IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, June 2024

Machine Learning Based IoT Intrusion Detection System

Dr. Srinivas Kanakala
Department of CSE
VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
srinivaskanakala@gmail.com

Abstract: *Today, the use of IoT devices has increased rapidly. Every home network has at least one IoT device. These devices communicate over the Internet through various additional Networking protocols, one of the main protocols is* ***The MQTT Protocol****. This protocol is one of the hot layers for attacks on IoT devices. Various unauthorized devices or authorized devices with malicious intent try to connect to the network and compromise the devices. Hence, it's very important to detect the Intruder in our network. For the detection of Intruders, we use the* ***IDS (Intrusion Detection system)****. This system should be trained to identify the MQTT attacks and get them to the notice of the Network Administrator. We'll be using ML techniques to train the IDS in identifying the attacks. We'll be using datasets that include unidirectional data flow, bidirectional data flow, and packet-based data flow information. Thus, through this project, an IDS with enough Accuracy in detecting the attacks will be developed. To train our IDS, we will be using MQTT-IoT-IDS2020 Dataset. We will be training our IDS with this dataset and will be analyzing the above dataset with our networking knowledge and will be making respective alterations in the dataset according to the attacks to improve the accuracy.*

Keywords: MQTT Protocol

I. INTRODUCTION

In recent years, a great number of Internet of Things (IoT) devices and networks have been used for a variety of applications. Healthcare, smart cities, supply chain management, and farming are some of the use cases. With the increased use of IoT, new protocols are being implemented. MQTT is a popular new protocol for machine-to-machine communication. MQTT is one of the protocols used in IoT networks. We investigate the security issues connected with using MQTT. There are 53,396 MQTT devices that are publicly open and accessible. Their work emphasises the importance of robust detection approaches for MQTT attacks in order to overcome security flaws. IoT Intrusion Detection Systems (IDS) have particular requirements due to the unique nature of the usage situations involved.

IoT IDSs must be versatile, expandable, and built with real or simulated traffic appropriate for their intended use. However, the number of publicly available IoT datasets is restricted, which restricts IoT IDS development. As a result, we are using the newly published IoT-MQTT Dataset to create an IDS that assesses attack scenarios. We, as a team, worked on the following: Analysing a new MQTT dataset that contains both benign and attack scenarios. Evaluating the importance of using high-level (flow-based) features to construct the IDS. Six different machine learning approaches will be used to evaluate the suggested model. Examining the differences between MQTT-based and general threat detection, which emphasises the unique setup and thus the demands of MQTT (IoT) networks

II. LITERATURE REVIEW

In the existing systems, [1] [6] the authors have created a novel Dataset for detecting the attacks on the IoT devices. The dataset is delivered to make it use to other future works. There are many papers published to work on the MQTT protocol [2]. The papers have given the basic explanation of what exactly a MQTT protocol[5][6][7] is, what are the variable fields present in an mqtt [2]packet and the system architecture of the MQTT Network[5]. The pages [7] [8] have demonstrated how the publish subscribe mechanism works and the authentication of the clients with the broker[5]. The basic components and the working of the MQTT broker[5] are explained. The possible vulnerabilities are explained

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, June 2024

through the packet analysis[6][7]. There are no publishing's which have integrated the knowledge of mqtt protocol[5] and the MQTT_IOT2020 Dataset[1]. The fields that can raise a possibility of creating an attack on the MQTT Broker are explored[7] with the help of publicly available MQTT Broker.

2.1 Proposed system

A conceptual model represents the structure, views, and behaviour of a system. An architecture description is a formal explanation and depiction of a system that is organised in such a way that it allows for reasoning about the system's structures and behaviours. A system architecture is made up of system components and sub-systems that work collectively to execute the entire system.

