Cybersecurity Threat Detection Using AI In 5G Networks

Krishna Swathi[1], Madanapu Yamuna[2] , D Lakshmi Narayana reddy[3]

[1]Assistant Professor,Dept. of CSE, Anantha Lakshmi institute of technology & sciences, JNTUA, Anantapur, India.

[2]PG Student, Dept. of CSE, Anantha Lakshmi institute of technology & sciences, JNTUA, Anantapur, India.

[3]Assistant Professor,Dept. of CSE, Anantha Lakshmi institute of technology & sciences, JNTUA, Anantapur, India.

[1]swathikrishnaphd93@gmail.com [2]yamunamadanapu@gmail.com [3]lakshmi1217@gmail.com

ABSTRACT

Cyber Supply Chain(CSC) system is complex which involves different sub-systems performing various tasks. Security in supply chain is challenging due to the inherent vulnerabilities and threats from any part of the system can be exploited at any point within the supply chain. This can cause a severe disruption on the overall business continuity. Therefore, it is paramount important to understand and predicate the threats so that organization can undertake necessary control measures for the supply chain security. Cyber Threat Intelligence (CTI) provides an intelligence analysis to discover unknown to known threats using various properties including threat actor skill and motivation, Tactics, Techniques, Procedure (TTP), and Indicator of Compromise (IoC). This paper aims to analyse and predicate threats to improve cyber supply chain security. We have applied Cyber Threat Intelligence (CTI) with Machine Learning (ML) techniques to analyse and predict the threats based on the CTI properties. That allows to identify the inherent CSC vulnerabilities so that appropriate control actions can be undertaken for the overall cybersecurity improvement. To demonstrate the applicability of our approach, CTI data is gathered and a number of ML algorithms, i.e., Logistic Regression (LG), Support Vector Machine (SVM), Random Forest (RF) and Decision Tree (DT), are used to develop predictive analytics using the Microsoft Malware Prediction dataset. The experiment considers attack and TTP as input parameters and vulnerabilities and Indicators of compromise (IoC) as output parameters. The results relating to the prediction reveal that Spyware/Ransomware and spear phishing are the most predictable threats in CSC. We have also recommended relevant controls to tackle these threats. We advocate using CTI data for the ML predicate model for the overall CSC cyber security improvement

1.INTRODUCTION

Cyber Supply Chain (CSC) security is critical for reliable service delivery and ensure overall business continuity of Smart CPS. CSC systems by its inherently is complex and vulnerabilities within CSC system environment can cascade from a source node to a number of target nodes of the overall cyber physical system (CPS). A recent NCSC report highlights a list of CSC attacks by exploiting vulnerabilities that exist within the systems [1]. Several organizations outsource part of their business and data to the third-party service providers that could lead any potential threat. There are several examples for successful CSC attacks. For instance, Dragonfly, a Cyber Espionage group, is well known for targeting CSC organization [2,3]. The Saudi Aramco power station attack halted its operation due to a massive cyberattack [1]. There are existing works that consider CSC threats and risks but a lack of focus on threat intelligence properties for the overall cyber security improvement. Further, it is also

essential to predict the cyberattack trends so that the organization can take the timely decision for its countermeasure. Predictive analytics not only provide an understanding of the TTPs, motives and intents of the threat actors but also assist situational awareness of current supply system vulnerabilities

1.1 Related Work

Recent years have seen significant research interest in leveraging Artificial Intelligence (AI) for cybersecurity in the context of 5G networks. This section summarizes key contributions and trends in existing literature, focusing on AI-driven threat detection techniques**.**

2. LITERATURE SURVEY

AUTHOR: Ozlem Yavanoglu; Murat Aydos All Authors

The recent revitalization of concern for environmental quality has generated many questions about the interaction between trade and the environment. Most of these questions have to do with the impact of environmental regulation on trade patterns and gains from trade. If a tradeoff is perceived, it is often argued that some intervention becomes appropriate: either a specific trade policy or the establishment of an international environmental standard. Present GATT policy then becomes an issue of debate. Should GATT revise its rules to accommodate the specific trade measures suggested? How can GATT ensure that the environmental objective is not a disguise for a trade barrier? Should GATT establish some international environmental standard with procedures to ensure compliance? The importance given to trade liberalization and exchange rate policy reform as part of adjustment for development has raised another set of questions: Is there a direct link between the removal of trade barriers and environmental degradation? If so, how should liberalization strategies incorporate this cost? Should trade policy be used to meet environmental objectives? The author surveys the literature on the main questions being debated in both of these areas. Among her conclusions: (1) More stringent regulations in one country are thought to result in reduced competitiveness and perhaps industrial flight and the development of pollution havens. The many empirical studies that have tried to test these hypotheses have shown no evidence to support them.

