SUGISAKA MASANORI

Journal of Robotics, Networking and Artificial Life
Vol. 9(3); December (2022), pp. 303–308
ISSN (Online):2352-6386; ISSN (Print):2405-9021
https://alife-robotics.org/jrnal.html

Research Article

# Data-Balancing Algorithm Based on Generative Adversarial Network for Robust Network Intrusion Detection

I-Hsien Liu[1], Cheng-En Hsieh[1], Wei-Min Lin[1], Jung-Shian Li[1], Chu-Fen Li[2]

[1]*Department of Electrical Engineering / Institute of Computer and Communication Engineering, National Cheng Kung University, No.1, University Rd., East Dist., Tainan City 701401, Taiwan*

[2]*Department of Finance, National Formosa University, No.64, Wunhua Rd., Huwei Township, Yunlin County 632301, Taiwan*

ARTICLE INFO

*Article History*
Received 20 October 2021
Accepted 04 October 2022

*Keywords*
Anomaly traffic detection
Machine learning
IDS dataset
GAN
Performance analytics

ABSTRACT

With the popularization and advancement of digital technology and network technology in recent years, cyber security has emerged as a critical concern. In order to defend against malicious attacks, intrusion detection systems (IDSs) increasingly employ machine learning models as a protection strategy. However, the effectiveness of such models is dependent on the algorithms and datasets used to train them. The present study uses five different supervised algorithms (Naïve Bayes, CNN, LSTM, BAT, and SVM) to implement the IDS machine learning model. A data-balancing algorithm based on a generative adversarial network (GAN) is proposed to mitigate the data imbalance problem in the IDS dataset. The proposed method, designated as GAN-BAL, is applied to the CICIDS 2017 dataset and is shown to improve both the recall rate and the accuracy of the trained IDS models.



## 1. Introduction

The Internet brings unparalleled advantages in communication, convenience, knowledge sharing, entertainment, and so on. However, with the increasing reliance of individuals, enterprises and government agencies on the Internet, the risks and impacts of malicious network attacks have scaled proportionally. The effects of cyber-attacks range from simple inconvenience, such as a temporary lack of access to a website or service, to issues of national strategic importance, such as disruptions to national infrastructures, the theft of sensitive government data, the malfunction of military systems, and so on. On April 15, 2021, the US National Security Agency (NSA), US Cybersecurity and Infrastructure Security Agency (CISA), and Federal Bureau of Investigation (FBI) issued a joint report, in which they advised that Russian military intelligence routinely conducts malicious cyber activities against government and private sector networks in the US and its global partners [1]. Cyber security, therefore, is increasingly regarded as a matter of national strategic importance.

As artificial intelligence (AI) and big data collection technology continue to advance, AI has found use in a wide range of applications, such as healthcare, transportation, education, research, and so on. AI has also emerged as a powerful tool for realizing intrusion detection systems (IDSs) for protecting sensitive data networks from malicious attack. However, the performance of the machine learning models used in such systems is dependent not only on the machine learning algorithm used to train the model, but also the quality of

*Corresponding author's E-mail jsli@mail.ncku.edu.tw URL: www.ncku.edu.tw, www.nfu.edu.tw*

the dataset used for the training process, i.e., the extent to which the data instances in the dataset are correctly labeled and represent the target data which are to be detected by the IDS. Machine learning research on IDS systems generally uses public datasets for classification model training. However, some of these datasets are incompatible with current network attacks since they are outdated and lack sufficient diversity. Consequently, more robust datasets are required for the training and evaluation of IDS machine learning models.

Accordingly, the present study commences by training the IDS machine learning model using five different algorithms, namely Naïve Bayes, CNN, LSTM, BAT, and SVM. In general, an imbalance between the number of normal data instances and the number of malicious data instances in the dataset degrades the performance of the IDS model. Thus, a data-balancing algorithm based on a generative adversarial network (GAN), designated as the GAN-BAL algorithm, is thus proposed to mitigate the data imbalance problem in the IDS dataset. It is shown that the application of the GAN-BAL algorithm to the CICIDS 2017 dataset improves both the recall rate and the accuracy of the IDS machine learning models.

## 2. Related Work

Since 1998, more than thirty machine learning traffic datasets have been used to evaluate the ability of machine learning algorithms to achieve malicious traffic detection [2]. For example, Zhao et al. [3] proposed the TSA-AdaBoost algorithm to learn and judge abnormal traffic, and verified its performance through KDD-99 and UNSW-NB15 datasets. In conducting IDS research, it is essential to select a dataset which properly matches the research requirement. This section introduces three of the most commonly used machine learning traffic datasets, namely NSL-KDD. UNSW-NB15 and CICIDS 2017.

