UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Using Artificial Intelligence (AI) and Machine Learning (ML) Techniques to Solve the Complex Cyber Attribution Challenge

Supervisor: Dr Moses Dlamini

Student number: u23912589

Clinton Mgoduswa

September 4, 2024

Contents

Abstract

Cyber attribution is crucial for identifying the source of cyber-attacks and improving cybersecurity strategies. Traditional methods for attribution are time-consuming and may not have provided accurate results, especially for sophisticated cyberattacks. The study aimed to explore the effectiveness of Artificial Intelligence (AI) and Machine Learning (ML) models in accurately identifying the source of cybercriminals. The research utilized AI (DeepLog) and ML (Autoencoder-based Anomaly Detection) models, and public datasets of known cyber-attacks were collected for evaluation. The study employed quantitative research approaches and assessed the models' performance using metrics such as accuracy, precision, recall, and F1 score.

The research also investigated the impact of factors like dataset size, model complexity, and training techniques on the predictive performance of the models. The results provided insights into the capabilities of AI and ML in cyber attribution, comparing their accuracy and effectiveness. By leveraging these advanced technologies, organizations could enhance their defense strategies and prevent future cyber-attacks.

Chapter One
Introduction & Problem Statement

1 Chapter One

1.1 Introduction

Cyber attribution refers to the process of identifying the origin of a cyber-attack or other malicious exploit (Cuzzocrea et al., 2022). This enables cybersecurity professionals to gain visibility of the threat source, enhance cyber defense controls, perform offensive security, and ultimately improve their future cybersecurity strategy. In today's world, it's important for organizations to continually improve their defense strategies since cybercriminals are continually refining and adapting their strategies. Additionally, malware kits available on the dark web make it possible even for common criminals without technology knowledge to employ highly sophisticated methods of intrusion.

According to the results of the 2023 State of Email Security survey (SOES), a majority of the respondents (59%) identified the biggest difficulty they face as dealing with the highly advanced and complex nature of the cyber-attack they encounter (Mimecast, 2023). These cyber-attacks entail criminal aspects such as the recent report by Toro-Alvarez (2023) where about four billion US Dollars ($4 million) was stolen through crypto hacking. Also, a separate report by Guly´as and Kiss (2023) mentioned an increase in crypto hacking, stating that about $3.3 billion was stolen in 2021. Therefore, this gives the impression that cybercrime is rapidly growing, with the cost of cybercrime predicted to hit $8 trillion in 2023 and to grow to $10.5 trillion by 2025 (Alshaikh et al., 2023). The increasing number of data breaches and cyber-attacks has led to a rise in identity fraud, resulting in huge financial losses for both businesses and individuals. In the United States alone, identity fraud losses reached $52 billion in 2022, affecting 42 million adults whose sensitive information was stored in vulnerable databases (Smith et al., 2023).

The volume of reported cybercrime continues to rise, with South African Banking Risk Information Centre (SABRIC) reporting that there were 25,187 reported incidents of cybercrime in South Africa in the year 2020, with losses amounting to R1.8 billion (Sabric, 2021). This includes various types of cybercrime such as phishing, malware attacks, and ransomware. The cost of cybercrime increased by 32% in 2020 compared to the year 2019, with the average cost of a cyber-attack being R2.6 million (approximately $175,000 USD) per company. The COVID-19 pandemic was the cause of the increased adoption of remote work, and the rise of ransomware attacks. In the midst of this, South Africa ranks 17th globally in terms of the number of attempted cyberattacks (Alawid et al., 2022).

Business email compromise (BEC) attacks where criminals use email to impersonate a company executive and trick or steal sensitive information, also increased in 2020 (Atlam et al., 2022). This resulted in significant financial losses and damage to a company's reputation. Global ransomware attacks have been a consistent form of cyberattack on banks in recent years. In 2020, Bank of America suffered a data breach that affected an unknown number of customers, and this was due to the breach attributed to vulnerability in the company's mobile app that allowed unauthorized access to customer accounts (Despotovi ́c et al., 2023). Also, in 2018, Banco de Chile suffered a cyberattack that caused significant disruption to its operations. The attack was carried out using malware to steal login credentials and gain access to the bank's systems (Hoheisel, 2022).

Inherent internal weaknesses in the configuration and implementation of a computer system and network render them susceptible to cyberattacks and threats (Mahor et al., 2022). Incorrect configuration, lack of adequate procedures, inexperienced or untrained personnel are examples of vulnerability-causing sources in building a computer network system. These vulnerabilities increase the chances of cyberattacks and threats within and outside a network.

Quite a number of people from different fields are becoming dependent on cyber networks. Cybersecurity is the practice of protecting the integrity of data, networks, and programs from threats in cyberspace (Akinola and Afonja, 2022). Cyber attribution is essential in cybersecurity to detect the intent of the cyber attacker and to take necessary measures to prevent future cyberattacks (Amini and Bozorgasl, 2023). The complexity of cyber attribution makes it difficult to identify the source of a cyberattack, particularly in cases where the attacker uses sophisticated techniques to conceal their identity, such as using proxies or Virtual Private Networks (VPNs) (Riggs et al., 2023).

