JOURNAL OF BASIC SCIENCE AND ENGINEERING

NEW FRAMEWORK FOR AUTOMATIC CYBER THREAT SYSTEM USING ARTIFICIAL INTELLIGENCE PROCESSING

[1]Srimanthula Vijay Kumar [2]Dr. Ratna Raju Mukiri [3]S. Amarnath Babu
[1]M. Tech Scholar, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
[2]Associate Professor, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
[3]Associate Professor, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
e-mail: mukiriratnaraju001@gmail.com

Abstract: Critical infrastructure (CI) typically refers to the essential physical and virtual systems, assets, and services that are vital for the functioning and well-being of a society, economy, or nation. Cyber threats pose significant risks to organizations and individuals, necessitating the development of advanced systems for threat intelligence modeling and identification. Cyber strategists at a national level require AI-based decision support systems for deciding a country's cyber posture or preparedness. Theproposes an AI-based solution that autonomously collects multidimensional cyber-attack data on social media posts on cyber-related outcry. The proposed system provides critical analytical capability in the cyber-threat spectrum and uses sophisticated AI based algorithms for anomaly detection, prediction, sentiment analysis, location detection, translation, A wide variety of Cyber Threat Information (CTI) is used by Security Operation Centres (SOCs) to perform validation of security incidents and alerts. Security experts manually define different types of rules and scripts based on CTI to perform validation tasks. We have performed experiments on real healthcare ecosystems in Fraunhofer Institute for Biomedical Engineering, considering in particular three different healthcare scenarios, namely implantable medical devices with the purpose of demonstrating the feasibility of our approach. s. We also cover how these techniques can address diverse cybersecurity concerns such as threat detection, mitigation, prediction, diagnosis for root cause findings, and so on in different CI sectors, such as energy, defence, transport, health, water, agriculture,This approach offers more insight into the nature of the danger and suggests possible ways to address it.

Index Terms: Cyber threat intelligence, Cyber threat prediction, Decision support system, Cyber anomaly detection, Cyber-attack dashboard, Social-media analysis, Security Automation, Machine Learning, Artificial Intelligence, Natural Language Processing.

1. INTRODUCTION

Identifying and analyzing Cyber Threat Information (CTI) is an important part of validating security alerts and incidents [1]. Any piece of information that helps organizations identify assess, and monitor cyber threats is known as CTI. On the other hand the requirement of Artificial Intelligence (AI) based sophisticated analytical algorithms have been demonstrated in [2]. While AI can be used to conduct sophisticated cyber-attacks AI can also be used in detecting cyber-attack as demonstrated in none of these existing literatures provide country-level historical cyber statistics, country level cyber prediction, and anomaly detections with AI [3]. Cyber threat

intelligence is crucial for understanding and mitigating potential cyber threats. It involves the collection, analysis, and interpretation of information about existing and potential cyber-attacks enabling organizations to make informed decisions and bolster their defenses [4]. In this context, the development of a robust cyber threat intelligence system becomes imperative for maintaining the security and integrity of digital infrastructures the will be underpinned by a homogeneous information network [5]. Cyber threat intelligence is also available on informal sources, such as public blogs dark webs, forums, and social media platforms. Informal sources allow any person or entity on the Internet to publish in real-time the threat information in natural language format[6]. Recent advances in artificial intelligence (AI) such as data science (DS) modeling and machine learning (ML) techniques have drastically changed we analyze data and use the extracted knowledge for automation and intelligent decision-making in various real-world application domains, including cyber security applications [7].

