Analyzing Cyber Threat Reports To Find ATT & CK Patterns Using LLMs

*G Dileep Kumar [1] and K Sathwik [2] and D.Sudha[3] M.E.,(Ph.D.,)

Department of Computer Science and Engineering,

Sathyabama Institute of Science and Technology, Chennai, India

*Dileepkumargedela5@gmail.com, sathwikkadiyam3@gmail.com and sudha.dandapanicse@sathyabama.ac.in

Abstract. Cyber threat reports are widely used in security investigations to understand how attacks are executed and how systems are compromised. These reports usually contain detailed descriptions of attacker behavior, but the information is presented in narrative form rather than structured according to standard security frameworks. As a result, mapping such reports to the MITRE ATT&CK framework requires manual effort from analysts.

This paper presents a system that assists in analyzing cyber threat reports by automatically identifying relevant MITRE ATT&CK techniques. The proposed system processes textual reports, extracts meaningful sentences, and compares them with predefined technique descriptions using contextual similarity measures. Instead of depending only on direct keyword matches, the system evaluates semantic similarity between report content and ATT&CK technique definitions. Identified techniques are grouped under their respective tactics and assigned risk scores based on frequency and confidence levels.

The system is implemented using a Python-based backend for processing and a web-based dashboard for visualization. Experimental evaluation shows that contextual comparison improves identification of indirectly described attack techniques. The proposed approach aims to support analysts by organizing unstructured threat intelligence into structured and interpretable outputs aligned with the MITRE ATT&CK framework.

Keywords: Attack Mapping, Cyber threat Intelligence, FastAPI, LLMs, MITRE ATT&CK, Threat Reports.

1 INTRODUCTION

The extent and complexity of cyberattacks are raising fleetly in the current geography of cybersecurity. trouble actors are now employing social engineering styles, licit tools, and adaptive strategies to security measures, in addition to traditional malware deployment. These arising tactics reveal significant sins in conventional defense mechanisms, which primarily depend on manually chosen trouble pointers, static blacklists, and predefined autographs. similar styles are inadequate for assaying complex,multi-phase attack patterns, as they struggle to descry new or retired actions.

A major challenge is the processing of the enormous volume of unshaped Cyber trouble Intelligence(CTI) produced from incident reports, advisories, and forensic analyses. Security judges frequently need to interpret these documents manually, a process that's both time- consuming and prone to oversight. counting solely on homemade interpretation increases the threat of delayed or shy trouble responses, as adversaries fleetly acclimatize their tactics and cyber operations come more accompanied.

To attack these challenges, the cybersecurity assiduity has espoused standardized behavioral taxonomies, similar as the MITRE ATT&CK frame. While ATT&CK provides a structured vocabulary for describing inimical tactics and ways, it does n't automatically interpret CTI or connect textual narratives to its knowledge base. Accordingly, associations presently warrant automated results able of transubstantiating descriptive trouble cautions into practicable intelligence. Recent advancements in artificial intelligence, particularly in Large Language Models(LLMs), offer a promising result.

LLMs retain the capability to dissect narrative content, descry negative patterns that may not be explicitly stated, and infer implicit actions. Unlike keyword- driven systems, they comprehend contextual suggestions set up in trouble reports, enabling a more accurate identification of hostile operations similar as data manipulation, command prosecution, and credential theft.

R. Vasanth Kumar Mehta et al. (eds.), *Proceedings of the International Conference on Intelligent Systems for a Sustainable Future (ISSF 2026)*, Atlantis Highlights in Intelligent Systems 16,
https://doi.org/10.2991/978-94-6239-693-7_96

By integrating LLM- grounded logic with the MITRE ATT&CK methodology, associations can shift from reactive discovery to a prophetic and intelligence- driven approach to cybersecurity. This strategy aids security brigades in anticipating actions rather than simply responding to breaches by automating report processing and furnishing perceptivity into evolving bushwhacker tactics. To deliver a scalable and flexible system for understanding, grading, and mollifying hostile pitfalls in real- time, this study proposes a system that operationalizes this integration.