### 2.1. *NSL-KDD*

In 1998, the MIT Lincoln Laboratory collected the network packets from a simulated US Air Force LAN over a nine-week period. The packets were then converted into connection records to create a traffic dataset for IDS benchmarking and evaluation purposes. In 1999, the dataset was used in the Knowledge Discovery and Data Mining (KDD) Cup for competition purposes, and hence came to be known generally as the KDD 99 dataset from then on [4]. However, in 2009, Tavallaee et al. [5] discovered that the dataset contained redundant duplicate records and focused on specific record learning characteristics. Thus, the authors improved the dataset by removing the duplicate records and conditional random sampling records in order to create a new enhanced dataset designated as the NSL-KDD dataset.

### 2.2. *UNSW-NB15*

In 2015, Moustafa and Slay [6] recorded the network traffic at the Australian Centre for Cyber Security (ACCS) for 31 hours and converted the collected packet log file into a traffic dataset using the IXIA tool. The dataset, known as UNSW-NB15, contains a total of 257,673 data records with 49 features, and is applicable to nine common network attack methods, namely Fuzzer, Analysis, Backdoor, DoS, Exploit, Generic, Reconnaissance, Shellcode, and Worm.

### 2.3. *CICIDS 2017*

In 2017, Sharafaldin et al. [7] generated an IDS dataset known as CICIDS 2017 containing a total of 2,136,470 flows relating to benign traffic and malicious traffic flows generated by eight common attack modes, namely Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DDoS. 80 traffic features were extracted from the dataset using the CICFlowmeter tool [8]. The optimal feature set required to detect each attack mode was then identified using the Random Forest algorithm. In addition, the quality of the CICIDS dataset was evaluated using the 11 criteria listed in the framework proposed by Gharib et al. [9] in 2016. Overall, the results indicated that the dataset was the only dataset among all the publicly available datasets published since1998 to meet all 11 criteria.

## 3. System Structure

Public IDS evaluation datasets typically contain far more benign data samples than malicious data samples, and the resulting data imbalance problem degrades the training and detection performance of the machine learning model. Accordingly, this study first used a sandbox system to generate malicious samples and then constructed a more balanced IDS dataset using these samples and a newly-proposed data-balancing algorithm based on a generative adversarial network designated as GAN-BAL. Fig. 1 shows the basic system architecture

employed in the present study. First, it is necessary to collect data to establish a traffic dataset, and then use the GAN-BAL algorithm to balance the proportion of samples in the dataset.
The details of the malicious flow creation process and GAN-BAL algorithm are described in the following sub-sections.

### 3.1. *Data Collection and Malicious Flow Creation*

True flow traffic data were collected from a real network of research center TWISC@NCKU [10] and transformed from a pcap file to a traffic statistics csv file using

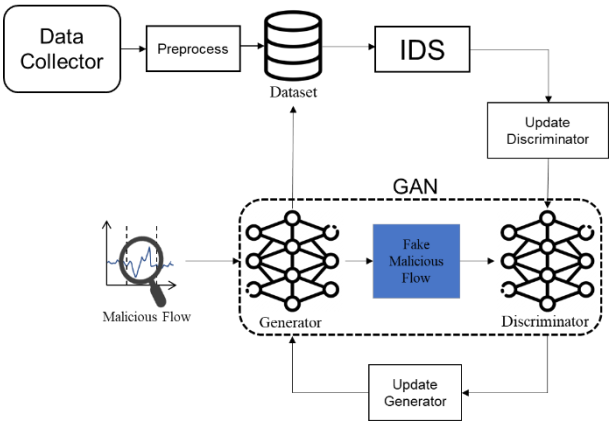


Fig. 1. System architecture.

CiCFlowmeter [8].
A Cuckoo sandbox system was used to produce malicious program samples with which to mitigate the imbalance problem in the IDS dataset (see Fig. 1). As shown in Fig. 2, the Cuckoo system triggered malicious program samples through the network on the client side, and Agent.py on the client side recorded the various behaviors of these samples and sent them back to the host on the user side. The traffic generated by the malicious samples was analyzed and found to contain very little background traffic since the Cuckoo system started to record the traffic flow only after the samples were triggered. The background flow was judged to be sufficiently small to be ignored. Thus, the entire flow recorded by the Cuckoo system was regarded as malicious flow for the purposes of the current study.

