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: 7.301
Volume 3, Issue 3, June 2023

Brute Force Attacks Detection on IoT Networks using Deep Learning Techniques

Shubham A. Shirodkar
Late Bhausaheb Hiray S.S. Trust's Institute of Computer Application, Mumbai, India
University of Mumbai, Mumbai, India
shubhamshiro373@gmail.com

Abstract: *The Internet of Things (IoT) sector is expanding quickly, and its applications are becoming more prevalent in our day-to-day lives. Various protocols are used to control communication between IoT devices. The Message Queue Telemetry Protocol (MQTT), a simple and trustworthy communication protocol, is a well-known illustration of these protocols. However, MQTT-IoT networks have been the target of cyberattacks, which highlights the need for an effective intrusion detection system for spotting such attempts. The brute force attack is a common sort of such attacks. We suggest deep learning in this study as a means of automatically identifying brute force assaults on MQTT-IoT networks. We train the deep learning model with a large number of instances and a flow-based feature using the MQTT-IoT-IDS2020 dataset. With more than 99% accuracy in differentiating between regular and brute force attacks, the classification model is quite accurate in detecting such attempts.*

Keywords: Internet of Things

I. INTRODUCTION

IoT is a network of things that can perceive, engage with, and talk to the physical environment. Our daily lives are impacted by IoT applications such as e-health, traffic management, smart homes, smart cities, and smart cars. By 2025, there will likely be more than 27 billion connections, reflecting the IoT's rapid growth.

IoT architecture typically comprises of three key elements: Hardware (sensor nodes), Middleware (data storage and analysis), and Presentation Layer (tools for data presentation to the end user) are the first three layers. Scalability, energy use, security, privacy, and other concerns can all have an impact on the IoT design. As a result, research into messaging protocols that allow IoT devices to exchange messages quickly and securely is ongoing.

In terms of message protocols, MQTT, CoAP, HTTP, and AMQP are all well-liked options. MQTT, which has several features like low memory needs, reduced packet loss, and support for communication in low bandwidth, is one of them and is frequently utilised in this field. A simple machine-to-machine (M2M) communication protocol is what MQTT is intended to be. Messages are published by a publisher (sending device) under a certain topic name. Following that, a broker (server), a central node that supports communication, distributes the message to all subscribers (receiving devices) associated with the topic

IoT networks are often seen to be extremely vulnerable to various threats relating to data protection, availability, unauthorized access, man-in-the-middle attacks, brute force assaults, and other types of attacks. Among the aforementioned assaults, brute force attacks on IoT networks are fairly widespread and cando significant harm. In these types of attacks, the attacker makes numerous attempts to log into a device, or edge device, using different passwords. This is usually done by utilizing special software that is made for creating password combinations. When an attacker takes control of an edge device, it automatically means they have access over all Internet-connected IoT devices because it collects data from numerous sorts of sensors.

In 2016, a malware called Mirai attacked and took control of a significant number of IoT devices, converting them into attack-ready bots. When Mirai finds a device responding to a standard port, it tries a combination of hardcoded usernames and passwords to log in as the administrator. If it functions, the unit disables its regular administrative functions and waits for further instructions from the control centre. Numerous gadgets became malicious agents as a result of the attacks, causing severe harm.

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: 7.301
Volume 3, Issue 3, June 2023

It is crucial to establish an intrusion detection system (IDS) on the network level for the automated detection of brute force attacks in order to combat the aforementioned threat. Building an intelligent IDS capable of detecting brute force intrusions and generating an alarm or taking action to directly halt the assault uses machine learning techniques that leverage supervised or unsupervised learning. Bayes-based methods, Support Vector Machines, Decision Trees, and Deep Neural Networks are common strategies in this field.

In this paper, we suggest using machine learning to detect brute force assaults across IoT networks effectively. The MQTT-IoT-IDS2020 huge data collection, which includes traits for both typical and aggressive behaviours, is used for this purpose. To create a classification model that may be utilised for precise prediction, the data set is prepared and fed into a deep learner. With the aid of several performance evaluation metrics, the suggested model is tested and assessed.

The remainder of this essay is structured as follows: Section 2 examines the relevant research in this field. The research technique is presented in Section 3 of this article. The findings of the experiment are covered in Section 4. Section 5 finishes the essay and makes ideas for additional research.

