

https://dspace.org/

Using ensemble learning for the network intrusion detection problem

Kalonji, Roland Mpoyi

2019-08-01
https://hdl.handle.net/10539/29057
Kalonji, Roland Mpoyi, (2019). Using ensemble learning for the network intrusion detection problem, University of the Witwatersrand, https://hdl.handle.net/10539/29057

Downloaded from a DSpace repository. https://dspace.org/

ABSTRACT

Nowadays, most organizations and platforms employ an intrusion detection system (IDS) to enhance their network security and protocol systems. The IDS has therefore become an essential component of any network system; it is a tool with several applications that can be tuned to specific content in a network by identifying various accesses (normal or attack). However, network intrusion detection system (NIDS) that focuses on revealing suspicious activities, is not effective in solving various problems such as identifying false IP packets and encrypted traffic. Hence, this work investigates the use of ensemble learning to solve these types of network intrusion detection problems (NIDPs). Random forest (RF), Decision Tree (DT) and Support Vector Machine (SVM) are introduced as classifiers based on Boruta and Principal Component Analysis (PCA) algorithms. In general, the main difficulties in using ensemble for the intrusion problem are to minimize false alarms and to maximize detection accuracy (Anuar et al., 2008). Additionally, the NIDP is divided into five categories, namely the detection of probe attacks, denial of service, remote to local, user to root and normal instances. Each problem is examined by one of the three aforementioned classifiers. In tackling these problems, the three classifiers achieved competitive results comparing to the works conducted by Balon-Perin (2012), Zainal et al. (2009) and Kevric et al. (2017). The results revealed that ensemble learning achieved more than 99% accuracy in demarcating attacks from normal connections. Particularly, RF, DT and SVM allowed to safeguard the NIDS from known and unknown attacks by developing reliable techniques. The KDD99 and NSL KDD datasets have been used to implement and measure the system performance (Fan et al., 2000; Dhanabal and Shantharajah, 2015).

Keywords: ensemble learning, network intrusion detection problem, random forest, decision tree, support vector machine.