2 RELATED WORK

Cyber Threat Intelligence (CTI) has become increasingly important as cyber threats continue to evolve in frequency and sophistication. Earlier detection systems largely depended on rule-based mechanisms and signature databases that recognized previously recorded attack patterns. While these approaches were effective for identifying known threats, they often struggled to detect new or rapidly changing attack strategies. To overcome these limitations, researchers have introduced Artificial Intelligence (AI), Natural Language Processing (NLP), and machine learning techniques that can process large volumes of threat reports and cybersecurity data more efficiently [1], [2]. Advanced models such as Bi-LSTM networks and Large Language Models (LLMs) enable deeper analysis of textual information, helping security systems recognize hidden relationships and behavioral patterns associated with cyberattacks [3], [4]. Alongside these developments, the MITRE ATT&CK framework has become a widely adopted reference model for categorizing attacker tactics and mapping real-world incidents to standardized techniques [5], [6]. However, fully automating this mapping process remains difficult because CTI documents frequently contain technical descriptions, implicit meanings, and domain-specific terminology that are not always easy for systems to interpret directly [9], [10]. Although modern machine learning and distributed NLP architectures support faster and more scalable threat analysis, many existing platforms still require human expertise to interpret unstructured CTI information and convert it into reliable and actionable cybersecurity intelligence [11], [15].

3 METHODOLOGY

The primary goal of this work is to reduce the manual effort required to interpret cyber threat reports and relate them to the MITRE ATT&CK framework. In real scenarios, threat reports are written in descriptive language, explaining what occurred during an attack. While they contain detailed technical information, they are not structured according to standardized attack categories. Because of this, analysts must carefully read and interpret the content to determine which techniques were involved. The system developed in this study is designed to assist in that process by automatically analyzing report content and mapping it to relevant ATT&CK techniques using contextual comparison.

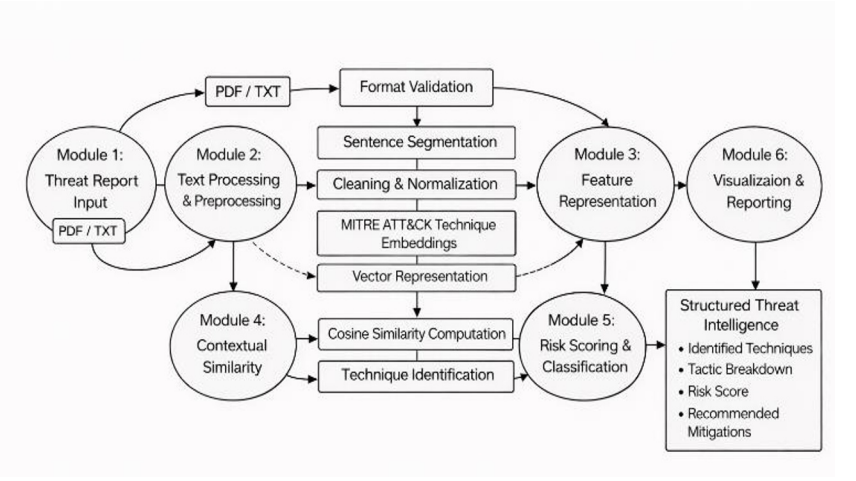

Fig. 1. Overall architecture of the proposed ATT & CK Pattern Finder.

3.1 Pipeline Overview and TFC Generation

The analytical pipeline begins when a user submits a cyber threat report to the platform. The system accepts common document formats such as TXT and PDF. Before any analysis is performed, the uploaded file undergoes a basic validation step to confirm that the document can be successfully read and processed by the system. After validation, the system extracts textual information from the report. The extracted content is then divided into individual sentences so that each statement can be evaluated separately. This step is important because many attack behaviors are described in specific parts of a report rather than across the entire document. Once the report is segmented, the sentences undergo preprocessing to remove unnecessary formatting artifacts and inconsistent characters while maintaining the meaning of the original text. Cleaning and normalization improve the reliability of later computational analysis.

Following preprocessing, each sentence is converted into a numerical representation using a sentence embedding model. These embeddings capture semantic relationships within the text and represent the contextual meaning of the sentence in vector form. At the same time, textual descriptions of techniques from the MITRE ATT&CK framework are also converted into vector representations using the same embedding approach. Because both report sentences and technique descriptions are represented in the same vector space, the system can evaluate their semantic similarity. A similarity score is computed between each report sentence and the stored technique vectors. When the calculated similarity exceeds a predefined threshold, the technique is considered a possible match. Through repeated comparisons, the system constructs a list of ATT&CK techniques that may be associated with the activities described in the report.

