International Journal of Science and Technology Research Archive

ISSN: 0799-6632 (Online)

Journal homepage: https://sciresjournals.com/ijstra/

(RESEARCH ARTICLE)

Check for updates

Advance persistent threat prediction using knowledge graph

Nagendrababu NC *, Samyama Gunjal GH and Himabindhu N

Department of computer science, University Visvesvaraya College of Engineering, Bangalore University, Bengaluru, India.

International Journal of Science and Technology Research Archive, 2024, 06(02), 071–082

Publication history: Received on 05 March 2024; revised on 08 May 2024; accepted on 10 May 2024

Article DOI: https://doi.org/10.53771/ijstra.2024.6.2.0047

Abstract

Advanced persistent threats (APTs) are a major threat to cybersecurity, and they are typically attributed to nation-state actors or well-organized groups with sophisticated capabilities. This knowledge graph is intended to help you understand and attribute APT organizations by providing a framework for understanding their characteristics, attributing challenges, attributing clues, attributing methodologies, and attributing limitations. By understanding APT organizations and attributing challenges, clues, methodologies, and attribution limitations, you can gain valuable insights and methods for unraveling the mystery surrounding APT organizations. The graph highlights the difficulties and intricacies associated with attribution, such as false flags, use of proxies, cooperation between APTs and the evolving tactics employed by threat actors. State- sponsored attribution is based on government statements or intelligence agency reports; private sector attribution is based on cybersecurity firms' reports or threat intelligence sharing; and academia and independent research is based on academic and non-academic sources. The graph serves as a resource for cybersecurity professionals, analysts and researchers looking for a systematic framework to improve their understanding and ability to attribute cyberattacks to attack actors. It offers in-depth analysis and practical advice to navigate the complex landscape of APP attribution in today's rapidly changing cybersecurity landscape.

Keywords: Advanced persistent threats; Cybersecurity; Machine learning; Knowledge graph

1. Introduction

In an ever-connected and digital world, the world of cybersecurity is facing a formidable and unpredictable enemy: advanced persistent threat organizations (APT). These highly advanced, well- funded, and long-term threat actors present a unique set of challenges to any organization, government, or individual. To attribute cyberattacks to an APT organization is like trying to solve a jigsaw puzzle, often with layers of deceit and anonymity. In the world of cybersecurity, "cyber" is a key term because it emphasizes the need to defend digital assets, information, and systems against a variety of online threats, such as hacking, malware attacks, data breaches and other malicious activity. As technology advances, the term 'cyber' has become increasingly prominent in discussions about digital security and privacy, as well as the overall security of the online environment. What is digital attribution? Digital attribution is the process of attributing and assigning liability to a specific person, group, organization or nation-state behind a cyberattack, intrusion, or malicious activity in the digital world. Cyber attribution is the process of determining the source, methods, and motivations of a cyber-threat in order to identify who is responsible. Cyber attribution insight is a collection of knowledge, techniques, and insights associated with the attribution of cyberattacks to a specific person, group, or entity responsible for the attack. Cyber attribution is a critical area of research and practice in cybersecurity and digital forensic science. It involves identifying who is responsible for a cyber-incident, understanding their motivation, and Tracking their digital activities. APT stands for advanced persistent threat. An APT attack is a highly sophisticated and sophisticated type of cyber-attack in which a highly skilled and well-funded adversary (e.g. nation-state, organized cybercriminal groups, etc.) gains unmonitored access to an organization's network or system over a long period of time.

* Corresponding author: Nagendrababu NC

Copyright © 2024 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

Cyber threat organizations (also known as cybercriminal groups, hacking collectives, etc.) are groups or individuals that carry out malicious cyber activities for money, political purposes, or any other malicious purpose. These groups can differ in sophistication, tactic, and goals. Here are some examples of cyber threat organizations or groups: Advanced Persistent Threat (APT) Groups: These are usually state-sponsored, long-term, cyber espionage groups. For example, APT28 (Fancy Bear) or APT29(Cozy Bear) are related to Russian intelligence agencies. Lazarus Group: This is a state-sponsored group based out of North Korea. They carry out cyberattacks on financial institutions and cryptocurrency exchanges, as well as various political targets. DarkTequila: This is a cybercriminal group that specializes in financial theft. It targets banks and financial institutions, particularly in Latin America. REvil (Ransomware As A Service (RaaS): This is a group that specializes in ransomware attacks. It targets large corporations and demands high ransoms. APT34 is a state-sponsored Iranian hacking group that targets Middle East and international organizations for cyber espionage. FIN7 is a group of financially motivated cybercriminals that targets hospitality and restaurant sectors. Magecart is a group of cybercriminals that specialize in cyberskimming attacks on online credit card websites. Magecart injects malicious code into websites to steal customer payment card information. Silence is a group of Russian-speaking hackers that targets banks for financial fraud. DarkSide is a ransomware group that targets critical infrastructure, large corporat16ions, and other high-profile targets. They often demand high ransoms. Anonymous is a group of hacktivist hacktivists that launch DDoS attacks and cyberprotests against government, corporate, and other organizations.

