WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing

[1]Thota Indu, [2]Alakanti Ankith Reddy, [3]Ramagiri Ayush, [4]S. Manish Kumar, [5]Mrs. N. Bhargavi

[1,2,3,4] Student, [5]Assistant Professor
Dept of CSE, Siddhartha Institute of Technology and Sciences
[1] 21TQ1A6714, 21tq1a6714@siddhartha.co.in, [2] 21TQ5A6740, 21tq5a6740@siddhartha.co.in, [3] 21TQ1A6715, 21tq1a6715@siddhartha.co.in
[4] 21TQ1A6736, 21tq1a6736@siddhartha.co.in, [5]bhargavi.cse@siddhartha.co.in

ABSTRACT

In the contemporary cybersecurity landscape, the window between vulnerability disclosure and exploitation is shrinking. This research introduces a comprehensive framework leveraging Natural Language Processing (NLP) and Machine Learning (ML) for the real-time detection and profiling of emerging cyber threats using Twitter as a primary OSINT source. The proposed system is designed to automatically identify and profile cyber threats by continuously monitoring social media feeds. It identifies potential threat names, classifies their intents using a dual-layered ML model, and generates alerts enriched by the MITRE ATT&CK framework. By mapping discovered threats to known adversarial tactics, techniques, and procedures (TTPs), the system offers a deeper understanding of evolving threats. The framework achieved an F1-score of 77% in profiling, indicating its effectiveness in accurately identifying and classifying threats early in their lifecycle.

Index Terms - Audio Feature Extraction, Emotion Classification, Sentiment analysis, Speech signal processing .

Introduction

The internet's growing role in business, governance, and society has increased exposure to a multitude of cyber threats. As digital transformation accelerates, organizations are more vulnerable to cyber-attacks that exploit weaknesses in software, human behavior, and infrastructure. Cyber Threat Intelligence (CTI), encompassing both structured and unstructured data sources, has emerged as a critical component of defense strategies. Structured CTI includes databases like CVE and NVD, while unstructured sources include blogs, forums, and social media platforms. Twitter, in particular, has proven to be an effective early warning system for cyber threats. This project aims to harness the power of NLP to process unstructured data from such platforms, enabling early detection and profiling of emerging cyber threats.

PROBLEM DEFINITION :

Traditional cybersecurity solutions often rely on signature-based detection and manually curated threat feeds, which can lag behind real-time threat evolution. As attackers increasingly leverage zero-day vulnerabilities and social engineering tactics, security teams need tools that can detect emerging threats proactively. The core problem lies in the inability to process the high volume of unstructured threat intelligence available online in real-time. This leads to delayed responses and unmitigated vulnerabilities. Our project addresses this issue by building an automated system capable of extracting actionable intelligence from unstructured OSINT sources using NLP and ML.

OBJECTIVE OF THE PROJECT :

- To develop an end-to-end automated framework for the identification and profiling of emerging cyber threats.
- To collect and process OSINT data, primarily from Twitter, using robust NLP pipelines.
- To employ ML algorithms for identifying potential threat indicators and classifying their tactics.
- To integrate the MITRE ATT&CK framework for comprehensive threat profiling and contextual analysis.
- To design a user-friendly interface for presenting threat alerts and analytics to cybersecurity professionals.

.SCOPE OF THE PROJECT:

This project is scoped around cyber threat detection through OSINT, focusing on real-time processing of social media data. It aims to support Security Operations Centers (SOCs) and cybersecurity analysts by automating the collection, processing, and classification of threat intelligence. The

system is designed to be scalable, allowing for the inclusion of additional OSINT sources like Reddit, dark web forums, and threat blogs in future iterations. While the current focus is on Twitter, the architecture supports modular extension. The output includes alerts, visual analytics, and downloadable threat profiles.

I. LITERATURE SURVEY

TITLE: NLP-Driven Cyber Threat Identification and Mitigation

AUTHOR: Chen, X., and Roberts, M.