Nevertheless, Artificial Intelligence (AI) and Machine Learning (ML) can help identify the group or individual responsible for a new cyber-attack, predict and prevent future cyber-attacks by analyzing past cyber-attacks and identifying patterns. Furthermore, AI algorithms can help cybersecurity teams identify cyber vulnerabilities and proactively address them before they can be exploited by cyber attackers. This can be especially helpful in cases where the cyber attacker is trying to disguise their identity or use new tactics. Artificial intelligence fundamentally simulates human intelligence processes using computer systems and is a rapidly growing field that is transforming various industries, such as cybersecurity.

In the following chapter, we will delve into the challenges and complexities associated with cyber attribution. We'll explore the intricate issues surrounding the identification of cyber attackers and the role of AI and ML in addressing these problems

1.2 Problem Statement

Traditional methods for cyber attribution, such as Internet Protocol (IP) address tracing and digital forensics, are time-consuming and may not provide accurate results. Spoofed cyber-attacks do not make the situation any better, as identifying the source of an attack through traditional methods is difficult. Therefore, this leads to gaps between the high rising rate of cybercrime and the always low rate of successful prosecution of the perpetrators.

Given the rapidly increasing cybercrime which makes daily headlines, the expectation would be that more cybercriminals would also be getting caught and sent to jail (Lim and Thing, 2022). Yet, available literature reporting prosecution gives the impression that as much as the incidents of cyberattacks are reported, the literature is scanty on cybercriminals getting caught and brought to book (Alawida et al., 2002).

Available literature falls short of being conclusive on the subject matter of reporting prosecution of cybercriminals. Firstly, could it be that there is not enough incriminating digital evidence? Or the digital forensic investigators are failing to locate the origin or source of the cybercrimes? Secondly, could it be that the justice system is failing to prosecute cybercriminals? Or could it be that there are not enough prosecutors that are technologically equipped to be able to present water-tight cybercrime cases to the courts? Though, the situation may not be as clear as we would like to see. But one thing that remains true and crystal clear is that the issue of attributing cybercrimes is complex and hard. The complexity of cyber attribution is a global concern across different organizations, and identifying the source or origin of a cyber-attack is a critical issue that requires urgent attention to ensure that those who are involved are brought to book.

Therefore, this study aims to use Artificial Intelligence and Machine Learning (AI and ML) as alternatives to unpack the complex cyber attribution challenge. The author hypothesizes that Machine Learning and Artificial Intelligence tools would be able to track the true sources or origins of cyberattacks that use different tactics such as IP spoofing to hide.

1.3 Research Questions

This has led to the following research questions:

- Given the increasing number of cybercrimes, is ML and AI able to provide an accurate source of the cybercriminals?
- Are ML and AI comparable in providing the source of cybercriminals with the same degree of accuracy?
- Does the dataset size, model complexity, and training techniques affect the predictive performance of ML and AI models in identifying the source of cybercrime?

  Now, let's proceed to the research objectives.

1.4 Research Objectives

This section outlines the research objectives guiding our investigation

- Investigate the statics of cybercrimes and the associated number of successful prostitution.
- Investigate and determine if AI and ML are comparable in providing the origin of cyber-attacks with the same degree of accuracy.
- Investigate to determine if the dataset size, model complexity, and training techniques affect the predictive performance of ML and AI models in identifying the origin of cyber-attacks.

In conclusion, the complex nature of cyber attribution, exacerbated by the rise of cybercrime and challenges in traditional methods, necessitates innovative solutions. This study aims to utilize Artificial Intelligence and Machine Learning to address these challenges and accurately identify cybercriminal sources. Chapter Two explores the background and existing literature in this field to provide insights into the current state of cyber attribution.

Chapter Two
Background & Literature

2 Chapter Two

2.1 Background

To trace cybercriminals, organizations can leverage Artificial Intelligence (AI) and Machine Learning (ML) based cyber attribution systems that enable real-time identification of the source and culprits behind cyber-attacks. These systems should be design to manage to find cyber-attacks, analyze data, and aid organization to whether implement possible intervention or prosecution laws against cybercriminals.

Hence, the necessary importance of providing complete justice to victims of cybercrime through compensatory remedies and punishing offenders with the highest type of punishment (Arandjelovic, 2023). This necessitates strict statutory laws to regulate criminal activities in cyberspace along with the use of AI and ML tools to detect and analyze the true source of criminal activities.

Artificial intelligence and ML algorithms can be trained to recognize patterns and anomalies in network traffic and identify suspicious behavior, such as a phishing email, IP spoofing, or a DDoS attack. This can help organizations quickly respond to threats and prevent damage to their systems.

Traditional cybersecurity methods such as IP address tracing and digital forensics have several limitations, reviewed elsewhere (alazab et al., 2023), and for this reason, these may not always provide accurate results. However, AI and ML tools can be used to improve the accuracy and efficiency of cybersecurity. Furthermore, the use of these tools to analyze and assimilate cyberattacks and threat data, organizations can rapidly and accurately identify the source and perpetrators of a cyber-attack. Artificial intelligence and ML-based cybersecurity systems can provide organizations with advanced threat detection capabilities, helping to protect against the significant financial and reputational damage caused by cybercrime.

The integration of AI and ML in cyber attribution systems is essential for rapid source identification in cyberattacks, enhancing cybersecurity, and countering traditional investigative limitations. The literature review will delve into these advancements.

2.2 Literature Review

Artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools in the field of cybersecurity, enabling organizations to enhance their defense against cyber threats. These technologies automate the identification of cyberattacks in real-time, enabling swift and effective identification of the true source of cyberattacks. Several studies have demonstrated the potential of AI and ML in various cybersecurity domains, such as threat detection, source identification, and attack mitigation. However, there are limited comparative studies on this subject matter.