### 3.2. *GAN-BAL Algorithm*

Generative adversarial networks (GANs) use two neural networks (a generator network and a discriminator network, see Fig. 1) to produce new synthetic data instances which are sufficiently similar to the input data to pass for original training samples. The synthesis

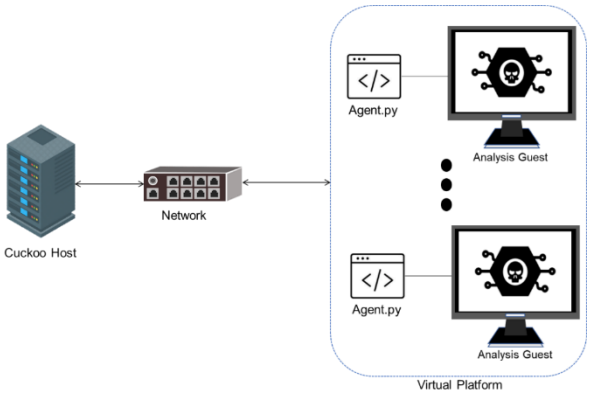


Fig. 2. Structure of Cuckoo system used to generate malicious traffic samples.

process is adversarial in the sense that the generator Fig. 2 continuously attempts to deceive the discriminator with the samples it produces, while the discriminator continuously attempts to keep from being fooled. Once the generator has been adequately trained, it can be used to create new plausible samples on demand.
GANs have been used in many science, entertainment, fashion, and research applications in recent years, and have recently also been applied to network intrusion detection. For example, Wang et al. [11] proposed a semi-supervised learning encryption traffic classification method based on a generative adversarial network (GAN), called ByteSGAN, and conducted experiments through public datasets ISCX2012 VPN-non VPN and Cross-market. The experimental results show that ByteSGAN can effectively Improve traffic classifier performance. Huang et al. [12] developed an IDS system based on an imbalanced generative adversarial network (IGAN-IDS) to deal with the imbalance problem inherent in IDS datasets. The present study builds upon the work conducted in [12] to construct a GAN-based data-balancing algorithm for generating new malicious traffic instances based on the traffic flows induced by the malicious samples produced in the Cuckoo system. The workflow and pseudo code of the proposed algorithm, designated as the GAN-BAL algorithm, are presented in Fig. 3. After the generative model in the GAN generates traffic, the generated traffic is sent to the discriminant model to get its discriminating result of the forged traffic. The generative model will improve its generative strategy through the discriminating results of the

discriminant model. Then, the generative model will use the new generation strategy to generate traffic, and send the generated traffic to the intrusion detection system to judge the authenticity of the forged traffic. The discriminant model will improve its discriminative strategy through the judgment results of the intrusion detection system. After the above steps are repeated, the generative model iteratively improves its "fake" ability from the feedback from each discriminative model.

## 4. Experimental Result

This section describes the effectiveness of the proposed GAN-BAL algorithm in improving the performance of five common machine learning algorithms.

The GAN-BAL algorithm was first applied to address the imbalance problem in the CICIDS 2017 dataset. The original CICIDS dataset and balanced CICIDS dataset were then used to train and evaluate five common supervised algorithms, namely CNN, LSTM, BAT, SVM and NB. The results showed that the balanced CICIDS dataset increased the recall and accuracy rates of the CNN model by 20% and 4%, respectively. Similarly, for the LSTM model, the recall and accuracy increased by 10% and 2%. Finally, for the BAT, SVM, and NB models, the recall and accuracy increased by 16% and 3%, 15% and 4%, and 1% and 1%, respectively. However, for the CNN, LSTM, and BAT models, the precision rate decreased by 3%, 5%, 3%. For SVM and NB models, the precision rate remains the same.

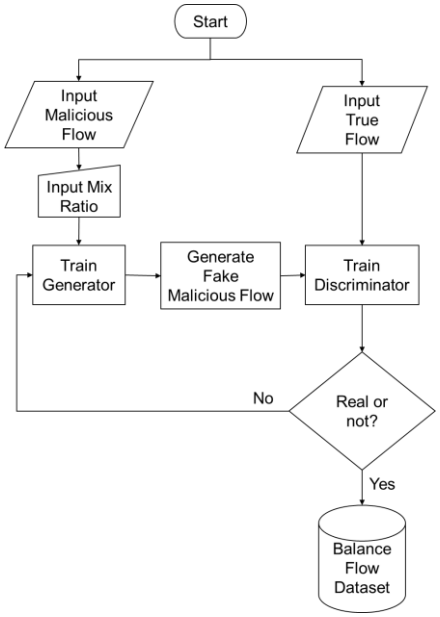


Fig. 3. Flowchart of GAN-BAL algorithm.

The original GAN algorithm is also used in the above experiments to further compare the performance of the GAN-BAL algorithm with that of the original GAN algorithm. As shown in Fig. 4, the GAN-BAL algorithm improved the recall rate, precision, F1-Score and accuracy rate by 3% in every case.

Overall, the experimental results show that although the GAN-BAL algorithm causes a loss of precision, the precision is still better than using the original GAN algorithm. In other words, the results confirm the effectiveness of the proposed GAN-BAL algorithm in improving the malicious intrusion detection performance of supervised machine learning models.

