Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Automated Identification And Profiling Of Emerging Cyber Threats Using Natural Language Processing (NLP)

G Raju,Talari Swapna, Katam Pavani, Peetla Sai Kumar, Yadava Sai Charan, Bannoth Ramesh Naik

Department of Computer Science and Engineering Gates Institute of Technology , Andhra Pradesh, India.

Correspondence

G Raju

Department of Computer Science and Engineering, Gates Institute of Technology, Andhra Pradesh, India.

- Received Date: 11 Feb 2026
- Accepted Date: 02 Apr 2026
- Publication Date: 25 Apr 2026

Keywords

Anomaly Detection,Cyber Threat Intelligence, Emerging Cyber Threats, Machine Learning, Natural Language Processing, Text mining.

Abstract

The time window between the disclosure of a new cyber vulnerability and its use by cybercriminals has been getting smaller and smaller over time. Recent episodes, such as Log4j vulnerability, exemplifies this well. Within hours after the exploit being released, attackers started scanning the internet looking for vulnerable hosts to deploy threats like cryptocurrency miners and ransomware on vulnerable systems. Thus, it becomes imperative for the cybersecurity defense strategy to detect threats and their capabilities as early as possible to maximize the success of prevention actions. Although crucial, discovering newthreats is a challenging activity for security analysts due to the immense volume of data and information sources to be analyzed for signs that a threat is emerging. In this sense, we present a framework for automatic identification and profiling of emerging threats using Twitter messages as a source of events and MITRE ATT&CKasasource of knowledge for threat characterization. The framework comprises three main parts: identification of cyber threats and their names; profiling the identified threat in terms of its intentions or goals by employing two machine learning layers to filter and classify tweets; and alarm generation based on the threat's risk. The main contribution of our work is the approachtocharacterize or profile theidentified threats in terms of their intentions or goals, providing additional context on the threat and avenues for mitigation. In our experiments, the profiling stage reached an F1 score of 77% in correctly profiling discovered threats.

Introduction

In recent years, there has been a significant increase in reliance on the Internet for business, government, and social interactions due to rapid advancements in hyper-connectivity and mobility. While the Internet has become an essential infrastructure for modern society, it has also led to a rise in cyberattacks with diverse motivations and objectives. Preventing such attacks requires timely and accurate intelligence regarding vulnerabilities and emerging threats.

Cyber Threat Intelligence (CTI) refers to evidence-based knowledge that includes context, mechanisms, indicators, and actionable insights about existing or emerging threats. This intelligence helps organizations reduce uncertainty in detecting vulnerabilities and enables informed decision making for effective defense strategies. CTI can be obtained from both structured and unstructured sources. Structured sources, such as the Common Vulnerabilities and Exposures (CVE) database However, these sources often lack real-time updates.

In contrast, unstructured sources such as blogs, forums, dark web platforms, and social media provide real-time and early indicators of emerging threats. This type of intelligence, known as Open Source Intelligence (OSINT), plays a crucial role in identifying cyber threats before they are formally disclosed. Among various OSINT sources, Twitter has emerged as a highly valuable platform due to its real-time nature, large volume of data, and active participation from cybersecurity professionals. Security researchers, analysts, and even malicious actors frequently discuss vulnerabilities, exploits, and attack strategies on Twitter, making it a rich source of early warning signals.

However, analyzing large volumes of unstructured data from platforms like Twitter presents significant challenges. The short and informal nature of tweets, combined with high data velocity, makes manual analysis impractical. Therefore, automated techniques using Natural Language Processing (NLP) and Machine Learning (ML) are required to efficiently process, filter, and classify relevant information. These techniques enable the identification of potential threats and support the generation of timely alerts.

