, identify attack patterns, and enhance threat detection. While such methods improve automation and accuracy, they often require large datasets, complex model training, and significant computational resources. Additionally, many of these systems focus only on detection and do not provide clear explanations or guidance for defensive actions.

Several existing systems have implemented simulated environments and security tools to demonstrate cyber attacks such as intrusion detection and vulnerability scanning. Although these systems offer practical exposure, they are often complex to configure and may not be suitable for beginners. Moreover, they primarily emphasize attack identification and lack the ability to explain the attack process in a simple and understandable manner or provide step-by-step defense strategies.

With the advancement of Generative AI, Large Language Models (LLMs) have shown strong capabilities in understanding and explaining complex cybersecurity concepts. These models can analyze attack behavior, interpret system responses, and generate human-readable explanations along with recommended defense measures. This makes them more effective compared to traditional machine learning approaches, especially in educational and interactive environments.

However, challenges such as system integration, performance efficiency, and real-time analysis still exist. To address these issues, this project proposes a cybersecurity simulation system integrated with an LLM that automates attack execution, analysis, and explanation. The system provides both red team and blue team perspectives, along with user-friendly insights and recommendations, thereby improving learning efficiency and reducing the need for expert guidance.

III. EXISTING FRAMEWORK

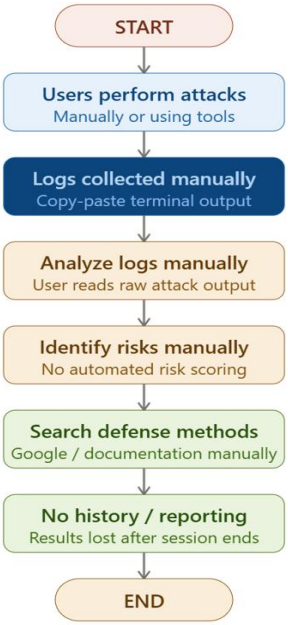

Fig 1: Manual threat analysis process

In current digital environments, various applications and systems are increasingly exposed to cybersecurity threats such as unauthorized access, data breaches, and injection attacks. Existing learning and analysis methods mainly rely on theoretical concepts or basic security tools, which provide limited insight into real-time attack scenarios. These approaches are often used only for demonstration or detection purposes, without offering deeper understanding or intelligent analysis of cyber threats.

At present, the study and analysis of cyber attacks are mostly carried out manually by learners or security professionals. This involves observing system behavior, identifying vulnerabilities, and interpreting technical outputs from security tools. Since attack patterns and system responses can be complex and dynamic, this process requires strong technical knowledge and experience, making it difficult for beginners to fully understand the concepts.

After identifying a security issue, the root cause is usually determined through manual investigation, followed by applying appropriate defense mechanisms. However, this approach is time-consuming and may delay the learning or response process, especially when dealing with multiple types of attacks. Additionally, traditional methods do not provide structured guidance or clear explanations for understanding how attacks occur and how they can be prevented.

Furthermore, manual analysis increases the chances of errors, confusion, and incomplete understanding during the learning process. Overall, the existing system lacks automation, does not provide clear attack explanations, and fails to offer intelligent defense suggestions. As a result, cybersecurity learning remains less effective, less interactive, and difficult to apply in practical real-world scenarios.

IV. PROPOSED FRAMEWORK

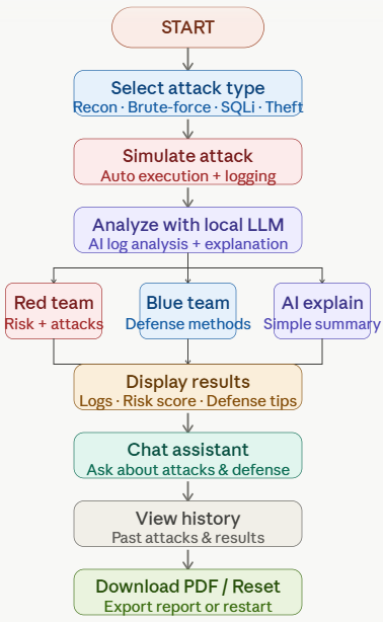

Fig 2: cyber defense

The proposed system introduces an intelligent and automated cybersecurity simulation framework using Large Language Models (LLMs). Unlike traditional approaches, it minimizes manual effort by automatically executing and analyzing various cyber attacks in a controlled environment. The system provides a practical and interactive way to understand both attack techniques and defense mechanisms in an efficient manner.

