THE DEVELOPMENT OF CYBER THREAT INTELLIGENCE SYSTEM FRAMEWORK FOR THE MINING INDUSTRY

Proposed by:
MUHAMMAD UMER FAROOQ
2925331

Supervisors:

Dr. Helen Robertson

Dr. Ahsan Mahboob

November 5, 2024

A proposal submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science (Dissertation) in Computer Science

School of Computer Science & Applied Mathematics
University of the Witwatersrand

Abstract

The mining industry's growing reliance on digital technologies has significantly increased its vulnerability to cyber threats, which can lead to severe operational, financial, and environmental consequences. This research aims to develop a comprehensive Cyber Threat Intelligence (CTI) system tailored specifically for the mining sector, utilizing advanced data analytics and machine learning to enhance cyber threat detection and mitigation capabilities. A mixed-methods approach combining systematic literature reviews, empirical testing, and qualitative analyses ensures a robust framework that addresses both technical and ethical considerations. The proposed CTI system will gather, aggregate, and analyze data from diverse sources, such as network logs and open-source intelligence feeds, employing machine learning techniques to identify patterns and anomalies indicative of cyber threats. This research also critically examines the ethical implications of deploying such systems, ensuring compliance with industry standards and regulations. By integrating stakeholder feedback throughout the development and implementation phases, the study ensures that the CTI system is practical, effective, and aligned with the mining industry's specific needs. The anticipated outcomes include improved resilience against cyber threats, minimized risk of operational disruptions, and enhanced protection of sensitive data and infrastructure. This research contributes to the broader field of cybersecurity in critical infrastructure, providing valuable insights into the application of CTI systems within industry-specific contexts.

Declaration

I, Muhammad Umer Farooq, hereby declare the contents of this research proposal to be my own work. This proposal is submitted for the Master of Science by Dissertation in Computer Science at the University of the Witwatersrand. This work has not been submitted to any other university, or for any other degree.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Dr. Helen Robertson from the School of Computer Science and Mathematics, and Dr. M. Ahsan Mahboob, the Head of Sibanye Stillwater Digital Mining Laboratory (DigiMine), for their invaluable guidance and support throughout my research journey. I would like to thatnk and acknowledge the financial support provided by the Sibanye-Stillwater Digital Mining Laboratory (DigiMine), Wits Mining Institute (WMI).

Dr. Robertson's expertise and insightful feedback were instrumental in refining my work, and her encouragement was a constant source of motivation. I am equally grateful to Dr. Mahboob for his mentorship and for providing me with the opportunity to collaborate within the DigiMine lab, where his leadership and profound knowledge in the field greatly enhanced my research experience.

Their unwavering support, constructive feedback, and encouragement have been crucial to the successful completion of this proposal. I am deeply appreciative of their time and efforts, and I feel privileged to have worked under their guidance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Mining Industry is as important part of the global economy. It depends on digital technologies to improve operations, safety, and sustainability. However, this reliance on technology exposes the industry to significant cyber threats that can disrupt operations, compromise safety, and cause financial loss and environmental harm. Given these risks, the need for effective cyber threat detection in mining is more urgent than ever. Traditional and reactive cybersecurity measures often fail to address the complex and evolving threats specific to the mining sector. This has led to the growing importance of Cyber Threat Intelligence (CTI) System for the mining sector, which proactively detect, analyze, and respond to potential threats using intelligence feeds. These systems collect and analyze data from multiple sources, helping organizations identify and mitigate risks before they cause damage. Despite potential threats, the urgency for effective cyber threat detection in mining has never been greater. Traditional reactive cybersecurity measures used in the sector often fall short of addressing the sophisticated and evolving threats unique to the mining sector. This research aims to fill these gaps by developping tailored Cyber Threat Intelligence System Framework specifically for the mining industry. The CTI Framework will use advanced data analytics and machine learning algorithms to detect patterns and anomalies in threats datam and it will improve the industry's ability to defend against cyber-attacks. Additionally, this study will examine the ethical implications of implementing such a system, ensuring it meets both technical and ethical standards. By designing a CTI system tailored to the mining sector, this research will address a critical need and contribute to the broader field of cybersecurity. In Figure: 1.1, World Economic Forum Global Risks Perception Survey 2023-2024 indicating a wide subset of the global population to potential digital and physical exploitation.