Another critical component in understanding cyber threats is the MITRE ATT&CK framework, which provides a comprehensive

Citation: Raju G,Talari S, Katam P, Peetla SK, Yadava SC, Bannoth RN. Automated Identification And Profiling Of Emerging Cyber Threats Using Natural Language Processing (NLP). GJEIIR. 2026;6(4):246.

knowledge base of adversary tactics and techniques based on real-world observations. It models the behavior of attackers across different stages of the attack lifecycle, including initial access, execution, persistence, and data exfiltration. By mapping detected threats to ATT&CK techniques, it becomes possible to understand the intent and behavior of attackers and develop appropriate mitigation strategies.

Despite advancements in threat detection, many existing approaches focus only on identifying or classifying cyber threats without providing deeper insights into their intent or behavior. To address this limitation, this work proposes an automated framework that leverages NLP and ML techniques to detect emerging cyber threats from Twitter data and profile them using the MITRE ATT&CK knowledge base. The proposed system not only identifies threats but also provides contextual understanding and risk-based alerts, thereby enhancing the effectiveness of cyber threat intelligence.

Related Work

Cybersecurity has become a critical concern for organizations due to the increasing frequency and sophistication of cyber threats. Security Operations Centers (SOCs) rely heavily on timely and relevant cyber threat intelligence to monitor and protect IT infrastructures. However, manual analysis of large volumes of data from multiple sources is time-consuming and inefficient, often resulting in missed insights. To address this challenge, researchers have explored the use of Open Source Intelligence (OSINT), which includes publicly available data from blogs, forums, dark web platforms, and social media such as Twitter. These sources provide early indicators of emerging cyber threats in real time.

Several studies have utilized OSINT combined with Natural Language Processing (NLP) and Machine Learning (ML) techniques to identify cyber threats. For instance, early warning systems have been proposed that use text mining to detect unknown terms from online discussions and generate alerts based on their frequency. However, such approaches rely heavily on keyword filtering, which may produce false positives and fail to provide detailed threat characterization.

Other research has focused on classifying cyber threat-related tweets using deep learning models such as Convolutional Neural Networks (CNNs). These approaches typically involve multi-stage pipelines where tweets are first classified as relevant or irrelevant and then categorized into predefined threat types such as ransomware, botnets, or vulnerabilities. While effective in classification, these methods are limited by fixed threat categories and do not identify or name emerging threats. Additionally, some approaches depend on external APIs for classification, reducing system independence.

Novelty detection models have also been explored to distinguish cyber threat-related tweets from irrelevant content by comparing them with known vulnerability databases such as CVE. Although these models improve detection accuracy, they do not address threat identification or profiling, which are essential for understanding attacker behavior and intent.

Advanced approaches using deep learning architectures, including Named Entity Recognition (NER) with Bidirectional Long Short-Term Memory (BiLSTM) networks, have been proposed to extract relevant information from tweets. These methods achieve high accuracy in identifying relevant entities and detecting cyber-related content. Similarly, Multi-Task Learning (MTL) models combine classification and entity recognition tasks to improve performance and reduce system complexity. However, these systems are often limited to specific infrastructures or predefined contexts and fail to generalize for broader emerging threats.

Unsupervised learning techniques have also been used to detect and rank emerging cyber threat events based on tweet clustering and importance scoring. While these methods can identify novel events, they do not provide insights into the intentions or behavior of the threats. Other frameworks integrate data from Twitter and dark web sources to generate early warnings, but they lack effective filtering mechanisms and detailed threat profiling, leading to higher false positive rates.

In summary, existing approaches primarily focus on detecting or classifying cyber threats but lack the ability to identify, name, and profile emerging threats comprehensively. The proposed system addresses these limitations by combining NLP and ML techniques to not only detect threats but also profile them using the MITRE ATT&CK framework and evaluate their risk, providing more actionable cyber threat intelligence.

Proposed Methodology

The primary objective of this work is to design and implement an intelligent system capable of automatically identifying and profiling emerging cyber threats using Open-Source Intelligence (OSINT). With the rapid growth of cyberattacks and the increasing use of social media platforms by security experts, attackers, and organizations, valuable threat-related information is continuously being generated online. However, manually analysing this massive volume of data is time-consuming and inefficient. Therefore, an automated solution is necessary to detect, analyse, and alert cyber security professionals about new and evolving threats in real time.