Open Source Cyber Threat Intelligence (OPT) refers to cybersecurity threat and vulnerability intelligence that is available to the general public and can be accessed and disseminated freely. OpT intelligence comes from a wide range of sources, such as security researchers and government agencies, as well as cybersecurity organizations and the broader cybersecurity community. OPT plays an essential role in improving the cybersecurity posture of organisations and individuals. APT (Advanced Persistent Threat) organizations are sophisticated, often state-backed or well-coordinated cyber threat groups, that carry out targeted and long-term cyberattacks against targeted targets, including government, corporate, critical infrastructure and research institutions. Advanced persistent threat (APT) groups are characterized by sophisticated capabilities, substantial resources, and long-term goals. APT1 is believed to have ties to the Chinese government and has gained notoriety through its cyber espionage activities against a variety of industries, such as aerospace and defense APT28 is associated with the Russian government and is known to have engaged in high-level cyber operations, such as the targeting of political organizations, government entities, and other targets APT29 is also associated with Russian government and has engaged in cyber espionage activities, targeting government agencies, and critical infrastructure. APT32(OceanBuffalo) APT32 is a Vietnamese APT group that specializes in cyber espionage against private sector and government organizations, especially in Southeast Asia. APT41(Winnti Group) APT41 is a Chinese APT group that focuses on espionage, financial theft and cybercrime. It targets gaming, healthcare and telecommunications sectors. Equation Group (NSA) Equation Group is one of the most well-known APT groups in the world. It is known for its advanced cyber capabilities, such as the development of sophisticated malware. Understanding APT organizations and their characteristics and activities is essential for organizations and governments in order to strengthen their cybersecurity defenses and effectively respond to cyber threats.

2. Literature review

Following are general overview of Cyber Attribution Insights and their components: Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

- APTs pose significant challenges to organizations as they employ sophisticated techniques to breach security defenses and remain undetected for prolonged periods. This literature survey provides an overview of relevant studies and research papers that focus on knowledge graph- based approaches for APT organization attribution. APTs are sophisticated, targeted attacks launched by skilled adversaries with the intent to compromise systems and gain unauthorized access to sensitive information. In recent years, the development of cybersecurity knowledge graphs has emerged as a promising approach to improve threat intelligence, enhance situational awareness, and support effective decision-making in cybersecurity. This literature review aims to provide an overview of relevant studies and research papers related to CAI. Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- The paper [b7] CyberRel is to automate the extraction of such entities and relations to help security professionals and organizations make sense of vast amounts of textual data, including news articles, reports, and other documents. By identifying entities and their relationships, CyberRel can assist in threat detection, situational awareness, and decision- making in the field of cybersecurity.

- The paper [b2] by Z. Tian is a challenging but essential task for cybersecurity professionals and organizations. These threats, often referred to as Advanced Persistent Threats (APTs), are designed to evade traditional security measures and operate stealthily over an extended period.
- The work [b3] focus on evaluating system related to Cyber Threat Intelligence (CTI) that is built on a Heterogeneous Information Network (HIN) model. Heterogeneous Information Networks are a specialized data structure that can be particularly useful in capturing and analyzing complex relationships and information in various domains, including cybersecurity.
- The paper [b5] that suggests a specific technique or method for identifying and categorizing named entities within the field of cybersecurity. In natural language processing (NLP) and text analysis, Named Entity Recognition (NER) is a critical task that involves identifying and classifying entities, such as names of people, organizations, locations, and more, in a given text. Self-attention mechanisms are often used in NLP and deep learning to model contextual relationships between words or tokens in a sentence or document.
- The paper [b17] focuses on involves the process of identifying and extracting structured information and relationships from unstructured text sources, such as reports, articles, and documents. Linked data, in this context, typically refers to data that is interconnected and can be represented using standardized formats and ontologies, making it machine-readable and suitable for further analysis.
- The work [b18] of specific knowledge graph designed for the field of cybersecurity. Knowledge graphs are structured representations of data and information, often organized in a graph format that connects entities and their relationships. They are valuable tools for organizing, querying, and making sense of complex and interconnected information in various domains, including cybersecurity.
- The paper [b8] focuses on the development of a system called TTPDrill, which aims to automatically and accurately extract threat actions from unstructured text in Cyber Threat Intelligence (CTI) sources. CTI sources often contain large volumes of unstructured information, such as security reports, blogs, and forums, which makes it challenging to extract actionable threat information efficiently.

Table 1 Referred Paper Summary table

Paper Summary			
Paper	**Type of detection**	**Technique**	**Features**
15	Static Analysis	Naive Bayes, Support Vector Machines, Decision Trees and their boosted versions.	Lack of readable descriptions useful for computer-forensic experts. Lack of generalization in results.
16	Comparative Analysis	Synthesize the Findings, Snowballing, Datamining.	Signature-Based, Heuristic- Based, Model Checking, Deep Learning, Cloud-Based Malware Detection
18	machine learning deep learning	K-Nearest Neighbors, Neural Networks, Support Vector Machine.	Accuracy rate, recall rate, precision rate and F-measure.
19	machine learning	Neural Network, Deep Learning, Hidden Markov Model, Transfer Learning	Android security, malware detection, feature extraction, classifier evaluation
20	Android Security Repackaged Malware.	Support Vector Machine, K-nearest neighbor, Decision Tree, (R.Forest) in non-repackaged malware classification	User Interaction Features, fuzzy hashing technique, approximate the class-level dependence, Inferring class- level call dependence.
21	Fine-tuned deep learning,machine learning.	conventional learning-based, visualization techniques	Extracts deep features from the color image.
22	Deep learning, machine learning.	neural network with multiple layers, feature learning, Windows malware classification	deep learning to extract N- gram, Byte unigram features, Entropy-based features,

