A Review on the Analysis of Cyber Threat Detection Systems

T. C. Swetha Priya[1], Nallula Shruthi Goud[2], Myakala Shreya[3], Munigala Vaishnavi[4]

[1]Department of Information Technology, Stanley college of Engineering and technology for women, India
E-mail: tcswethapriya@stanley.edu.in
[2]Department of Information Technology, Stanley college of Engineering and technology for women, India
E-mail: nallulashruthi@gmail.com
[3]Department of Information Technology, Stanley college of Engineering and technology for women, India
E-mail: shreya.myakala@gmail.com
[4] Department of Information Technology, Stanley college of Engineering and technology for women,India
E-mail: vaishnavimunigala@gmail.com

1. Abstract

With the advancement of digital world, Cybercrimes like hacking or ransomware are causing big problems ranging from money loss to invading national security. So, we need Cybersecurity for protecting computers and networks from malicious attackers. It keeps networks, systems, and programs safe from digital attacks that want to steal, change, or destroy important information. So, Cybersecurity is like a digital guard that stops unauthorized access, spots potential threats, and deals with problems quickly. It's important for individuals, businesses, and governments to keep the data safely. People use strong passwords, update software, and use firewalls to protect themselves. Cybersecurity experts also create smart tools and plans to prevent new threats by making the digital world safe to everyone. This paper presents a brief overview of various threat detection systems in Cyber world.

2. Keyword:

cyber security, threat, attack, machine learning, network

3. Introduction

Our Cyber Threat Detection System is designed to comprehensively address the dynamic landscape of computer security by identifying potential threats and vulnerabilities that could lead to adverse impacts on computer systems and applications. A primary focus of our project is the mitigation of cyber attacks, which includes countering any malicious attempts to gain unauthorized access and cause damage to computers, computing systems, or networks. This paper aims to create an adaptable cyber threat detection system utilizing Python, HTML, and CSS. Employing a modular architecture ensures flexibility and scalability, allowing seamless integration of new threat detection modules. The system's primary features include precise detection of DoS, and Intrusion attempts. The user-friendly interface incorporates a real-time dashboard for thorough monitoring, facilitating quick responses to potential threats.

The project's objective is to offer organizations a flexible and efficient cybersecurity solution against dynamic threats. This paper focuses on various Cyber threat detection systems and also analyses the advantages and disadvantages of various mechanisms. A crucial component of our defense strategy is network intrusion prevention, where our system proactively identifies and repels unauthorized penetrations into enterprise networks or individual machines within assigned domains, thereby securing the digital perimeter against evolving threats.

4. Literature Survey

A Performance Evaluation Perspective in [18] focuses on the escalating dependence on cyberspace in daily life, leading to a rise in cyber threats. Conventional methods struggle against evolving cybercriminal techniques. The study evaluates the effectiveness of three prominent machine learning techniques—deep belief networks, decision trees, and support vector machines—in detecting cyber threats.

The authors in [7] focuses on creating a reliable Cyber Attack Detection Model (CADM) to protect users and assist network operators. The CADM analyzes network data patterns to classify cyber-attacks, using the ensemble classification method. LASSO is employed for feature extraction, capable of handling large datasets with enhanced visualization. Gradient Boosting and Random Forest algorithms are utilized to classify network traffic data, creating an ensemble method.

The authors in [8] discusses the importance of protecting network systems from potential

damage, which can manifest as various threats such as viruses, direct attacks, and phishing attempts by hackers seeking information. This approach operates in a controlled sandbox environment to prevent attackers from compromising the system. The focus is on utilizing machine learning to enhance cybersecurity.

The authors in [10] discuss the increasing significance of cybersecurity in our digitally-oriented lives. In this paper, a cyber threat detection system leveraging Machine Learning is introduced for the analysis of webpage content, enabling the differentiation between legitimate and malicious pages. Experimental findings underscore the model's efficacy in accurately identifying malicious web pages, thereby fortifying overall cybersecurity.

The authors in [12] addresses the challenge by introducing an AI-based system using artificial neural networks. The proposed technique transforms collected security events into individual profiles and employs deep learning methods (FCNN, CNN, LSTM) for enhanced detection. The AI-SIEM system focuses on distinguishing true and false positive alerts, aiding rapid response to cyber threats. Experiments with benchmark datasets (NSLKDD, CICIDS2017) and real-world data show superior performance compared to traditional machine learning methods (SVM, k-NN, RF, NB, DT), establishing the effectiveness of the proposed approach for network intrusion detection.