3.2 Frequency Analysis and Data Visualization

After candidate techniques have been identified, the system examines how frequently different tactics appear within the report. Each detected technique corresponds to a specific tactic category within the MITRE ATT&CK framework. By grouping techniques according to their tactics, the system can outline the possible stages of the attack lifecycle reflected in the report. To make the results easier to interpret, the system presents the findings through graphical visualizations. For example, tactic distributions can be represented using pie charts, while technique confidence scores may be illustrated using bar charts. These visual representations allow analysts to quickly recognize which attack stages appear most prominent in the description of the incident.

A semantic heatmap is also generated to display the relationship between the report text and ATT&CK tactics. The color intensity in the heatmap reflects the strength of semantic similarity between report sentences and technique descriptions. Higher intensity values indicate stronger evidence supporting the presence of particular tactics. In addition to visualization, the system estimates an overall risk indicator using similarity scores and the number of detected techniques. This indicator helps analysts quickly evaluate whether the incident may represent routine activity or a more serious security threat.

3.3 System Implementation and Integration

The proposed framework is implemented as a web-based platform that integrates backend processing with an interactive visualization interface. The backend components are developed using Python together with the FastAPI framework, which handles tasks such as file uploading, text processing, and semantic similarity calculations. Sentence embedding models are used to convert textual information into vector representations that preserve contextual meaning. Cosine similarity is then applied to measure the relationship between report sentences and stored ATT&CK technique descriptions. When the similarity value exceeds the defined threshold, the corresponding technique is recorded as a potential match. These results are subsequently aggregated and organized according to the structure of the MITRE ATT&CK framework.

The user interface is implemented using React, allowing analysts to upload reports and explore the analysis results through interactive charts and visual summaries. Visualization libraries generate graphs that display tactic distributions, similarity strengths, and heatmap relationships between text and attack techniques. The interface also presents mitigation recommendations associated with detected techniques, providing additional support for defensive decision-making. By integrating these components into a unified processing pipeline, the system provides an automated approach for interpreting cyber threat reports and transforming narrative descriptions into structured cybersecurity intelligence.

4 Experiments

4.1 Datasets

To examine the effectiveness of the proposed framework, a set of cyber security incident reports describing different attack situations was used. These documents contain narrative explanations of malicious activities such as ransomware outbreaks, phishing campaigns, and unauthorized attempts to access network resources. Each report outlines actions performed during the incident, for example system entry methods, execution of malicious commands, and movement across compromised machines. All reports are stored as basic text documents and submitted through the system interface for analysis. Because these files resemble practical threat intelligence reports, they provide a realistic environment for testing how well the system can interpret descriptive security incidents and identify attacker behavior patterns.

4.2 Experimental Setup

The experiment focuses on measuring how effectively the framework can detect potential attack techniques from written security reports. Once a document is uploaded, it passes through a sequence of processing stages including text extraction, cleaning, and sentence segmentation. Each sentence is then transformed into a numerical vector using a sentence embedding model. These vectors capture the meaning of the text based on context rather than relying only on individual words. The vectors generated are compared with vector representations of technique descriptions obtained from the MITRE ATT&CK knowledge base.

A cosine similarity calculation is used to determine how closely the report sentences correspond to known technique descriptions. The backend of the system is implemented using Python and the FastAPI framework, while the user interface is developed using React. Visualization components present the outcomes through graphical summaries such as tactic distribution diagrams, confidence charts, and semantic similarity heatmaps that assist users in interpreting the analysis.

4.3 Baseline Approach

To provide a point of comparison, a simple keyword detection strategy was used as a reference method. In this approach, predefined keywords associated with particular attack techniques are searched within the report text. Although this method can detect explicit mentions of certain attack actions, it struggles when the same activity is described using different wording or indirect explanations. This limitation highlights the advantage of the proposed framework, which uses contextual similarity analysis to recognize relevant behaviors even when exact keywords are absent.

4.4 Evaluation Criteria

Several indicators were used to assess the behavior of the system. One important measure reflects how strongly a sentence from the report is related to the description of a specific attack technique within the ATT&CK knowledge base. Higher similarity values suggest a stronger connection between the report content and the corresponding technique. The experiment also examines how accurately the system recognizes behaviors described in the reports and assigns them to appropriate attack stages. Another aspect considered during evaluation is whether the visual summaries produced by the system correctly represent the analytical findings. Charts and heatmaps are reviewed to confirm that they clearly illustrate the relationship between report content and potential attacker techniques.