2.2.1 Cyber Threat or CyberAttack Detection

In the realm of cybersecurity, the evolution of detection algorithms takes center stage, driving transformative changes. Zainel and Kocak. (2022) stand at the forefront, pioneering research that underscores the prowess of deep learning algorithms and Convolutional Neural Networks (CNNs). Their work revolves around dissecting network traffic with precision, promptly flagging intrusions. These AI-based techniques outshine traditional methods, boasting detection rates that soar beyond 98%. The integration of deep learning marks a significant milestone, ushering in a new era characterized by enhanced accuracy and efficiency in cyber threat detection.

Amidst the unpredictable landscape of zero-day exploits, Saharkhizan et al. (2020) embark on a journey marked by adaptability. Leveraging Recurrent Neural Networks (RNNs), they delve into the analysis of system call sequences. Their model excels at identifying emerging cyberattack patterns, shedding light on AI's potential to mitigate the ever-evolving threats that loom on the horizon. This innovative approach instills hope in the face of constantly shifting cyberattack landscapes.

In the intricate realm of insider cyber threats, Meira et al. (2022) rise to the challenge by developing a hybrid model that merges Anomaly Detection and Natural Language Processing (NLP) techniques. Their focus turns to human interactions and behaviors, particularly scrutinizing employee communications. The outcome is a model that excels at identifying suspicious behaviors indicative of data exfiltration and unauthorized activities. This fusion of AI and NLP encapsulates the multifaceted nature of effective threat detection, recognizing that the human element is integral to the cybersecurity puzzle.

The looming specter of adversarial attacks on AI-powered systems necessitates fortified defenses, and Wang et al. (2023) heed the call. They delve into the domain of Adversarial Machine Learning for intrusion detection, proposing an architecture that identifies both conventional attacks and adversarial manipulations. Their development work underscores the critical need for resilient security systems capable of withstanding the intelligent cyber onslaught that threatens our digital landscapes.

Expanding the horizons of artificial intelligence's impact, Katzef et al. (2022) set their sights on safeguarding critical infrastructures, notably industrial control systems (ICS). Their approach employs Generative Adversarial Networks (GANs) for anomaly detection within ICS networks. By learning the normative behaviors of these pivotal systems, their GAN-based framework demonstrates exceptional sensitivity to deviations, showcasing its potential in fortifying critical infrastructures.

The synergy of AI and big data analytics emerges as a potent strategy, as exemplified by Pazho et al. (2022). They harness graph-based deep learning to construct a comprehensive model adept at detecting intricate cyberattack patterns within expansive networks. This synergy captures nuanced relationships between entities, resulting in enhanced detection accuracy and amplified threat identification.

2.2.2 Detection of the Source or Origin of Cyber Attacks

In the dynamic field of cybersecurity, artificial intelligence and machine learning have emerged as formidable tools for specialized cyber threat detection. Mohmand et al. (2022) have made significant contributions, showcasing the effectiveness of AI-powered frameworks. Their work delves into uncovering the origins of Distributed Denial-of-Service (DDoS) attacks and isolating malware within network traffic. Employing a spectrum of AI techniques, from Bayesian networks to Support Vector Machines (SVMs), they achieve impressive precision tailored to specific cyberattack classifications.

Moving forward, Mahmud and Tari (2022) have pioneered a machine learning-based cyber attribution framework. Their approach leverages network traffic features to identify the sources of cyberattacks, boasting an impressive accuracy rate of 91.6%. This highlights the potential of machine learning in attributing cyber activities to their origins, a crucial aspect of cyber threat detection. However, in the realm of phishing detection, Ojewumi et al. (2022) shine a spotlight on a machine learning model utilizing a Random Forest Classifier. Their method achieves a remarkable accuracy rate of 97% in detecting phishing emails, underscoring the capability of machine learning to trace the origins of deceptive online activities.

Expanding our horizons further, Ahsan et al. (2022) embark on a comprehensive exploration, employing a spectrum of machine learning and statistical methods. Deep learning and Bayesian classification are among the techniques they employ. Their research underscores the efficacy of these methods in accurately pinpointing the origin of cyberattacks, highlighting the value of machine learning in the attribution process.

Advancing into the realm of advanced persistent threats (APTs), AL-Aamri et al. (2023) introduce a hybrid model that amalgamates unsupervised clustering and supervised learning. By integrating diverse data sources, their approach excels in associating APTs with their likely sources, making a significant contribution to the attribution process.

Addressing the unique challenges posed by cyberattacks in cloud environments, Mishra and Tyagi (2020) employ an ensemble learning approach that leverages cloud-specific features. Their model, incorporating network and log data, showcases enhanced accuracy in tracing the origins of cloud-based attacks, which exploit dynamic infrastructures.

The spotlight then turns to nation-state cyber-attacks in Kida and Olukoya's (2022) study. They employ natural language processing (NLP) techniques to attribute attacks to specific linguistic groups. This innovative approach involves analyzing linguistic patterns within attack-related code and documents, shedding light on potential threat actors behind state-sponsored activities.

With the escalation of cyber-attack sophistication, Albasheer et al. (2022) propose a hybrid model that intertwines deep learning and graph-based analysis. This approach proves effective in dissecting intricate attack campaigns, demonstrating remarkable accuracy in tracing convoluted attack paths and identifying likely orchestrators.