The authors in [4] discuss the emergence of smart cities powered by AI and IoT, as the world increasingly embraces AI. In response, the authors propose an Intrusion Detection System (IDS) utilizing Machine Learning (ML) to monitor network traffic. The study focuses on comparing various ML algorithms. Notably, ADA Boost demonstrated outstanding accuracy at 98.3%, underscoring its effectiveness in fortifying intrusion detection within the dynamic context of smart city environments.

The authors in [8] discusses the evolution of educational testbeds, particularly in the context of cybersecurity training. The paper proposes a design life cycle for developing cybersecurity testbeds, aiming to make them more accessible to. The results validate the proposed design life cycle, emphasizing the importance of aligning the testbed's technology with the intended challenges.

The authors in [19] discuss the application of Hidden Markov Models (HMM) in predicting cybersecurity attacks through the analysis of large network datasets. Leveraging the success of HMM in diverse domains, the study employs this model for its predictive and probabilistic properties. The methodology involves the utilization of Fuzzy K Mean clustering, subsequent manual labeling, and HMM state-based analysis. The obtained results exhibit effectively discerning cybersecurity attacks, uncertainties, and the absence of attacks. This approach proves more encouraging compared to traditional anomaly detection methods.

The authors in [6] talks about making affordable educational testbeds for cybersecurity training. However, existing testbeds often need maintenance and resources, limiting their use beyond experts. The paper suggests a step-by-step method to create educational cybersecurity testbeds. A case study demonstrates building one with open-source tech. The goal is to make cybersecurity training accessible to a wider audience. Future plans include automating and expanding the testbed.

The authors in [5] discuss the intricate nature of the cyber domain and its far-reaching impact on interconnected systems and diverse domains, exemplified by its potential to disrupt healthcare through a cyber attack on an electricity system. The paper underscores knowledge as a critical resource. It highlights the role of cyber security training and exercises, drawing on the military adage "You Fight Like You Train." Introducing the term "cyber arena" as the next-generation cyber range, the authors highlight the imperative for increasingly complex infrastructure in the continually evolving digital landscape.

The authors in [11] discuss in their paper that numerous implementations of IT-security management systems frequently fail to meet expectations, resulting at times in increased integrated costs attributed to insufficient cyber risk management practices. Their focus is on a specific case that involves a consulting company responsible for implementing a cyber-security system. The case highlights challenges encountered in the selection of IT-security metrics and the utilization of management tools, particularly within the framework of consulting services for holding companies.

The authors in [12] discusses the importance of intrusion detection systems (IDS) in network security and the use of machine learning techniques to enhance their accuracy. Emphasis is placed on improving performance by reducing false alarms and increasing detection rates. The proposed approach employs classification methodologies, specifically Support

Vector Machine (SVM) and Naïve Bayes, to analyze extensive network traffic data using the NSL-KDD dataset. The evaluation indicates that SVM outperforms Naïve Bayes in terms of accuracy and misclassification rates. The study aims to provide a comparative analysis of these effective classification methods.

The authors in [13] discuss the increasing demand for intrusion detection systems (IDS) in response to the widespread utilization of computer networks. The emphasis is placed on the imperative to safeguard network availability, integrity, and confidentiality. The text delineates two primary methods of detection: signature-based, which involves matching against established attack patterns, and anomaly-based, centered around identifying deviations from normal behavior. The paper furnishes a comprehensive overview of research endeavours aimed at constructing efficient IDS, employing single, hybrid, and ensemble machine learning classifiers. These approaches are systematically evaluated across seven datasets, and the ensuing results are thoroughly deliberated upon and compared. The insights gleaned from the discussion provide valuable considerations for prospective endeavours in the field.

The authors in [14] discuss Intrusion Detection Systems (IDS) in this text, presenting them as tools designed to scrutinize networks or systems for potential malicious activity. Two primary approaches are explored: Knowledge-Based Intrusion Detection (KBID), which involves matching against established signatures, and Anomaly-Based Intrusion Detection (ABID), employing Machine Learning to construct behaviour models. The paper's objective is to delve into the strengths and weaknesses of both methodologies, draw comparisons between existing IDS and Intrusion Prevention Systems, and introduce a novel concept UTPE.