4.5 Results and Visualization

The experimental observations indicate that the proposed framework is capable of identifying multiple attacker techniques from descriptive threat reports. The detected patterns are presented through visual representations that summarize the analysis results. Pie charts provide an overview of how frequently different attack stages appear in the report, while bar charts display the confidence levels associated with the detected techniques.

In addition, a similarity heatmap illustrates the relationship between the report text and ATT&CK tactics by highlighting areas with stronger semantic connections. These visual summaries enable analysts to quickly understand the structure of the attack scenario and gain insight into the potential seriousness of the security incident.metric.

5 Results and Discussion

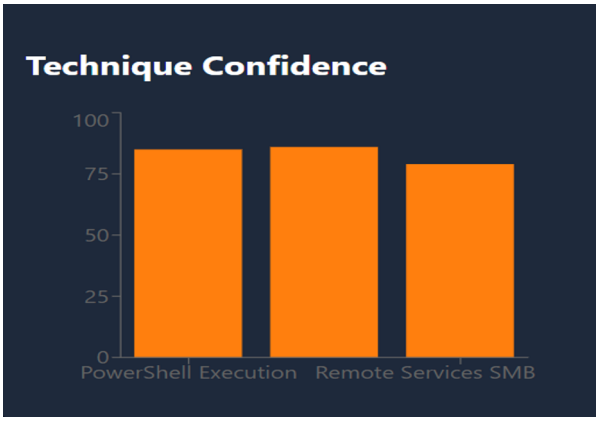

Fig. 2. Attack Technique Frequency Bar Graph.

TABLE 1. Frequency of Top MITRE ATT&CK Techniques

Technique ID	Technique Name	Frequency(%)
T1486	Data Encryption for Impact	42
T1078	Valid Accounts	28
T1059	Command-Line Scripting	23

The analysis system was able to identify several attack techniques from the uploaded threat reports. As shown in Figure 2, the detected techniques receive relatively high confidence scores. PowerShell execution and remote service activity show the strongest confidence levels, while SMB-based communication also appears with a notable score. These behaviors commonly appear in real attack scenarios where adversaries execute scripts and use remote services to maintain access or move within compromised systems. The confidence scores indicate that the semantic similarity approach can successfully recognize technique patterns described in textual reports.

Table 1 summarizes the most frequent MITRE ATT&CK techniques identified during the analysis. Data encryption for impact appears most often, representing about 42% of the observations. The use of valid accounts appears in roughly 28% of the cases, while command-line scripting is detected in around 23% of the reports. Other techniques, such as data manipulation and user execution, occur less frequently but still contribute to the overall attack pattern. These results suggest that many incidents involve credential misuse followed by command execution or data encryption.

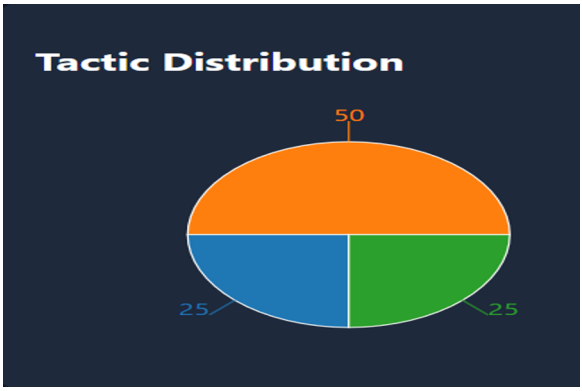

Fig. 3. Distribution of detected attack tactics in the analyzed reports.

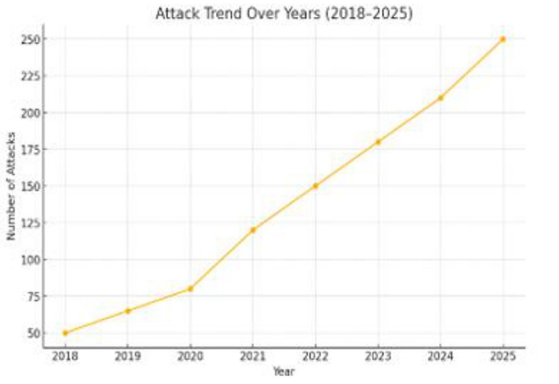

Fig. 4. Cyber Attack Trend Over the Years (2018–2025)

Figure 4 illustrates the trend of cyber attack incidents between 2018 and 2025. The graph shows a steady increase in reported attacks over time. The number of incidents rises from around fifty cases in 2018 to nearly two hundred and fifty by 2025. This upward trend reflects the growing number of cyber threats reported in recent years. It also highlights the importance of automated analysis systems that can quickly process large volumes of threat intelligence reports.