Finally, Rashid et al. (2023) venture into a novel avenue by harnessing user behavior analytics (UBA) alongside machine learning to attribute attacks to specific individuals. Their model focuses on behavioral patterns, effectively identifying insider threats and individual-specific malicious activities, thereby significantly aiding source identification.

2.3 Research Gaps

The current research on AI and ML-based cyber attribution systems still has several gaps. Firstly, there is a lack of studies comparatively investigating the accuracy of ML and AI models in effectively identifying the source of the cyber-attack. While network traffic data is commonly used, incorporating other relevant data sources could enhance the accuracy and efficiency of cyber attribution.

Secondly, limited research has been conducted on the real-time capabilities of AI and ML-based cyber attribution systems. Real-time detection and response are crucial in minimizing the potential damages caused by cyber-attacks. Further exploration of the real-time capabilities of these systems is necessary to ensure their effectiveness in dynamic and rapidly evolving cybersecurity landscapes.

In summary, the literature review underscores the significant impact of AI and ML in enhancing cybersecurity, particularly in threat detection and source attribution. However, comparative studies are limited, and real-time capabilities require further exploration. Chapter Three delves into the methodology and discussion, addressing these gaps and advancing our understanding of AI and ML in cyber attribution.

Chapter Three
Methodology & Discussion

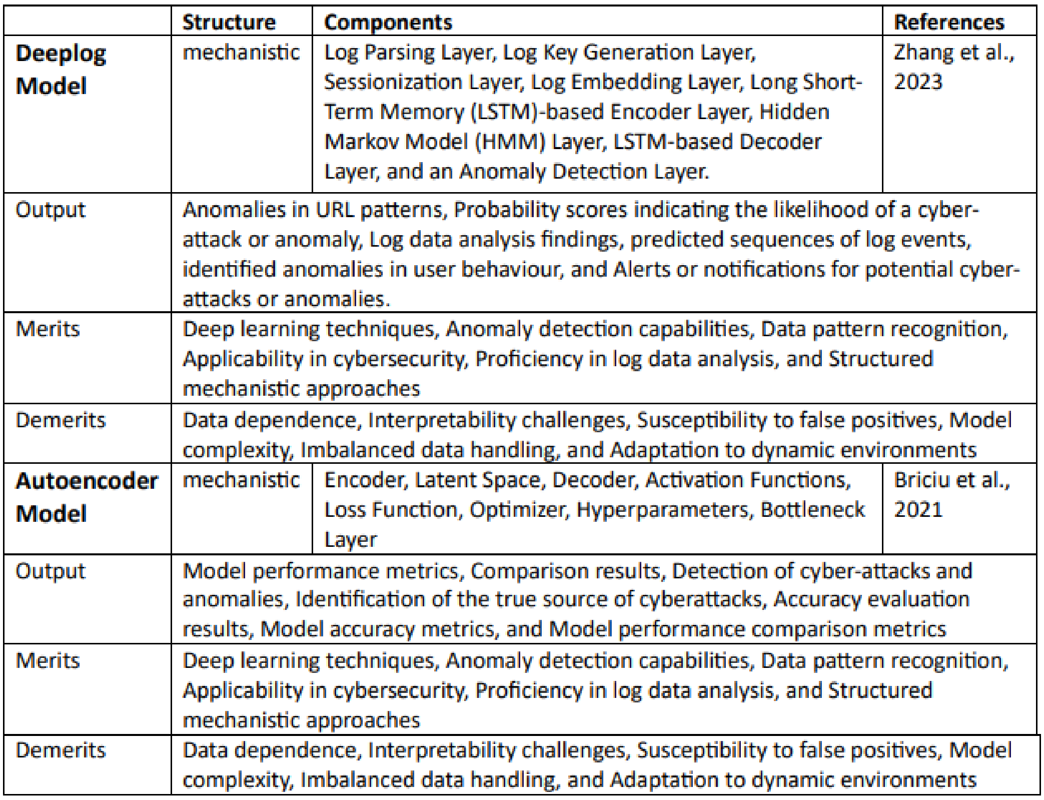
3 Chapter Three

3.1 Methodology

3.1.1 Data collection and processing

Models in Table 1 were identified and selected based on the literature describing their ability to detect cyber-attacks and anomalies in cyber attribution. They are encompassing a variety of crucial features, including deep learning techniques, anomaly detection capabilities, data pattern recognition, applicability in cybersecurity, proficiency in log data analysis, and structured mechanistic approaches. These models were selected with a specific purpose and scope in mind: to identify the true source of cyberattacks through the utilization of provided datasets, model training, accuracy evaluation, and performance comparison.

They make use of essential input variables, including URL features, content features, label information, and model parameters, delivering crucial output in the form of model performance metrics and comparison results. However, it's worth acknowledging their limitations, encompassing factors like data dependence, interpretability challenges, susceptibility to false positives, model complexity, imbalanced data handling, and adaptation to dynamic environments. These models were meticulously selected after an exhaustive review of their input variables, outputs, limitations, and associated coding.



