proceedings

Proceedings

Anomaly Detection in IoT: Methods, Techniques and Tools [†]

Laura Victoria Vigoya Morales *, Manuel López-Vizcaíno, Diego Fernández Iglesias and Víctor Manuel Carneiro Díaz

Department of Computer Science, University of A Coruña, 15071 A Coruña, Spain

* Correspondence: l.v.vigoya@udc.es

† Presented at the 2nd XoveTIC Conference, A Coruña, Spain, 5–6 September 2019.

Published: 22 July 2019

Abstract: Nowadays, the Internet of things (IoT) network, as system of interrelated computing devices with the ability to transfer data over a network, is present in many scenarios of everyday life. Understanding how traffic behaves can be done more easily if the real environment is replicated to a virtualized environment. In this paper, we propose a methodology to develop a systematic approach to dataset analysis for detecting traffic anomalies in an IoT network. The reader will become familiar with the specific techniques and tools that are used. The methodology will have five stages: definition of the scenario, injection of anomalous packages, dataset analysis, implementation of classification algorithms for anomaly detection and conclusions.

Keywords: Intrusion Detection Systems; analysis; metric; algorithm design; computer network management

1. Introduction

The anomalies in a network cannot always be categorized as an attack, but they give important insights into the traffic behavior they may have. Although they are not always harmful elements, they can help identify important and critical information in various applications [1]. In the detection of anomalies, the scarcity of data with the appropriate characteristics makes it necessary to carry out systematic analysis that allows parameterizing the data to avoid problems in the accuracy of the results. To analyze such data and find a relationship or predict known or unknown data, data mining techniques are used. In these we find, clustering, classification, and techniques based on machine learning. The purpose of this work is to present the methodology developed to detect traffic anomalies, based on a real IoT system, and after its application determines the efficiency and effectiveness of the proposed methodology, using specific techniques and tools. The methodology will have five stages: set the scenario, inject of anomalous packages, dataset analysis, implement of classification algorithms for anomalies detection and develop the conclusions, recommendations, and implications.

2. Network Distribution and Dataset Generation

The first phase of the methodology consists in generating a labeled dataset that allows injecting selected characteristics in classification and prediction algorithms. To develop our scenario, we rely on the installed system of the CPD, located in the CITIC. The data was obtained after performing the mathematical modeling of a real IoT system. This consists of a network of 16 temperature sensors, 4 InRows with two devices responsible for sensing the temperature of the airflow and two for the coolant fluid of the unit, one for input and one for output, respectively. The data from the sensors were sent to the monitoring system every 5 min for a week. During the data analysis, we found that the data complied with seasonality conditions. This assertion allowed us to use time series and forecasting for analysis and prediction of series behavior [2].

For the simulation and generation of the labeled dataset, a virtualized system was implemented. The pilot comprises five virtual machines, where four of these nodes represent the controllers of the sensors of each of the simulated InRows and the fifth machine consists in the MQTT broker that allows to publish and subscribe to topics. Every node (ubuntu 18.04) has an NTP client that allows controlling the temporal synchrony between all the nodes and the broker, while the broker contains an MQTT v3.1 mosquitto server [3] that can handle requests in port 1883 MQIsdp (MQSeries SCADA protocol). Once the traffic has been captured it is possible to modify most of the fields of the different TCP/IP protocols: Ethernet, IP, TCP or MQTT, with python-scapy 2.3.3-3 library [4]. This allows not only to modify any field of the captured packets but also to reinsert these modified packets into the network, simulating anomalous situations such as interception, duplication or removal of packages. The dataset finally contains a seven days network activity labeled with attacks distributed over this period with diverse intrusion scenarios.

3. Dataset Analysis

The purpose of the dataset analysis is to determine the most relevant characteristics that can affect the classification, to be incorporated into the machine learning algorithms. To interpret the data, the traffic is collected using sniffing tools in the broker. For the analysis, we export the pcap to a csv using tshark and study what is involved in the process [5]. In this stage, to generate models that determine the behavior of anomalous traffic in the network, a statistical analysis of conditions such as protocol, source and destination IP addresses, source and destination ports, flags, time, duration, mean bytes, number of packets and weekday was performed. MQTT protocol is an application layer protocol, so it is on top of TCP/IP heap. Therefore, both the client and the broker need to have a TCP/IP stack and the analysis must be done taking this into account [6].

4. Identification of Machine Learning Algorithms and Optimization of Early Detection

Once the marked dataset has been generated and the main characteristics have been determined, machine learning algorithms are used to classify packages. For this task we can make use of logistic regression, LDA, QDA, K-nearest neighbors (KNN) method, Tree-Based methods, Support Vector Machines (SVM), etc.[7]. The results can be evaluated using different metrics: the traditional precision, recall and F1, and Early Risk Detection Error (ERDE) [6].

5. Conclusions

This work presented a methodology to develop a systematic approach to dataset analysis for detecting traffic anomalies in an IoT network. Having a methodology allows to standardize, structure and organize the work and in this way generate efficiency and effectiveness in the realization of the project.

Author Contributions: All authors have equally contributed to this article.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Agrawal, S.; Agrawal, J. Survey on Anomaly Detection using Data Mining Techniques. *Procedia Comput. Sci.* **2015**, *60*, 708–713.
2. Forecasting: Principles and Practice. Available online: http://OTexts.com/fpp2 (accessed on 10 July 2019).
3. An Open Source MQTT Broker. Available online: https://mosquitto.org/ (accessed on 10 July 2019).
4. Scapy. Available online: https://libraries.io/pypi/scapy/2.3.3 (accessed on 10 July 2019).
5. Tshark: Terminal-based Wireshark. Available online: https://www.wireshark.org/docs/man-pages/tshark.html (accessed on 10 July 2019).

6. Fernandez, D.; Vigoya, L.; Cacheda, F.; Novoa, F.J.; Lopez-Vizcaino, M.F.; Carneiro, V. A Practical Application of a Dataset Analysis in an Intrusion Detection System. In Proceedings of the 2018 IEEE 17th International Symposium on Network Computing and Applications (NCA), Cambridge, MA, USA, 1–3 November 2018.
7. An Introduction to Statistical Learning. Available online: https://doi.org/10.1007/978-1-4614-7138-7 (accessed on 10 July 2019).