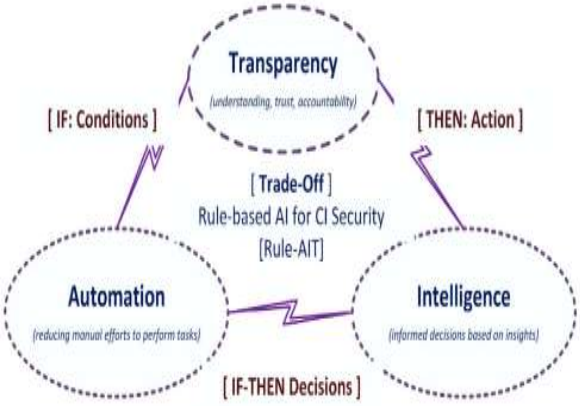

Figure. 1.human understanding and decision explanation

It shows the results of the experiments performed in a real-world healthcare ecosystem scenario from Institute for Biomedical Engineering (IBMT) with three assets. implantable medical devices wear ables and bio bank assessing the applicability and the usefulness of the proposed work [8].

2. RELATED WORK

The cyber-attacks have also targeted medical devices such as infusion healthcare services such as medicine delivery of the healthcare system [9]. For instance implantable cardiac devices get security features associated with the system architecture, which uses device-to-device [10].These models are also capable of adapting to new and emerging threats by continuously analyzing recent data and learning from patterns which makes them more resilient to dynamic cyber security trends. With a better understanding of context and relationships within data these models can reduce false positives and negatives eventually helps to build a more powerful model according to today's need [11]. While the Internet has become an indispensable infrastructure for businesses governments and societies there is also an increased risk of cyber-attacks with different motivations and intentions. Preventing organizations from cyber exploits needs timely intelligence about cyber vulnerabilities and attacks referred to as threats [12]. Existing literature demonstrates that there

has been a growing interest using AI-based Anomaly detection and other deep learning methodologies for detecting cyber-attack within these studies anomaly detection has been used on Network traffic data for detecting intrusions [13]. Most detectors produce alerts upon detecting malicious activity that require a security team to act on it. These alerts require validation before analyzing them for decision validator performs a task related to prioritizing and identifying the relevance [14].

3. SYSTEM ARCHITECTURE

Threat actors constantly target to obtain patient-sensitive information and hacking is considered one of the main causes that discloses patient sensitive healthcare data [15]. These models are also capable of adapting to new and emerging threats by continuously analyzing recent data and learning from patterns, which makes them more resilient to dynamic cyber security trends [16]. With a better understanding of context and relationships within data these models can reduce false positives and negatives which eventually helps to build a more powerful model according to today needs [17]. The proposed approach presented in this paper automatically acquires multidimensional data on cyber-attack covering dimensions such as Spam, Ransom ware, Web Threats, Network Attacks, Malicious Mails, Local Infections, Exploits, and On Demand Scans [18].

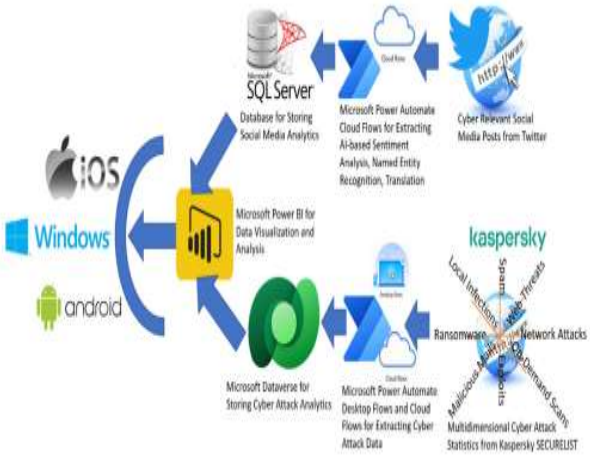

Figure.2. System Architecture

4. PROPOSE SYSTEM

Cyber threats can also be diagnosed transparently by examining the dependencies and relationships between entities in rules derived from data discovered knowledge and rule-based AI modeling facilitates key features like automation intelligence and transparency according to today's need for cyber security modeling [19].

• **Automation** - The capability to perform tasks without manual intervention the need for human intervention through executing the generated rules.

• **Intelligence** - The capability of rule-based modeling for informed decision-making based on the discovered knowledge and automatic learning patterns and useful insights extracted from security data.

