CYBER ATTACKS DETECTION USING MACHINE LEARNING

ABSTRACT

We like to have simple and automated solutions, but these simple and automated solutions in technology could also contains risks if not deal properly. Due to no international standard of compatibility for IoT, security and privacy concerns are there which needs to be focus. There can be multiple types of attack on IoT networks which can damage the device or steal the sensitive information. Therefore, artificial intelligence (AI) techniques has an ability to detect and classify an unknown network behaviour by learning the network attacks patterns based on large volumes of historical data. We considered Aposemat IoT-23 which is a labelled dataset and created in the Avast laboratory. Basically, the goal of this large dataset is to provide labelled and real IoT attacks. In this paper, we used this dataset, considered the relevant workings, investigate the background and implement the machine learning algorithms such as Decision Tree, Random Forest and Naive Bayes. We also compared the accuracy among these machine learning algorithms on the IoT-23 dataset and showed the most efficient machine learning algorithm is Random Forest as per results by using Aposemat IoT-23 dataset, as well as showed feature engineering techniques to preprocess the mentioned dataset for detection and classification of IoT network attacks.

INTRODUCTION

Today, we are living in the society where many things are going to be automated and digitalize. Technology is now involving in our daily life and there are many simple examples for that such as mobile phones, personal computers etc. Converting things to smart devices and making these processes automated, IoT is one of the technology which plays an important role for that purpose. So we can say that it is one of the most important technologies for businesses as well as for our daily life. But, it is important to remember that as the technology increases there are also a number of issues increases related with that technology. Similarly, as the number of devices connected it means the more information is sharing between these devices and if there is any type of bug in the sharing system, there is a chance that each connected device could corrupt and confidential information could steal by the hacker.

There should be an international standard for compatibility of IoT here which is not yet, therefore it is very difficult for devices which are manufactured from different companies to communicate with each other. Also there are many IoT devices which requires and ask to input user personal information such as name, location and contact as well as data which are important to hackers such as social media information. Therefore, the information sharing between IoT devices needs to be secured. Also IoT privacy and security are cited as major concerns. There are number of attacks on IoT including malware. Malware can be defined as a malicious software or bug which is designed to gain access and damage your device, device could be computer or IoT device.

IoT devices are vulnerable to network attacks therefore, malware and network attack detection in IoT is the focus of research in recent years. There are many workings are there to address the issue and detect network attacks. In comparison, ML and DL which can be defined as machine learning and deep learning in artificial intelligence has the power to detect unknown network behavior by automatically learning the networks attacks and malware patterns based of large datasets. In this paper we will focus on the security aspect of networks of IoT by understanding the use of machine learning based algorithms in artificial intelligence for the detection of network attacks and malwares. For this purpose, we will consider Aposemat IoT-23 which is a labeled dataset and created in the Avast laboratory. This dataset also provides benign IoT traffic which is helpful to develop or implement machine learning based algorithms in artificial intelligence.

SYSTEM ANALYSIS

EXISTING SYSTEM

The existing system focuses on addressing the security and privacy concerns in IoT networks, recognizing the absence of international standards for compatibility in the IoT landscape. The project utilizes the Aposemat IoT-23 dataset, a labeled dataset created in the Avast laboratory, designed specifically to provide real-world examples of IoT attacks. The primary objective is to leverage artificial intelligence techniques to detect and classify unknown network behaviors based on historical data patterns. The machine learning algorithms employed include Decision Tree, Random Forest, and Naive Bayes. Through a comparative analysis, the results indicate that Random Forest proves to be the most efficient algorithm for detecting and classifying IoT network attacks on the Aposemat IoT-23 dataset.

LIMITATIONS OF EXISTING SYSTEM

Dependency on Labeled Datasets: The existing system relies on the Aposemat IoT-23 dataset for training and testing machine learning algorithms. However, this dependency on a single dataset may limit the system's adaptability to new and emerging types of IoT attacks not covered in the provided dataset.

Static Machine Learning Models: The use of Decision Tree, Random Forest, and Naive Bayes machine learning algorithms implies a static approach to network attack detection. These models might struggle to adapt to dynamic and evolving attack strategies, potentially leading to a decreased accuracy in detecting novel threats.

Limited Generalization: The effectiveness of the system may be constrained by its ability to generalize across diverse IoT network environments. Factors such as network scale, device types, and communication protocols may vary, affecting the system's performance in real-world scenarios that differ from the Aposemat IoT-23 dataset.

