

Adversarial Attack against DNN-based DDoS Intrusion Detection System

Mariama Mbow *

Hiroshi Koide †

Kouichi Sakurai ‡

Keywords: Adversarial Machine Learning, Evasion Attack, DDoS detection, DNN, Network Intrusion Detection System

Abstract

Nowadays, deep learning (DL) is a popular method for implementing a network intrusion detection system (NIDS). However, studies have shown that deep learning algorithms can be vulnerable to adversarial samples: inputs that are intentionally crafted to cause the model to make wrong decisions. In this paper, we investigate an evasion attack which aims to bypass DNN-based DDoS intrusion detection system. The work presented is two folds: (1) we implement a deep learning approach for intrusion detection system using Deep Neural Network (DNN); and (2) we perform an evasion attack against the built DNN for NIDS using two well known adversarial sample methods Fast Gradient Sign Method (FGSM), and Deepfool. The experimental results on CICIDS2017 benchmark dataset show that both model evasion attacks can successfully decrease the accuracy of the NIDS, *i.e.*, can influence the DDoS detector to misclassify the attack traffic as benign.

* Department of Informatics Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan, (mbow.mariama.076@s.kyushu-u.ac.jp)

† Research Institute for Information Technology, Cyber Security Center, Kyushu University Fukuoka, Japan

‡ *