ABSTRACT: In an era of increasing cyber threats, traditional methods of threat detection and mitigation are often inadequate due to the sheer volume and complexity of data. This paper presents an innovative approach to cyber threat identification and mitigation leveraging Natural Language Processing (NLP). By analyzing vast amounts of unstructured text data from sources such as threat reports, social media, forums, and dark web communications, our system can detect emerging threats in real-time.

We introduce a multi-layered NLP framework that includes entity recognition, sentiment analysis, and topic modeling to extract actionable intelligence. The framework utilizes machine learning algorithms to classify and prioritize threats based on their potential impact. Furthermore, we incorporate advanced NLP techniques such as contextual embeddings and transformer models to enhance the accuracy of threat detection.

Our evaluation demonstrates that the NLP-driven system significantly outperforms traditional keyword- based approaches, reducing false positives and improving detection speed. Case studies of real-world cyber incidents show that the system can not only identify threats early but also provide comprehensive profiles of threat actors, enabling more effective mitigation strategies.

By integrating automated response mechanisms, the system offers proactive threat mitigation, issuing alerts, and recommending security measures tailored to the specific threat profile. This research highlights the critical role of NLP in enhancing cyber threat intelligence and underscores its potential to transform cybersecurity practices.

TITLE: Profiling Cyber Threat Actors with Natural Language Processing Techniques

AUTHOR: Wang, Y., and Patel, R

ABSTRACT: The increasing frequency and sophistication of cyber threats necessitate advanced methods for identifying and profiling cyber threat actors. This paper proposes a comprehensive approach leveraging Natural Language Processing (NLP) techniques to profile cyber threat actors effectively. The methodology involves several key steps: data collection, preprocessing, text analysis, and profiling. Data is sourced from diverse unstructured text sources, including social media, hacker forums, and dark web marketplaces, which are rich in information about threat actors' behaviors, motivations, and tactics. Preprocessing steps include tokenization, normalization, and entity recognition to structure the raw text data. NLP techniques such as sentiment analysis, topic modeling, and text classification are then applied to extract meaningful patterns and insights.

The profiling process focuses on identifying the characteristics and behaviors of threat actors, including their preferred attack methods, target selection criteria, and communication patterns. Sentiment analysis helps in understanding the threat actors' psychological states and motivations, while topic modeling identifies recurring themes and topics in their communications. Text classification aids in categorizing different types of threat actors based on their actions and intents.

TITLE: Leveraging Natural Language Processing for Cyber Threat Intelligence

AUTHOR: Lee, H., Kim, S., and Park, J.

ABSTRACT: The increasing sophistication and frequency of cyber threats necessitate advanced methodologies for effective threat identification and response. This paper explores the application of Natural Language Processing (NLP) techniques to enhance Cyber Threat Intelligence (CTI). By leveraging NLP, unstructured text data from diverse sources such as threat reports, social media, forums, and the dark web can be systematically analyzed to extract valuable intelligence. The proposed framework employs entity recognition to identify key actors, malware, and vulnerabilities, while relationship extraction elucidates connections between these entities. Topic modeling and sentiment analysis are utilized to detect emerging threat trends and gauge the intent behind cyber threats. Additionally, text classification techniques categorize threat information, aiding in rapid prioritization and response. The integration of NLP into CTI not only automates the labor-intensive process of threat analysis but also improves the accuracy and timeliness of threat detection. Case studies demonstrate the effectiveness of the approach in real-world scenarios, highlighting its potential to transform cybersecurity operations by providing actionable insights and enhancing situational awareness.

II. EXISTING SYSTEM

Existing systems often use rule-based or keyword-based approaches to identify threats in text. While this provides a baseline capability, it suffers from low accuracy, high false positives, and limited adaptability. These systems usually do not perform deep semantic analysis or context-based profiling. Moreover, most do not leverage knowledge bases such as MITRE ATT&CK for understanding the nature and behavior of threats. Another limitation is

the absence of threat naming or identification, which is essential for tracing threat campaigns. As a result, existing systems fall short in providing a holistic view of the threat landscape.