Scalability Challenges: The system's scalability might be a limitation when dealing with large-scale IoT networks. As the number of devices and the complexity of network architectures increase, the computational demands of the chosen machine learning algorithms may become a bottleneck, affecting real-time detection capabilities.

Overhead Due to Feature Engineering: While feature engineering is employed to enhance the dataset's preprocessing, it introduces additional computational overhead. This may impact the system's efficiency, particularly in resource-constrained IoT devices, where computational resources are limited, and real-time processing is crucial.

PROPOSED SYSTEM

The proposed system aims to overcome the limitations of the existing approach by introducing several enhancements to strengthen IoT network attack detection using artificial intelligence. Firstly, the system proposes the incorporation of a more diverse set of labeled datasets, beyond the Aposemat IoT-23 dataset, to ensure a comprehensive understanding of evolving attack patterns. This expansion enables the system to generalize better and recognize novel threats that may not be covered by a single dataset.

Secondly, the proposed system advocates for the integration of dynamic and adaptive machine learning models that can evolve with the changing nature of IoT attacks. This may involve exploring deep learning techniques or other advanced algorithms capable of capturing intricate patterns and adapting to emerging threats in real-time.

Thirdly, the proposed system emphasizes the development of a more scalable architecture, considering the increasing scale and complexity of IoT networks. This could involve the implementation of distributed computing techniques or lightweight algorithms suitable for resource-constrained IoT devices, ensuring efficient and effective network attack detection at scale.

Additionally, the proposed system suggests leveraging anomaly detection methods alongside traditional classification approaches to enhance the detection of previously unseen attacks. Anomaly detection can identify deviations from normal network behavior, providing a proactive defense against emerging threats not explicitly defined in the training dataset.

Lastly, the proposed system aims to optimize feature engineering processes to minimize computational overhead. This involves refining preprocessing techniques to strike a balance between improving detection accuracy and ensuring efficient resource utilization, particularly in the context of IoT devices with limited computational capabilities. Overall, the proposed system seeks to advance the state-of-the-art in IoT network attack detection by addressing existing limitations and embracing more dynamic, scalable, and adaptable approaches.

ADVANTAGES OF PROPOSED SYSTEM

Improved Detection Accuracy: The proposed system leverages a more diverse set of labeled datasets and advanced machine learning models, contributing to enhanced detection accuracy. This ensures a comprehensive understanding of various IoT attack patterns and improves the system's ability to recognize and classify both known and emerging threats.

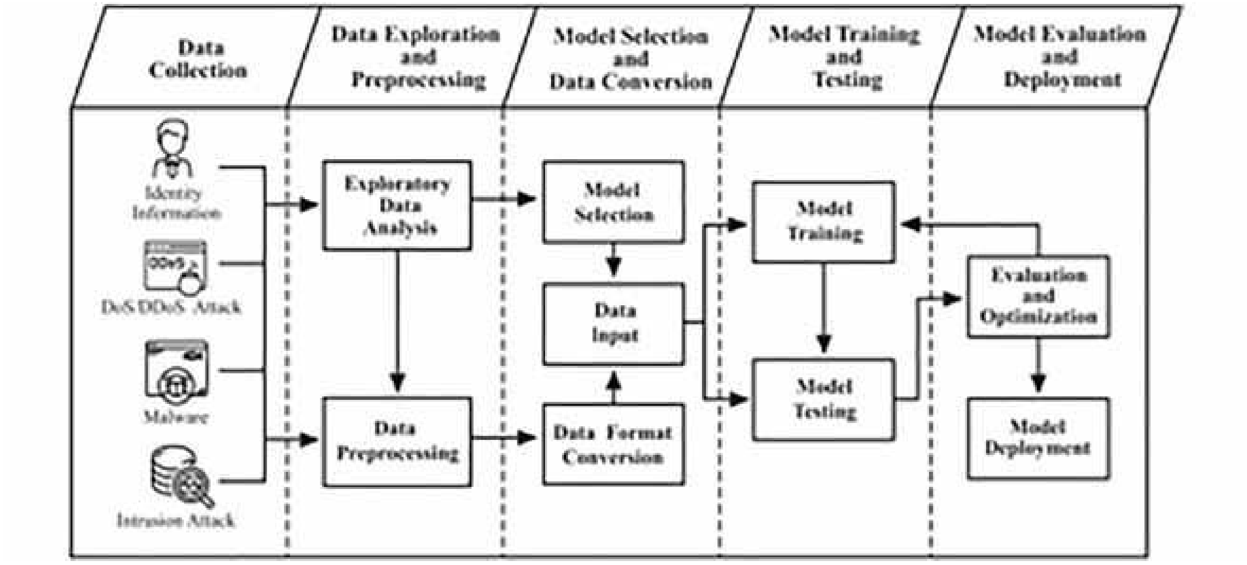
Adaptability to Dynamic Threats: By incorporating dynamic and adaptive machine learning models, the proposed system can better adapt to the evolving nature of IoT attacks. This adaptability allows the system to recognize and respond to new and previously unseen threats in real-time, providing a more proactive defense mechanism.