2.1 Proposed System

Networks always face security challenges with different types of attacks. Some are permanent, while others are non-persistent. APT (advanced persistent attack) remains in the network permanently. Most of the research on cyber threat intelligence focuses on automating threat entities extraction from public attack events. However, this is not feasible. In this paper, I propose using Knowledge Graph on the APT attack dataset. OSCTI (Open Source Cyber Threat Intelligence) is becoming increasingly influential in obtaining real-time network security information. The main goal of the cyber security knowledge graph is to change expression of threat knowledge to allow security researchers to accurately and efficiently obtain different types of threat information to make pre-decision-making. The attribution technology not only helps security analysts to detect advanced persistent threats, The same threat can also be identified from different attack events, so it is important to track the attack threat actor Proposed paper Apply knowledge graph technology Take into account the latest research in cyber threat attack attribution Study key related technologies Study key theories in the development and application of advanced persistent threat knowledge graph (APT) from OSCTI Designing CAI based on knowledge graph Inspired by ontology theory, CAI was built as an APP knowledge graph model Based on real APP attack scenarios Designing an APP threat knowledge Extraction algorithm for completion and update of the knowledge graph Using deep learning using GRU layers and expert knowledge.3

2.1.1 Architecture for Proposed System

The design of an Advanced Persistent Threat (APT) detection system utilizing a knowledge graph requires the integration of different elements to gather, assess, and display information pertaining to APT operations. The architecture can vary depending on the specific approach and techniques used, but here is a generalized overview of the key components.

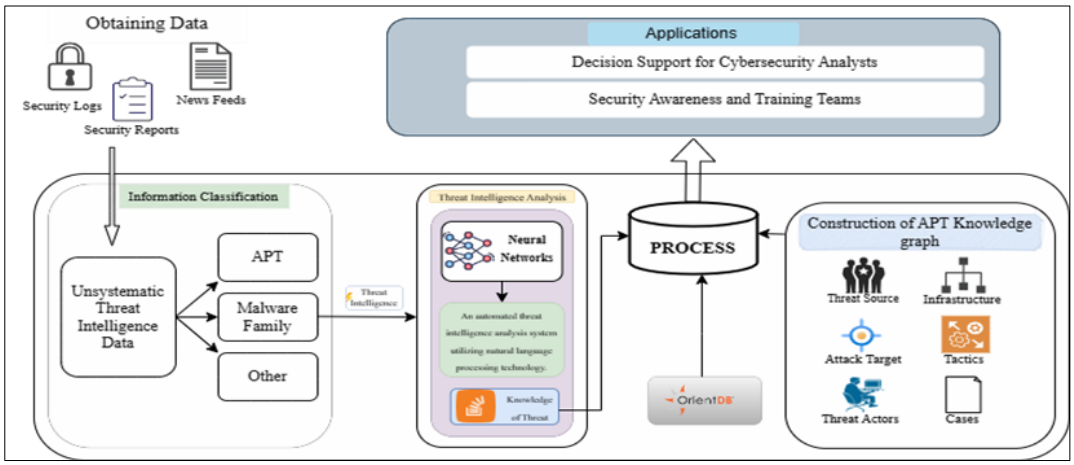

Figure 1 Framework of APT Knowledge Graph

- Data Acquisition Layer: Integration of Log Sources Gather information from a variety of sources including security logs, network traffic, endpoint logs, and threat intelligence feeds. API Integration allow for the seamless integration of external threat intelligence platforms, antivirus solutions, and various other security tools, thereby enhancing the dataset with valuable information.
- Enhancing and Analyzing Data: Data normalization and parsing are essential processes to ensure consistency in the collected data. By standardizing and normalizing the data, we can eliminate any inconsistencies or discrepancies that may exist. Furthermore, enriching the raw data with additional context from various sources such as threat intelligence feeds, vulnerability databases, and historical incident data can provide valuable insights and enhance the overall quality of the data. This enrichment process adds depth and relevance to the collected information, enabling better analysis and decision-making.
- Knowledge Graph Construction: Extraction of Entities Detect and retrieve various entities such as IP addresses, domains, and file hashes from enriched data. Mapping of Relationships Establish connections between entities by analyzing observed patterns, attack tactics, techniques, and procedures (TTPs). Utilization of Graph Database Employ a graph database like Neo4j to store and depict the knowledge graph, effectively capturing the interconnectedness of entities related to Advanced Persistent Threats (APTs).

- Feature Selection/Dimensionality Reduction: In order to improve efficiency and reduce noise, feature selection or dimensionality reduction techniques may be applied. This helps to identify the most informative and relevant features for malware detection.
- Model Evaluation and Update: The performance of the detection system is evaluated using metrics such as accuracy, precision, recall, and F1 score. Feedback from the evaluation is used to refine and update the models to improve their effectiveness in detecting new and evolving malware threats.
- Integration of Threat Intelligence: Implement Continuous Feed Integration by subscribing to real-time threat intelligence feeds in order to keep the knowledge graph up-to-date with the most recent information regarding APT groups, campaigns, and indicators. Utilize Indicator Correlation to match threat intelligence indicators with the current knowledge graph, enabling the detection of possible APT-related patterns.

