Preliminary analysis of malware detection in opcode sequences within IoT environment

ABSTRACT

With the technological development and means of communication, the Internet of Things (IoT) has become an essential role in providing many services in daily life through millions of heterogeneous but interconnected devices and nodes. This development is opening to many security and privacy challenges that can cause complete network breakdown, bypassed access control or the loss of critical data. This paper attempts to provide a preliminary analysis for malware detection within data generated by IoT-based devices and services in the form of operational codes (Opcode) sequences. Three machine learning algorithms are evaluated and compared for accuracy, precision, recall and F-measure. The results showed that the Random Forest (RF) achieved the best accuracy of 98%, followed by SVM and k-NN, both with 91%. The results are further analyzed based on the Receiver Operating Characteristic (ROC) curve and Precision-Recall curve to further illustrate the difference in performance of all three algorithms when dealing with IoT data.

Keyword: Machine learning; Malware detection; Operation codes