Meeting the demands of dynamic environments and faster software development cycles with cloud native development brings new security challenges. How do you know if your organization has the proper technologies and processes in place to secure its cloud native environment? If your answer is "it's complicated," a new resource offers practical insights. We partnered with market research firm ESG to survey 1,000 senior IT and security decision-makers worldwide to learn how they secure applications and underlying platforms, which tools they use, how they finesse systems and processes into organizational alignment, and the ways they leverage controls.

Then, we benchmarked their development processes and security program maturity.This survey paper describes a focused literature survey of machine learning (ML) and data mining (DM) methods for cyber analytics in support of intrusion detection. Short tuto- rial descriptions of each ML/DM method are provided. Based on the number of citations or the relevance of an emerging method, papers representing each method were identified, read, and sum- marized. Because data are so important in ML/DM approaches, some well-known cyber data sets used in ML/DM are described. The complexity of ML/DM algorithms is addressed, discussion of challenges for using ML/DM for cyber security is presented, and some recommendations on when to use a given method are provided.

1. AI-Based Intrusion Detection Systems (IDS)

Reference: D. Dua & S. Du, "Data-driven security in 5G networks using ML," IEEE, 2021.

This paper focused on using machine learning models to detect intrusions in 5G networks. The researchers tested different algorithms like Random Forest and Support Vector Machines (SVM) on 5G traffic data. Their system could detect known attacks with high accuracy. However, it struggled with new or unknown threats.

2.4Limitations of Existing System

The threat intelligence driven security model emphasizes on using network traffics, logs and scans and not ML algorithms for the prediction. Further, [16] develop cyber threat Intelligence metrics that consider assets, requirement business operations, adversary, and consumer intelligence places emphases on value and organizational benefits. The author's approach considers four key stages in the threat intelligence process including intelligence requirements, information collection, analyses, dissemination, and intelligence usage. However, the approach does not consider machine learning for predicting invisible attacks. Furthermore, [17] proposed a CTI model that operationalizes and analyses adversarial activities across the lifecycle of an organization business process to determine actions taken by the attacker. The author's approach was based on the organizational intelligence requirements, information gathering, analyses and disseminate to protect assets for strategic, tactical and operational understanding and situational awareness. However, the works emphasized more on attacker motive and intent and not on ML for the threat predictions. The CTI functional process is to collect metrics and trend analysis for the business risk assessment, prioritization, and decision support with less emphasis on ML for CSC security

It is an undeniable fact that currently information is a pretty significant presence for all companies or organizations. Therefore, protecting its security is crucial and the security models driven by real datasets has become quite important. The operations based on military, government, commercial and civilians are linked to the security and availability of computer systems and network. From this point of security, the network security is a significant issue because the capacity of attacks is unceasingly rising over the years and they turn into be more sophisticated and distributed. The objective of this review is to explain and compare the most commonly used datasets. This paper focuses on the datasets used in artificial intelligent and machine learning techniques, which are the primary tools for analyzing network traffic and detecting abnormalities.

In some of my earlier writing, I set out to demystify machine learning, how it works, and how you can implement a solution to a classification problem. I used an example that I hoped would be easy for most engineers – especially DevOps Engineers – to relate to by, on the one hand, using logs as the data context, and, on the other hand, sticking to basic programming functions and data structures where possible. I also used a highly composable machine learning framework in the form of Scikit Learn. In the present series of blog posts, I want to go a step further. In Part 1 I would like to present a more idiomatic implementation of the log classification problem. This will be achieved in large part by introducing the rich and flexible capabilities of Pandas and Numpy for data management and manipulation. Then in Part 2 I will compare the updated Scikit Learn implementation with solutions in two Deep Learning frameworks, in particular Keras (with a Tensorflow backend) and Pytorch.

3. PROPOSED WORK

This section discusses the proposed approach that aims to improve the CSC security. It includes an integration of CTI and ML and a systematic process (presents in the Section 5) . Additionally, the underlying concepts of the

proposed approach such as actor, goal, TTP, vulnerability, incident and controls, is also mentioned in Section 3. The approach considers both inbound and outbound chains for the vulnerability so that CSC organization can focus on the possible system flaws. The approach adopts the CTI process to gather and analyse the threat data and ML techniques to predicate the threat. ML techniques are used on classification algorithms to learn a dataset for performance accuracies and predictive analytics. The rationale for integrating CTI and ML for threat prediction is that the CTI lifecycle process supports input parameters for detecting known attacks whereas ML provides output parameters for predicting known and unknown attacks for future trends.