• **Transparency and Trust** - Transparency typically refers to model visibility such as rule structure and understanding decisions through the generated rules which is the foundation of explainable and responsible AI development [20].

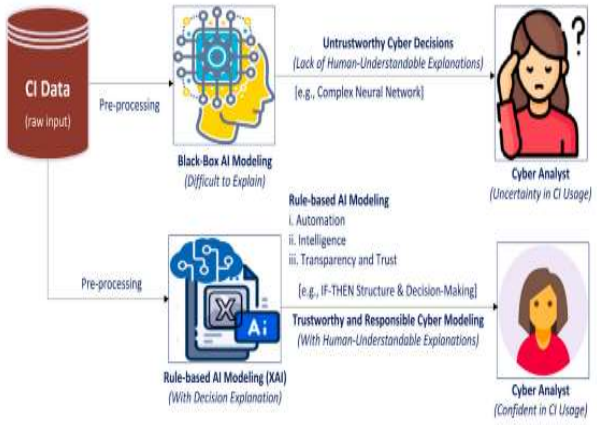

Figure. 3. Rule-based AI modeling vs Black-Box solutions from the perspective of a CI cyber

Proposed Framework

Our proposed framework Smart Validator that automates the identification and classification of CTI for validation of alerts. It comprises of three layers: (i) threat data collection layer, (ii) threat data prediction model building layer and (iii) threat data validation laye [21]. A parser utilizes various information processing techniques to extract information from the output of a scraper and organize the data into a structured and language-agnostic format [22]. For forums or blog posts written in natural language a parser is required to extract threat information from sentences. NLP tools and techniques is used to build a parser based on the structure of a document and information required by security team [23]. The collector can also query external data sources to find out missing information about available threat data.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

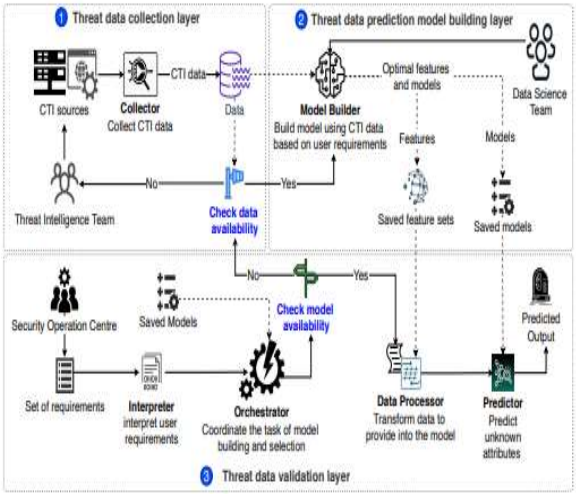

Figure 4: An overview of Smart Validator identification and validation of cyber threat data

The domain name is text based preprocessed data is passed to a feature engineering module where data is transformed into features which are used as input for the ML algorithms [24].

5. ALGORITHIM

STEP1: Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models which are typically decision trees.

STEP2: Logistic regression Classifiers Logistic regression analysis studies the association between a categorical dependent variable and a set of independent variables. The name logistic regression is used when the dependent variable has only two values such as 0 and 1 or Yes and No.

STEP3: SVM In classification tasks a discriminant machine learning technique aims at finding, based on an independent and identically distributed training dataset discriminant function that can correctly predict labels for newly acquired instances.