III.DRAWBACKS OF EXISTING SYSTEM

Inability to identify previously unseen threat names or tactics.

- *Lack of integration with established cybersecurity frameworks.*
- *Limited semantic understanding due to keyword reliance.*
- *High volume of false positives leading to analyst fatigue.*
- *Manual effort required for validation and response.*
- *No support for real-time alert generation or threat scoring*

IV. Proposed Algorithm

The overall goal of this work is to propose an approach to automatically identify and profile emerging cyber threats based on OSINT (Open Source Intelligence) in order to generate timely alerts to cyber security engineers. To achieve this goal, we propose a solution whose macro steps are listed below.

1) Continuously monitoring and collecting posts from prominent people and companies on Twitter to mine unknown terms related to cyber threats and malicious campaigns;

2) Using Natural Language Processing (NLP) and Machine Learning (ML) to identify those terms most likely to be threat names and discard those least likely;

3) Leveraging MITRE ATT&CK techniques' procedures examples to identify most likely tactic employed by the discovered threat;

4) Generating timely alerts for new or developing threats along with its characterization or goals associated with a risk rate based on how fast the threat is evolving since its identification

Key Components and Functionality:

1. **Twitter Stream Collector**
    - *Function:* Captures real-time tweets from cybersecurity-related sources using the Twitter API. Acts as the primary input module for data acquisition.
2. **Text Preprocessor**
    - *Function:* Performs cleaning, tokenization, stop-word removal, and normalization to prepare text for analysis. Ensures data consistency and quality for downstream processing.
3. **Named Entity Recognizer (NER)**
    - *Function:* Extracts specific threat-related entities such as malware names, vulnerability codes (e.g., CVEs), IP addresses, and attack types. Provides structured representations of unstructured text.
4. **Binary Classifier**
    - *Function:* Identifies whether a tweet is cybersecurity-relevant using supervised machine learning models (e.g., SVM, CNN, BERT). Filters irrelevant content early in the pipeline.
5. **Multi-Class Classifier**
    - *Function:* Categorizes relevant tweets into specific threat types (e.g., phishing, ransomware) and associates them with MITRE ATT&CK tactics. Supports detailed threat profiling.
6. **Risk Scoring Engine**
    - *Function:* Computes the severity of identified threats based on frequency, novelty, and trustworthiness of the source. Used to prioritize alerts and support mitigation decisions.
7. **Database Management System**
    - *Function:* Stores raw data, processed outputs, user information, and historical threat profiles. Supports real-time queries and long-term analysis.

8. **Visualization Dashboard**
    - *Function:* Displays real-time analytics, threat maps, trend graphs, and alerts. Allows analysts to navigate data and interact with results.
9. **Alert Notification Module**
    - *Function:* Issues warnings and updates to security analysts based on predefined thresholds. Enables proactive incident response.
10. **User Access Control**
    - *Function:* Manages login, authentication, and user privileges. Ensures secure and role-based access to system features.

Advantages :

- **Automation of Threat Detection:** The system eliminates the need for manual monitoring of large volumes of data by automatically analyzing tweets and identifying potential threats. This leads to increased efficiency and allows analysts to focus on high-priority incidents.
- **Early Warning Capabilities:** By leveraging social media platforms like Twitter, which are often the first places where cyber vulnerabilities and exploits are mentioned, the system provides early alerts, enabling quicker response and mitigation.
- **Context-Aware Threat Profiling:** Using the MITRE ATT&CK framework, the system profiles threats by mapping them to known tactics, techniques, and procedures (TTPs). This contextualization helps analysts understand the threat's behavior and potential impact.
- **Reduction in False Positives:** Traditional systems that rely on keyword matching generate a high number of false positives. The use of machine learning classifiers reduces irrelevant alerts and increases the precision of threat detection.
- **Scalability and Extensibility:** The architecture is modular and scalable, making it suitable for organizations of various sizes. Additional data sources such as Reddit, threat blogs, and dark web forums can be integrated seamlessly in future enhancements.
- **Enhanced Decision-Making:** The structured output provided by the system supports better decision-making by security teams. Threats are not only detected but also categorized based on their severity and intent, guiding analysts on appropriate mitigation strategies.
- **Resource Optimization:** By reducing the workload of security analysts and automating repetitive tasks, the system optimizes human resource utilization. It ensures that human expertise is applied where it is most needed—evaluating and responding to validated threats.
- **Real-Time Alerting:** The system's capability to generate alerts in real time ensures that critical threats are communicated without delay. This supports a proactive defense posture and minimizes potential damage.
- **Data-Driven Insights:** The analytics dashboard provides visualizations and metrics that support long-term strategic decisions, helping organizations to track trends, identify common attack vectors, and plan preventive actions accordingly.
- **Improved Collaboration and Communication:** With structured threat profiles and clear classification of risks, cross-functional teams including IT, security, and compliance can communicate more effectively, ensuring unified incident response efforts.