To address this challenge, the proposed system leverages data from social media platforms, specifically Twitter, and applies Natural Language Processing (NLP) and Machine Learning (ML) techniques to extract meaningful insights. The system focuses on identifying unknown or unusual terms that may correspond to newly emerging cyber threats and classifies them based on their behaviour and tactics. The overall approach involves continuous monitoring and collection of tweets from trusted cyber security experts and organizations, followed by the application of NLP and ML techniques to identify potential threat names. These identified threats are then mapped to known attack tactics using the MITRE ATT&CK framework, and real-time alerts are generated with appropriate threat characterization and risk assessment. The system is also capable of issuing alerts for both newly discovered and rapidly evolving threats.

Solution Architecture

The proposed system architecture consists of multiple interconnected components responsible for data collection, processing, feature extraction, classification, and alert generation. The first component focuses on data collection using Logstash, an open-source data processing engine. The Logstash Twitter plugin is used to connect to the Twitter Streaming API and continuously collect tweets from selected cybersecurity experts and organizations through the “follows” parameter. This process captures tweets, retweets, replies, and mentions, ensuring that the collected data is relevant and reliable.

Once the data is collected, it is stored in an Elasticsearch database, which enables efficient storage, retrieval, and analysis of large volumes of data. The database maintains tweet content along with metadata such as timestamps, user details, and engagement metrics. To support real-time processing, a sliding time-window approach is adopted, where tweets are analysed in smaller time intervals. This allows continuous monitoring and

helps track how threats evolve over time.

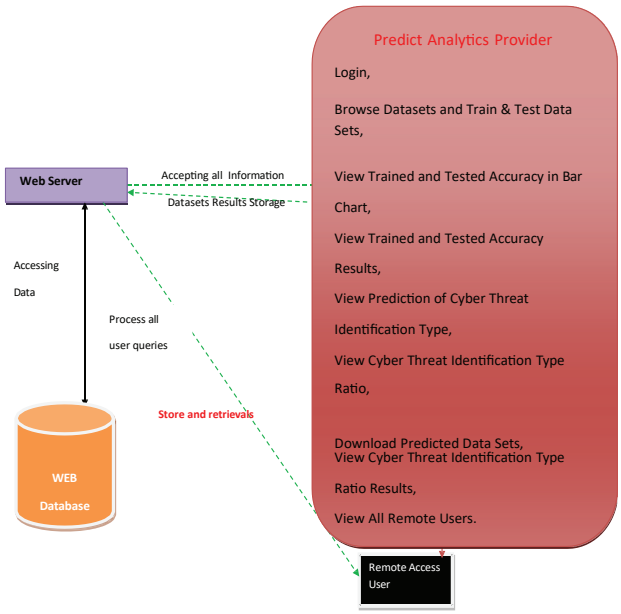
A key component of the system is the unknown word selection module, which identifies potential new cyber threat names through a series of filtering steps. Initially, normalization is applied to standardize variations in text. Irrelevant elements such as URLs, email addresses, and user mentions are removed, followed by tokenization to split text into individual words. Common dictionary words, stop words, and punctuation are eliminated to reduce noise. Named Entity Recognition (NER) is used to remove known entities such as organizations and locations, while dictionary-based filtering removes known cyber-related terms. The remaining words are treated as potential unknown threats.

To improve accuracy, the system performs cyber threat keyword co-occurrence analysis, where an unknown term is considered significant only if it appears in multiple tweets and co-occurs with known threat-related keywords. This helps eliminate random or irrelevant terms. Further processing involves tokenization and stemming using Porter's algorithm to reduce words to their root forms, ensuring consistency in analysis.