STEP4: Convolutional Neural Network (CNN) A Convolutional Neural Network (CNN) is a type of deep learning algorithm specifically designed for image processing and recognition tasks. They excel at assigning importance to various objects and features within the images through convolution layers which apply filters to detect local patterns

Threat Data Validation Layer

The instead of setting a fixed value we consider providing security teams the flexibility to set the confidence score based on their application needs. We design Algorithm describing the key steps of the threat data validation layer. These steps are coordinated and orchestrated by the orchestrator. SOC preference is the input of Algorithm. The interpreter receives the SOC requirements and extracts observed attribute unknown attribute and confidence scores

ALGORITHM

Algorithm Model building with orchestrator in threat data validation layer

1: Input: AS , confidence score
2: Output: predictedData
3: Interpret (AS, confidence score)
4: IsModels= CheckModel(AS, confidence score)
5: if IsModels true then
6: model, featureEng = getModel(AS,confidence score)
7: processedData = transformData(featureEng, AS)
8: predictedData = predictOutput(model,processedData)
9: else
10: IsData = CheckData(AS)
11: if IsData true then
12: CTIData = RetrieveData(CTI, AS)
13: model = buildModel(CTIData, AS, confidence score)
14: if model is built then
15: go to step 6
16: else
17: go to step 19
18: else 1
9: RequestData(AS)
20: return NotApplicable
21: return predicted Data

Our proposed framework, Smart Validator streamlines gathering, identification and classification of CTI. Smart Validator allows a security team to swiftly make a response about incoming alerts. As most information is generated in a structured way, it can be easily pre-processed to share through a CTI platform such as MISP or Collective Intelligence Framework (CIF) to benefit diverse security teams [25].

6. EXPERIMENTAL RESULTS

The web search results provide valuable insights into the tools and methodologies used in cyber threat intelligence and analysis, underscoring the importance of these practices in modern cyber security. The presented information underscores the need for organizations to leverage advanced tools and methodologies to proactively identify, analyze, and mitigate cyber threats in an increasingly complex threat landscape. An inbuilt function from the sci-kit-learn library standard Scaler() was used to standardize the data. The function transformed data into a normalized distribution to remove outliers from the data, allowing for building more accurate prediction models

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Figure 5: Country Wise cyber threat analysis

7. CONCLUSION AND FUTURE WORK

We propose a novel framework Smart Validator to build an effective and efficient validation tool using CTI that automates the validation of the security alerts and incidents based on SOC's preferences. Different from the manual approaches, Smart Validator is designed in a way so that SOCs can add their requirements without worrying about collecting CTI and using CTI to build a validation model. The proposed solution furnishes an all-encompassing comprehension of cyber threats across the globe. Using a set of dashboards available in all platforms a strategic decision maker can perform. Further mitigating the possible contribution of false positives and negatives found by the ML NLP module to the threat level in addition to a constant increasing of the NL corpora size we could integrate the threat prioritization obtained by our method also with the information available in the CS KBs, correcting the threat level with a weight extracted from the KBs. Overall, our study on knowledge discovery and rule-based AI modeling opens a promising path for next-generation CI security modeling and can be used as a reference guide for CI researchers, professionals, and policy makers

8. REFERENCES

[1] B. D. Le, G. Wang, M. Nasim, and A. Babar, ''Gathering cyber threat intelligence from Twitter using novelty classification,'' 2019, arXiv:1907.01755

[2] M. Malatji, A.L. Marnewick, S. Von Solms, Cybersecurity capabilities for critical infrastructure resilience, Inf. Comput. Secur. 30 (2) (2022) 255–279.

[3] R. Baskerville, P. Spagnoletti, J. Kim, Incident-centered information security: Managing a strategic balance between prevention and response, Inf. Manag. 51 (1) (2014) 138–151.

[4] I.H. Sarker, Multi-aspects AI-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview, Secur. Priv. (2023) e295.