The authors in [15] discuss the escalating issue of malware, specifically targeting Android devices, with a particular emphasis on the identification of zero-day malware through the application of machine learning classifiers. In this study, the proposed system scrutinizes Android applications (APKs) and categorizes them as either legitimate or malicious. Leveraging a dataset featuring 27 distinct features (CICMalDroid2020), the methodology, notably employing the Random Forest classifier, attains an impressive accuracy rate of 98.6%, surpassing the performance of alternative machine learning classifiers. The research underscores the efficacy of utilizing machine learning for the detection and protection against Android malware.

The authors in [2] discuss a comprehensive survey in their paper, delving into deep learning methodologies applied to the realm of cyber security intrusion detection. The focal point of their investigation revolves around seven distinct categories of datasets. Within the survey, they meticulously examine intrusion detection systems that leverage deep learning techniques, scrutinizing seven specific models. The evaluation of these models extends to their performance metrics in both binary and multiclass classification scenarios, employing authentic traffic datasets. The study relies on crucial indicators like accuracy, false alarm rate, and detection rate to gauge the efficacy of diverse methods.

The authors in [2] addresses the escalating cyber threats by proposing a hybrid learning approach, combining swarm intelligence and evolutionary methods, specifically PSO-GA (PSO-based GA), for feature selection in an intrusion detection framework. The study focuses on IoT-based systems, emphasizing the need for robust security measures. The proposed model, evaluated using ELM-BA with bootstrap resampling, demonstrated a 100% accuracy in detecting PortScan, SQL injection, and brute force attacks. The results suggest the effectiveness of the hybrid model in cybersecurity applications.

The authors in [17] discuss the escalating trend of increasing internet users, coupled with a rising demand for web services, mobile web applications, and desktop web applications. Alongside this growth, the potential for system hacks is on the rise. Web applications, accessible globally, store data in backend databases for result retrieval. As a consequence, the threat of SQL injection attacks has become a prominent security concern for web applications, posing risks of confidential information theft. This paper delves into a method for detecting SQL injection attacks, focusing on the removal of parameter values from the SQL query, and presents the obtained results.

The authors in [1] discuss injection attacks, which involve the insertion of malicious code into a network to retrieve all data for the assailant. This type of attack poses a significant challenge to internet security. Detecting a SQL injection attack remains difficult, as administrators may not be aware of the ongoing attack until there is a modification in the database content. The algorithm assigns the highest weight

to the value of each weak tree, creating a strong model by updating weights at each step during dataset training. The process involves adding the input of each layer by calculating the average of previous outputs. The results indicate that the proposed algorithm and program can accurately detect injection attacks more effectively than the initial neural techniques, which degrade with the increasing number of intermediate layers in the program.

The authors in [13] discuss the pervasive usage of the internet by over 4 billion people, emphasizing the consequential socio-technical threat posed to both the Government and the public due to the widespread adoption of mobile technology and the onset of the digital era. This surge in digital advancements has given rise to illicit opportunities, particularly in the realm of cyber crime. Cyber crime, defined as the unlawful use of digital media either as a tool, a target, or both, has witnessed a concerning escalation, especially in the context of the ongoing COVID-19 situation. Notably, phishing attacks and various other forms of cyber crime have proliferated. This paper delves into the different types of cyber crime, focusing on contemporary attacks related to Phishing, Artificial Intelligence, Cloud technology, and Blockchain. The primary objective of the work centers on the exploration of how Machine Learning can be effectively employed in detecting diverse cyber crime fields. Additionally, the authors delve into the application of various Machine Learning models in predicting, identifying, and mitigating complex cyber threats.

5. Methodology

5.1. Algorithms

1. Autoencoder Algorithm

- Purpose: Detects anomalies by learning efficient data representations and reconstructing data to identify outliers.

- Functionality: Trains an Autoencoder neural network to compress and reconstruct data, learning patterns to detect deviations.

- Output: Reconstruction error and encoded features for further processing.

2. Decision Tree with PCA

- Purpose: Uses PCA for dimensionality reduction and a Decision Tree for classification, improving efficiency and performance.

