engineering proceedings

MDPI

Proceeding Paper

Detection of DoS Attacks in an IoT Environment with MQTT Protocol Based on Intelligent Binary Classifiers †

Álvaro Michelena [1,*,‡], Francisco Zayas-Gato [2,‡], Esteban Jove [1,2,‡] and José Luis Calvo-Rolle [1,2,‡]

[1] Campus de Elviña, Universidade da Coruña, CITIC, 15071 A Coruña, Spain; esteban.jove@udc.es (E.J.); jlcalvo@udc.es (J.L.C.-R.)
[2] Departamento de Ingeniería Industrial, Universidade da Coruña, CTC, 15405 Ferrol, Spain; f.zayas.gato@udc.es
* Correspondence: alvaro.michelena@udc.es
† Presented at the 4th XoveTIC Conference, A Coruña, Spain, 7–8 October 2021.
‡ These authors contributed equally to this work.

Abstract: The present work deals with the problem of detecting Denial of Service attacks in an IoT environment. To achieve this goal, a dataset registered in an MQTT protocol network is used, applying dimension reduction techniques combined with classification algorithms. The final classifiers presents successful results.

Keywords: MQTT; IoT; DoS; logistic regression; KNN; decision trees; deep neural networks

Citation: Michelena, A.; Zayas-Gato, F.; Jove, E.; Calvo-Rolle, J.L. Detection of DoS Attacks in an IoT Environment with MQTT Protocol Based on Intelligent Binary Classifiers. *Eng. Proc.* **2021**, *7*, 16. https://doi.org/10.3390/engproc2021007016

Academic Editors: Joaquim de Moura, Marco A. González, Javier Pereira and Manuel G. Penedo

Published: 9 October 2021

1. Introduction

Use of the IoT (Internet of Things) paradigm has increased during recent years; this technology has become an essential pillar for a wide variety of processes, in industrial, home, and telecommunications applications, among others. This new concept contributes to encourage connectivity between physical devices, such as controllers, sensors, and actuators, looking for a greater flexibility and process optimisation [1].

However, the significant increase in the flow of communications has resulted a rise in vulnerability, caused by different attacks that put at risk the system's integrity. According to the nature of each attack, different consequences are possible, such as appearance of malware that may harm the equipment, unauthorised access to network information, or DoS (Denial of Service) attacks [2].

In this context, the implementation of algorithms capable of detecting these attacks plays a significant role to ensure the integrity of an IoT environment. Accordingly, this work proposes the use of different intelligent techniques to face the task of detecting DoS attacks in an MQTT network. This document is structured as follows: After the present section, the description of the dataset is carried out in the case of study section. Then, the used techniques are detailed, followed by the experiments and results section. Finally, the conclusions are exposed in the last section.

2. Materials and Methods

2.1. Dataset Description

The MQTT (Message Queuing Telemetry Transport) protocol works at the application level of the TCP (Transmission Control Protocol). This environment is one of the most used in IoT systems [3]. It is based on a star architecture, which pivots on a central broker that manages the network messages. The message procedure follows a publication/subscription approach, where the messages are characterised as a string implementing a nested structure.

To generate the dataset, a server with an Aedes library acted as broker. An ESP 8266 device was in charge of establishing a connection with the several sensors and actuators.

However, the broker was vulnerable to DoS attacks through port MQTT 1883. An MQTT-malaria program was in charge of performing these operations.

The traffic registered during the experiments contained a total number of 94624 samples, with 65 variables containing network information and a label indicating whether the instance is "normal" or "attack". After an initial analysis of the original dataset, the repeated samples were removed, and the constant variables deleted. Furthermore, the categorical variables were transformed following a natural coding criteria. Finally, the data presented 39 variables, 49910 normal instances, and 9429 attacks.

2.2. Used Techniques

2.2.1. Principal Component Analysis

This dimension reduction technique aims to find the directions of higher variability in a dataset, known as principal components [4]. This is performed through the calculation of the eigenvalues of the correlation matrix. Then, using the eigenvectors, the initial set can be linearly transformed into lower dimension space.

2.2.2. Classification Techniques

Logistic Regression

The Logistic Regression (LR) classification technique makes use of a sigmoid function to calculate the class membership probability, whose values are fitted following a gradient descent criteria [5].

K Nearest Neighbours

This classification method uses the data density to label a new instance. To estimate the class membership, it evaluates the K Nearest Neighbours (KNN) and counts the number of samples of each class [5].

Decision Trees

A Decision Tree (DT) algorithm is implemented by repeatedly splitting the dataset using a criteria that maximises the sample separation. At each split, the entropy decrease should be maximised due to the own split [5].

Deep Neural Networks

The Deep Neural Networks (DNN) are based on an architecture made of multiple layers, whose neurons are connected with the neurons of adjacent layers. The weight of each connection, and the parameters of activation functions are tuned during the training process following a minimising error criteria [5].

3. Experiments and Results

3.1. Experimental Setup

Different experiments were carried out to obtain the best classifier. First, with the aim of minimising the computation times and improve the classifier performance, a dimension reduction was carried out using PCA. In this case, two types of reduction were considered: two components and five components. A 10-fold cross-validation was developed, measuring the accuracy, F1 score, precision, recall, specificity, and the Area Under the Receiving Operating Curve (AUC) [6]. This last measure is the one selected to determine the best classifier, because it is nonsensitive to class distribution.

3.2. Results

First, an initial analysis of the PCA result was conducted. From the results achieved in Figure 1, the number of components selected were two and five. With this configuration, the four classification techniques were tested, leading to the final results shown in Figure 2.