II. RELATED WORK

IoT devices typically have limited CPU, memory, and battery capacities, which leaves them open to cyberattacks since they frequently do not use high-end security measures. Security and real-time speed are trade-offs that IoT developers must constantly make. However, protection of these systems from various types of assaults, including viruses, spoofing attacks, denial of service attacks, brute force attacks, and others, is essential. Researchers from all around the world have recently focused on this in order to create intelligent IDS to combat the numerous and constantly changing types of cyberattacks on IoT systems.

For the purpose of detecting assaults (invalid data) on IoT systems, the authors of this study constructed a neural network. To simulate edge devices, they employed an Arduino Uno. A temperature sensor is attached to each Arduino Uno. To create a gateway, ten devices are linked to a Raspberry Pi model. Every two seconds, the edge device sent the gateway the temperature reading. Valid and invalid data are simulated, respectively. Additionally, a data set with features for describing the data is created. The neural network model demonstrated accuracy in incorrect data detection

The effectiveness of ensemble learning (XGBoost) and recurrent networks for categorizing attacks on Internet of Things devices that use the MQTT protocol was recently tested by the authors. The three types of assaults being researched are intrusion, man-in-the-middle, and denial of service. These three different assault types were simulated by the authors, who then gathered data. The best features are then chosen after the data has been normalized and balanced between classes. The learning algorithms are fed with the data, and their performance is assessed. Overall, the authors' results for accuracy were very high.

In their research, the authors tested a classifier for the identification of MQTT-based IoT network assaults. They created an environment with IoT sensors and a centralized broker for this purpose. They then performed various denial of service attacks and created a data set based on typical and malicious behaviours. There are 28 features that define the data set. Three algorithms—Naive Bayes, Decision Trees, and Neural Networks—are trained on the data set. The detection of attacks on the IoT network has a high accuracy rate of up to 99.0%, according to the authors.

In their study, the scientists put a classifier to the test for recognizing attacks on MQTT-based IoT networks. For this, they developed a setting complete with Internet of Things sensors and a centralized broker. After that, they carried out a number of denial-of-service assaults and compiled data on both normal and harmful activities. The data set is defined by 28 features. The data set is used to train three algorithms: Neural Networks, Decision Trees, and Naive Bayes. According to the authors, there is a high accuracy rate of up to 99.0% for the detection of assaults on the IoT network.

The authors of a recent paper proposed the MQTTest data set, which focuses on MQTT communications. They made use of IoT-Flock, a tool that simulates IoT devices and MQTT-based network communication. The authors ran simulations of both regular traffic and other attacks, including as floods, denial-of-service attacks, and brute-force attacks. 33 features pertaining to the TCP and MQTT layers were extracted. Neural networks and random forests gave the best accuracy scores on the data set when a variety of machine learning techniques were evaluated against it.

Similar to the authors' work, the MQTT-IoT-IDS2020 data set is proposed. It comprises of four different forms of attacks, including brute force attacks, and five recorded situations that imitate typical behaviours. Twelve MQTT sensors, a broker, a device that simulates a camera, and an attacker make up the fictitious Internet of Things network.

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: 7.301
Volume 3, Issue 3, June 2023

TCP and MQTT layers are characterized by features in the data set. The authors used six machine learning methods to test the effectiveness of the suggested feature set, and they reported good accuracy rates for the classification issue of the five classes: normal class and four attack classes.

In general, academics have given creating methods for detecting attacks on IoT networks a lot of attention. The majority of current methods are based on machine learning techniques, which are quite effective at identifying new assaults. We employ deep learning technique in conjunction with a current and massive data set, MQTT- IoT-IDS2020, for accurate detection of brute force assaults. This method is similar to these approaches, but with a focus on brute force attacks on IoT networks that implement MQTT protocol. We demonstrate how the suggested method is highly effective in identifying such attacks using performance metrics.

III. METHODOLOGY

The classifier-building methodology is provided in this section.

Data Collection

MQTT-IoT-IDS2020 is the data set being investigated. It comprises of five recorded cases, one of which is connected to normal behaviour and the others to four different forms of attacks: aggressive scanning, UDP scanning, Sparta SSH brute forcing, and MQTT brute forcing. Tcpdump is used to gather the data. Twelve MQTT sensors, a broker, a machine that simulates a camera feed, and an attacker make up the network architecture. By leveraging randomized messages from the sensors, normal behaviour is produced. A tool called MQTT-PWN is used to create brute-force assaults. We decided to test our methodology with two abstract levels of features in this data set: Bi-flow features and Uni-flow features. Table 1 shows the quantity of features and instances associated with each of the feature sets stated above. As we are primarily interested in the detection of brute force attacks, we explain the numbers connected to the normal class and the MQTT_brute force class. The work of the feature sets are preprocessed as follows, with a detailed explanation of the data set provided: As these data do not truly indicate discrimination between the classes being studied, we first remove features that define the source and destination IP addresses, time stamps, and MQTT flags. The data collection is next subjected to normalization, where feature values are normalized in the range [0-1]. As a result, we arrived at the improved feature set shown in Table 2.