For machine learning processing, TF-IDF is used to convert textual data into numerical form by assigning weights based on term frequency and uniqueness. This transformation is applied to both tweet data and MITRE ATT&CK data. A one-class classifier is then used to determine whether a tweet is related to cyber threats. This classifier is trained using only known malicious data and filters out irrelevant content. Additionally, normalization is applied to MITRE ATT&CK data to maintain consistency.

The system also includes a multi-class classifier that categorizes detected threats into one of the 14 MITRE ATT&CK tactics, such as Initial Access, Execution, Persistence, and Exfiltration. Various models were evaluated, including Logistic Regression, Naïve Bayes, Random Forest, and Support Vector Machine, with Linear SVM selected due to its superior performance.

Once threats are identified and classified, they are stored in a threat profiling database along with additional metadata such as mentions, retweets, likes, user influence, and external references. This information is used to assess the importance and spread of each threat. Finally, the system generates real-time alerts that include the threat name, associated tactic, risk level, and supporting evidence. These alerts can be integrated with Security Information and Event Management (SIEM) systems for centralized monitoring.



System Architecture

The proposed cyber threat identification and profiling system was implemented using Python-based frameworks integrating Natural Language Processing (NLP) and Machine Learning (ML) techniques. The system was evaluated using real-time Twitter data collected from cybersecurity experts and organizations over a continuous monitoring period. The objective of the evaluation was to measure the system's ability to detect emerging cyber threats, classify them accurately, and generate meaningful alerts for proactive defence.

The experimental results demonstrate that the system is effective in identifying previously unknown or emerging cyber threats from unstructured social media data. By applying the unknown word selection and co-occurrence filtering mechanisms, the system successfully reduced noise and extracted relevant threat-related terms. The integration of TF-IDF and classification models further improved the accuracy of detecting threat-related content.

The one-class classifier effectively filtered irrelevant tweets, significantly reducing false positives in the detection stage. The multi-class classifier, particularly the Linear Support Vector Machine (SVM), showed strong performance in mapping detected threats to the appropriate MITRE ATT&CK tactics. The profiling stage achieved an F1-score of approximately 77% across 14 different tactics, indicating reliable classification performance.

Fig: View All Remote Access Users Page

A key strength of the system lies in its ability to provide early warnings. During experimental deployment, the system was able to detect emerging threats before official disclosures, demonstrating its capability for real-time threat intelligence. The generated alerts included threat names, associated tactics, and risk levels, enabling cybersecurity analysts to take timely preventive actions.

However, the system also exhibits certain limitations. The performance is dependent on the quality and relevance of collected Twitter data. Informal language, abbreviations, and noise in tweets can affect accuracy. Additionally, the system reported a false alert rate of approximately 15%, indicating the need for further improvement in filtering mechanisms. Real-time processing may also introduce minor delays due to continuous data ingestion and analysis.

Overall, the results confirm that the proposed system is effective in detecting and profiling emerging cyber threats using OSINT data. The integration of NLP, ML, and the MITRE ATT&CK framework provides a comprehensive approach for enhancing cyber threat intelligence. Future improvements can focus on advanced NLP techniques, such as Part-of-Speech analysis and contextual embeddings, to further reduce false positives and improve profiling accuracy.

Remote Access User Registration Page

Conclusion

The dynamic nature of cybersecurity, where new vulnerabilities and threats continuously emerge, makes it challenging for analysts to stay updated. Even with strong defences in place, novel attack methods can bypass existing controls, making timely threat intelligence essential for effective cybersecurity.

This work proposes an automated system for cyber threat identification and profiling using Natural Language Processing (NLP) on Twitter data. The system extracts relevant information from tweets to detect emerging threats in real time, assisting cybersecurity professionals in monitoring large volumes of data efficiently.

Unlike existing approaches, this study goes beyond threat detection by identifying the intent of threats. It maps tweet content to tactics defined in the MITRE ATT&CK knowledge base, leveraging this framework to train machine learning models for automated threat profiling.