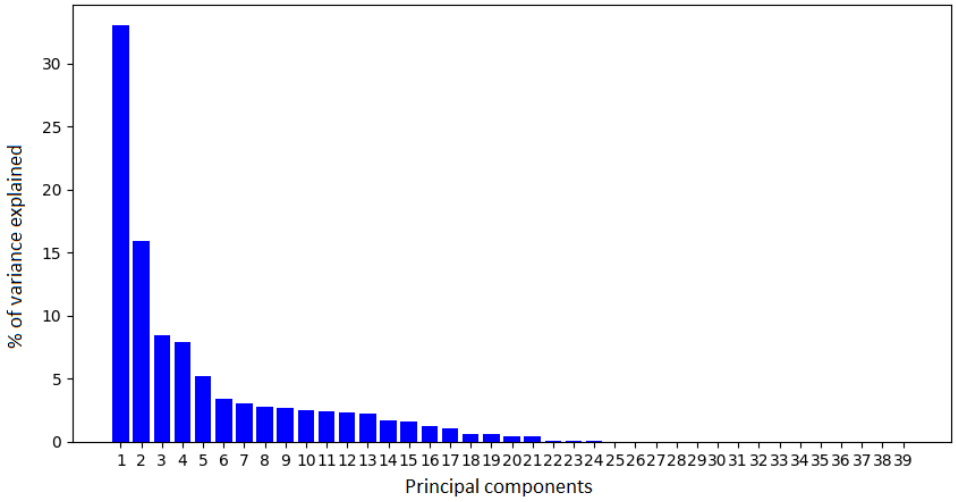

Figure 1. Result of PCA.

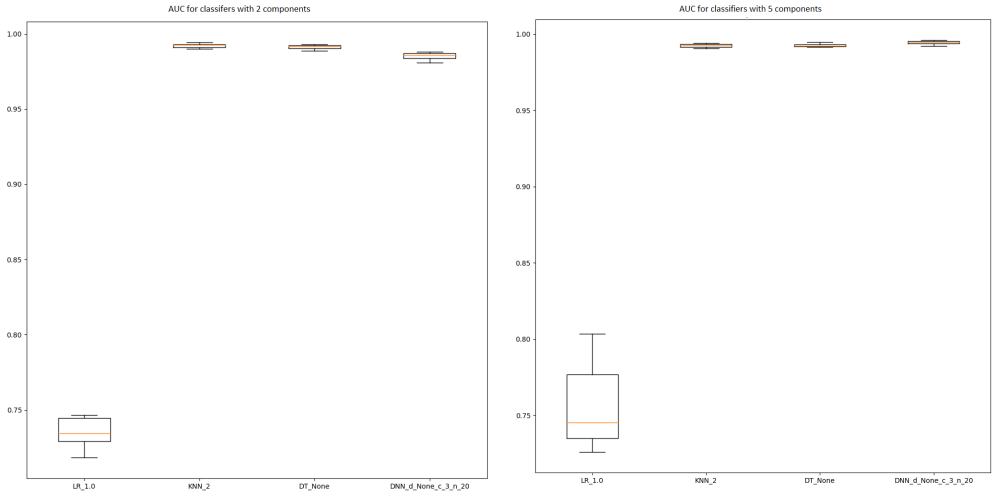

Figure 2. Boxplot representing AUC results for 2 and 5 components.

4. Conclusions

The present papers deals with the detection of DoS attack by means of intelligent classifiers. LR classifiers do not achieve as good a performance as the rest of the techniques. Furthermore, using two and five components does not affect significantly the classifiers performance. The implementation of this approach could entail significant benefits for IoT environments with MQTT protocols.

Acknowledgments: CITIC, as a Research Center of the University System of Galicia, is funded by Consellería de Educación, Universidade e Formación Profesional of the Xunta de Galicia through the European Regional Development Fund (ERDF) and the Secretaría Xeral de Universidades (Ref. ED431G 2019/01).

References

1. Lee, J.; Kao, A. Industry 4.0 Factory in Big Data Environment. tec. News. *HARTING's Technol. Newsl.* **2014**, *26*, 8–9.
2. Zhang, X.; Upton, O.; Beebe, N.L.; Choo, K.K.R. IoT Botnet Forensics: A Comprehensive Digital Forensic Case Study on Mirai Botnet Servers. *Forensic Sci. Int. Digit. Investig.* **2020**, *32*, 300926. [CrossRef]
3. Liu, J.; Kantarci, B.; Adams, C. Machine learning-driven intrusion detection for Contiki-NG-based IoT networks exposed to NSL-KDD dataset. In Proceedings of the 2nd ACM Workshop on Wireless Security and Machine Learning, Linz, Austria, 13 July 2020; pp. 25–30.
4. Martinez, A.M.; Kak, A.C. Pca versus lda. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 228–233. [CrossRef]

5. Dreiseitl, S.; Ohno-Machado, L. Logistic regression and artificial neural network classification models: A methodology review. *J. Biomed. Informat.* **2002**, *35*, 5–6. [CrossRef]
6. Jove, E.; Gonzalez-Cava, J.M.; Casteleiro-Roca, J.L.; Méndez-Pérez, J.A.; Reboso-Morales, J.A.; Pérez-Castelo, F.J.; de Cos Juez, F.J.; Calvo-Rolle, J.L. Modelling the hypnotic patient response in general anaesthesia using intelligent models. *Log. J. IGPL* **2019**, *27*, 189–201. [CrossRef]