V. System Architecture

The system architecture for automated cyber threat identification and profiling is designed to enable real-time data ingestion, natural language processing, threat classification, and visualization. It is structured into five core layers to support scalability, modularity, and efficient computation:

A. Data Acquisition Layer

- *Twitter Stream Collector:* This module connects to the Twitter API and continuously retrieves tweets from selected cybersecurity-related accounts, hashtags, and keywords. Tweets serve as the primary source of open-source threat intelligence.
- *Data Buffering:* For high-frequency data ingestion, optional integration with buffering systems like Apache Kafka can ensure fault tolerance and throughput regulation.

B. Preprocessing and NLP Layer

- *Text Cleaning:* Each tweet undergoes preprocessing steps such as tokenization, lowercasing, punctuation removal, stop-word elimination, and lemmatization to prepare it for analysis.
- *Named Entity Recognition (NER):* Employs NLP techniques to extract threat-related entities such as malware names, software systems, IP addresses, and vulnerability identifiers (e.g., CVEs).

C. Classification and Profiling Layer

- *Binary Classification Model:* A supervised ML classifier (e.g., logistic regression, CNN, or BERT) determines whether a tweet is cybersecurity-relevant.

- *Multi-Class Classification Model:* Relevant tweets are passed to a second classifier that identifies the threat category (e.g., ransomware, phishing) and maps its activity to tactics within the MITRE ATT&CK framework.
- *Threat Risk Scoring:* A risk engine evaluates threat severity based on indicators such as term frequency, novelty, temporal spread, and the credibility of sources.

D. Storage and Database Layer

- *Processed Data Storage:* Structured data—including raw tweets, identified entities, classification outcomes, and risk scores—are stored in a relational or NoSQL database (e.g., MySQL, MongoDB).
- *Threat History Archive:* Enables longitudinal analysis and trend monitoring by storing historic data on identified threats and actor behavior.

E. Visualization and Alerting Layer

- *Web Dashboard:* A browser-based dashboard built with Django and frontend frameworks displays threat intelligence in real-time. It includes bar charts, timelines, entity lists, and MITRE tactic maps.
- *Alert Generation:* Based on risk scores, the system generates timely notifications or alerts for cybersecurity analysts, sent via the dashboard or email.

User Roles:

- *Predictive Analytics Provider (Administrator):*
    - Role: Oversees the system and manages its operations.
    - Responsibilities:
        - Manage user roles and access.
        - Train and update ML models.
        - Validate data quality and threat detection accuracy.
        - Maintain system configurations.
- *Remote Access User (Security Analyst):*
    - Role: Investigates and responds to detected threats.
    - Responsibilities:
        - Monitor real-time threat alerts on the dashboard.
        - Analyze threat details and MITRE mappings.
        - Download threat reports and take mitigation action.
        - Provide feedback on false positives or missed detections.

Fig.1 Remote Access User Login Page

This represents the login screen of the cyber threat detection system. The interface allows registered users to securely access their dashboards. It includes a username and password field, with an option for new users to register. Security analysts and administrators log in here to interact with features such as real-time threat monitoring, alert management, and system analytics. The thematic keywords above the login form highlight the system's focus on botnet detection, DDoS defense, and filtering mechanisms.