Table 1. Number of instances per class for the Bi-flow and Uni-flow feature sets.

	# Features	Classes	# Instances
Bi-flow	31	Normal	88160
		MQTT_bruteforce	14544
Uni-flow	18	Normal	176041
		MQTT_bruteforce	28874

Table 2. Number of instances per class for the Bi-flow and Uni-flow feature sets after preprocessing

	# Features	Classes	# Instances
Bi-flow	21	Normal	88160
		MQTT_bruteforce	14544
Uni-flow	11	Normal	176041
		MQTT_bruteforce	28874

Classification

When the data set is ready, a deep learning supervised machine learning algorithm is given it in order to train and construct the classification model. A group of machine learning algorithms known as "deep learning" employ numerous layers of non-linear information processing for supervised learning. It's crucial to include numerous layers when modelling intricate data interactions

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: 7.301
Volume 3, Issue 3, June 2023

We utilize the Deeplearning4j Java library, which is included as a package for the WEKA workbench, to train our model. The package makes it easier to create a variety of complex neural network topologies. We employ a forward and backward prorogation architecture with an input layer, two hidden levels, and an output layer. A collection of neurons is implemented at each layer to capture the data's structure. The pre-processed feature set is used by the input layer. The output layer makes use of a softmax activation function, and it feeds the information to two hidden layers with ReLU activation.

Performance Evaluation

We use two validation techniques to assess the classifier's performance. In the first approach, we employ hold-out testing, in which we construct the classifier using 66% of the data set and evaluate its performance using the remaining data set (a distinct data set from the training data set). In the second validation method, we employ 5-fold cross validation. In this technique, the data set is split into five disjoint sets, four of which are utilised for training and the fifth for testing. The performance metrics and approach are averaged over the five studies.

We utilize five metrics to assess the effectiveness of classification: accuracy, false positive rate, precision, recall, and F-measure. On the basis of the tested data, classification accuracy is computed as the proportion of correctly identified samples to all tested samples. The false positive rate (FPR), which measures the rate of false positives and is determined as the number of false positives divided by the total number of true negatives and false positives, is another performance evaluation metric. The ratio of true positives to the total of true positives and false negatives is used to calculate recall.The precision is calculated by dividing the total number of true positives by the total number of true positives and false positives. The ratio between the double multiplication of precision and recall and the summation of precision and recall is used to determine the F-measure.

IV. EXPERIMENTS AND ANALYSIS

The outcomes of the trials are reported in this section; first, we present the hold-out testing results for classification. The classification outcomes are then shown using 5-fold cross validation.

Experiment 1: Experiments using hold-out testing

We use deep learning to develop the classifier in this experiment, using 66% of the data set for training and the remaining 24% for testing the training model and computing other metrics. The classification measurements for the two feature sets, Bi-flow and Uni-flow, are shown in Table 3.

Table 3. Deep learning classification results using hold-out testing

	Accuracy (%)	FPR (%)	Precision (%)	Recall (%)	F-Measure (%)
Bi-flow	99.56	0.03	99.60	99.60	99.60
Uni-flow	99.67	0.02	99.70	99.70	99.70

Table 3 makes it abundantly evident that the suggested method is highly accurate and produced high classification measures for the two feature sets. This demonstrates the capability of deep learning to distinguish between the two classes of data: normal class and class of bruteforce attacks. The confusion matrices for the two feature sets are shown in Tables 4 and 5 for further study of the classification performance results. The columns represent the anticipated classes, and the rows the ground truth classes.

Table 4. Confusion matrix for deep learning classifier using Bi-flow features

a	b	← classified as
30042	5	a = normal
149	4723	b = brute force attack

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: 7.301
Volume 3, Issue 3, June 2023

The diagonals of both tables show that a higher percentage of normal instances than brute force instances are accurately categorized. The tables make it evident that there are many true positives and few false alarms.