2.2 Data set

A dataset of Advanced Persistent Threats (APTs) usually includes both structured and unstructured data concerning cyber threats from advanced and persistent attackers. These datasets are crucial for cybersecurity professionals to analyze and comprehend the strategies used by sophisticated adversaries. Yet, acquiring authentic APT datasets can be difficult because of the confidential nature of the information.

- Signs of Compromise (SoCs): IP addresses, domain names, cryptographic hashes of files, and various other markers linked to malevolent actions.
- Network traffic logs record communication patterns, anomalies, and potentially harmful traffic detected on the network.
- The APT Attack Dataset module plays a crucial role in our cybersecurity application by allowing users to effortlessly incorporate and analyze datasets linked to Advanced Persistent Threats (APTs). Users have the ability to upload datasets that include a wide variety of APT- related data, including Indicators of Compromise (IoCs), network traffic logs, endpoint data, and other relevant information.
- The Dataset Knowledge Graph module is specifically created to convert unprocessed cybersecurity datasets into a well-organized knowledge graph. This enables users to visually interpret and grasp the connections between network attributes and cyber threats. By utilizing graph algorithms, this module constructs an integrated framework that enhances the depth of analysis for security incidents related to networks.
- The "Preprocess Dataset" plays a vital role in both machine learning and deep learning pipelines. Its purpose is to transform raw datasets into a format that is suitable for efficient model training and evaluation. This module encompasses several preprocessing steps, such as managing missing values, randomizing the data, normalizing the values, and dividing the dataset into separate subsets for training and testing purposes.
- The outlined procedure includes numerous crucial stages in preprocessing and implementing a deep learning algorithm for cybersecurity, particularly utilizing Bidirectional Long Short-Term Memory (BI-LSTM) in conjunction with Gated Recurrent Unit (GRU).
- Generating a comparison graph is an essential part of assessing the effectiveness of various algorithms or models. In your specific situation, you aim to compare the accuracy and possibly other metrics of the suggested BI-LSTM with GRU algorithm against alternative models.

3. Results and Discussion

APT dataset" typically refers to a dataset that contains information related to Advanced Persistent Threats (APTs). APTs are sophisticated cyber-attacks conducted by well-resourced adversaries, often with specific objectives such as espionage, sabotage, or financial gain. These attacks are characterized by their stealthy nature, persistence, and use of advanced techniques to evade detection.

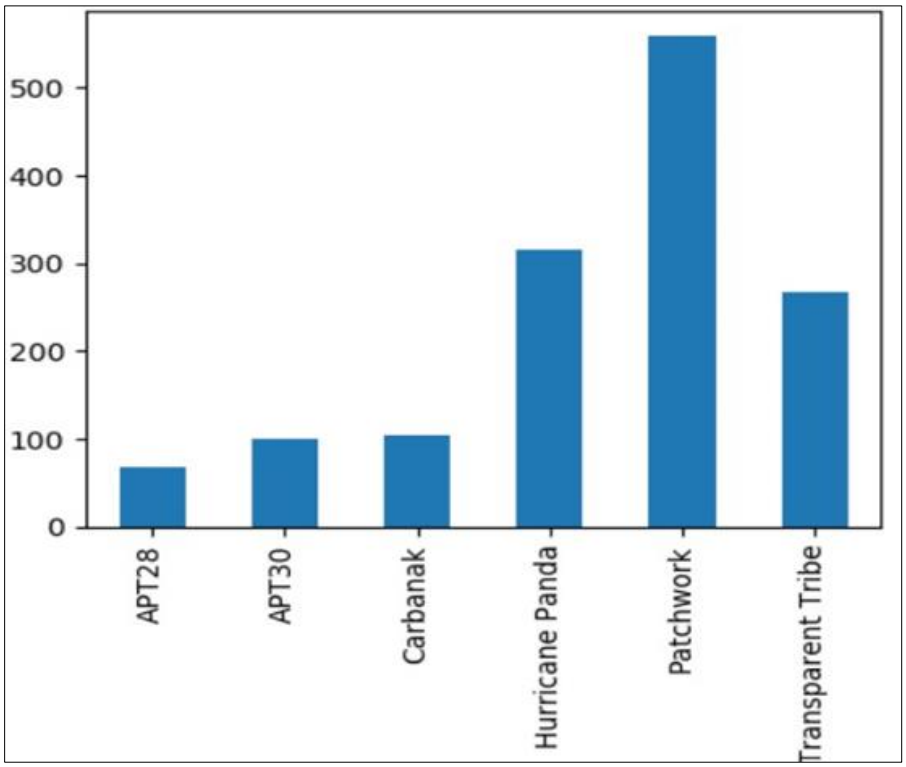

Figure 2 APT Attacks Found in Dataset

Using Natural Language Processing (NLP) techniques to detect Advanced Persistent Threats (APTs) involves analyzing text data, such as security reports, news articles, threat intelligence feeds, and other sources, to identify indicators of APT activity. Natural Language Processing (NLP) techniques play a crucial role in analyzing and understanding text data, including cybersecurity- related content.