- Functionality: PCA reduces feature dimensions; a Decision Tree classifier uses these to classify data based on signatures.

- Output: Classification accuracy and reduced feature set.

3. Deep Neural Network (DNN) Algorithm

- Purpose: Enhances classification accuracy by learning complex data patterns, building on Decision Tree predictions.

- Functionality: Trains a DNN on labeled data from the Decision Tree, using multiple neuron layers for hierarchical data representation.

- Output: Classification accuracy, precision, and recall metrics.

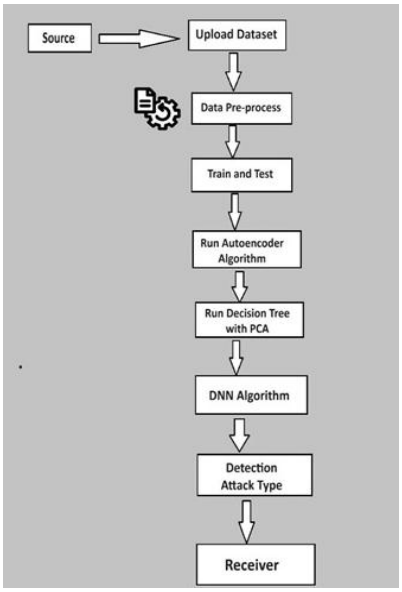
4. Detection & Attribute Attack Type

- Purpose: Identifies and attributes attack types for unlabeled data using the trained DNN model.

- Functionality: Uses the trained DNN to predict attack types from a new test dataset for real-time detection and classification.

- Output: Predicted attack types for timely threat identification and mitigation. modules

5.2. Flowchart



5.3. Modules

TensorFlow: Open-source library for dataflow and machine learning, developed by Google.

Numpy: Essential package for scientific computing with high-performance array processing.

Pandas: Open-source library for data manipulation and analysis.

Matplotlib: 2D plotting library for creating publication-quality figures.

Scikit-learn: Provides various machine learning algorithms with a consistent interface.

Python: High-level, interpreted programming language known for readability and versatility.

6. Results

Algorithm Name	Accuracy	Precision	Recall	FSCORE
AutoEncoder	90.09191764705882	73.16825122443 1	74.12077505827506	73.5919607643021 4
Decision Tree with PCA	90.45935882352941	86.04246236800727	74.23521503576344	73.8925021820118 2
DNN	99.96323529411765	74.56615968110227	75.0	74.7792986788443 7

7. Conclusion

This article proposes using AI to detect application layer attacks. A graph-based division method and dynamic programming create examples (PCRE standard expressions) for the model. These expressions help illustrate app behavior and identify digital attacks. Results show that this method effectively detects such attacks.

Future Enhancements:

- The system can be upgraded to adapt to new technologies as they emerge.
- Security can be enhanced to address future issues, such as implementing single sign-on.

8. Acknowledgement

We would like to express our sincere gratitude to our guide Mrs. T C Swetha Priya, Asst Professor, IT Department and our HOD, Dr. Srinivasu Badugu for their valuable suggestions and support during the research and in writing this paper.

9. References

[1]. A. Sivasangari, J. Jyotsna and K. Pravalika, "SQL Injection Attack Detection using Machine Learning Algorithm," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2021, pp. 1166-1169, doi: 10.1109/ICOEI51242.2021.9452914.

[2]. Alatawi, Mohammed Naif, Najah M. Alsubaie, Habib Ullah Khan, Tariq Sadad, Hathal Salamah Alwageed, Sajid Ali and Islam Zada. “Cyber Security against Intrusion Detection Using Ensemble-Based Approaches.” Security and Communication Networks (2023): n. page.

[3]. Ayesha S. Dina, D. Manivannan, Intrusion detection based on Machin e Learning techniques in computer networks, Internet of Things, Volume16, 2021, 100462, ISSN 25426605,https://doi.org/101016/j.iot.2021.10046

[4]. Chohan, Maria & Haider, Usman & Ayub, Muhammad Yaseen & Shoukat, Hina & Bhatia, Tarandeep & Hassan, Muhammad. (2023). Detection of Cyber Attacks using Machine Learning based Intrusion Detection System for IoT Based Smart Cities. EAI Endorsed Transactions on Smart Cities. 7. 10.4108/eetsc.3222.