1.1 Problem Statement

Mining Industry plays a crucial role in global economic development. Mining operations are adopting digital technologies rapidly to enhance productivity, safety, and efficiency. These are becoming more vulnerable to cyber threats. These threats can disrupt operations, can compromise sensitive data, can cause financial loss, and even it

Figure 1.1: World Economic Forum Global Risks Analysis Survey 2023-2024

can endanger lives and the environment. Evolution Mining is an Australian gold mining company. It suffered a ransomware attack in August 2024 on its IT systems, which, though swiftly contained, underscored the industry's susceptibility to cyber threats. A month earlier, in July 2024, Sibanye-Stillwater also faced a cyberattack, leading to temporary IT system outages and manual processing in some of its operations. Norsk Hydro is a major metals and mining company. It experienced a ransomware attack in 2020 that caused significant operational shutdowns and millions in losses. An attack on a South American in 2017 on a mining firm compromised sensitive geological data. An Australian mining company suffered unauthorized access in 2019 which disrupted automated processes and risked worker safety. These incidents highlight the urgent necessity for robust, proactive cybersecurity measures to protect the industry from evolving threats.

1.2 Research Questions

The following questions are formulated to guide the research for development of a Cyber Threat Intelligence (CTI) system tailored for the mining industry:

1. How can a framework for a Cyber Threat Intelligence (CTI) system be developed specifically for the mining industry?

2. What strategies and technologies can be utilized to efficiently collect, aggregate, and analyze cyber threat data from sources like network logs and open-source intelligence feeds for the mining industry?

3. How can advanced data analysis and machine learning techniques be applied to detect and analyze patterns and anomalies in cyber threat data specific to the mining sector?

4. What are the ethical challenges associated with implementing a CTI system in the mining industry, and how can these be effectively addressed to align with ethical standards and regulations?

1.3 Research Aims and Objectives

The adoption of digital technologies in the mining industry is essential for advancing productivity and safety. But it has also led it to significant cyber vulnerabilities. Cyber threats continuing to evolve. So, the need for a proactive approach to identify, analyze, and mitigate risks has become essential. The research aims and objectives focus on developing a robust CTI system Framework.

1.3.1 Research Aims

The primary aim of this research is to develop an effective Cyber Threat Intelligence (CTI) framework for the mining industry, designed to enhance cybersecurity resilience. This research aims to mitigate the increasing cyber threat risks within the mining industry, which has seen a sharp rise in digital vulnerabilities as it adopts more advanced technologies by developing a CTI Framework tailored to mining industry. This aim involves creating a specialized CTI system that can detect and respond to cyber threats, ensuring the safety, security, and continuity of mining operations in an increasingly digitalized environment.

1.3.2 Research Objectives

The objectives of this research are as follows:

- Develop a framework for a Cyber Threat Intelligence (CTI) System tailored specifically to the mining industry.
- To collect, aggregate, and analyze threat data from various sources, including network logs and open-source intelligence feeds.
- To apply advanced data analysis and machine learning techniques to identify patterns and anomalies within the collected threat data.
- To examine and address the ethical considerations and implications of implementing a CTI framework within the mining industry.

1.4 Limitations

The basic limitation of this research is the generalizability of its findings across the diverse operational environments within the mining industry. While the study aims to develop a robust Cyber Threat Intelligence (CTI) Framework tailored to mining, it relies on specific datasets and threat scenarios that may not comprehensively reflect all real-world contexts. Differences in mining operations, network architectures, and data characteristics across various geographical and organizational settings could impact the performance and applicability of the CTI framework. Additionally, the effectiveness of the proposed data analysis and machine learning techniques in detecting threats and enhancing model interpretability may vary depending on the complexity and heterogeneity of cyber threat data across mining entities. Consequently, while the research aspires to make valuable contributions to the mining sector's cybersecurity capabilities, its application to the broader landscape of mining operations may be constrained.