[5] Dr. Ratna Raju Mukiri, Estimate Requirement of Online Report Analysis using Random Forest Regression,Vol. 10, 04-Special Issue.
[6] I.H. Sarker, AI-Driven Cybersecurity and Threat Intelligence: Cyber Automation, Intelligent Decision-Making and Explainability, Springer, 2024.
[7] M. Touhiduzzaman, S.N.G. Gourisetti, C. Eppinger, A. Somani, A review of cybersecurity risk and consequences for critical infrastructure, 2019 Resil. Week (RWS) 1 (2019) 7–13.
[8] I. Stellios, P. Kotzanikolaou, M. Psarakis, C. Alcaraz, J. Lopez, A survey of iot-enabled cyberattacks: Assessing attack paths to critical infrastructures and services, IEEE Commun. Surv. Tutor. 20 (4) (2018) 3453–3495.
[9] H. Kayan, M. Nunes, O. Rana, P. Burnap, C. Perera, Cybersecurity of industrial cyber-physical systems: a review, ACM Comput. Surv. 54 (11s) (2022) 1–35.
[10] M.A. Husnoo, A. Anwar, R.K. Chakrabortty, R. Doss, M.J. Ryan, Differential privacy for IoT-enabled critical infrastructure: A comprehensive survey, IEEE Access 9 (2021) 153276–153304.
[11] D. Bhamare, M. Zolanvari, A. Erbad, R. Jain, K. Khan, N. Meskin, Cybersecurity for industrial control systems: A survey, Comput. Secur. 89 (2020) 101677.
[12] A.M. Koay, R.K.L. Ko, H. Hettema, K. Radke, Machine learning in industrial control system (ICS) security: current landscape, opportunities and challenges, J. Intell. Inf. Syst. 60 (2) (2023) 377–405.
[13] S. Nazir, S. Patel, D. Patel, Assessing and augmenting SCADA cyber security: A survey of techniques, Comput. Secur. 70 (2017) 436–454.
[14] L. Das, S. Munikoti, B. Natarajan, B. Srinivasan, Measuring smart grid resilience: Methods, challenges and opportunities, Renew. Sustain. Energy Rev. 130 (2020) 109918.
[15] E.M. Wells, M. Boden, I. Tseytlin, I. Linkov, Modeling critical infrastructure resilience under compounding threats: a systematic literature review, Prog. Disaster Sci. (2022) 100244
[16] K.-C. Lee, C.-H. Hsieh, L.-J.Wei, C.-H. Mao, J.-H. Dai, and Y.-T. Kuang, ‘‘Sec- buzzer: Cyber security emerging topic mining with open threat intelligence retrieval and timeline event annotation,’’ Soft Comput., vol. 21, no. 11, pp. 2883– 2896, Jun. 2017.
[17] A. Ritter, E. Wright, W. Casey, and T. Mitchell, ‘‘Weakly supervised extraction of computer security events from Twitter,’’ in Proc. 24th Int. Conf. World Wide Web, May 2015, pp. 896–905.
[18] A. Queiroz, B. Keegan, and F. Mtenzi, ‘‘Predicting software vulnerability using security discussion in social media,’’ in Proc. Eur. Conf. Cyber Warfare Secur., 2017, pp. 628–634.
[19] A. Bose, V. Behzadan, C. Aguirre, and W. H. Hsu, ‘‘A novel approach for detection and ranking of trendy and emerging cyber threat events in Twitter streams,’’ in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2019, pp. 871–878.
[20] Kaspersky, Cyber threat statistics, 2023, [Online]. Available: Accessed32023.
[21] Kaspersky, Daily ransomware cyber threat statistics, 2023, [Online]. Available: Accessed32023.

[22] Gao, Y., Li, X., Li, J., Gao, Y., Guo, N., 2018. Graph mining-based trust evaluation mechanism with multidimensional features for largescale heterogeneous threat intelligence, in: 2018 IEEE International Conference on Big Data (Big Data), IEEE. pp. 1272–1277.
[23] Gibert, D., Mateu, C., Planes, J., 2020. The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. Journal of Network and Computer Applications 153, 102526.
[24] Ibrahim, A., Thiruvady, D., Schneider, J., Abdelrazek, M., 2020. The challenges of leveraging threat intelligence to stop data breaches. Front. Comput. Sci. 2: 36. doi: 10.3389/fcomp .
[25] Islam, C., Babar, M.A., Nepal, S., 2019. A multi-vocal review of security orchestration. ACM Computing Surveys (CSUR) 52, 1–45.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930