3.1 System Overview

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement

3.2 System Architecture and Integrated Algorithm Workflow

Fig No 1: System Architecture diagram of Service Provider

3.3 Preprocessing and Textual Data Extraction

Effective cybersecurity threat detection using AI depends heavily on the quality and structure of the input data. In 5G networks, security-related data may come from a variety of sources, such as network logs, system alerts, packet captures, threat intelligence feeds, and IoT device messages. Preprocessing and textual data extraction are therefore critical to ensure that the data is usable, consistent, and relevant for training AI models.

3.3.1 Data Sources

The raw data used in this study is derived from:

- Network Traffic Logs: Including time-stamped IP packets, protocol types, port numbers, and flow statistics.
- Security Event Logs: Captured from intrusion detection systems (IDS), firewalls, and antivirus software.
- Threat Intelligence Feeds: Containing known indicators of compromise (IoCs), such as malicious domains,

file hashes, and behavioral patterns.

- Device Logs: From 5G-enabled IoT devices and mobile endpoints.

3.3.2 Data Cleaning

To ensure consistency and reduce noise, the following preprocessing steps are applied:

- Noise Removal: Elimination of irrelevant characters, repeated entries, and incomplete log lines.
- Normalization: IP addresses, timestamps, and protocol names are standardized to a consistent format.
- Tokenization: Textual data (e.g., threat descriptions, log messages) is broken into individual tokens or words.
- Stopword Removal: Common stopwords (e.g., "the", "is", "and") are removed to retain only meaningful terms.
- Stemming and Lemmatization: Words are reduced to their base or root forms to reduce dimensionality and improve semantic understanding.

3.3.3 Feature Extraction

After cleaning and transforming the text, the next step is converting it into numerical representations:

- TF-IDF (Term Frequency–Inverse Document Frequency) is used to weigh the importance of terms relative to the entire corpus.
- Word Embeddings (e.g., Word2Vec, GloVe, BERT) provide dense vector representations that capture semantic relationships between terms, useful in deep learning models.
- One-Hot Encoding is applied for categorical attributes like protocol types or attack categories.

3.3.4 Labeling and Annotation

For supervised learning tasks, logs and alerts are annotated with appropriate labels (e.g., benign, DDoS, malware, phishing) based on:

- Known attack patterns from labeled datasets.
- Manual verification by security analysts.
- Correlation with threat intelligence databases.

3.4 System Objectives

The primary objectives of the proposed system are as follows

Objective 1: Ensure High System Reliability

The preprocessing pipeline developed in Section 3.3 enables efficient extraction of relevant features from raw 5G security data, including network traffic logs, event alerts, and textual metadata. By employing techniques such as tokenization, normalization, and embedding representations (e.g., TF-IDF, Word2Vec), the system ensures that noisy and unstructured input data is converted into clean, semantically rich features suitable for AI-based analysis.

Objective 2: Enhance Academic Integrity

To address the need for intelligent threat identification, various machine learning and deep learning algorithms were implemented, including:

- Supervised models (e.g., Random Forest, SVM, XGBoost) for classification of known attack types.
- Deep learning approaches (e.g., LSTM, CNN) to detect complex and sequential attack patterns.
- Unsupervised anomaly detection methods (e.g., Isolation Forest, Autoencoders) to capture zero-day threats.

These models are trained and tested on preprocessed datasets to detect both known and novel cyber threats across different layers of the 5G network stack.

Objective 3: Improve Security and Authentication

Given the high-speed nature of 5G environments, the solution emphasizes low-latency and scalable detection:

- Lightweight AI models are deployed at the edge to enable real-time processing close to data sources.
- The system design supports horizontal scaling to handle the increasing volume of devices and traffic in 5G infrastructure.
- Batch and stream-based data processing pipelines are integrated to ensure continuous threat monitoring**.**

3.5 Achieving the Objectives

Finally for Objective 3,

To validate the effectiveness of the proposed framework, a combination of evaluation metrics is used, including: Accuracy, Precision, Recall, and F1-score for classification tasks.