1.5 Overview

Chapter 2

Background and Literature Review

2.1 Introduction

This chapter discusses the background and prior research related to Cyber Threat Intelligence (CTI) systems in detail. This section provides the foundational understanding of CTI frameworks and situates the current research within the context of existing literature. The review covers the evolution of cybersecurity challenges in the mining industry, key components of CTI frameworks, and the use of data analytics and machine learning techniques to identify cyber threats. The chapter also discusses the ethical considerations and regulatory requirements relevant to developing a CTI system in a highly specialized industrial environment.

2.2 Background

The mining industry's integration into the digital ecosystem has resulted in increased automation, improved safety measures, and enhanced operational efficiency. However, this dependence on digital technology has simultaneously made mining operations susceptible to sophisticated cyber threats. These threats are often complex, ranging from ransomware attacks that can halt operations to Advanced Persistent Threats (APTs) aimed at stealing valuable intellectual property.

Industry 4.0 and Digitalization

The concept of Industry 4.0 has brought a digital transformation across sectors, including mining, through the use of IoT, machine learning, cloud computing, and big data. In mining, technologies like real-time monitoring systems, autonomous mining equipment, and predictive maintenance are becoming the norm. However, this transformation exposes critical infrastructure to cyber vulnerabilities [Wang and Lu 2013; Sajid *et al.* 2016].

Cyber Threat Landscape in Mining

The mining sector faces unique cyber threats due to the critical nature of its operations. Examples of past cyber incidents include ransomware attacks and data breaches targeting geological data or disrupting automated processes. High-profile cases like the ransomware attack on Evolution Mining in 2024 and Norsk Hydro's incident in 2020 illustrate the potential impact on productivity and financial stability.

SCADA Systems and Vulnerabilities

Supervisory Control and Data Acquisition (SCADA) systems are widely used in mining to monitor and control physical processes. However, traditional SCADA systems lack the security measures necessary to counter the vulnerabilities introduced by cloud and IoT integrations. The need for an advanced CTI system becomes evident, as these vulnerabilities can lead to catastrophic failures in mining operations [Wang and Lu 2013].

Cybersecurity Challenges Unique to Mining Operations

- **Remote and Harsh Environments:** Mining operations are often located in remote areas with limited connectivity, making it challenging to deploy and maintain comprehensive cybersecurity solutions.
- **Legacy Systems and Modernization:** Many mining facilities still use outdated systems that are difficult to secure, and updating these without causing operational disruptions is a significant challenge.
- **Supply Chain Vulnerabilities:** The mining industry relies heavily on third-party vendors and contractors, which introduces additional security risks and necessitates robust supply chain security measures.

2.3 Related Work

2.3.1 Cyber Threat Intelligence (CTI) Frameworks

CTI frameworks have emerged as crucial solutions for enhancing the proactive capabilities of cybersecurity defenses. They leverage data collection, analysis, and dissemination to provide actionable insights into emerging threats. Various researchers have proposed different components for CTI frameworks, each tailored to address specific cybersecurity needs.

Existing Framework Comparison

CTI Framework Adoption in Other Critical Infrastructures

Components of a CTI Framework

CTI Data Collector *Role and Functionality:* The data collector component is responsible for aggregating raw cyber threat data from multiple sources. These include OSINT feeds, network logs, vendor-provided threat intelligence, and even data from the dark web. The data collected must be diverse and comprehensive to ensure accurate threat detection.

Literature Examples:

- [Lee and Shon 2016] : Developed an OSINT-focused data collection strategy that ensures timely and relevant threat information. Their framework emphasizes preparing and implementing an OSINT plan before gathering and analyzing data from open sources.
- [Ryandy *et al.* 2020]: Outlined a systematic approach to data collection, emphasizing the importance of processing, analysis, and evaluation to ensure that the collected data is usable for threat intelligence purposes.
- [Tundis *et al.* 2022]: Focused on feature selection and OSINT source identification, demonstrating that a well-designed data collection system can significantly impact the quality of threat intelligence.