Scalability for Large IoT Networks: The proposed system addresses scalability challenges by introducing a more scalable architecture. This ensures efficient performance in large-scale IoT networks, accommodating the increasing number of devices and complexities in network structures without compromising the system's ability to detect and respond to security threats.

Comprehensive Defense with Anomaly Detection: The integration of anomaly detection methods alongside traditional classification approaches adds a layer of comprehensive defense. Anomaly detection enables the system to identify deviations from normal network behavior, offering a proactive defense against novel and unexpected attacks that may not be explicitly defined in the training dataset.

Optimized Resource Utilization: The proposed system optimizes feature engineering processes to minimize computational overhead. This optimization ensures efficient resource utilization, making the system more suitable for deployment in resource-constrained IoT devices. By balancing detection accuracy and computational efficiency, the proposed system maximizes its effectiveness without putting undue strain on device resources.

SYSTEM ARCHITECTURE



MODULES

Data Preprocessing Module: This module focuses on preparing and refining the input data for the machine learning algorithms. It involves tasks such as data cleaning, handling missing values, normalization, and transforming the raw data from the Aposemat IoT-23 dataset into a format suitable for effective learning by the algorithms.

Machine Learning Algorithm Implementation Module: This module encompasses the implementation of machine learning algorithms, including Decision Tree, Random Forest, and Naive Bayes. Each algorithm is configured and trained on the preprocessed data to learn and recognize patterns indicative of IoT network attacks.

Dataset Integration Module: The system integrates multiple labeled datasets beyond the Aposemat IoT-23 dataset to ensure a more comprehensive understanding of diverse attack patterns. This module facilitates the combination of various datasets, promoting a broader and more adaptable detection capability.

Dynamic Model Adaptation Module: To address the dynamic nature of IoT threats, this module introduces dynamic and adaptive machine learning models. These models continuously evolve and learn from ongoing network behaviors, allowing the system to adapt to emerging threats in real-time and enhance its ability to detect novel attack patterns.

Scalability and Optimization Module: This module focuses on enhancing the scalability and efficiency of the system. It includes strategies for accommodating large-scale IoT networks, optimizing resource utilization, and incorporating lightweight algorithms suitable for deployment on resource-constrained IoT devices. The goal is to ensure the system performs effectively across diverse network environments while efficiently managing computational resources.

HARDWARE REQUIREMENTS

MINIMUM (Required for Execution)		MY SYSTEM (Development)
System	Pentium IV 2.2 GHz	i3 Processor 5th Gen
Hard Disk	20 Gb	500 Gb
Ram	1 Gb	4 Gb

SOFTWARE REQUIREMENTS

Operating System	Windows 10/11
Development Software	Python 3.10
Programming Language	Python
Domain	Image Processing & Cloud Computing
Integrated Development Environment (IDE)	Visual Studio Code
Front End Technologies	HTML5, CSS3, Java Script
Back End Technologies or Framework	Django
Database Language	SQL
Database (RDBMS)	MySQL
Database Software	WAMP or XAMPP Server
Web Server or Deployment Server	Django Application Development Server

Design/Modelling	Rational Rose

REFERENCES

[1] Vibekananda Dutta , Michał Chora´s, Marek Pawlicki and Rafał Kozik, "A Deep Learning Ensemble for Network Anomaly and Cyber-Attack Detection", Sensors, August 2020.

[2] Quoc-Dung Ngo, Huy-Trung Nguyen, Van-Hoang Le, Doan-Hieu Nguyen, "A survey of IoT malware and detection methods based on static features", ICT Express, December 2020.