	Structure	Components	References
Deeplog Model	mechanistic	Log Parsing Layer, Log Key Generation Layer, Sessionization Layer, Log Embedding Layer, Long Short-Term Memory (LSTM)-based Encoder Layer, Hidden Markov Model (HMM) Layer, LSTM-based Decoder Layer, and an Anomaly Detection Layer.	Zhang et al., 2023
Output	Anomalies in URL patterns, Probability scores indicating the likelihood of a cyber-attack or anomaly, Log data analysis findings, predicted sequences of log events, identified anomalies in user behaviour, and Alerts or notifications for potential cyber-attacks or anomalies.		
Merits	Deep learning techniques, Anomaly detection capabilities, Data pattern recognition, Applicability in cybersecurity, Proficiency in log data analysis, and Structured mechanistic approaches		
Demerits	Data dependence, Interpretability challenges, Susceptibility to false positives, Model complexity, Imbalanced data handling, and Adaptation to dynamic environments		
Autoencoder Model	mechanistic	Encoder, Latent Space, Decoder, Activation Functions, Loss Function, Optimizer, Hyperparameters, Bottleneck Layer	Briciu et al., 2021
Output	Model performance metrics, Comparison results, Detection of cyber-attacks and anomalies, Identification of the true source of cyberattacks, Accuracy evaluation results, Model accuracy metrics, and Model performance comparison metrics		
Merits	Deep learning techniques, Anomaly detection capabilities, Data pattern recognition, Applicability in cybersecurity, Proficiency in log data analysis, and Structured mechanistic approaches		
Demerits	Data dependence, Interpretability challenges, Susceptibility to false positives, Model complexity, Imbalanced data handling, and Adaptation to dynamic environments		

Table 1

3.1.2 Data for evaluating models

Before evaluating the models, a thorough examination of the data was conducted to ensure there was no overlap between the data sets used for model development and those utilized for model evaluation. As a result, any data used in the model's development phase was excluded from the data sets employed for model evaluation. The evaluation of the models, focus on their ability to detect cyber-attacks and anomalies in cyber attribution, involved categorizing strategies related to various aspects, including URL features, content features, label information, and model parameters. The data collected from this evaluation process provided essential insights into model performance, accuracy, and comparative analysis. This evaluation considered factors such as model complexity, susceptibility to false positives, handling of imbalanced data, adaptability to dynamic environments, and the potential for adoption by stakeholders in the field of cyber attribution, extending beyond the research scope.

3.1.3 Model Evaluation

The criterion used for evaluation of models for cyber attribution was based on two criteria. Firstly, the Mean Square Prediction Error (MSPE) was utilized, defined as:

$$MSPE = \frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2 \tag{1}$$

where O_i represents the observed value, P_i denotes the predicted value, and n signifies the total number of observations, combining both observed and predicted values. The Root Mean Square Prediction Error (RMSPE) was calculated, and models with the lowest RMSPE were considered as the best-performing ones.

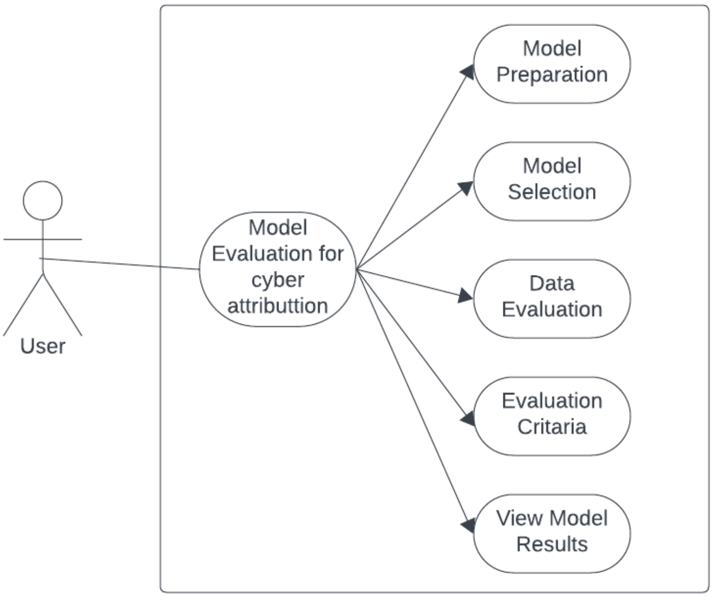
In this section, the methodology for model selection, data collection, and evaluation was outlined. Models were selected based on their ability to detect cyber-attacks and anomalies in cyber attribution. The next section, the Discussion, will delve into the Process Flow for Cyber Attribution Analysis.

3.2 Discussion

3.2.1 Process Flow for Cyber Attribution Analysis

In this section, we present various diagrams that illustrate key aspects of our cyber attribution analysis process:

Find the Use Case diagram below

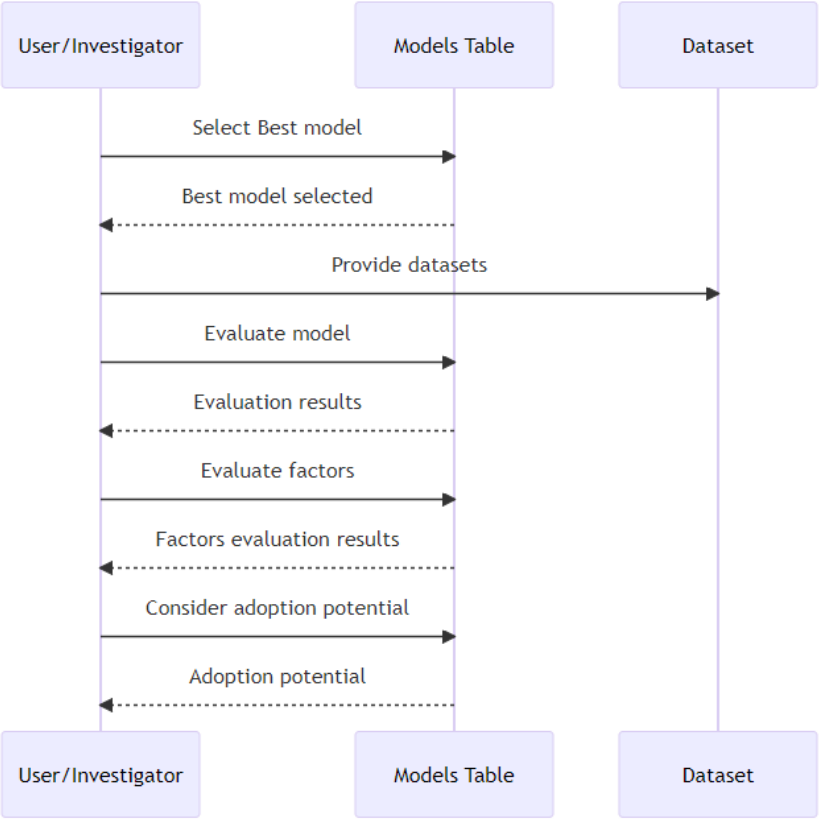


Use Case diagram

The User interacts with the system through the following use cases:

- Model Selection: Selecting a model from the available options based on specific criteria.
- Model Preparation: Preparing the selected model for the cyber attribution task.
- Model Evaluation for Cyber Attribution: Evaluating the performance and effectiveness of the model in detecting cyber attacks and anomalies in cyber attribution.
- Data Evaluation: Examining the data to ensure there is no overlap between the data sets used for model development and model evaluation.
- Evaluation Criteria: Defining and applying criteria to evaluate the models, including aspects such as URL features, content features, label information, and model parameters.
- View Model Results: Viewing the performance metrics, accuracy, and comparative analysis results of the evaluated models.

Find the Sequence Diagram below



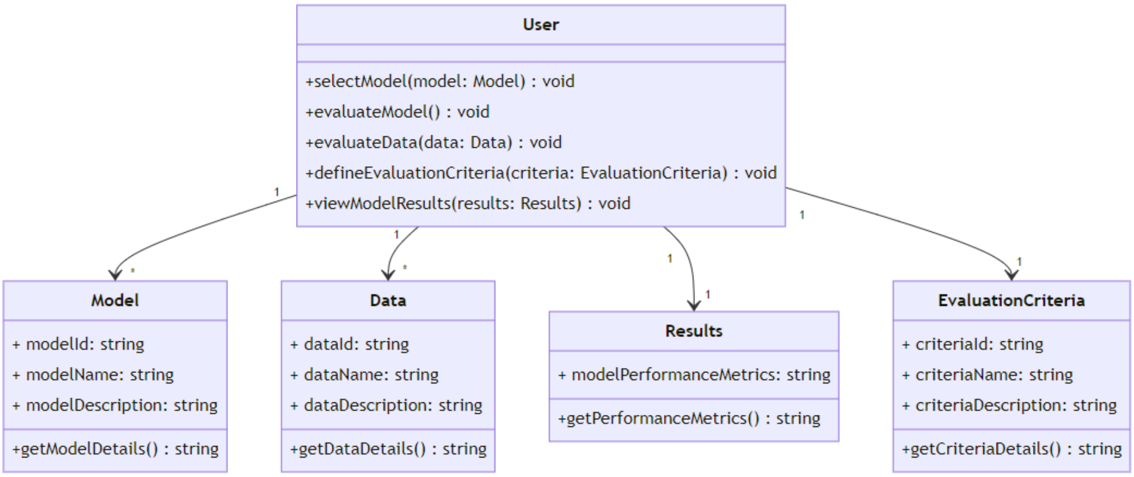
Sequence Diagram

Here is a step-by-step explanation of the actions and interactions depicted in the revised sequence diagram:

- The User/Investigator initiates the interaction by selecting the Model from the Models Table.
- The Models Table receives the request and acknowledges the selection by providing the Model.
- The User/Investigator proceeds to provide the necessary datasets to the Models Table for evaluation.
- The User/Investigator communicates with the Models Table, initiating the evaluation of Model using the provided datasets.

- The Models Table performs the evaluation and responds with the evaluation results to the User/Investigator.
- The User/Investigator further interacts with the Models Table to evaluate additional factors related to the Best model's performance or suitability.
- The Models Table processes the request and provides the evaluation results of these additional factors to the User/Investigator.
- Finally, the User/Investigator considers the adoption potential of the Model by communicating with the Models Table.
- The Models Table responds with the adoption potential information, indicating whether the Model is suitable for adoption or deployment.

Find the Class Diagram below


User
+selectModel(model: Model) : void
+evaluateModel() : void
+evaluateData(data: Data) : void
+defineEvaluationCriteria(criteria: EvaluationCriteria) : void
+viewModelResults(results: Results) : void
1
*
1
*
1
1
1
1
Model
+ modelId: string
+ modelName: string
+ modelDescription: string
+getModelDetails() : string
Data
+ dataId: string
+ dataName: string
+ dataDescription: string
+getDataDetails() : string
Results
+ modelPerformanceMetrics: string
+getPerformanceMetrics() : string
EvaluationCriteria
+ criteriaId: string
+ criteriaName: string
+ criteriaDescription: string
+getCriteriaDetails() : string

Class Diagram

This class diagram represents a functioning model in the context of the cyber attribution system.