3.1 Bi-LSTM with GRU Layers

A combination of Bidirectional Long Short-Term Memory (Bi-LSTM) and Gated Recurrent Units (GRU) forms a recurrent neural network architecture. This architecture effectively merges the bidirectionality of Bi-LSTM layers with the efficiency of GRU units. It is widely employed in various sequential data processing tasks, including natural language processing (NLP), time series prediction, and sequence classification. Bidirectional LSTMs involve the utilization of two LSTMs that process sequential input in both forward and backward directions. These two networks are identical and share the same hyperparameters during training. The only distinction lies in the fact that one network receives input from the start of a sentence and progresses forward, while the other network receives input from the end and moves backward.

Algorithm for implementing a Bi-LSTM with GRU model for Advanced Persistent Threats (APTs).

Algorithm: BiLSTM_With_GRU_for_APTs Input: -Training dataset (X_train, y_train)

- Testing dataset (X_test, y_test)
- Parameters: max_sequence_length, vocab-size, num-classes
- Model hyperparameters: lstm-units, gru-units, embedding-dim, dense-units
- Training hyperparameters: epochs, batch-size, learning-rate.

Output: - Trained Bi-LSTM with GRU model

- Initialize Bi-LSTM with GRU model
  - Add a Bidirectional LSTM layer with units=lstm_units, return_sequences=True
  - Add a GRU layer with units=gru_units
- Compile the model

    - Compile the model with an appropriate optimizer (e.g., Adam), loss function (e.g., categorical_crossentropy), and metrics (e.g., accuracy)
- Train the model
    - Train the model on the training dataset (X_train, y_train) using model.fit
    - Set the number of epochs, batch size, and learning rate based on hyperparameters
- Evaluate the model
    - Evaluate the trained model on the testing dataset (X_test, y_test) using model.evaluate
    - Obtain evaluation metrics such as accuracy, precision, recall, and F1-score
- Save or return the trained model
    - Save the trained model for future use or return it for inference on new data End Algorithm

Advanced Persistent Threat (APT) types, the actual labels and the predicted labels from a classification model. Then, we can compare these labels to compute the counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for each APT type. Once we have these counts, we can visualize the confusion matrix to understand the model's performance in classifying different APT types Transparent Tribe (46 instances): Transparent Tribe, also known as APT36 or Mythic Leopard, is a cyber-espionage group known to target government and military organizations in South Asia, particularly India. This group is known for using spear-phishing emails and malware such as Crimson RAT to infiltrate target networks.

Patchwork (109 instances): Patchwork, also known as Dropping Elephant or Chinastrats, is a cyber-espionage group with links to China. This group has been observed targeting organizations in various sectors across multiple countries, including government, defense, and telecommunications. Patchwork is known for using a variety of tactics, including phishing emails, watering hole attacks, and malware such as FakeM and CHINACHOPPER.

Hurricane Panda (67 instances): Hurricane Panda, also known as APT27 or Emissary Panda, is a Chinese cyber espionage group believed to be associated with the Chinese military. This group has targeted organizations in various sectors, including defense, technology, and government, with a focus on stealing sensitive information and intellectual property. Hurricane Panda is known for using a range of sophisticated tactics and tools, including custom malware such as China Chopper and HTTP Browser.

Carbanak (20 instances): Carbanak, also known as FIN7 or Anunak, is a cybercriminal group known for targeting financial institutions worldwide. This group is notorious for its sophisticated attacks, including malware-based campaigns targeting point-of-sale systems and financial networks. Carbanak is responsible for stealing millions of dollars through techniques such as spear- phishing, malware, and social engineering.

APT30 (25 instances): APT30, also known as Scarlet Mimic, is a cyber-espionage group believed to be associated with the Chinese government. This group has targeted governments, military organizations, and defense contractors across Southeast Asia, particularly Vietnam, with a focus on political and military intelligence gathering. APT30 is known for its long-term, persistent campaigns and the use of custom malware such as BACKSPACE and NetTraveler.

APT28 (14 instances): APT28, also known as Fancy Bear or Sofacy, is a cyber-espionage group believed to be associated with the Russian government. This group has been linked to various high- profile attacks, including the 2016 Democratic National Committee email leak and the targeting of government agencies, military organizations, and critical infrastructure worldwide. APT28 is known for its advanced techniques, including spear-phishing, zero-day exploits, and the use of sophisticated malware such as Sednit and XAgent.