Table 5. Confusion matrix for deep learning classifier using Uni-flow features

a	b	← classified as
59901	8	a = normal
219	9543	b = brute force attack

Experiment 2: Experiments using 5-fold cross validation testing

We use 5-fold cross validation to test the performance of the proposed approach in order to further investigate its robustness. The categorization metrics for the two feature sets are shown in Table 6: both Uni-flow and Bi-flow

Table 6. Deep learning classification results using 5-fold cross validation testing

	Accuracy (%)	FPR (%)	Precision (%)	Recall (%)	F-Measure (%)
Bi-flow	99.56	0.03	99.60	99.60	99.60
Uni-flow	99.67	0.02	99.70	99.70	99.70

Table 6 makes it obvious that the suggested strategy performs well when evaluated using a five-fold cross-validation. Instead of just one set, as in the first experiment, the deep learning methodology exhibits strength in classification over five distinct sets with this validation method.

Tables 7 and 8 give the confusion matrices for the two feature sets for further investigation of the classification performance results. These two tables show that the 5-fold cross validation procedure produced good classification measures.

Table 7. Confusion matrix for deep learning classifier using Bi-flow features

a	b	← classified as
88134	26	a = normal
429	14115	b = brute force attack

Table 8. Confusion matrix for deep learning classifier using Uni-flow features

a	b	← classified as
175994	47	a = normal
644	28230	b = brute force attack

We display the accuracy results for both feature sets over the five folds in figures 1 and 2 to further illustrate the deep learning algorithm's performance stability across various folds of the 5-fold cross validation test. These graphs make it very evident that the deep learning classifier is stable, with little variation between folds.

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: 7.301
Volume 3, Issue 3, June 2023

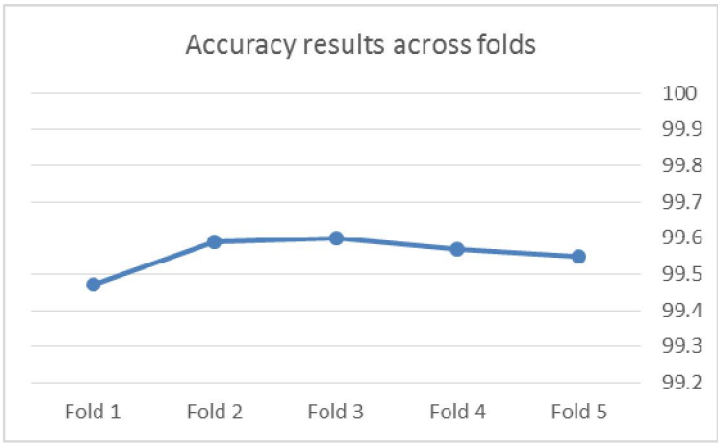

Fig.1.Accuracy across folds for Bi-flow feature

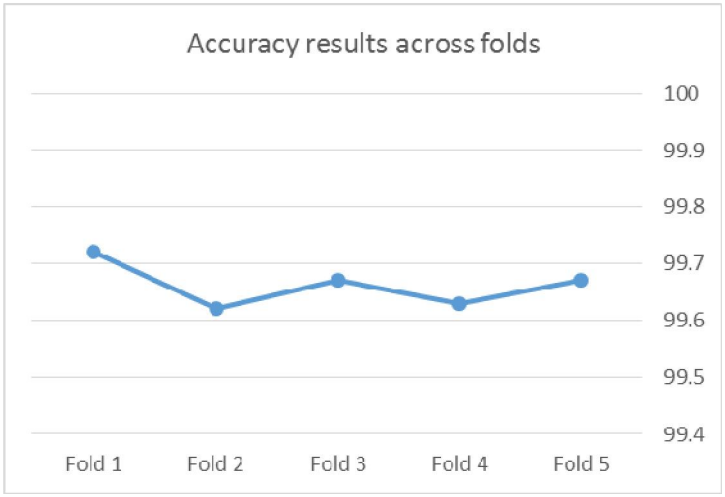

Fig.2. Accuracy across folds for Uni-flow features

V. CONCLUSION

The challenge of identifying brute force assaults on IoT networks was the focus of this article. We used the MQTT-IDS-2020 data set, a prominent and current data set in this field that mimics a real-world MQTT IoT network in both regular operation and four different types of assaults.

Since brute force attacks are frequent and can seriously harm IoT networks, we decided to focus on two scenarios: conventional and MQTT_brute force attacks. Bi-flow features and Uni-flow features were both put into practice. With 102754 instances in the Bi-flow feature set and 204915 instances in the Uni-flow feature set, the number of instances is relatively large for both feature sets. The data sets are preprocessed, and extraneous features are eliminated, resulting in a feature vector with 21 dimensions for the Bi-flow feature set and 11 dimensions for the Uni-flow feature set.