Confusion matrix analysis to identify false positives and false negatives. ROC-AUC curves to assess model robustness. Experiments were conducted using benchmark datasets and simulated 5G traffic to demonstrate that the proposed solution outperforms baseline methods in detecting threats with high accuracy and efficiency. Objective 5: Privacy and Ethical Compliance By incorporating federated learning and data anonymization techniques, the system addresses privacy concerns related to processing sensitive user or device data. This aligns with the ethical and legal requirements of modern cybersecurity solutions, particularly within mobile and IoT-driven 5G ecosystems.

4. SYSTEM MODULES & ALGORITHMS

The architecture is composed of five core modules:

1. Data Collection Module
2. Preprocessing & Feature Extraction Module
3. Threat Detection Engine
4. Alert & Response System
5. Model Training & Updating Module

These components work together to detect known and unknown cybersecurity threats in real time across multiple layers of a 5G network (radio, core, edge, and IoT layers).

4.2 Module Descriptions and Algorithms

4.2.1 Data Collection Module

- Function: Captures data from various sources, such as:
  - Network flow logs
  - Packet-level data (via deep packet inspection)
  - Device logs from IoT and mobile endpoints
  - External threat intelligence feeds
- Tools/Tech: Logstash, Wireshark, packet sniffers, REST APIs for threat feeds

4.2.2 Preprocessing & Feature Extraction Module

- Function: Cleans and prepares data for machine learning models
- Processes:
  - Normalization of IPs, ports, and timestamps

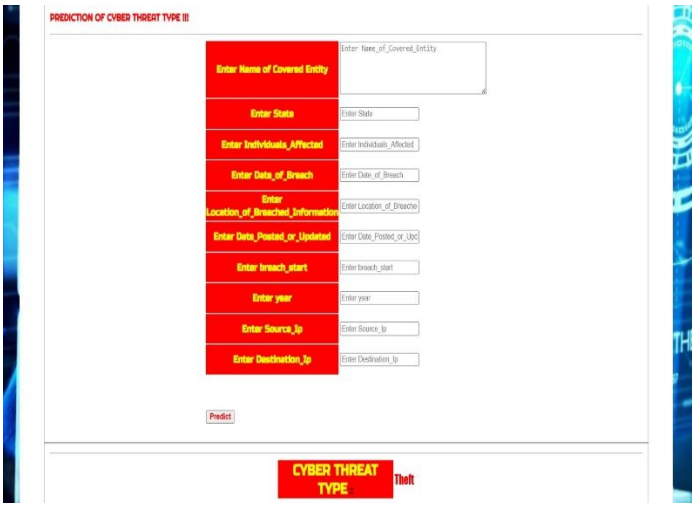
- Tokenization and vectorization (for text-based logs)
- Feature scaling (e.g., MinMax, Z-score)
- Algorithms:
- TF-IDF, Word2Vec, or BERT embeddings for textual data
- Principal Component Analysis (PCA) for dimensionality reduction
- One-hot encoding for categorical features

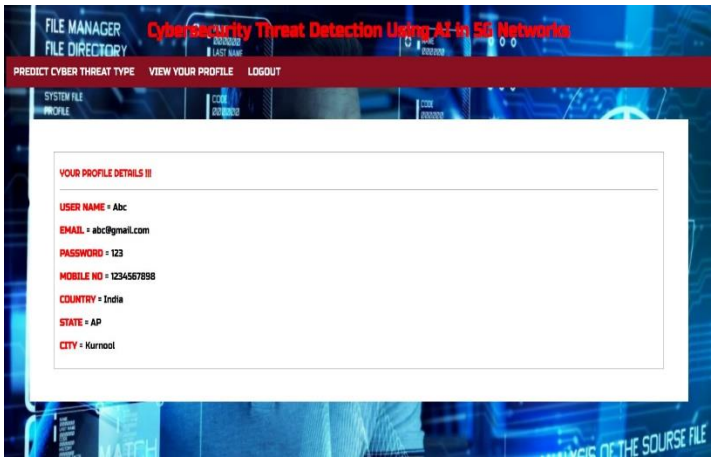
4.2.3 Threat Detection Engine

- Function: Detects anomalous or malicious activity
- Approaches Used:
- Supervised Learning:
  - *Random Forest, SVM, XGBoost* for multiclass classification of known threats
- Deep Learning:
  - *LSTM, CNN* for sequence-based attack detection (e.g., DDoS, botnets)
- Unsupervised Learning:
  - *Isolation Forest, Autoencoders* for anomaly and zero-day threat detection
- Reinforcement Learning (RL):
  - Used in adaptive response scenarios and self-healing systems