The proposed System, System architecture starts with detecting the face mask of a person, extracting the required features of the person, and identifying features in the database. After detecting it checks the percentage of mask present on the face. The system first scans the person's face, if there are multiple persons it gives multiple results by preprocessing the multiple faces. After preprocessing, the system extracts the features from the face, then classifies the features and compares them with the features which are trained and stored in the database. After recognizing the face mask, it displays the percentage. In this research, we used the MQTT IoT-IDS Dataset. The dataset was built using a simulated MQTT network architecture. The network consists of twelve sensors, a broker, a simulated camera, and an attacker. Five scenarios are documented: (1) regular operation, (2) aggressive scan, (3) UDP scan, (4) Sparta SSH brute-force, and (5) MQTT brute-force attack. The raw pcap files are saved, and the features are extracted. The raw pcap data are analysed at three levels of abstraction: (a) packet features, (b) unidirectional flow features, and (c) bidirectional flow characteristics. Initially we imported all libraries which are required like pandas, matplotlib seaborn etc. Data set is in the CSV format. Data consists of both categorical and numerical data. Some columns have null values which as to be preprocessed. Choosing model. To choose which model to use we must see and compare the performance of different models. In our project we compared different models namely Linear Regression, k-Nearest Neighbors, Decision Tree, SVM, Random Forest and Gaussian Naïve Bayes. As there is only one data set to know the performance correctly, the K-Fold method is useful. In the K-fold cross validation method we divide limited data sets into K number of train data and test data, For each train and test data used to find accuracy of model from which we get k number of accuracies.

III. RESULTS

Table 1 says that, for the Normal flow of traffic, Fig 1 shows *Overall detection accuracy trend using different ML technique*. Figure 2 shows Benign Class Trends. **Decision Trees** have given the highest Intrusion detection accuracy . In the case of Unidirectional flow, **Random Forest and Decision Trees** have given the highest accuracy and for Bidirectional Flow, **KNN Algorithm** stands on top with **99.9%** accuracy.

Table 1: showing Overall detection accuracy

	Features		
	Packet	Unidirectional	Bidirectional
LR	78.87%	98.23%	99.44%
k-NN	69.13%	99.68%	99.9%
DT	88.55%	99.96%	99.95%
RF	65.39%	99.98%	99.97%
SVM (RBF Kernel)	77.4%	97.96%	96.61%
NB	81.15%	78%	97.55%
SVM (Linear Kernel)	66.69%	82.6%	98.5%