Analysis Medium *Role and Functionality:* This component transforms raw data into actionable intelligence through pre-processing, correlation analysis, pattern detection, and anomaly recognition. The analysis medium uses algorithms and heuristics to classify and prioritize threats.

Literature Examples:

- [Kim *et al.* 2016]: Emphasized the importance of structured data analysis, correlation techniques, and the use of YARA rules for malware detection. Their research shows how the analysis medium can derive meaningful insights from large datasets.
- [Noor *et al.* 2019]: Proposed an analysis approach that incorporates cyber threat attribution, helping organizations understand the origin and intent behind cyberattacks.
- [Islam *et al.* 2022]: Integrated network data with a threat detector and alert validation system, highlighting the necessity of real-time data analysis to reduce false positives and improve threat response.

Information Platform *Role and Functionality:* This component acts as the user interface and decision-support system for disseminating analyzed intelligence. It allows for threat information sharing, real-time monitoring, and supports collaborative efforts among stakeholders.

Literature Examples:

- [Böhm *et al.* 2018]: Designed a CTI information platform with features like data filtering, mapping, rendering, and user interaction. Their focus was on making threat intelligence more accessible and actionable for security teams.

- [Kim *et al.* 2016]: Focused on the conversion of data into security rules, making the information platform a critical part of cybersecurity operations. This framework highlighted the use of automated security updates based on the analyzed intelligence.

- [Papastergiou *et al.* 2021]: Proposed a comprehensive information-sharing model that integrates deep and dark web mining, live monitoring, and data protection orchestrators. Their approach, which included the HybridNet and ShareNet components, emphasizes the importance of robust and secure information dissemination.

2.3.2 Advanced Data Analytics and Machine Learning in CTI

The application of advanced data analytics and machine learning in CTI frameworks is becoming increasingly prevalent. These technologies enhance the capability to detect patterns, classify threats, and predict future cyber-attacks.

Machine Learning Techniques

Algorithms such as decision trees, support vector machines (SVMs), and neural networks are commonly used for threat detection. Unsupervised learning methods, like clustering, help in identifying unknown threat patterns, while supervised learning assists in classifying known threats.

Data Fusion and Anomaly Detection

Techniques like time-series analysis and real-time data fusion are essential for identifying anomalies in network behavior. For instance, Islam *et al.* [2022] utilized data from both network and business operations to validate alerts and ensure that only genuine threats are flagged.

Case Studies in Industrial Settings

Review studies that have successfully implemented machine learning in critical infrastructure protection, such as detecting anomalies in SCADA systems or preventing ransomware attacks. Discuss how these methods can be adapted to the mining industry.

Collaborative Cybersecurity Initiatives and Intelligence Sharing

- **Industry Partnerships and Alliances:** The role of collaborations among mining companies and cybersecurity organizations in improving threat intelligence capabilities.
- **Information Sharing Platforms:** The significance of platforms like ISACs (Information Sharing and Analysis Centers) that facilitate threat intelligence sharing across the industry.
- **Global Cybersecurity Alliances:** Participation in global threat intelligence networks to stay updated on emerging cross-border cyber threats.

Future Trends and Emerging Threats

- **Zero Trust Architecture:** Exploring the adoption of zero trust principles in securing industrial networks.
- **Blockchain Technology:** The potential of blockchain for securing data and communications in mining operations.
- **Quantum Computing Risks:** How advancements in quantum computing could pose new cybersecurity threats to the mining industry.

Chapter 3

Research Methodology

3.1 Research design

3.2 Methods

The research will follow a comprehensive, multi-phase approach tailored to address the unique cybersecurity challenges encountered by the mining industry. The process will begin with a requirements analysis, which will involve collecting stakeholder opinions and reviewing relevant literature to identify specific needs and gaps in current cybersecurity frameworks. Then, during the framework development phase, a customized architecture will be developed, integrating data from various sources, such as network logs and open-source intelligence feeds. Following this, real-time threat data will be collected and aggregated, which will then be analyzed using advanced machine learning techniques like to detect patterns, classifying and clustering the logs and anomalies. Ethical considerations will be carefully examined to ensure compliance with industry regulations and data privacy standards. A prototype of the CTI System will be developed and tested in a controlled environment to evaluate its effectiveness. Then, the system's performance will be assessed, and detailed documentation will be prepared, offering insights and recommendations for practical implementation within the mining industry.