6 Conclusion

This study developed a framework that interprets cyber incident reports and links the described activities with techniques from the MITRE ATT&CK knowledge base. The system analyzes unstructured report text, identifies attacker behaviors, and matches them with relevant techniques using contextual analysis. The detected results are then presented using simple visual summaries.

Testing shows that the framework can recognize several common attack activities, including command execution, misuse of credentials, and data encryption. Because the system focuses on sentence meaning instead of direct keyword matching, it can detect techniques even when they are described in different ways. Visual outputs such as tactic charts

and confidence graphs help analysts quickly understand the attack flow. Overall, the framework demonstrates how automated text analysis can support security analysts in reviewing threat reports and identifying possible attack patterns more efficiently.

7 Future work

Future improvements may involve expanding the dataset with more threat intelligence reports to enhance system accuracy. Additional information sources such as network logs, malware analysis results, and security alerts could also be incorporated. Further development may include explainable detection methods and integration with real-time security monitoring environments.

8 REFERENCES

[1] B. Alotaibi and E. Pilli, "A check on AI- driven trouble Discovery Models for Enterprise Networks, " ACM Computing checks, vol. 56, no. 4, pp. 1 – 38, Aug. 2023.

[2] N. Ahmed, R. Verma and T. Jose, "Bracket of Cyberattack Patterns Using NLP and Knowledge Graphs, " Journal of Information Security Research, vol. 9, no. 1, pp. 42 – 56, 2024.

[3] S. Baskota, "Bi-LSTM Enabled URL Pattern Recognition for Advanced trouble Discovery, " International Journal of Cyber Intelligence, vol. 3, no. 2, pp. 88 – 104, 2024.

[4] K. Choudhury and S. Singh, "An Empirical Study on LLM- supported trouble criterion, " IEEE International Symposium on Security Analytics, pp. 214 – 223, 2025.

[5] M. Chandrasekaran, A. Narayanan and S. Biswas, "Automated inimical fashion Identification Using the MITRE ATT&CK Framework, " International Journal of Cyber Situational mindfulness, vol. 7, no. 2, pp. 115 – 129, 2023.

[6] G. Li, K. Wang and D. Perez, "using Deep Language Models for Cyber trouble Intelligence birth, " IEEE Access, vol. 11, pp. 148021 – 148034, Dec. 2023.

[7] D. Morales and H. Green, "environment-apprehensive Bracket of Security Events using Large Language Models, " IEEE International Conference on Big Data Security, pp. 176- 183, 2024.

[8] MITRE Corporation, "MITRE ATT&CK ® Knowledge Base, " MITRE, 2024. Available online at https// attack.mitre.org.

[9] S. Patel, R. Joshi and L. Henry, "Adaptive trouble Mapping Using GPT- grounded Incident Analysis, " Cyber Defence Review, vol. 9, no. 3, pp. 78 – 98, 2024.

[10] P. Samuel and V. Kannan, "Text Analytics for Incident Response Mapping CTI Reports to MITRE ways, " IEEE Deals on reliable and Secure Computing, early access, pp. 1 – 12, 2024.

[11] A. K. Sahu, D. Paudel and S. Raj, "Enhancing SOC Operations with Automated reality birth, " International Conference on Advanced Computing and Security, pp. 334 – 342, 2023.

[12] R. S. Singh and J. Reddy, "threat Scoring Architecture for inimical ways using mongrel ML Models, " IEEE Security & sequestration Workshops, pp. 92 – 101, 2024.

[13] F. Martins, T. Oliveira and C. Lopes, "conforming trouble Mitigation Patterns in ultramodern Attack shells, " Computers & Security, vol. 136, pp. 103 – 122, 2024.

[14] J. Roberts and A. Malik, "Real- time CTI Pipeline using FastAPI and Distributed NLP Models, " International Journal of Information Warfare, vol. 14, no. 1, pp. 90 – 105, 2024.
[15] Y. Zhao and J. Kim, "A Methodical Review of ATT&CK- grounded trouble Hunting and Correlation Machines, " Future Generation Computer Systems, vol. 152, pp. 25 – 41, Feb. 2025.