The classes in this diagram include:

- **User:** Represents a user of the system. selecting a model, evaluating the model's performance, evaluating data, defining evaluation criteria, and viewing model results.
- **Model:** Represents a specific model available for selection. It has attributes such as "modelId," "modelName," and "modelDescription"

to store information about the model. The class includes a method to retrieve the details of the model.

- **Data:** Represents a dataset used for evaluation. It has attributes such as ”dataId,” ”dataName,” and ”dataDescription” to store information about the data. The class includes a method to retrieve the details of the data.

- **Results:** Represents the results of model evaluation. It has attributes such as ”modelPerformanceMetrics to store relevant information. The class includes methods to retrieve the performance metrics.

- **EvaluationCriteria:** Represents the criteria used for evaluating the models. It has attributes such as ”criteriaId,” ”criteriaName,” and ”criteriaDescription” to store information about the evaluation criteria. The class includes a method to retrieve the details of the criteria.

Associations:

- **User** has relationships with **Model**, **Data**, **Results**, and **EvaluationCriteria**, indicating that a user can interact with multiple models, data sets, evaluation criteria, and results.

In this Discussion section, we presented the process flow for our cyber attribution analysis, using diagrams to illustrate key interactions and components. These visual aids provide a clear overview of our methodology. In Chapter Four, we delve into the Conclusion and Future Research.

Chapter Four
Conclusion & Future Research

4 Chapter Four

4.1 Conclusion

In conclusion, the research study explored the challenges associated with cyber attribution and investigated the applicability of AI and ML in addressing these challenges. The findings indicated that traditional methods of cyber attribution were time-consuming and may not have provided accurate results, especially in the case of spoofed cyber-attacks. However, the integration of AI and ML techniques in cyber attribution systems showed great potential in improving accuracy, efficiency, and real-time capabilities.

The research study highlighted the effectiveness of AI and ML-based approaches in identifying the source of cyber-attacks, such as DDoS attacks, malware, and phishing. These techniques achieved high accuracy rates and outperformed traditional methods. Moreover, the integration of multiple data sources and the use of advanced analytics could further enhance the accuracy and efficiency of cyber attribution systems. Furthermore, further research was needed to address the gaps in the current literature and explore the full potential of AI and ML in cyber attribution.

4.2 Future Research

Future research in this area could focus on the development of more advanced AI and ML models that can handle more complex and sophisticated cyber attacks. Additionally, research could explore the use of AI and ML techniques for predicting and preventing cyber attacks before they occur. Another area of research could be the development of more comprehensive and diverse datasets that can be used to train and evaluate attribution models. Finally, research could focus on the ethical implications of using AI and ML for cyber attribution, including issues related to privacy, bias, and accountability.

References

1. Cuzzocrea, A., Fadda, E., & Mumolo, E. (2022). Cyber-attack detection via non-linear prediction of IP addresses: an innovative big data analytics approach. *Multimedia Tools and Applications*, 1-19.

2. Mimecast. (2023). ”The State of Email Security Report 2023.” Mimecast. Retrieved from `https://www.mimecast.com/state-of-email-security/download-hub/`

3. Toro-Alvarez, M. M. (2023). Hacking. In *Handbook on Crime and Technology*, 334.

4. Gulyás, O., & Kiss, G. (2023). Impact of cyber-attacks on financial institutions. *Procedia Computer Science*, 219, 84-90.

5. Alshaikh, H., Hefny, H. A.,& Darwish, N. R. (2023). Crypto-Ransomware Detection and Prevention Techniques and Tools: A Survey. *International Journal of Computing and Digital Systems*, 13(1), 1-1.

6. Smith, R. G., Sarre, R., Chang, L. Y. C., & Lau, L. Y. C. (Eds.). (2023). *Cybercrime in the Pandemic Digital Age and Beyond.* Springer Nature.

7. South African Banking Risk Information Centre (SABRIC). (2021). ”sabric-crime-stats-2021.” SABRIC. Retrieved from `https://www.sabric.co.za/media-and-news/press-releases/sabric-annual-crime-stats-2021/`

8. Alawida, M., Omolara, A. E., Abiodun, O. I., & Al-Rajab, M. (2022). A deeper look into cybersecurity issues in the wake of Covid-19: A survey. *Journal of King Saud University-Computer and Information Sciences.*

9. Atlam, H. F., & Oluwatimilehin, O. (2022). Business Email Compromise Phishing Detection Based on Machine Learning: A Systematic Literature Review. *Electronics*, 12(1), 42.

10. Despotovi´c, A., Parmakovi´c, A., & Miljkovi´c, M. (2023). Cybercrime and Cyber Security in Fintech. In *Digital Transformation of the Financial Industry: Approaches and Applications* (pp. 255-272). Cham: Springer International Publishing.

11. Hoheisel, R. E. (2022). Ransomware as a tool for diversion and coverup: A possible modus operandi for advanced persistent threats? A (practical) investigation into the use of ransomware as masquerade and distraction (Master’s thesis, University of Twente).

12. Mahor, V., Garg, B., Telang, S., Pachlasiya, K., Chouhan, M., & Rawat, R. (2022, June). Cyber Threat Phylogeny Assessment and Vulnerabilities Representation at Thermal Power Station. In *Proceedings of International Conference on Network Security and Blockchain Technology: ICNSBT 2021* (pp. 28-39). Singapore: Springer Nature Singapore.

