Feature Selection for Improving Deep Neural Network-based Intrusion Detection System

Anh Sy DO[†a)], ***Student Member***, **Akiko MANADA**[†b)], ***and*** **Kohei WATABE**[††c)], ***Members***

1. Introduction

Network Intrusion Detection Systems (NIDSs) are the primary defense against unauthorized access and threats by continuously monitoring network traffic. In particular, learning-based NIDSs utilize machine learning algorithms to improve detection capabilities. However, the effectiveness of these systems is often hampered by the sheer volume of data they must process, which can include many redundant or irrelevant features. To address this challenge, a feature selection has emerged as a critical step in optimizing NIDS performance. By selecting the most relevant features from network flow data, it is possible to enhance the efficiency of intrusion detection while minimizing computational overhead.

2. Proposed Methodology

In this study, we introduce an advanced feature selection technique derived from the modified Cuckoo Search Algorithm (CSA), known as Mutation Cuckoo Fuzzy (MCF). This method combines the exploratory strengths of the CSA with the mutation operator and Fuzzy C-Means (FCM) logic to better handle the complexities of network data and ensure the diversity of the solution population. The selected features are used as inputs for the designed multi-layered Deep Neural Network (DNN)-based IDS to evaluate the effectiveness in improving performance and resource utilization.

3. Experiment and Results

Our study is applied to the CIC-IDS2017 dataset [1]. From preliminary tests, MCF parameters for the initial population of solutions were set to 10, where a discovery rate P_α for the worst solutions and the maximum number of iterations were set to 0.25 and 200, respectively. For FCM logic, the centroid was set to 2, fuzziness to 2.0, and the error threshold to 1×10^{-5}. The mutation operator combines the best and worst solutions to ensure the diversity of the solutions. Our feature selection method reduces the number of features of the CIC-IDS2017 dataset from 78 to 29. We designed a multi-layered DNN for intrusion detection, with an input layer for the selected features, three hidden layers (1024, 768, 512 neurons), and an output layer for binary classification [2]. The CIC-IDS2017 dataset comprises 225745 balanced samples, and they are divided into 165730 samples for training and 60015 samples for testing.

Table 1 Performance comparison of feature selection methods

Feature selection	Selected features	Acc. (%)	Pre. (%)	Recall (%)	F-score (%)	E.T. (s)
MCF	29	**99.45**	**99.32**	99.59	**99.45**	**331.52**
naïve-MCF	26	98.92	98.68	99.15	98.91	392.95
RF	33	99.38	99.10	**99.67**	99.38	4659.45
Chi-Square	24	95.84	96.40	95.22	95.81	351.47
RFE	**13**	97.31	96.19	98.50	97.33	1367.72

Table 1 compares the performance of feature selection techniques for DNN-based IDS on the CIC-IDS2017 dataset. Our proposed MCF method, utilizing 29 selected features, achieved the highest accuracy (99.45%), precision (99.32%), and F-score (99.45%). As for the Recall, the DNN-RF achieved the highest value (99.67%), but the execution time of our proposed method is significantly shorter than that of the DNN-RF.

4. Conclusion

This study introduced the MCF feature selection technique for a DNN-based IDS. This technique reduced the CIC-IDS2017 dataset features from 78 to 29, which significantly improved IDS efficiency. Future work will address efficiency on larger datasets and performance across different IDSs.

Acknowledgments

This work was partly supported by JSPS KAKENHI Grant Number JP23H03379.

References

[1] I. Sharafaldin, A.H. Lashkari, A.A. Ghorbani, *et al.*, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," ICISSP, vol.1, pp.108–116, 2018.

[2] A. Thakkar and R. Lohiya, "Fusion of statistical importance for feature selection in deep neural network-based intrusion detection system," Information Fusion, vol.90, pp.353–363, 2023.

[†]The author is with the Graduate School of Engineering, Nagaoka University of Technology, Nagaoka, Niigata, Japan.

[††]The author is with the Graduate School of Science and Engineering, Saitama University, Saitama-city, Saitama, Japan.

a) E-mail: s213158@stn.nagaokaut.ac.jp

b) E-mail: amanada@vos.nagaokaut.ac.jp

c) E-mail: kwatabe@mail.saitama-u.ac.jp