[5]. Droos, A. Al-Mahadeen, T. Al-Harasis, R. Al-Attar and M. Ababneh, "Android Malware Detection Using Machine Learning," 2022 13th International Conference on Information and Communication Systems (ICICS), Irbid,

Jordan, 2022, pp. 36-41, doi: 10.1109/ICICS55353.2022.9811130.

[6]. Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 149-155, doi: 10.1109/ICOSEC49089.2020.9215333.J. Shah, "Understanding and study of intrusion detection systems for various networks and domains," 2017 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2017, pp. 1-6, doi: 10.1109/ICCCI.2017.8117726.

[7]. F. Hossain, M. Akter and M. N. Uddin, "Cyber Attack Detection Model (CADM) Based on Machine Learning Approach," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), DHAKA, Bangladesh, 2021, pp. 567-572, doi: 10.1109/ICREST51555.2021.9331094.

[8]. K. R. Dalal and M. Rele, "Cyber Security: Threat Detection Model based on Machine learning Algorithm," 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2018, pp. 239-243, doi: 10.1109/CESYS.2018.8724096.

[9]. K. R. Dalal and M. Rele, "Cyber Security: Threat Detection Model based on Machine learning Algorithm," 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2018, pp. 239-243, doi: 10.1109/CESYS.2018.8724096.

[10]. Koçyiğit, Emre & Korkmaz, Mehmet & Sahingoz, Ozgur & Diri, Banu. (2021). Real-Time Content-Based Cyber Threat Detection with Machine Learning. 10.1007/978-3-030-71187-0_129.

[11]. Lee, Jonghoon & Kim, Jonghyun & Kim, Ikkyun & Han, Kijun. (2019). Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2953095.

[12]. Livshitz, P. A. Lontsikh, N. P. Lontsikh, E. Y. Golovina and O. M. Safonova, "The Effects of Cyber-security Risks on Added Value of Consulting Services for IT-security Management Systems in Holding Companies," 2020 International Conference Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS), Yaroslavl, Russia, 2020, pp. 119-122, doi: 10.1109/ITQMIS51053.2020.9322883.

[13]. M. Arshey and K. S. Angel Viji, "Thwarting Cyber Crime and Phishing Attacks with Machine Learning: A Study," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 353-357, doi: 10.1109/ICACCS51430.2021.9441925.

[14]. M. Frank, M. Leitner and T. Pahi, "Design Considerations for Cyber Security Testbeds: A Case Study on a Cyber Security Testbed for Education," 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSci Tech), Orlando, FL, USA, 2017, pp. 38-46, doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.23.

[15]. M. Karjalainen and T. Kokkonen, "Comprehensive Cyber Arena; The Next Generation Cyber Range," 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Genoa, Italy, 2020, pp. 11-16, doi: 10.1109/EuroSPW51379.2020.00011.

[16]. Mohamed Amine Ferrag, Leandros Maglaras, Sotiris Moschoyiannis, Helge Janicke,Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study,Journal of Information Security and Applications,Volume 50,2020,102419,ISSN 2214-2126,https://doi.org/10.1016/j.jisa.2019.102419.

[17]. R. A. Katole, S. S. Sherekar and V. M. Thakare, "Detection of SQL injection attacks by removing the parameter values of SQL query," 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2018, pp. 736-741, doi: 10.1109/ICISC.2018.8398896.

[18]. Shaukat Dar, Kamran & Luo, Suhuai & Chen, Shan & Liu, Dongxi. (2020). Cyber Threat Detection Using Machine Learning Techniques: A Performance Evaluation Perspective. 10.1109/ICCWS48432.2020.9292388.

[19]. T. T. Teoh, Y. Y. Nguwi, Y. Elovici, N. M. Cheung and W. L. Ng, "Analyst intuition based Hidden Markov Model on high speed, temporal cyber security big data," 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin,

China, 2017, pp. 2080-2083, doi: 10.1109/FSKD.2017.8393092.

[20]. U. S. Musa, M. Chhabra, A. Ali and M. Kaur, "Intrusion Detection System using Machine Learning Techniques: A Review," 2020 International Conference on Smart