After the data sets are prepared, the deep learning algorithm uses them to develop the training model, which is then validated using the hold-out testing and 5-fold cross-validation testing methods. Using the hold out validation approach, the classification results were very good for both feature sets, with 99.6% accuracy for the Bi-flow feature set and 99.7% for the Uni-flow feature set. The 5-fold cross validation test showed that the results were reliable and that they performed about as well as hold-out testing.The results are generally positive, and the application of such a classification model can be quite effective in identifying brute force assaults on MQTT-IoT networks. We intend to investigate deep learning for the identification of various attacks on the MQTT protocol and other types of protocols in the future.

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: 7.301
Volume 3, Issue 3, June 2023

REFERENCES

[1] Tournier, J., Lesueur, F., Le Mouël, F., Guyon, L., & Ben-Hassine, H. (2021). A survey of IoT protocols and their security issues through the lens of a generic IoT stack. Internet of Things,16,100264.
[2] https://iot-analytics.com/number-connected-iot-devices/,accesses1-5-2022.
[3] Shafique, K., Khawaja, B. A., Sabir, F., Qazi, S., Mustaqim, M. (2020). Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT scenarios. Ieee Access, 8, 23022-23040.
[4] Al-Masri, E., Kalyanam, K. R., Batts, J., Kim, J., Singh, S., Vo, T., Yan, C. (2020). Investigating messaging protocols for the Internet of Things (IoT). IEEE Access, 8, 94880-94911.
[5] Naik, N. (2017, October). Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP. In 2017 IEEE international systems engineering symposium (ISSE) (pp. 1-7). IEEE.
[6] Khan, M. A., Khan, M. A., Jan, S. U., Ahmad, J., Jamal, S. S., Shah, A. A., ... Buchanan, W. J. (2021). A Deep Learning-Based Intrusion Detection System for MQTT Enabled IoT. Sensors, 21(21), 7016.
[7] Singh, M., Rajan, M. A., Shivraj, V. L., Balamuralidhar, P. (2015, April). Secure mqtt for internet of things (iot). In 2015 fifth international conference on communication systems and network technologies (pp. 746-751). IEEE.
[8] Alani, M. M. (2018, December). IoT lotto: Utilizing IoT devices in brute-force attacks. In Proceedings of the 6th International Conference on Information Technology: IoT and Smart City (pp. 140-144).
[9] Najafabadi, M. M., Khoshgoftaar, T. M., Kemp, C., Seliya, N., Zuech, R. (2014, November). Machine learning for detecting brute force attacks at the network level. In 2014 IEEE International Conference on Bioinformatics and Bioengineering (pp. 379-385). IEEE.
[10] Perrone, G., Vecchio, M., Pecori, R., Giaffreda, R. (2017, April). The Day After Mirai: A Survey on MQTT Security Solutions After the Largest Cyber-attack Carried Out through an Army of IoT Devices. In IoTBDS (pp. 246-253).
[11] Hindy, H., Bayne, E., Bures, M., Atkinson, R., Tachtatzis, C., Bellekens, X. (2020, September). Machine learning based IoT intrusion detection system: an MQTT case study (MQTT-IoT-IDS2020 dataset). In International Networking Conference (pp. 73-84). Springer, Cham.
[12] Xiao, L., Wan, X., Lu, X., Zhang, Y., Wu, D. (2018). IoT security techniques based on machine learning: How do IoT devices use AI to enhance security?. IEEE Signal Processing Magazine, 35(5), 41-49.
[13] Canedo, J., Skjellum, A. (2016, December). Using machine learning to secure IoT systems. In 2016 14th annual conference on privacy, security and trust (PST) (pp. 219-222). IEEE.
[14] Alaiz-Moreton, H., Aveleira-Mata, J., Ondicol-Garcia, J., Muñoz-Castañeda, A. L., García, I., Benavides, C. (2019). Multiclass classification procedure for detecting attacks on MQTT- IoT protocol. Complexity, 2019.
[15] Syed, N. F., Baig, Z., Ibrahim, A., Valli, C. (2020). Denial of service attack detection through machine learning for the IoT. Journal of Information and Telecommunication, 4(4), 482-503.
[16] Ciklabakkal, E., Donmez, A., Erdemir, M., Suren, E., Yilmaz, M. K., Angin, P. (2019, October). ARTEMIS: An intrusion detection system for MQTT attacks in Internet of Things. In 2019 38th Symposium on Reliable Distributed Systems (SRDS) (pp. 369-3692). IEEE.