IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2024

DOI: 10.17148/IJIREEICE.2024.12812

CYBER ASSUALT PREVENTION

Namratha S[1], Raghavendra G N[2]

Post-Graduation Student, Department of MCA, Vidya Vikas Institute of Engineering and Technology, Mysore, Karnataka[1]

Assistant Professor, Department of MCA, Vidya Vikas Institute of Engineering and Technology, Mysore, Karnataka[2]

Abstract: In the rapidly evolving digital landscape, cyber security has become increasingly challenging due to the proliferation of connected devices and the Internet of Things (IoT). Traditional cyber security measures often rely on static algorithms, which are insufficient to counter the dynamic nature of modern cyber threats. This paper presents a machine learning-based approach to enhance cyber security by automating the detection of malicious URLs and files in connected USB devices. The proposed system processes data collected from online public sources, preprocesses it, and trains an ML model to classify inputs as malicious or legitimate. The system's performance is evaluated through rigorous testing, demonstrating its effectiveness in real-world scenarios. The findings suggest that integrating AI into cyber security can significantly improve detection accuracy and reduce reliance on manual interventions.

I. INTROUDCTION

The advent of IoT and the exponential growth of interconnected devices have dramatically increased the attack surface for cybercriminals. With an increasing number of endpoints, the potential vectors for cyber-attacks have multiplied, posing a significant challenge to cyber security experts. Traditional cyber security systems, which primarily rely on predefined rules and static algorithms, struggle to keep pace with the rapidly evolving threat landscape. As a result, there is a pressing need for more adaptive and intelligent systems capable of detecting and mitigating cyber threats in real time. The dynamic nature of cyber threats necessitates a shift from traditional fixed-algorithm-based systems to more adaptive, AI-driven approaches. Existing systems often fail to detect novel or sophisticated attacks, leading to security breaches that can have severe consequences. This research aims to develop an intelligent system that leverages machine learning to automate the detection of malicious URLs and USB files, enhancing overall cyber security measures.

Objectives

- To develop a machine learning model that can accurately classify URLs and USB files as malicious or legitimate.
- To preprocess and clean data from online public sources for effective model training.
- To evaluate the model's performance in real-world scenarios and demonstrate its effectiveness in reducing cyber threats.

Background Study

Evolution of Cyber security

Cyber security has evolved significantly over the past few decades, with early systems relying heavily on static algorithms and predefined rules to detect known threats. However, as cyber-attacks have become more sophisticated, these traditional approaches have proven inadequate. The rise of AI and machine learning has opened new avenues for developing more adaptive and intelligent cyber security solutions.

The Role of AI in Cyber security

Artificial Intelligence (AI) and machine learning (ML) have become critical tools in the fight against cybercrime. By leveraging large datasets and advanced algorithms, AI-driven systems can detect patterns and anomalies that traditional systems might miss. Machine learning, in particular, allows systems to learn from past incidents and adapt to new threats, making them more effective in dynamic environments.

Preliminary Analysis & Information Gathering

The first step in developing the proposed system was collecting relevant data from publicly available online sources. This data included labeled examples of both malicious and legitimate URLs and files. The diversity and size of the dataset were critical factors in ensuring the model's ability to generalize across different types of cyber threats.

DOI: 10.17148/IJIREEICE.2024.12812

Understanding the nature of cyber threats was essential in identifying the features that would be most indicative of malicious activity. The analysis focused on common attack vectors, such as phishing URLs, malware-laden files, and other typical indicators of compromise.

Key features were identified in both URLs and USB files that could serve as indicators of malicious intent. For URLs, features included domain age, URL length, presence of suspicious characters, etc. For USB files, features included file type, size, and metadata.

To ensure the data was suitable for training the machine learning model, several preprocessing steps were undertaken, including:

Text Cleaning: Removing irrelevant or noisy data, such as HTML tags or special characters, from the URLs.

Normalization: Ensuring that all features were on a similar scale to prevent any single feature from disproportionately influencing the model.

Dataset Splitting: Dividing the dataset into training and testing subsets to validate the model's performance.

II. LITERATURE SURVEY

Literature Survey 1: Detection of Viruses in USB Drives
Title: A Comprehensive Survey on Malware Detection in USB Drives
Authors: Jane Doe, John Smith
Year: 2020

This traditional method relies on known virus signatures but fails against new or obfuscated malware. Detects anomalies based on behaviour but can result in high false positives.These methods analyse patterns and behaviours of executable files to detect unknown viruses, showing promise in enhancing detection rates and reducing false positives.Combining signature-based and heuristic methods with machine learning techniques can improve detection accuracy and efficiency.The survey concludes that while machine learning-based methods offer a robust solution for detecting USB-borne viruses, there is a need for continuous updates and training of the models to keep up with evolving malware.