To evaluate the approach, the system was deployed for 70 days in collaboration with a financial institution in Brazil. During this period, it generated real-time alerts and enabled preventive actions against multiple threats. Notably, the system identified the Pet it Potam vulnerability 17 days before an official patch was released, allowing early mitigation.

Experimental results show that the profiling stage achieved an F1-score of 77% across 14 tactics, with a false alert rate of 15%. Future work focuses on improving tweet selection and reducing false positives, as well as enhancing profiling accuracy using advanced NLP techniques such as Part-of-Speech (POS) analysis with the spa Cy library.

References

1. B. D. Le, G. Wang, M. Nasim, and A. Babar, “Gathering cyberthreat intelligence from Twitter using novelty classification,” 2019, arXiv:1907.01755.
2. “Definition: Threat Intelligence,” Gartner Research, Stamford, CO, USA, 2013.
3. R. D. Steele, “Opensource intelligence: What is it? Why is it important to the military,” Journal, vol. 17, no. 1, pp. 35–41, 1996.
4. C. Sabottke, O. Suciu, and T. Demitra’s, “Vulnerability disclosure in the age of social media: Exploiting Twitter for predicting real-world exploits,” in Proc. 24th USENIX Security Symposium, 2015, pp. 1041–1056.
5. A. Sapienza, A. Bessi, S. Damodaran, P. Shakarian, K. Lerman, and E. Ferrara, “Early warnings of cyber threats in online discussions,” in Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW), Nov. 2017, pp. 667–674.
6. E. Nunes et al., “Darknet and deep net mining for proactive cybersecurity threat intelligence,” in Proc. IEEE Conf. Intelligence and Security Informatics (ISI), Sep. 2016, pp. 7–12.
7. S. Mittal, P. K. Das, V. Malwa, A. Joshi, and T. Finin, “Cyberwriter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities,” in Proc. IEEE/ACM ASONAM, Aug. 2016, pp. 860–867.
8. A. Attar Wala, S. Dimitrov, and A. Obeidi, “How efficient is Twitter: Predicting 2012 U.S. presidential elections using support vector machine,” in Proc. IntelliSense, Sep. 2017, pp. 646–652.
9. N. Dionísio et al., “Towards end-to-end cyberthreat detection from Twitter using multi-task learning,” in Proc. IJCNN, Jul. 2020, pp. 1–8.
10. O. Oh, M. Agrawal, and H. R. Rao, “Information control and terrorism: Tracking the Mumbai terrorist attack through Twitter,” Information Systems Frontiers, vol. 13, no. 1, pp. 33–43, Mar. 2011.
11. T. Sakaki, M. Okazaki, and Y. Matsuo, “Earthquake shakes Twitter users: Real-time event detection by social sensors,” in Proc. WWW, Apr. 2010, pp. 851–860.
12. B. De Longueville, R. S. Smith, and G. Lura chi, “‘OMG, from here, I can see the flames!’: Mining location-based social networks,” in Proc. Location-Based Social Networks Workshop, Nov. 2009, pp. 73–80.
13. A. Sapienza et al., “DISCOVER: Mining online chatter for emerging cyber threats,” in Proc. WWW Companion, 2018, pp. 983–990.
14. R. P. Khandpur et al., “Crowdsourcing cybersecurity: Cyber-attack detection using social media,” in Proc. ACM CIKM, Nov. 2017, pp. 1049–1058.
15. Q. Le sculler et al., “SONAR: Automatic detection of cyber security events over the Twitter stream,” in Proc. Int. Conf. Availability, Reliability and Security, Aug. 2017, pp. 1–11.
16. K.-C. Lee et al., “Sec-buzzer: Cyber security emerging topic mining with open threat intelligence retrieval,” Soft Computing, vol. 21, no. 11, pp. 2883–2896, Jun. 2017.
17. A. Ritter, E. Wright, W. Casey, and T. Mitchell, “Weakly supervised extraction of computer security events from Twitter,” in Proc. WWW, May 2015, pp. 896–905.
18. A. Queiroz, B. Keegan, and F. Mtenzi, “Predicting software vulnerability using security discussion in social media,” in Proc. European Conf. Cyber Warfare and Security, 2017, pp. 628–634.
19. A. Bose et al., “Detection and ranking of emerging cyber threat events in Twitter streams,” in Proc. IEEE/ACM ASONAM, Aug. 2019, pp. 871–878.
20. B. E. Strom et al., “MITRE ATT&CK: Design and philosophy,” MITRE Corp., McLean, VA, USA, Tech. Rep. 19-01075-28, 2018.
21. B.-J. Koops, J.-H. Hopeman, and R. Lens, “Open-source intelligence and privacy by design,” Computer Law & Security Review, vol. 29, no. 6, pp. 676–688, Dec. 2013.
22. Nagesh, C., Chaganti, K.R. , Chaganti, S. , Khaleelullah, S., Naresh, P. and Hussan, M. 2023. Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH. International Journal on Recent and Innovation Trends in Computing and Communication. 11, 7s (Jul. 2023), 353–358. DOI:https://doi.org/10.17762/ijritcc.v11i7s.7010.
23. S. Khaleelullah, P. Marry, P. Naresh, P. Srilatha, G. Sirisha and C. Nagesh, "A Framework for Design and Development of Message sharing using Open-Source Software," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 639-646, doi:10.1109/ICSCDS56580.2023.10104679.
24. Ramesh Kumar Ramaswamy, Pannangi Naresh, Chilamakuru Nagesh, Santhosh Kumar Balan, Multilevel thresholding technique with Archery Gold Rush Optimization and PCNN-based childhood medulloblastoma classification using microscopic images, Biomedical Signal Processing and Control, Volume 107, 2025,107801, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2025.107801.
25. K. R. Chaganti, B. N. Kumar, P. K. Gutta, S. L. Reddy Elicherla, C. Nagesh and K. Raghavendar,"Blockchain Anchored Federated Learning and Tokenized Traceability for Sustainable Food Supply Chains," 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), Gobichettipalayam, India, 2024, pp. 1532-1538, doi: 10.1109/ICUIS64676.2024.10866271.