13. Akinola, A., & Afonja, A. (2022). Introduction to Cyber-Security. ChudacePublishing.

14. Amini, M., & Bozorgasl, Z. (2023). A Game Theory Method to Cyber-Threat Information Sharing in Cloud Computing Technology. *International Journal of Computer Science and Engineering Research*, 11(4-2023).

15. Riggs, H., Tufail, S., Parvez, I., Tariq, M., Khan, M. A., Amir, A., Vuda, K. V., & Sarwat, A. I. (2023). Impact, Vulnerabilities, and Mitigation Strategies for Cyber-Secure Critical Infrastructure. *Sensors*, 23(8), 4060.

16. Lim, J. W., & Thing, V. L. (2022). Towards Effective Cybercrime Intervention. *arXiv preprint arXiv:2211.09524*.

17. Arandjelovi ́c, O. (2023). Crime and Punishment: A Rethink. *Philosophies*, 8(3), 47.

18. Alazab, A., Khraisat, A., & Singh, S. (2023). A Review on the Internet of Things (IoT) Forensics: Challenges, Techniques, and Evaluation of Digital Forensic Tools.

19. Zainel, H., & Ko¸cak, C. (2022). LAN intrusion detection using convolutional neural networks. *Applied Sciences*, 12(13), 6645.

20. Larabi-Marie-Sainte, S., Ghouzali, S., Saba, T., Aburahmah, L., & Almohaini, R. (2022). Improving spam email detection using deep recurrent neural network. *Indonesian Journal of Electrical Engineering and Computer Science*, 25(3), 1625-1633.

21. Syed, A. F., Atta, A., Nazeer, I., & Anwar, S. (2023). Cybercrimes and Print Media: A Critical Discourse Analysis of News Reporting in Pakistan. *Multimedia Tools and Applications*31(1.)

22. Saharkhizan, M., Azmoodeh, A., Dehghantanha, A., Choo, K.K.R. and Parizi, R.M., 2020. An ensemble of deep recurrent neural networks for detecting IoT cyber attacks using network traffic. *IEEE Internet of Things Journal*, 7(9), pp.8852-8859.

23. Meira, J., Carneiro, J., Bolón-Canedo, V., Alonso-Betanzos, A., Novais, P. and Marreiros, G., 2022. Anomaly detection on natural language processing to improve predictions on tourist preferences. *Electronics*, 11(5), p.779.

24. Wang, Y., Sun, T., Li, S., Yuan, X., Ni, W., Hossain, E. and Poor, H.V., 2023. Adversarial Attacks and Defenses in Machine Learning-Empowered Communication Systems and Networks: A Contemporary Survey. *IEEE Communications Surveys & Tutorials.*

25. Katzef, M., Cullen, A.C., Alpcan, T. and Leckie, C., 2022. Generative Adversarial Networks for anomaly detection on decentralised data. *Annual Reviews in Control*, 53, pp.329-337.

26. Pazho, A.D., Noghre, G.A., Purkayastha, A.A., Vempati, J., Martin, O. and Tabkhi, H., 2022. A Survey of Graph-based Deep Learning for Anomaly Detection in Distributed Systems. *arXiv preprint arXiv:2206.04149.*

27. Mohmand, M. I., Hussain, H., Khan, A. A., Ullah, U., Zakarya, M., Ahmed, A., Raza, M., Rahman, I. U., & Haleem, M. (2022). A machine learning-based classification and prediction technique for DDoS attacks. *IEEE Access*, 10, 21443-21454.

28. Mahmud, A., & Tari, Z. (Machine Learning to Ensure Data Integrity in Power System Topological Network Database Adnan Anwar, Abdun Mahmood, Piplob Ray, MD.)

29. Ahsan, M., Nygard, K. E., Gomes, R., Chowdhury, M. M., Rifat, N., & Connolly, J. F. (2022). Cybersecurity threats and their mitigation approaches using Machine Learning—A Review. *Journal of Cybersecurity and Privacy*, 2(3), 527-555.

30. Ojewumi, T.O., Ogunleye, G.O., Oguntunde, B.O., Folorunsho, O., Fashoto, S.G. and Ogbu, N.J.S.A., 2022. Performance evaluation of machine learning tools for detection of phishing attacks on web pages. *Scientific African*, 16, p.e01165.

31. AL-Aamri, A.S., Abdulghafor, R., Turaev, S., Al-Shaikhli, I., Zeki, A. and Talib, S., 2023. Machine Learning for APT Detection. *Sustainability*, 15(18), p.13820.

32. Mishra, S. and Tyagi, A.K., 2022. The role of machine learning techniques in internet of things-based cloud applications. *Artificial intelligence-based internet of things systems*, pp.105-135.

33. Kida, M. and Olukoya, O., 2022. Nation-State Threat Actor Attribution Using Fuzzy Hashing. *IEEE Access*, 11, pp.1148-1165.

34. Albasheer, H., Md Siraj, M., Mubarakali, A., Elsier Tayfour, O., Salih, S., Hamdan, M., Khan, S., Zainal, A. and Kamarudeen, S., 2022. Cyber-attack prediction based on network intrusion detection systems for alert correlation techniques: *a survey. Sensors*, 22(4), p.1494.

35. Rashid, K., Saeed, Y., Ali, A., Jamil, F., Alkanhel, R. and Muthanna, A., 2023. An Adaptive Real-Time Malicious Node Detection Framework Using Machine Learning in Vehicular Ad-Hoc Networks (VANETs). *Sensors*, 23(5), p.2594.