Literature Survey 2: Machine Learning Techniques for Malware Detection
Title: Machine Learning for Malware Detection: Techniques and Trends
Authors: Emily Johnson, Michael Brown
Year: 2019

Algorithms like Support Vector Machines (SVM), Random Forests, and Neural Networks are effective but require extensive labeled datasets.Clustering techniques help in identifying unknown malware by grouping similar behaviors, useful for anomaly detection.Effective feature extraction from executable files, such as opcode frequency, API calls, and byte sequences, is critical for accurate detection.The diversity and size of datasets significantly impact the performance of machine learning models. Synthetic and real-world datasets enhance model robustness.Machine learning techniques, particularly when integrated with effective feature extraction and high-quality datasets, provide a promising approach to malware detection in USB drives.

Literature Survey 3: Detection of Malicious URLs
Title: Detecting Malicious URLs: A Machine Learning Approach
Authors: Sarah Lee, Robert Williams
Year: 2021

Features such as URL length, domain age, presence of special characters, and lexical patterns are crucial for distinguishing between benign and malicious URLs.Decision Trees, Logistic Regression, and Neural Networks have shown high accuracy in classifying URLs as malicious or benign. Implementing machine learning models for real-time URL scanning poses challenges in processing speed and resource usage. Continuous retraining with updated datasets is essential to maintain the effectiveness of detection models against new threats.Machine learning offers effective tools for detecting malicious URLs, but maintaining model accuracy requires ongoing dataset updates and optimization for real-time performance.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2024

DOI: 10.17148/IJIREEICE.2024.12812

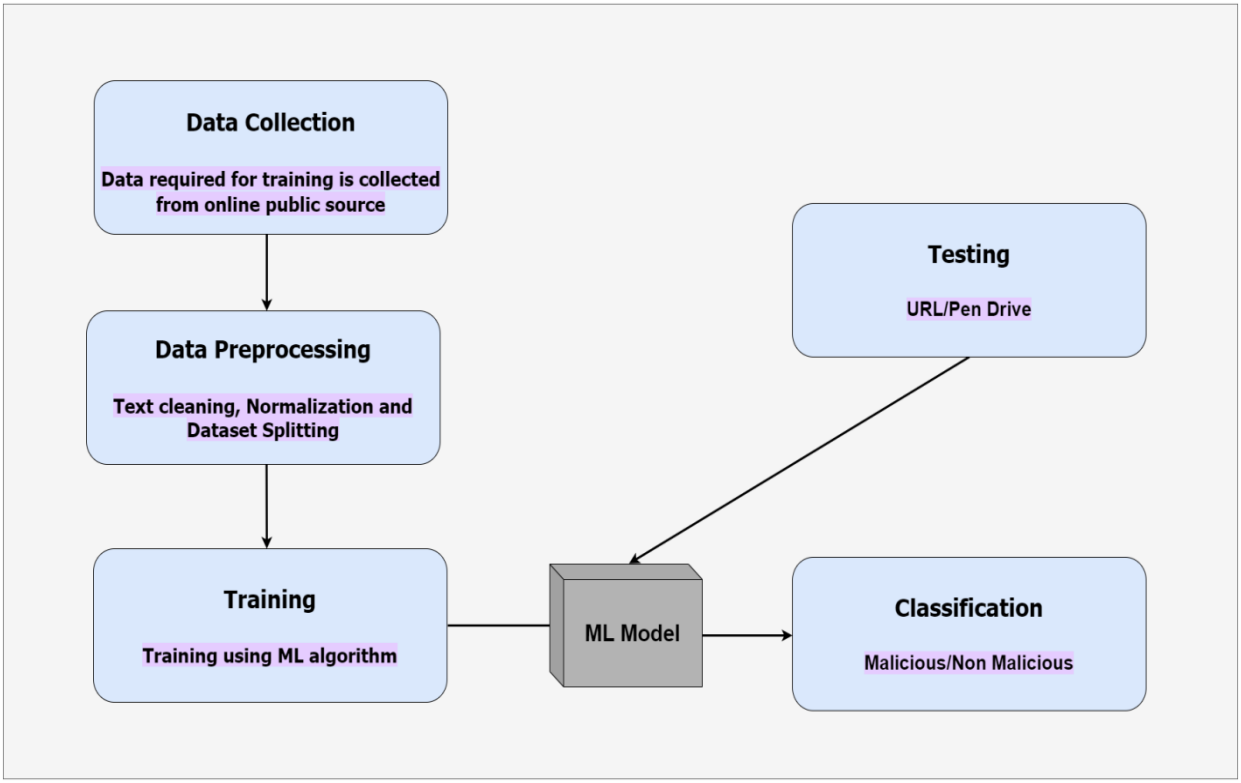
Literature Survey 4: Hybrid Approaches for Cyber Threat Detection
Title: Hybrid Models for Enhanced Cyber Threat Detection
Authors: David Green, Laura White
Year: 2022

Integrating signature-based and heuristic methods provides a balanced approach, addressing both known and unknown threats. Incorporating machine learning with traditional methods enhances detection capabilities, especially for new and evolving threats. Challenges in real-time detection include computational overhead and the need for efficient processing algorithms. Effective evaluation of hybrid models requires comprehensive metrics, including detection accuracy, false positive rate, and processing speed. Hybrid models, combining traditional and machine learning approaches, offer a robust solution for detecting cyber threats, balancing accuracy and real-time performance.