Fig.2 Remote Access User Registration Page

This illustrates the user registration form for the cyber threat detection and profiling system. The interface is designed to securely onboard new users by collecting essential information such as username, password, gender, contact details, and geographic location. The registration page ensures that only authenticated users can access the system's functionalities, including real-time threat alerts and profiling tools. This step is crucial for maintaining data integrity, enforcing user roles, and ensuring compliance with access control policies in the system.

Fig.3 **Remote Access User Prediction Cyber Threat operation page**

This figure displays the main interface for cyber threat prediction within the system. Users input relevant tweet metadata such as ID, text, timestamp, source, and associated technical attributes including IP addresses, protocols, and entity names. Once the data is submitted, the system processes it using Natural Language Processing (NLP) and multi-layered machine learning models to identify and classify the type of cyber threat. The predicted threat type is then displayed in real-time. This interface plays a critical role in converting unstructured social media content into actionable cybersecurity intelligence.

Fig.4 **View All Remote Access Users Page**

This figure presents the administrative interface for viewing all registered users of the system. Each entry includes user-specific details such as username, email address, gender, location, and mobile number. This centralized user management interface is accessible only to authorized administrators and is crucial for auditing access, maintaining user accountability, and ensuring role-based data visibility. The feature enhances transparency and allows system maintainers to validate and monitor active users contributing to or using the threat detection platform.

CONCLUSION

Given the dynamism of the cyber security field, with new vulnerabilities and threats appearing at any time, keeping up to date on them is a challenging but important task for analysts. Even following the best practices and applying the best controls, a new threat may bring an unusual way to subvert the defenses requiring a quick response. This way, timely information about emerging cyber threats becomes paramount to a complete cyber security system.

This research proposes an automated cyber threat identification and profiling based on the natural language processing of Twitter messages. The objective is exactly to cooperate with the hard work of following the rich source of information that is Twitter to extract valuable information about emerging threats in a timely manner.

This work differentiates itself from others by going a step beyond identifying the threat. It seeks to identify the goals of the threat by mapping the text from tweets to the procedures conducted by real threats described in MITRE ATT&CK knowledge base. Taking advantage of this evolving and collaborative knowledge base to train machine learning algorithms is a way to leverage the efforts of cyber security community to automatically profile identified cyber threats in terms of their intents.

To put in test our approach, in addition to the research experiment, we implemented the proposed pipeline and run it for 70 days generating online alerts for the Threat Intelligence Team of a big financial institution in Brazil. During this period, at least three threats made the team take preventive actions, such as the Petit Potam case, described in section V. Our system alerted the team making them aware of Petit- Potam 17 days before the official patch was published by Microsoft. Within this period, the defense team was able to implement mitigations avoiding potential exploits and, consequently, incidents..

References

[1] B. D. Le, G. Wang, M. Nasim, and A. Babar, ‘‘Gathering cyber threat intelligence from Twitter using novelty classification,’’ 2019, arXiv:1907.01755.

[2] Definition: Threat Intelligence, Gartner Research, Stamford, CO, USA, 2013.

[3] R. D. Steele, ‘‘Open source 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. Dumitras, ‘‘Vulnerability disclosure in the age of social media: Exploiting Twitter for predicting real-world exploits,’’ in Proc. 24th USENIX Secur. Symp. (USENIX Secur.), 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, A. Diab, A. Gunn, E. Marin, V. Mishra, V. Paliath, J. Robertson, J. Shakarian, A. Thart, and P. Shakarian, ‘‘Darknet and deepnet mining for proactive cybersecurity threat intelligence,’’ in Proc. IEEE Conf. Intell. Secur. Informat. (ISI), Sep. 2016, pp. 7–12.

[7] S. Mittal, P. K. Das, V. Mulwad, A. Joshi, and T. Finin, ‘‘CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities,’’ in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2016, pp. 860–867.