Initially, the system simulates common cyber attacks such as reconnaissance, brute-force login attempts, SQL injection, and data theft. These activities are carried out within a secure setup, and the generated outputs are collected as logs. The logs are then preprocessed to remove unnecessary data and converted into a structured format by extracting key details such as attack type, time of execution, response status, and system behavior.

The processed data is then analyzed using a locally deployed Large Language Model, which interprets patterns and identifies the nature of each attack. The model determines possible vulnerabilities, explains the attack process in simple terms, and provides both red team (attack perspective) and blue team (defense perspective) insights. Additionally, it suggests effective countermeasures and best practices to prevent similar threats in the future.

Finally, the results are displayed through an interactive web-based interface where users can easily view attack summaries, detailed analysis, risk levels, and recommended solutions. The system also includes features such as activity logs, a chat assistant for queries, and a history module for reviewing previous simulations. A report generation option is provided to download a complete analysis in PDF format, enabling better understanding and efficient cybersecurity learning.

V. METHODOLOGY

The proposed Cyber Defense with Secure LLM system is developed using a structured and modular methodology to automate cyber attack simulation and analysis in an efficient manner. The overall workflow is divided into multiple stages to ensure accurate interpretation, practical learning, and effective results.

Initially, the system simulates various common cyber attacks such as reconnaissance, brute-force login attempts, SQL injection, and data theft within a controlled environment. These activities generate system responses and logs, which are either automatically captured or stored for further processing. The generated data may include request details, response messages, timestamps, and system behavior during each attack.

In the next stage, preprocessing is performed where unnecessary or irrelevant information is filtered out, and important attributes such as attack type, execution time, response status, and error messages are extracted. This step organizes the data into a structured format, improving the quality and accuracy of further analysis.

After preprocessing, the structured data is provided to a locally deployed Large Language Model (LLM) using prompt-based techniques. The model analyzes the context of each attack, identifies patterns, and detects abnormal or malicious behavior. Based on this analysis, it classifies the type of attack and evaluates its potential impact on the system.

The next phase focuses on detailed analysis and interpretation, where the system determines the root cause of the attack, such as weak authentication, improper input validation, system misconfiguration, or data exposure vulnerabilities. Following this, the system generates clear explanations along with recommended defense strategies and best practices to prevent similar attacks in the future.

Finally, the analyzed results are presented through an interactive web-based interface, displaying attack summaries, explanations, severity levels, and suggested countermeasures in a simple and user-friendly format. Additional features such as logs, chat assistant support, history tracking, and PDF report generation enhance the overall usability. This complete methodology ensures automated, efficient, and intelligent cybersecurity learning using AI-based techniques.

VI. RESULTS AND DISCUSSIONS

The implemented web-based interface provides an interactive and user-friendly platform for displaying attack summaries, detailed analysis, and recommended defense strategies. This improves the overall usability of the system and makes cybersecurity concepts easier to understand, even for users with limited technical knowledge. The clear presentation of both red team and blue team perspectives helps users gain practical insight into how attacks occur and how they can be prevented.

The proposed system also demonstrates good performance while handling multiple simulated attacks within the controlled environment. It is capable of processing different types of attack scenarios such as reconnaissance, brute-force attempts, SQL injection, and data theft efficiently. However, the system performance may vary depending on the complexity of the attacks and the response time of the integrated Large Language Model (LLM) used for analysis and explanation.

Overall, the results indicate that the proposed approach significantly enhances cybersecurity learning by reducing manual effort, improving understanding of attack patterns, and providing faster analysis. The integration of Large Language Models with cybersecurity simulation proves to be an effective method for delivering intelligent, automated, and practical learning experiences in modern security environments.