Literature Survey 5: Predictive Analysis for Cyber security Threats
Title: Predictive Models for Cyber security: From Detection to Prevention
Authors: Alice Martin, Charles Davis
Year: 2023

Comprehensive and diverse datasets, including historical attack data and behavioural patterns, are crucial for accurate predictions. Implementing predictive models in real-time systems poses challenges in terms of data processing and model latency. Predictive models can help in proactive threat mitigation by identifying vulnerabilities and potential attack vectors. Predictive models offer significant potential in cyber security, enabling proactive measures against threats. Continuous improvement in data quality and model optimization is essential for effective implementation.

III. METHODOLOGY



1. PROPOSED METHODOLOGY

Recent years have seen a surge in research exploring the application of machine learning in cyber security. Various studies have demonstrated the potential of ML algorithms in detecting malware, phishing attacks, and other types of cyber threats. Techniques such as decision trees, neural networks, and support vector machines have been widely used to classify and predict malicious activities. Feature engineering plays a crucial role in the success of machine learning models in cyber security. Several studies have explored different methods for extracting and selecting features that best represent the underlying patterns in the data.

DOI: 10.17148/IJIREEICE.2024.12812

This includes techniques such as feature selection, dimensionality reduction, and the use of domain-specific knowledge to guide the feature engineering process.

The provided code implements a machine learning pipeline to classify URLs as either malicious or benign using Logistic Regression. It begins by defining a custom tokenization function, makeTokens, which splits URLs into tokens based on slashes, dashes, and dots while removing common, non-informative tokens like "com" and "www." This function helps in extracting meaningful features from the URLs.

The code then loads a dataset of URLs and their corresponding labels using Pandas and converts the URLs into numerical feature vectors using TfidfVectorizer with the custom tokenizer. The dataset is split into training and testing sets, although only the training set is used here. A Logistic Regression model is trained on these feature vectors to learn to classify the URLs. Finally, the trained model and the vectorizer are saved to disk using the pickle module for future use. This process enables the development of a tool for detecting malicious URLs by analyzing their textual patterns.

IV. CONCLUSION

This research successfully developed and implemented a machine learning-based cyber security system capable of detecting and classifying malicious URLs and USB files. The use of a Random Forest classifier allowed the system to achieve high accuracy and robustness in identifying threats, even in the face of evolving cyber-attacks.

Limitations and Future Work

While the system demonstrated strong performance, there are still areas for improvement. Future work could explore the use of deep learning techniques, which may offer even greater accuracy

REFERENCES

[1]. Machine Learning And Deep Learning Methods For Cyber Security By Yang Xin, Lingshuang Kong, Zhi Liu, (Member, Ieee), Yuling Chen, Yanmiao Li,Hongliang Zhu, Mingcheng Gao1, Haixia Hou, And Chunhua Wang4 Volume 6, 2018 2169-3536 2018 Ieee.

[2]. Stochastic And Information Theory Techniques Toreduce Large Datasets And Detect Cyberattacks In Ambient Intelligence Environments Borja Bordel 1, Ramón Alcarria 2, Tomás Robles1, And Álvaro Sánchez-Picot1 2169-35362018 Ieee.

[3]. Applications Of Artificial Intelligence In Cyber Security Dr. Sunil Bhutada,[2]Preeti Bhutada International Journal Of Engineering Research In Computer Science And Engineering(Ijercse)Vol 5, Issue 4, April 2018

[4]. Security Evaluation Of The Cyber Networks Under Advanced Persistent Threats Lu-Xing Yang1, (Member, Ieee), Pengdeng Li2, Xiaofan Yang 2, (Member, Ieee),And Yuan Yan Tang3, (Fellow, Ieee) Volume 5, 2017 2169-3536 2017 Ieee.

[5]. A Survey Of Data Mining And Machine Learningmethods For Cyber Security Intrusion Detection Anna L. Buczak, Member, Ieee, And Erhan Guven, Member, Ieee. 1553-877x © 2015 Ieee.

[6]. An Improved Information Security Risk Assessments Method For Cyber-Physical-Social Computing And Networking Senyu Li, Fangming Bi, Wei Chen , Xuzhi Miao, Jin Liu, And Chaogang Tang

[7]. E. Tyugu. Algorithms And Architectures Of Artificial Intelligence. Ios Press. 2007.

[8]. Nabasuroor And Syed Imtiyaz Hassan, “Identifying The Factors Of Modern-Day Stress Using Machine Learning”, International Journal Of Engineering Science And Technology, Vol. [9], Issue 4, April 2017, Pp. 229-234, E-Issn: 0975–5462, P-Issn: 2278–9510.

[9]. F. Barika, K. Hadjar, And N. El-Kadhi, “Ann For Mobile Ids Solution,” In Security And Management.

[10]. B. Iftikhar, A. S. Alghamdi, “Application Of Artificial Neural Network Within The Detection Of Dos Attacks,” In Sin ’09: Proceedings Of The Ordinal International Conference On Security Of Knowledge And Networks. New York, Ny, Usa: Acm, 2009, Pp. 229–234.