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, June 2024

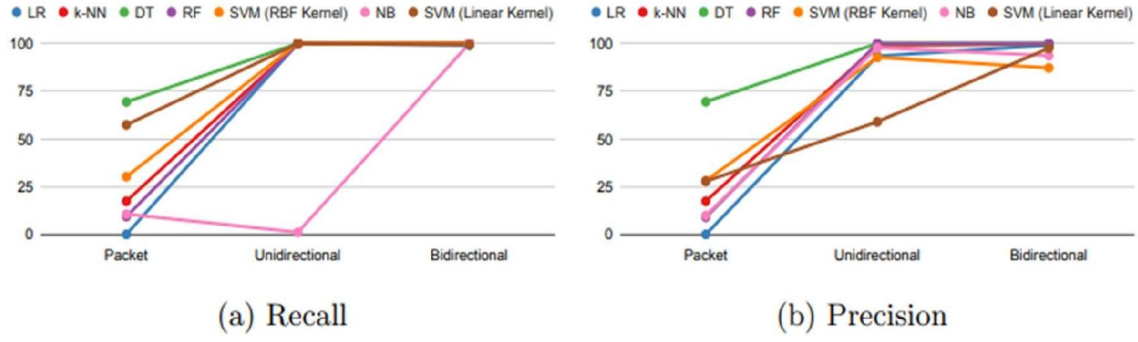

Fig 1: Overall detection accuracy trend using different ML technique

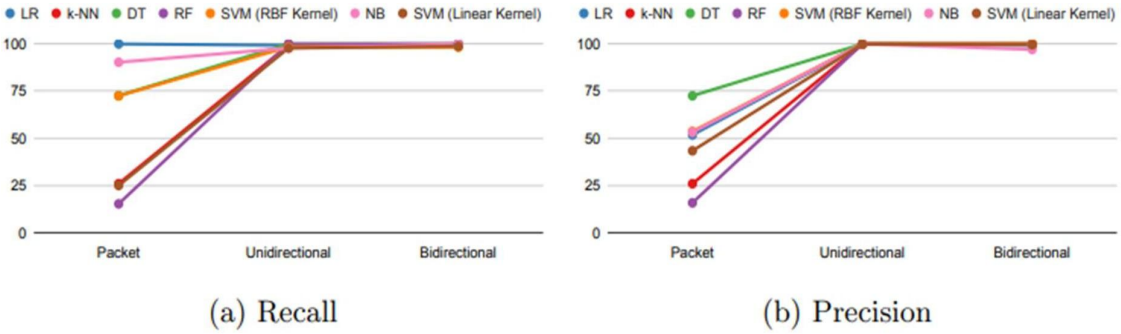

Fig 2 : Benign Class Trends

IV. CONCLUSION AND FUTURE WORK

The goal of this work is to investigate the many obstacles and requirements for developing IDS for IoT models, utilising a MQTT network as a case study. In this study, we examined six distinct machine learning approaches as attack classifiers. Data were collected using a simulated MQTT network. Using the dataset's raw pcap files, three feature levels were extracted: packet, unidirectional, and bidirectional. The research demonstrated that generic networking assaults are clearly distinguishable from regular operation due to their distinct behaviour and patterns when compared to the IoT configuration. However, MQTT-based attacks are more complex and can readily impersonate innocuous actions. The current research can be further extended in scope. An open webpage can be created which takes the features of the traffic as the input from the user and produces the type of traffic after classifying it by using the IDS which will be running in the backend part.

REFERENCES

[1]. Indy, H., Tachtatzis, C., Atkinson, R., Bayne, E., Bellekens, X.: MQTT-IOTIDS2020:
[2]. MQTT internet of things intrusion detection dataset. IEEE Dataport
[3]. Abeles, Daniel; Zioni, M.: MQTT-PWN, IoT exploitation & recon framework.
[4]. Barber, D.: Bayesian reasoning and machine learning. Cambridge University Press (2012)
[5]. Steinwart, I., Christmann, A.: Support vector machines. Springer Science & Business Media (2008) 21. VanderPlas, J.: Python data science handbook: Essential tools for working with data. "O'Reilly Media, Inc." (2016)
[6]. Soni, Dipa, and Ashwin Makwana. "A survey on mqtt: a protocol of internet of things (iot)." International conference on telecommunication, power analysis and computing techniques (ICTPACT-2017). Vol. 20. 2017.
[7]. Atmoko, R. A., R. Riantini, and M. K. Hasin. "IoT real time data acquisition using MQTT protocol." Journal of Physics: Conference Series. Vol. 853. No. 1. IOP Publishing, 2017.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, June 2024

[8]. Dinculeană, Dan, and Xiaochun Cheng. "Vulnerabilities and limitations of MQTT protocol used between IoT devices." Applied Sciences 9.5 (2019): 848.

[9]. Husnain, Muhammad, et al. "Preventing MQTT Vulnerabilities Using IoT-Enabled Intrusion Detection System." Sensors 22.2 (2022): 567.

[10]. Depren, Ozgur, et al. "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks." Expert systems with Applications 29.4 (2005): 713-722.

[11]. Ashoor, Asmaa Shaker, and Sharad Gore. "Importance of intrusion detection system (IDS)." International Journal of Scientific and Engineering Research 2.1 (2011): 1-4.

[12]. Jakkula, Vikramaditya. "Tutorial on support vector machine (svm)." School of EECS, Washington State University 37.2.5 (2006): 3.

[13]. Zhang, Shichao, et al. "Efficient kNN classification with different numbers of nearest neighbors." IEEE transactions on neural networks and learning systems 29.5 (2017): 1774-1785.