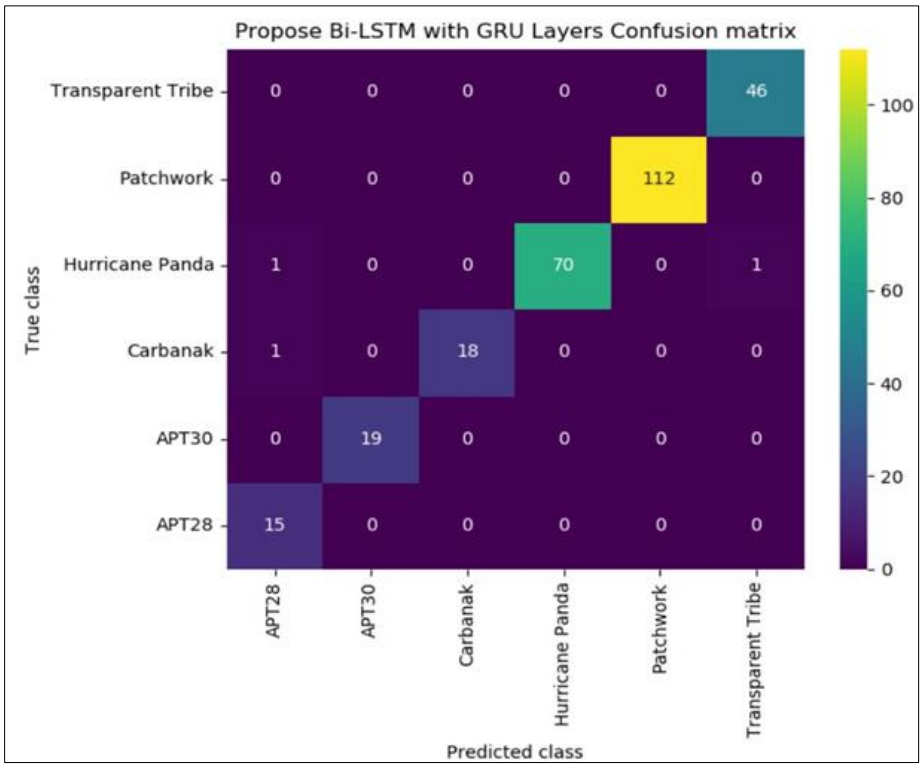

Figure 3 Bi-LSTM with GRU Confusion Matrix

3.2 Classification report

A Classification Report for Advanced Persistent Threat (APT) types would typically involve evaluating the performance of a classification model trained to identify and categorize different types of APTs. APTs are sophisticated cyber-attacks typically orchestrated by nation-states or highly organized cybercriminal groups with the intent to breach and persist within targeted networks over an extended period. These metrics are calculated based on the predictions made by the model and the actual labels. The most common metrics in a classification report include precision, recall, F1-score, and support.

Accuracy: To calculate the accuracy for Advanced Persistent Threat (APT) classification, you need to have a classification model that predicts whether a given instance belongs to a particular APT type or not, and you need a labeled dataset for evaluation.Precision: precision for Advanced Persistent Threat (APT) classification, you need to have a classification model that predicts whether a given instance belongs to a particular APT type or not. Precision measures the accuracy of the positive predictions made by the model for a specific class.

Recall (Sensitivity or True Positive Rate):Recall, also known as sensitivity or true positive rate, is another important metric for evaluating the performance of a classification model, especially in the context of Advanced Persistent Threat (APT) classification. Recall measures the ability of the model to correctly identify all instances of a particular class among all instances that truly belong to that class.

FMeasure: The FMeasure is the harmonic mean of precision and recall. It balances the trade- off between precision and recall and is a good overall measure of a model's performance. F1-Score is especially useful when you want to find a balance between false positives and false negatives.

Table 2 BI-LSTM with GRU report

	Accuracy	**precision**	**recall**	**FMeasure**
Propose Bi-LSTM with GRU Layers	98.93	97.68	98.65	98.09

3.3 Knowledge Graph

The knowledge graph visually represents the relationships and connections between different APT groups, helping to understand their characteristics, tactics, and historical events. This visualization aids cybersecurity analysts in identifying patterns, assessing threats, and strategizing defense measures effectively.

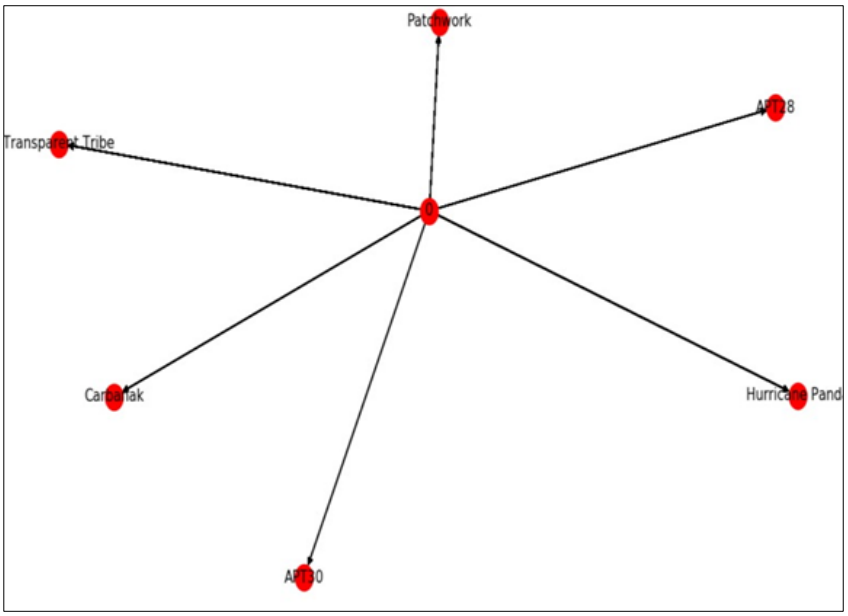

Figure 4 APT knowledge graph

Transparent Tribe (APT36, Mythic Leopard): Transparent Tribe, also known as APT36 or Mythic Leopard, is a cyber-espionage group known to target government and military organizations in South Asia, particularly India.This group is known for using spear-phishing emails and malware such as Crimson RAT to infiltrate target networks.