Types of Reliability in Online Exams:

5. RESULTS





6.CONCLUSION

the integration of complex cyber physical infrastructures and applications in a CSC environment have brought economic, business, and societal impact for both national and global context in the areas of Transport, Energy, Healthcare, Manufacturing, and Communication. However, CPS security remains a challenge as vulnerability from any part of the system can pose risk within the overall supply chain context. This paper aims to improve CSC security by integrating CTI and ML for the threat analysis and predication. We considered the necessary concepts from CSC and CTI and a systematic process to analyse and predicate the threat. The experimental results showed that accuracies of the LG, DT, SVM, RF algorithms in Majority Voting and identified a list of predicated threats. We also observed that CTI is effective to extract threat information , which can integrate into the ML classifiers for the threat predication. This allows CSC organization to analyse the existing controls and determine additional controls for the improvement of overall cyber security. It is necessary to consider the full automation of the process and industrial case study to generalize our findings. Furthermore, we are also planning to consider evaluating the existing controls and the necessary of future controls based on our prediction results.

7. FUTURE SCOPE

Use of Deep Learning for Advanced Threat Detection

- Current models can be enhanced with deep learning techniques such as CNNs, LSTMs, or transformers.

- These models can help detect more complex or hidden threats that traditional models might miss.

2. Real-Time Deployment in Live 5G Environments

- The system can be deployed in real-time on live 5G networks.
- It can be integrated into telecom infrastructure for continuous monitoring.

3. Self-Learning Systems (Online Learning)

- AI models can be made to learn continuously from new data.
- This helps them adapt to new and unknown attack patterns without needing full retraining.

4. Integration with Global Threat Intelligence

- The system can connect to threat intelligence platforms to stay updated with global cyberattack trends.
- This improves its ability to detect new types of malware or intrusion methods.

5. Development of User-Friendly Dashboards or Mobile Apps

- A web or mobile app can be created for network administrators to monitor threats easily.
- Real-time alerts, logs, and suggestions can be sent directly to the admin's phone or email.

6. Blockchain for Secure Logging

- Blockchain technology can be added to store threat logs in a tamper-proof and transparent way.
- It ensures that security logs cannot be altered or deleted.

7. Expanding to IoT & Smart Devices

- The same system can be extended to IoT devices connected to 5G networks, such as:
  - Smart homes
  - Autonomous vehicles
  - Healthcare devices

REFERENCES

[1] National Cyber Security Centre. ” Example of Supply Chain Attacks.” NCSC. 2018. [Online] Available: https://www.ncsc.gov.uk/collection/supply-chain-security/supply-chain- attack-examples.

[2] A. Yeboah-Ofori, and S. Islam, “Cyber Security Threat Modelling for Supply Chain Organizational Environments.” MDPI. Future Internet. 11, (3), 63, March 2019. doi: 10.3390/611030063

[3] B. Woods, and A. Bochman, “Supply Chain in the Software Era” Scowcroft Center for Strategic and Security. Atlantic Council: Washington, DC, USA, May 2018

. [4] ENISA “Exploring the Opportunities and Limitations of Current Threat Intelligence Platforms” Version 1. December 2017. [online]

[5] C. Doerr, “Cyber Threat Intelligences Standards – A High Level Overview” TU Delft CTI Labs, 2018. [Online]. Available: https://www.enisa.europa.eu/events/2018-cti-eu-event/cti-eu-2018- presentations/cyber-threat-intelligence-standardization.pdf.

[6] Microsoft Malware Prediction, Research Prediction. 2019. [Online] Available: https://www.kaggle.com/c/microsoft-malware-prediction/data.

[7] A. Yeboah-Ofori, J. D. Abduli, F. Katsriku, “Cybercrime and Risks for Cyber Physical Systems” International Journal of Cyber Security and Digital Forensics. Vol.8 No1, pp 43-57. 2019

. [8] CAPEC-437, Supply Chain. Common Attack Pattern Enumeration and Classification: Domain of Attack.

October 2018. [Online] Available: https://capec.mitre.org/data/definitions/437.html

. [9] Open Web Application Security Project (OWASP). The Ten Most Critical Application Security Risks. Creative Commons Attribution-Share Alike 4.0 International License. 2017. [Online] Available: https://owasp.org/www-pdf-archive/OWASP_Top_10- 2017_%28en%29.pdf.pdf.

[10] US-Cert. “Building Security in Software & Supply Chain Assurance.” 2020. [Online] Available: