Name: Braydon Reid

Date: 3/28/2025

Network Anomaly Detection

1. Introduction

Today is the age of networking and data used to increase workplace efficiency and safety. We have been doing this using “Industrial IoT,” which stands for Industrial Internet of Things [1]. Ultimately, this is an ecosystem of different devices, sensors, and applications that use networking equipment to collect, monitor, and analyze data from the workplace.[1][3] This analysis increases efficiency, reduces costs, and improves safety. However, the increased use of networking to communicate across devices also leaves room for various attacks over the network. Failure of Industrial IoT can have significant consequences, such as injury or death. The motivation behind this project is to increase the security of Industrial IoT by being able to distinguish between normal network behavior and potential attacks on the network.

The proposed solution will be able to aggregate and transform real-time data into meaningful features that can be used to detect patterns. By using machine learning, the system can distinguish between normal network behavior and potential attacks over the network. This will also be for early detection and quicker response time to ensure that any anomaly in the network can be addressed to prevent any disruption or safety hazards in the Industrial IoT environment.[2]

2. Prior Work

In the past, many efforts have been made to solve attack/anomaly detection through networks, most of which relied heavily on rule-based systems and statistical analysis of network traffic. However, these solutions never perfectly captured the dynamic behavior of modern IIoT (Industrial Internet of Things). [3]

However, recently, a study has come out called “BRUIIoT,” [2] which takes a different approach. They transform raw network data into aggregated features using Mutual Information

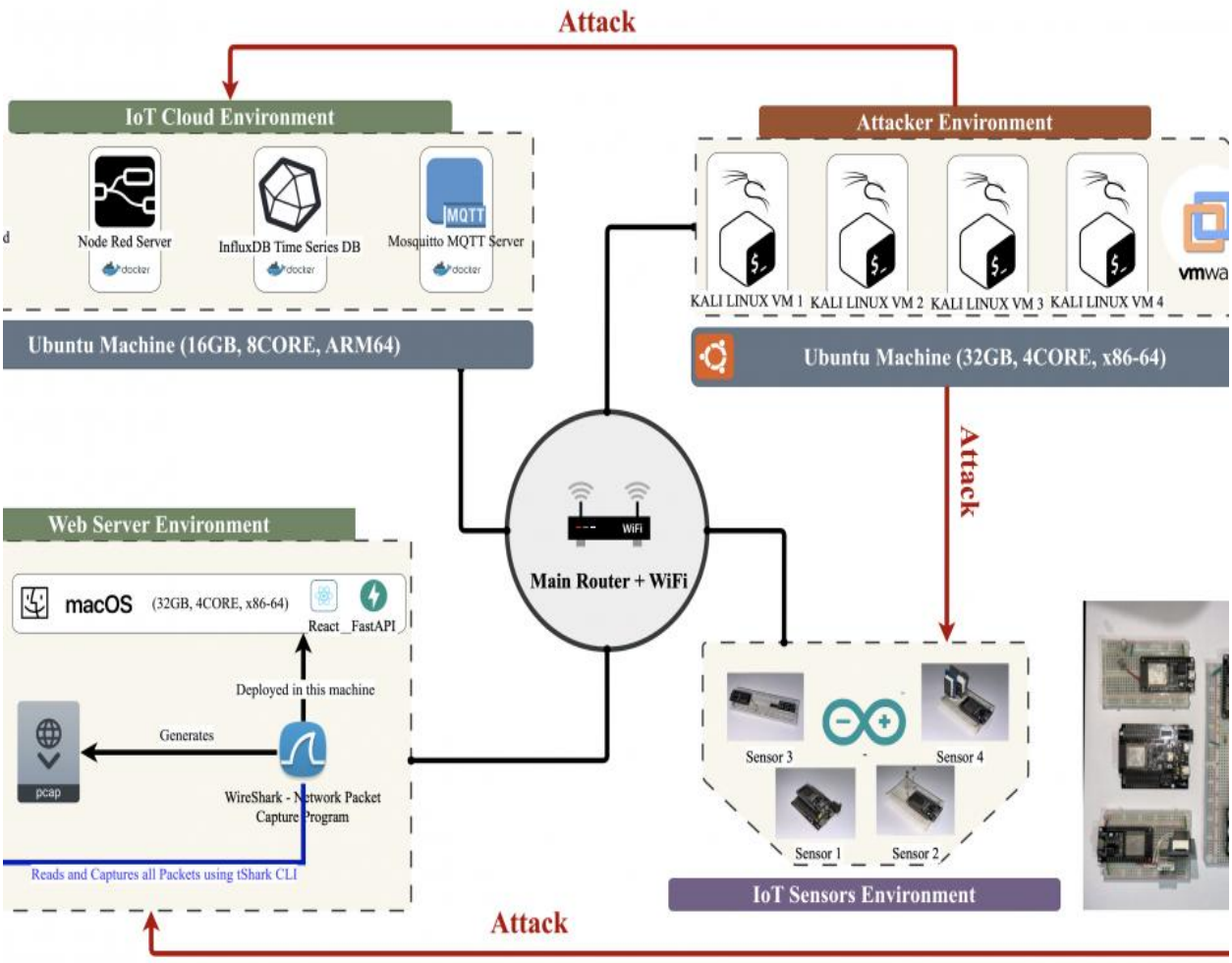
(MI)--based feature selection. The result is a refined set of 150 features, including IP counts, TCP acknowledgment patterns, and ICMP sequence ratios[2]. This allowed them to use machine learning to understand the patterns better and classify the attacks vs normal networking behavior [2][4].

3. Implementation

3.1 Understanding the attacks

To implement this classifier, we must first understand what an attack on these devices looks like. An IIoT typically has the following environment: sensors, a cloud environment, a Server for monitoring traffic, and a central router.[2][1][3] The attacker can exploit these environments by intercepting/altering packs through the transport layer, Distributed Denial of Service (DDoS), or by exploiting the protocol vulnerabilities in the network. Figure 1 can be used to visually represent how the IIoT looks, along with how an attacker might try to hack it.

Figure 1 [2]

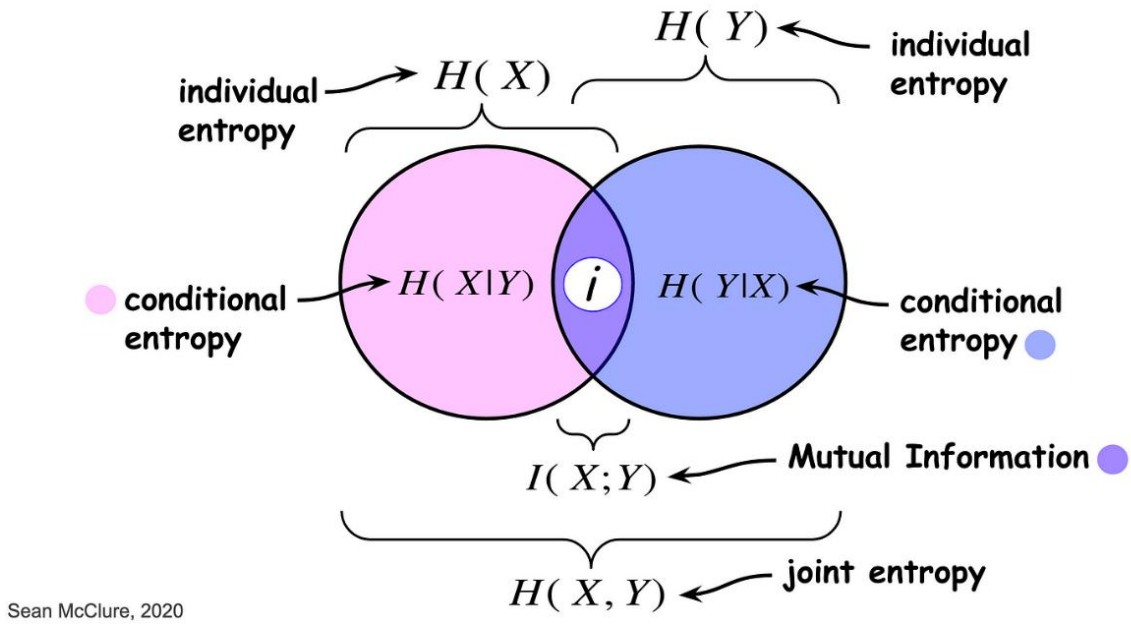


3.2 Collecting the Data

The primary issue of using machine learning to classify an attack vs. regular traffic is getting the data to train the model. To solve this problem, companies can capture traffic through the primary router, IoT cloud, and the Web Server Environment, giving them access to the data to train the model. They can then aggregate the data using Mutual Information (MI)-based feature selection[2][4][5]; this will help transform the raw network information into meaningful data that can capture patterns. Mutual information-based feature selection measures mutual dependence between two variables (random); in other words, it is the

amount of information that can be obtained about one variable through the other variable. This is done using a non-negative value, which indicates the degree of dependence between the two variables, with the concept that the higher the number, the greater the dependence. [4][5] This can be visualized in Figure 2.

Figure 2 [5]



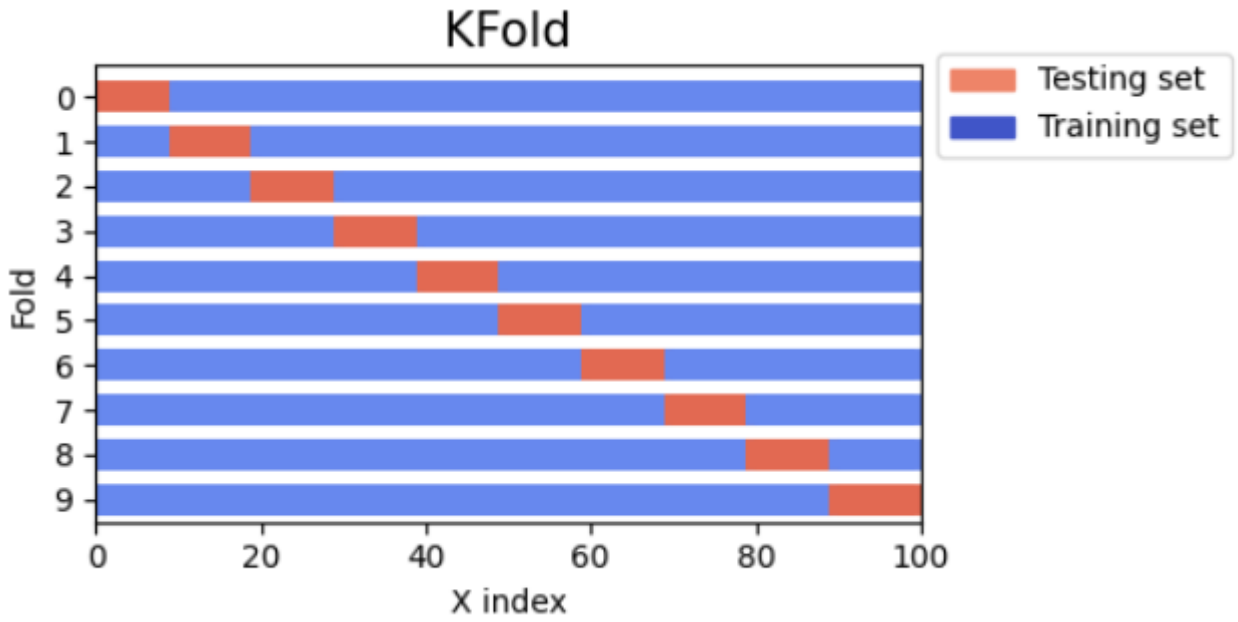
The dataset that I will be using for this project was pulled from IEEE Dataport, titled "BRURIIoT: A Dataset for Network Anomaly Detection in IIoT with an Enhanced Feature Engineering Approach. " This dataset contains 59 million network packets, which were condensed into 3 million records with 150 features. [2] This dataset will provide ample data to engage in various Machine Learning models.

3.3 The Approach

The approach I will be taking during this project is to use various Machine Learning Classifiers to accurately classify between normal network behavior and attacks on the network. These models include Gradient Boost, XGBoost, CatBoost, KNN, AdaBoost,

Random Forest, Extra Trees, and Support Vector Machine [2][7][8]. While training each of these models, the model will be evaluated using standard metrics such as accuracy, precision, recall, and F1-score. I will also use a confusion mat to visualize the accuracy of the models by analyzing true/false positives and true/false negatives. Along with this, I will also conduct K-Fold Cross Validation. K-fold validation is the process of dividing the data into the 'K' number of folds on which the model will be trained, and the remaining fold is what the model will be evaluated [6]. Figure 3 can be used to visualize this.

Figure 3 [6]



4. Conclusion

IIoT has been a massive game changer regarding workplace efficiency and safety; however, it has also brought security concerns to its use of the network to communicate. However, because it communicates over the network, it allows the owner to monitor the traffic coming in and out. This means the owner could detect the malicious actor on the network. This leaves them with the problem of how they can classify regular network traffic vs an attacker. This proposed solution will allow the owner of the IIoT to be able to accurately classify between regular network traffic and an

attacker, which will enable them to have early detection and a faster response time, which will help them to address the anomaly to prevent any disruption or safety hazards in the Industrial IoT environment.

5. Work Cited

[1] What Is Industrial IoT (IIoT)? Cisco. https://www.cisco.com/c/en/us/solutions/internet-of-things/what-is-industrial-iot.html#~role-of-it.

[2] FAHIM AL ISLAM, MD. SHAMSUZZAMAN, MD. SHOHANUR ISLAM, SHAHIDUL AHAD SAKIB, ABU SAYED MD. MOSTAFIZUR RAHAMAN, March 6, 2025, "BRURIIoT: A Dataset for Network Anomaly Detection in IIoT with an Enhanced Feature Engineering Approach", IEEE Dataport, doi: https://dx.doi.org/10.21227/fqqe-g413.

[3] Visconti P, Rausa G, Del-Valle-Soto C, Velázquez R, Donato Cafagna, Fazio RD. 2024. Machine Learning and IoT-Based Solutions in Industrial Applications for Smart Manufacturing: A Critical Review. Future Internet. 16(11):394–394. doi:https://doi.org/10.3390/fi16110394. https://www.mdpi.com/1999-5903/16/11/394.

[4] Information Gain and Mutual Information for Machine Learning. 2024 Apr 15. GeeksforGeeks. https://www.geeksforgeeks.org/information-gain-and-mutual-information-for-machine-learning/.

[5] McClure S. 2020 Nov 7. A Deep Conceptual Guide to Mutual Information - The Startup - Medium. Medium. https://medium.com/swlh/a-deep-conceptual-guide-to-mutual-information-a5021031fad0.

[6] GeeksforGeeks. 2024 May 27. Cross-Validation Using K-Fold With Scikit-Learn. GeeksforGeeks. https://www.geeksforgeeks.org/cross-validation-using-k-fold-with-scikit-learn/.

[7] Top 6 Machine Learning Classification Algorithms. 2024 Feb 26. GeeksforGeeks. https://www.geeksforgeeks.org/top-6-machine-learning-algorithms-for-classification/.

[8] Boosting in Machine Learning | Boosting and AdaBoost. 2019 May 3. GeeksforGeeks. https://www.geeksforgeeks.org/boosting-in-machine-learning-boosting-and-adaboost/.