The Development of Cyber Threat Intelligence System Framework for the Mining Industry

Subtitle

School of Computer Science & Applied Mathematics
University of the Witwatersrand

MUHAMMAD UMER FAROOQ
2925331

Supervised by Dr Helen Robertson

October 28, 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

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.

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, a vital part of the global economy, increasingly 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) systems, which proactively detect, analyze, and respond to potential threats. 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 often fall short of addressing the sophisticated and evolving threats unique to the mining sector. This research aims to fill these gaps by developping a comprehensive Cyber Threat Intelligence System Framework for the mining industry. The Framework will use advanced data analytics and machine learning to detect patterns and anomalies in threat data, improving 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 is a vital sector that plays a crucial role in global economic development. As mining operations are increasingly adopting digital technologies to enhance productivity, safety, and efficiency, they are becoming more vulnerable to cyber threats. These threats can disrupt operations, compromise sensitive data, cause financial loss, and even endanger lives and the environment. In August 2024, Evolution

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

Mining, an Australian gold mining company, suffered a ransomware attack 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. Similarly, in 2020, Norsk Hydro, a major metals and mining company, experienced a ransomware attack that caused significant operational shutdowns and millions in losses. Additionally, a 2017 attack on a South American mining firm compromised sensitive geological data, and in 2019, an Australian mining company suffered unauthorized access, disrupting automated processes and risking 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 research questions are formulated to guide the development of a comprehensive Cyber Threat Intelligence (CTI) system tailored for the mining industry:

1. **How can a comprehensive framework for a Cyber Threat Intelligence (CTI) system be developed specifically for the mining industry?**
    - What specific data sources, such as network logs and open-source intelligence feeds, are most relevant to the mining industry?
    - How can these data sources be effectively collected, aggregated, and processed to create a robust threat intelligence system?
2. **What advanced data analysis and machine learning techniques can be applied to identify patterns and anomalies in collected threat data?**
    - What machine learning models are most effective for detecting cyber threats in the context of mining industry operations?
    - How can the application of these techniques enhance threat detection and mitigation capabilities specific to the mining sector?

3. **What are the ethical considerations and implications of implementing a Cyber Threat Intelligence framework in the mining industry?**
    - What are the primary ethical concerns related to data privacy, security, and surveillance in the mining industry?
    - How can these ethical issues be addressed to ensure compliance with regulations and industry standards?

1.3 Research Aims and Objectives

Research Aims and Objectives...

1.3.1 Research Aims

Research Aims

1.4 Research Objectives

The objectives of this research are as follows:

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

1.5 Limitations

1.6 Overview

Chapter 2

Background and Literature Review

2.1 Introduction

2.1.1 Background

2.1.2 Related Work

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.

Figure captions are at the bottom. Table titles are at the top of the table as seen in Table 5.1 on the following page.

Table 5.1: Table Name

Col1	Col2
R0,C0	R0,C1
R1,C0	R1,C1

Chapter 6

Some Referencing Tricks

CleverRef and VarioRef are helpful:

- Normal Ref: See Figure 1.1
- CleverRef: See Figure 1.1 and Table 5.1
- CleverRef+VarioRef: See Figure 1.1 on page 2 and Table 5.1 on the facing page

Chapter 7

IDE/Editors

Overleaf has a great online editor for latex. Use it.

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

[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.