Patchwork (Dropping Elephant): Patchwork, also known as Dropping Elephant or Chinastrats, is a cyber-espionage group with links to China. This group has been observed targeting organizations in various sectors across multiple countries, including government, defense, and telecommunications. Patchwork is known for using a variety of tactics, including phishing emails, watering hole attacks, and malware such as FakeM and CHINACHOPPER.

Hurricane Panda (APT27): Hurricane Panda, also known as APT27 or Emissary Panda, is a Chinese cyber espionage group believed to be associated with the Chinese military. This group has targeted organizations in various sectors, including defense, technology, and government, with a focus on stealing sensitive information and intellectual property. Hurricane Panda is known for using a range of sophisticated tactics and tools, including custom malware such as China Chopper and HTTPBrowser.

Carbanak (FIN7): Carbanak, also known as FIN7 or Anunak, is a cybercriminal group known for targeting financial institutions worldwide. This group is notorious for its sophisticated attacks, including malware-based campaigns targeting point-of-sale systems and financial networks. Carbanak is responsible for stealing millions of dollars through techniques such as spear-phishing, malware, and social engineering.

APT30 (Scarlet Mimic): APT30, also known as Scarlet Mimic, is a cyber-espionage group believed to be associated with the Chinese government. This group has targeted governments, military organizations, and defense contractors across Southeast Asia, particularly Vietnam, with a focus on political and military intelligence gathering. APT30 is known for its long-term, persistent campaigns and the use of custom malware such as BACKSPACE and NetTraveler.

APT28 (Fancy Bear): APT28, also known as Fancy Bear or Sofacy, is a cyber-espionage group believed to be associated with the Russian government. This group has been linked to various high- profile attacks, including the 2016 Democratic National Committee email leak and the targeting of government agencies, military organizations, and critical infrastructure worldwide. APT28 is known for its advanced techniques, including spear-phishing, zero-day exploits, and the use of sophisticated malware such as Sednit and XAgent.

3.4 Comparison Metrics

Comparison Metrics graph for APT classification involves visualizing the performance of a classification model by displaying the counts of true positive, false positive, true negative, and false negative predictions for each APT type. This visualization provides insights into how well the model is performing for each class.

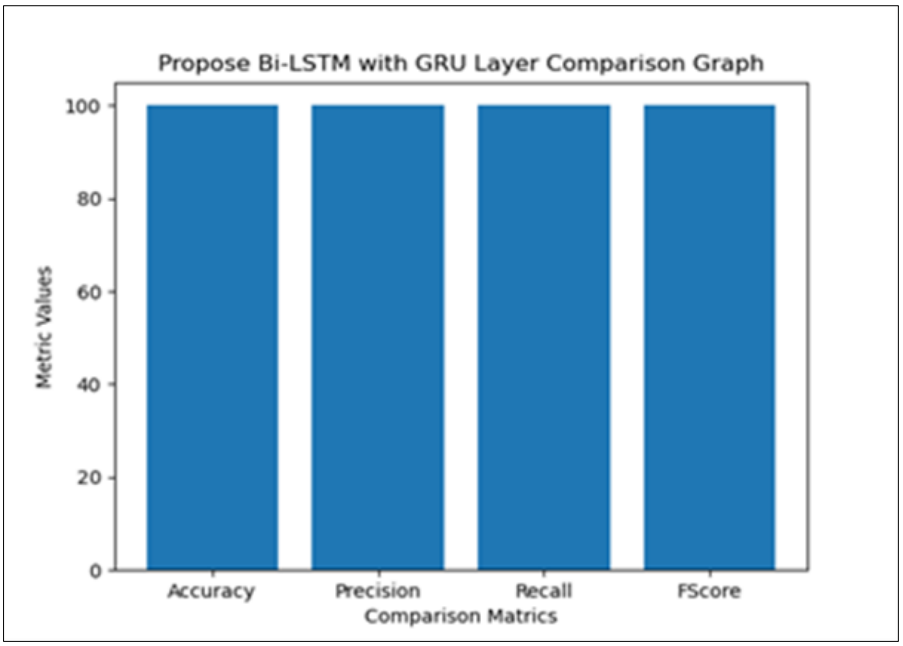

Figure 5 Comparison metrics

Accuracy Measures the overall correctness of the model's predictions:

$$\frac{T\square\square\square N\square\square\square\square\square\ \square\square\ P\square\square\square i\square\square i\square\square\square}{T\square\square eP\square\square i\square i\square e}$$

$$\square\square eci\square i\square\square \quad T\square\square eP\square\square i\square i\square e + Fal\square eP\square\square i\square i\square e \quad \text{.....(2)}$$

=

Recall provides insight into how effectively the model can identify instances of each APT type among all instances that truly belong to that class.

$$Recall = \frac{T\square\square eP\square\square i\square i\square e}{T\square\square eP\square\square i\square i\square e + Fal\square e\square ega\square i\square e} \quad \text{.....(3)}$$

$$F\square ea\square\square\square e = 2\frac{P\square eci\square i\square\square \times Recall}{P\square eci\square i\square\square \times Recall} = \frac{2TP + FP +}{} \quad \text{.....(4)}$$

$$\times \qquad F$$

FMeasure for each APT type and print out the results. Additionally, it will calculate the weighted average F1 Score across all APT types.