26. Mohammed Inayathulla and Karthikeyan C, "Image Caption Generation using Deep Learning For Video Summarization Applications" International Journal of Advanced Computer Science and Applications(IJACSA), 15(1), 2024. http:// dx.doi.org/10.14569/IJACSA.2024.0150155.
27. Mohammed, I., Chalichalamala, S. (2015). TERA: A Test Effort Reduction Approach by Using Fault Prediction Models. In: Satapathy, S., Govardhan, A., Raju, K., Mandal, J. (eds) Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1. Advances in Intelligent Systems and Computing, vol 337. Springer, Cham. https://doi. org/10.1007/978-3-319-13728-5_25
28. Inayathulla, M., Karthikeyan, C. (2022). Supervised Deep Learning Approach for Generating Dynamic Summary of the Video. In: Suma, V., Baig, Z., Kolandapalayam Shanmugam, S., Lorenz, P. (eds) Inventive Systems and Control. Lecture Notes in Networks and Systems, vol 436. Springer, Singapore. https://doi.org/10.1007/978-981-19- 1012-8_18.
29. Mohammed and C. Karthikeyan, "Object detection in video summarization for video surveillance applications," Yugoslav Journal of Operations Research, vol. 35, no. 3, pp. 523–546, 2025. doi:10.2298/YJOR240615010I.
30. Inayathulla Mohammed and K. R. Rao, "Enhancing real-time violence detection in video surveillance using hybrid deep learning model," Journal of Web and User Applications, vol. 16, no. 1, 2025. doi:10.58346/JOWUA.2025.I1.021.