S No	Parameter	Existing – Manual Process	Proposed – Cyber Defense LLM	Improvement (%)
1	Attack simulation time	15–30 minutes required to manually set up and run each attack	Automated simulation completed in under 30 seconds	~98% faster
2	Accuracy of analysis	65–75% accuracy — depends heavily on user skill and experience	85–90% accuracy using local LLM-based automated log analysis	+15–25% accuracy
3	Threat explanation effort	High — requires expert knowledge to interpret raw attack logs	Low — plain-language AI explanation auto-generated after each attack	~70% reduction
4	Defense suggestion speed	10–20 minutes manually searching documentation or online resources	Instant blue team defense methods provided after every simulation	~90% faster
5	Risk identification	No automated scoring — user manually judges severity of each attack	Red team risk level auto-classified and displayed per attack type	Fully automated
6	History & reporting	No logs saved — all results are lost once the session ends	Full attack history stored; complete PDF report downloadable anytime	Fully added
7	System reliability	Moderate — results vary based on user expertise and manual effort	High — consistent and repeatable AI-assisted output on every run	Strongly improved
8	Scalability	Limited — difficult to handle multiple attack types in one session	Efficient handling of all 4 attack types with chat assistant support	Fully scalable

Comparison Between Manual threat analysis and cyber defense with secure llm

VII. CONCLUSION

The proposed Cyber Defense with Secure LLM system provides an effective approach for understanding and analyzing cybersecurity threats in a practical and automated manner. By utilizing Large Language Models, the system simplifies the process of attack simulation and analysis, enabling automatic identification of attack types, detection of vulnerabilities, and generation of suitable defense strategies with minimal human effort.

The system improves the understanding of complex cybersecurity concepts by converting technical attack details into simple, human-readable explanations. This helps users quickly grasp how attacks occur and how they can be prevented, thereby enhancing learning efficiency and reducing the time required for analysis and response. The integration of features such as logs, chat assistance, and report generation further supports a comprehensive and interactive learning experience.

Overall, the project demonstrates the successful integration of Artificial Intelligence with cybersecurity practices and highlights its potential in creating intelligent, automated, and user-friendly security learning systems. It contributes to improving awareness of cyber threats and supports the development of effective defense mechanisms in modern digital environments.

VIII. FUTURE WORK

The system can be improved by enabling real-time monitoring of cyber attacks, allowing it to detect and analyze threats instantly as they occur. This enhancement will make the system more proactive and effective in preventing security issues at an early stage.

It can also be extended to support a wider range of attack types and different environments such as web, network, and IoT systems. This will increase the scalability of the project and make it applicable to real-world cybersecurity scenarios.

Future enhancements may include automated defense mechanisms and integration with advanced security tools. Additionally, improving the AI chat assistant will allow users to interact more easily and understand cybersecurity concepts in a simple and user-friendly manner.

REFERENCES

[1] Amazon Web Services, AWS Security Documentation. Available: https://docs.aws.amazon.com/security/
[2] Amazon Web Services, AWS CloudTrail User Guide. Available: https://docs.aws.amazon.com/cloudtrail/
[3] Amazon Web Services, AWS CloudWatch Monitoring Guide. Available: https://docs.aws.amazon.com/cloudwatch/
[4] OpenAI, OpenAI API Documentation. Available: https://platform.openai.com/docs
[5] OpenAI, Research on Large Language Models. Available: https://openai.com/research
[6] OWASP Foundation, OWASP Top 10 Web Application Security Risks. Available: https://owasp.org/www-project-top-ten/
[7] MITRE, ATT&CK Framework. Available: https://attack.mitre.org/
[8] IEEE Access, “Machine Learning for Cybersecurity Threat Detection,” 2022. Available: https://ieeexplore.ieee.org/
[9] ACM Digital Library, “AI-Based Cybersecurity Systems,” 2021. Available: https://dl.acm.org/
[10] NIST, Cybersecurity Framework. Available: https://www.nist.gov/cyberframework/
[11] Elastic, Log Monitoring and Observability. Available: https://www.elastic.co/what-is/log-monitoring
[12] Elastic, Elastic Stack Documentation. Available: https://www.elastic.co/guide/index.html
[13] SANS Institute, Cybersecurity Training and Research. Available: https://www.sans.org/
[14] Cisco, Network Security Basics and Threat Defense. Available: https://www.cisco.com/
[15] Microsoft Azure, Security and Monitoring Services Documentation. Available: https://learn.microsoft.com/azure/security/