4. Conclusion

The use of knowledge graphs for Advanced Persistent Threat (APT) prediction presents a promising and proactive approach to enhance cybersecurity defenses. Leveraging the interconnected nature of data through a knowledge graph allows organizations to gain valuable insights into the evolving threat landscape.

Knowledge graphs offer a comprehensive perspective on advanced persistent threats (APTs) by linking various fragments of data; including threat intelligence; indicators of compromise (IoCs); and past attack patterns. This all-encompassing comprehension plays a vital role in predicting and mitigating intricate security risks. The interconnected framework of knowledge graphs allows for the correlation of associations among entities; empowering security teams to unveil concealed links between threat actors; strategies; and compromised assets. This assists in attributing and identifying intricate attack patterns. Knowledge graphs provide security teams with the capability to detect potential APT activities in a proactive manner by identifying anomalies; patterns; and trends within the interconnected data. This proactive approach allows organizations to respond promptly and efficiently.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Advanced Persistent Threat. https://en.wikipedia.org/wiki/ Advanced persistent threat Information and communication technology; 2020;14 (06): 4-7

[2] Z. Tian; “Detection and traceability of high covert unknown threats in cyberspace;” Inf.Commun. Technol.; vol. 14; no. 06; pp. 4–7; 2020.

[3] Y. Gao; X. LI; H. PENG; B. Fang and P. Yu; ”HinCTI: A Cyber Threat Intelligence Modeling and Identification System Based on Heterogeneous Information Network;” in IEEE Transactions on Knowledge and Data Engineering; doi: 10.1109/TKDE.2020.2987019

[4] Y. Lin; P. Liu; H. Wang; W. Wang; and Y. Zhang; “Overview of network security threat intelligence sharing and exchange;” Comput. Res. Develop.; vol. 57; no. 10; 2020; pp2052- 2065.

[5] Li T; Guo Y; Ju A. A self-attention-based approach for named entity recognition in cybersecurity[C]//2019 15th International Conference on Computational Intelligence and Security (CIS). IEEE; 2019: 147-150.

[6] Zhu; Z.; Dumitras; T.: ChainSmith: automatically learning the semantics of malicious campaigns by mining threat intelligence reports. In: IEEE European Symposium on Security and Privacy (Euro S and P); vol. 2018; pp. 458–472. IEEE (2018) .

[7] Ghazi; Y.; Anwar; Z.; Mumtaz; R.; Saleem; S.; Tahir; A.: A supervised machine learning based approach for automatically extracting high-level threat intelligence from unstructured sources. In: 2018 International Conference on Frontiers of Information Technology (FIT); pp. 129–134. IEEE (2018)

[8] Yishuai Zhao; Bo Lang; and Ming Liu. Ontology-based unified model for heterogeneous threat intelligence integration and sharing. In 2017 11th IEEE International Conference on Anticounterfeiting; Security; and Identification (ASID); pages 11–15; 2017.

[9] Y. Guo; Z. Liu; C. Huang; J. Liu; W. Jing; Z. Wang; and Y. Wang; ”CyberRel: Joint entity and relation extraction for cybersecurity concepts;” in Proc. Int. Conf. Inf. Commun. Secur.; 2021; pp. 447–463

[10] Ghaith Husari; Ehab Al-Shaer; Mohiuddin Ahmed; Bill Chu; and Xi Niu. TTPDrill: Automatic and accurate extraction of threat actions from unstructured text ofCTI Sources. In ACM International Conference Proceeding Series; volume Part F1325; 2017.

[11] Zhenyuan Li and Jun Zeng and Yan Chen and Zhenkai Liang. AttacKG: Constructing Technique Knowledge Graph from Cyber Threat Intelligence Reports. arXiv: 2111.07093; 2021.

[12] H. Wang; G. Qi; and H. Chen; ”Knowledge Graph: Method; Practice and Application”; Beijing; China: Publishing House of Electronics Industry; 2019.

[13] Chinese. C. N. Li and S. A. Thompson; “Mandarin chinese: A functional reference grammar;” J. Asian Stud.; vol. 42; no. 3; pp. 10–12;1989.

[14] A. Alsaheel; Y. Nan; S. Ma; L. Yu; G. Walkup; Z.B. Celik; X. Zhang; D. Xu; “ATLAS: A sequence-based learning approach for attack investigation;”in Proc. 30th USENIX Secur. Symp.; 2021; pp3005- 3022

[15] MANDIANT. Sophisticated Indicators for the Modern Threat Landscape: An Introduction to OpenIOC. http://openioc.org/resources/An Introduction to OpenIOC.pdf; June 2017.

[16] A. Joshi; R. Lal; T. Finin and A. Joshi; ”Extracting Cybersecurity Related Linked Data from Text;” 2013 IEEE Seventh International Conference on Semantic Computing; 2013; pp. 252-259; doi: 10.1109/ICSC.2013.50.

[17] A. Joshi; R. Lal; T. Finin and A. Joshi; ”Extracting Cybersecurity Related Linked Data from Text;” 2013 IEEE Seventh International Conference on Semantic Computing; 2013; pp. 252-259; doi: 10.1109/ICSC.2013.50.

[18] E. Kiesling; A. Ekelhart; K. Kurniawan and F. Ekaputra; “The SEPSES knowledge graph: An integrated resource for cybersecurity;” in Proc. Int. Semantic Web Conf.; 2019; pp. 198–214