3.2.1 Pre-Modelling Phase

The pre-modelling phase focuses on the initial setup and preparation needed to build an effective Cyber Threat Intelligence (CTI) system. It includes data collection, data preprocessing, and defining the threat intelligence goals specific to the mining industry.

- **Data Collection**: Raw data is collected from various sources such as network logs, open-source intelligence (OSINT) feeds, and vendor-specific threat intelligence reports. In mining operations, sensor and operational data can also be leveraged to detect abnormal patterns in the network.

- **Data Preprocessing**: The collected data is cleaned and transformed into a usable format. This involves handling missing data, removing duplicates, and normalizing data formats. Feature extraction techniques such as Principal Component Analysis (PCA) can be applied to reduce noise.
- **Data Labeling**: For supervised learning, historical data is labeled by domain experts as normal or malicious based on past incidents.
- **Goal Definition**: Specific objectives are defined, such as detecting ransomware, Advanced Persistent Threats (APTs), or insider threats, which will influence the models and algorithms chosen in the next phase.

3.2.2 Modelling Phase

The modelling phase involves building machine learning models that predict and detect cyber threats based on the pre-processed data.

- **Model Selection**: Machine learning models such as decision trees, random forests, support vector machines (SVMs), or neural networks are selected based on the data type and threat scenarios.
- **Feature Engineering**: Relevant features are engineered to improve the model's ability to detect cyber threats. Time-series analysis may be used for real-time threat detection in operational environments.
- **Model Training**: Models are trained using supervised learning on labeled data or unsupervised learning on unlabelled data to detect anomalies. Cross-validation techniques are used to avoid overfitting.
- **Threat Classification**: The trained model is used to classify network activity or system behavior into predefined threat categories (e.g., malware, phishing, Denial of Service attacks).

3.2.3 Post-Modelling Phase

After building the models, the post-modelling phase focuses on evaluation, validation, and deployment.

- **Model Validation**: Models are validated using test datasets. Key performance metrics such as accuracy, precision, recall, and F1 score are used to measure the model's effectiveness.
- **Model Tuning**: Hyperparameter tuning (e.g., grid search or random search) is used to optimize the model. This process involves adjusting key parameters to improve performance.
- **Deployment Readiness**: Once validated, the model is integrated into existing systems (e.g., Security Information and Event Management systems) for real-time threat detection. The model undergoes stress testing to ensure it can handle real-time data streams.

3.2.4 Experimental Setup

The experimental setup defines how the CTI models are tested in a controlled environment before full deployment.

- **Test Environment Setup**: A simulated or controlled mining network is created, including virtual machines, network traffic simulators, and log generators to mimic realistic scenarios.
- **Data Injection**: Historical data and simulated attack data are injected into the system to evaluate its ability to detect and respond to various threats.
- **Performance Measurement**: The system's performance is measured using metrics such as detection rate, false alarm rate, and response time. Resource usage (CPU, memory) is also monitored to ensure scalability.
- **Comparison with Existing Systems**: The CTI framework's performance is compared to existing cybersecurity systems in the mining industry to assess improvements.

3.2.5 Optimization and Training Models

The optimization and training phase focuses on refining the models to maximize their performance and efficiency.

- **Hyperparameter Optimization**: Techniques such as grid search or Bayesian optimization are used to identify the best hyperparameters (e.g., learning rates, kernel functions) to improve detection accuracy.
- **Continuous Learning**: As new threats emerge, the model is retrained with updated data. Incremental learning or transfer learning methods may be used to update the model without complete retraining.
- **Efficiency Optimization**: The model is optimized for real-time performance through techniques such as model pruning, quantization, or the use of lightweight architectures.
- **Final Model Training**: The final optimized model is trained on the entire dataset to ensure robustness and accuracy. It is then ready for deployment into the CTI system.

