УДК 004.056

Port Scanning Detection Using Machine Learning Models

Anna A. Stepanian (ITMO University)
Scientific supervisor – tutor, Arthur R. Khanov
(ITMO University)

Introduction. Cybersecurity is the process of defending against malicious attacks on servers, mobile devices, electronic systems, and networks. Network security is the process of protecting a computer network from intruders, including malware and targeted attacks. As networks expand in size and complexity, IT experts must work to build a reliable network. Making networks extremely safe is a difficulty as many attackers attempt to use its flaws to compromise the availability, integrity, or confidentiality of the network. Port scanning is a kind of network reconnaissance used to find out which computer ports are open. Since certain programs listen on specific ports and respond to traffic in particular ways, this might allow the scanner to identify the apps that are currently operating on the system [1]. Machine learning models are used to increase port scanning detection accuracy as they learn from the data over time without being explicitly programmed to do so.

Main part. Detecting port scanning activities using machine learning techniques can be a complex task [2], but there are several steps to improve the accuracy of the model:

1. Collect high-quality data: Collecting accurate and diverse data is critical for any machine learning project [3]. When collecting data for port scanning detection, a range of different scanning techniques and sources can be included to ensure that the model can detect a wide variety of port scanning activity.
2. Feature engineering: Once the data have been collected, it is necessary to extract relevant features from it. Some features that could be useful for port scanning detection include the number of packets sent, the duration of the scan, the frequency of scanning, and the types of ports scanned.
3. Selecting appropriate algorithms: There are many machine learning algorithms that can be used for port scanning detection. Some popular options include Random Forest, Logistic Regression, Support Vector Machines, and Neural Networks. It is important to experiment with different algorithms to see which one works best for the specific use case.
4. Labeling data: Once the features have been extracted from the data, it is necessary to label it. This means to identify which instances of the data represent port scanning activities and which do not. There are manual labeling or automated labeling techniques to do this.
5. Training and testing: Start training the machine learning model with labeled data. Spliting the data into training and testing sets, and using the training set to train the model. Then using the testing set to evaluate the accuracy of the model.
6. Hyperparameter tuning: Hyperparameters are the settings for the machine learning model that are not learned during training. These settings can significantly affect the accuracy of the model, so it is important to experiment with different hyperparameters to find the best settings for the model.
7. Evaluation: Evaluation is the essential element to discover and choose the best machine learning algorithm suitable for port scanning detection. To resolve the task and determine the most accurate model Precision, Recall, Accuracy and F1 Score will be measured, and ROC curve will be illustrated.

Conclusion. Overall, improving the accuracy of port scanning detection using machine learning techniques requires a combination of good data, feature engineering, appropriate algorithm selection, labeling, training, and hyperparameter tuning. With careful attention to these factors, a highly accurate model for detecting port scanning activities can be built.

References:

1. Kumar, M. Satheesh, et al. "Artificial intelligence managed network defense system against

port scanning outbreaks." 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN). IEEE, 2019.

2. Balqees Ali Al-Ajmi, Amir Javed, Martin J. Chorley, “Creating a machine learning model based on network activity to detect attacks from a malicious webserver”. 2021 Department of Computer Science with Security and Forensics Cardiff University.

3. Dataset of Probing Attacks performed with nmap, unicornscan, hping3, zmap and masscan. – URL: https://zenodo.org/record/3558350. Publication date: 29.11.2019.

Anna A. Stepanian (author) Signature

Arthur R. Khanov (scientific supervisor) Signature