3.3 Limitations

3.4 Ethical Considerations

Chapter 4

Schedule of Work

Chapter 5

Conclusion

The CTI System for the Mining Industry aims to bridge the gap between the mining industry's unique operational demands and the growing need for robust cybersecurity measures. By developing a tailored CTI System, the study seeks to provide an architecture design of a CTI System. The development of the CTI System will enhance the ability to detect and respond to cyber threats proactively in the mining industry. The CTI System's integration of advanced data analytics and machine learning will offer more precise and real-time action against evolving threats. Additionally, this research will address ethical and regulatory concerns and ensure the proposed solution is effective and compliant with mining industry standards. The findings from this research will contribute significantly to both the mining industry and the broader field of cybersecurity. By providing a specialized system for cyber threat intelligence, this study will help safeguard critical mining operations, protecting both assets and the environment. The lessons learned and the methodologies developed can serve as a model for other industries facing similar cybersecurity challenges, marking a step forward in the ongoing effort to secure vital industrial infrastructure.

Appendix A

Extra Stuff

A.1 What is an appendix?

An appendix is useful when there is information that you need to include, but breaks the flow of your document, e.g. a large number of figures/tables may need to be shown, but maybe only one needs to be in the text and the rest are just included for completeness.

References

[Böhm *et al.* 2018] Fabian Böhm, Florian Menges, and Günther Pernul. Graph-based visual analytics for cyber threat intelligence. *Cybersecurity*, 1(1):16, 2018.

[Islam *et al.* 2022] Chadni Islam, M Ali Babar, Roland Croft, and Helge Janicke. Smart-validator: A framework for automatic identification and classification of cyber threat data. *Journal of Network and Computer Applications*, 202:103370, 2022.

[Kim *et al.* 2016] Daegeon Kim, JiYoung Woo, and Huy Kang Kim. "i know what you did before": General framework for correlation analysis of cyber threat incidents. In *MILCOM 2016-2016 IEEE Military Communications Conference*, pages 782–787. IEEE, 2016.

[Klein and Celik 2017] R. Klein and T. Celik. The Wits Intelligent Teaching System: Detecting student engagement during lectures using Convolutional Neural Networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2856–2860, Sep. 2017.

[Lee and Shon 2016] Seokcheol Lee and Taeshik Shon. Open source intelligence base cyber threat inspection framework for critical infrastructures. In *2016 Future Technologies Conference (FTC)*, pages 1030–1033. IEEE, 2016.

[Medoh and Telukdarie 2022] Chuks Medoh and Arnesh Telukdarie. The future of cybersecurity: a system dynamics approach. *Procedia Computer Science*, 200:318–326, 2022.

[Noor *et al.* 2019] Umara Noor, Zahid Anwar, Tehmina Amjad, and Kim-Kwang Raymond Choo. A machine learning-based fintech cyber threat attribution framework using high-level indicators of compromise. *Future Generation Computer Systems*, 96:227–242, 2019.

[Papastergiou *et al.* 2021] Spyridon Papastergiou, Haralambos Mouratidis, and Eleni-Maria Kalogeraki. Handling of advanced persistent threats and complex incidents in healthcare, transportation and energy ict infrastructures. *Evolving Systems*, 12(1):91–108, 2021.

[Ryandy *et al.* 2020] Ryandy, Charles Lim, and Kalpin Erlangga Silaen. Xt-pot: Exposing threat category of honeypot-based attacks. In *Proceedings of the 2020 International Conference on Engineering and Information Technology for Sustainable Industry*, pages 1–6, 2020.

[Sajid *et al.* 2016] Anam Sajid, Haider Abbas, and Kashif Saleem. Cloud-assisted iot-based scada systems security: A review of the state of the art and future challenges. *Ieee Access*, 4:1375–1384, 2016.

[Tundis *et al.* 2022] Andrea Tundis, Samuel Ruppert, and Max Mühlhäuser. A feature-driven method for automating the assessment of osint cyber threat sources. *Computers & Security*, 113:102576, 2022.

[Wang and Lu 2013] Wenye Wang and Zhuo Lu. Cyber security in the smart grid: Survey and challenges. *Computer networks*, 57(5):1344–1371, 2013.