## 5. Conclusion

The performance of IDS machine learning models is closely related to the datasets used for training and evaluation purposes. In particular, existing public IDS datasets typically contain an imbalance problem, wherein the number of malicious samples is significantly lower than that of normal samples. As a result, the classification accuracy and robustness of the trained models are seriously impaired. To address this problem, the present study has proposed a GAN-BAL algorithm for artificially increasing the proportion of malicious samples in the IDS dataset, thereby improving the modeling performance.

In contrast to the IGAN-IDS system proposed by Huang et al. [12], which takes noise as the input to the generative model, the GAN-BAL algorithm proposed in the present study takes malicious traffic induced by malicious samples as the input and considers a full range of

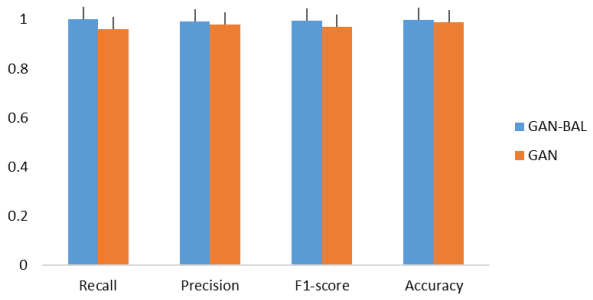


Fig. 4. Comparison of evaluation indicators between GAN-BAL algorithm and GAN algorithm.

malicious attack behaviors, including both traditional and modern.

The effectiveness of the GAN-BAL algorithm in improving the modeling performance of supervised machine learning algorithms has been confirmed via experiments with five common supervised models, namely Naïve Bayes, CNN, LSTM, BAT, and SVM.

When developing a system capable of IDS, whether it can operate in real-time will affect the effectiveness of network protection. Future work will combine the GAN-BAL algorithm with online learning, which can reduce the model construction time and shorten the time for machine learning model to detect malicious behavior.

## 6. Acknowledgements

This work was supported by the National Science and Technology Council in Taiwan under contract numbers 108-2221-E-006-110-MY3 and 111-2218-E-006-010-MBK.

## 7. References

1. NSA, CISA, FBI, & NCSC, "Russian GRU Conducting Global Brute Force Campaign to Compromise Enterprise and Cloud Environments," 15 4 2021. [Online]. Available: https://media.defense.gov/2021/Jul/01/2002753896/-1/-1/1/CSA_GRU_GLOBAL_BRUTE_FORCE_CAMPAIGN_UOO158036-21.PDF. [Accessed 2 7 2021].
2. M. Ring, S. Wunderlich, D. Scheuring, D. Landes and A. Hotho, "A survey of network-based intrusion detection data sets," Computers & Security, vol. 86, pp. 147-167, Sep. 2019.
3. Y. Zhao, G. Cheng, Y. Duan, Z. Gu, Y. Zhou and L. Tang, "Secure IoT edge: Threat situation awareness based on network traffic," Computer Networks, vol. 201, p. 108525, 2021.
4. LINCOLN LABORATORY, "1998 DARPA INTRUSION DETECTION EVALUATION DATASET," 2 1998. [Online]. Available: https://www.ll.mit.edu/r-d/datasets/1998-darpa-intrusion-detection-evaluation-dataset. [Accessed 1 6 2021].
5. M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, Canada, 8-10 July, 2009.
6. N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data set for Network Intrusion Detection systems," 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia, 10-12 Nov., 2015.
7. I. Sharafaldin, A. H. Lashkari, A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," ICISSP, pp. 108-116, 22-24 Jan 2018.
8. Brunswick, University of New, "Canadian Institute for Cybersecurity," Brunswick, University of New, 1 7 2017. [Online]. Available: https://www.unb.ca/cic/research/applications.html#CICFlowMeter. [Accessed 1 6 2021].
9. Amirhossein Gharib, Iman Sharafaldin, Arash Habibi Lashkari, A. Ghorbani, "An Evaluation Framework for Intrusion Detection Dataset," 2016 International Conference on Information Science and Security (ICISS), Pattaya, Thailand, 19-22 Dec., 2016.
10. TWISC@NCKU, "Taiwan Information Security Center," TWISC@NCKU, 1 4 2006. [Online]. Available: https://www.twisc.ncku.edu.tw/home. [Accessed 2 5 2021].
11. P. Wang, Z. Wang, F. Ye and X. Chen, "ByteSGAN: A semi-supervised Generative Adversarial Network for encrypted traffic classification in SDN Edge Gateway," Computer Networks, vol. 200, p. 108535, 2021.
12. S. Huang and K. Lei, "IGAN-IDS: An imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks," Ad Hoc Networks, vol. 105, p. 102177, 2020.

## Authors Introduction

Dr. I-Hsien Liu

He is a research fellow in the Taiwan Information Security Center @ National Cheng Kung University (TWISC@NCKU) and Department of Electrical Engineering, National Cheng Kung University, Taiwan. He obtained his PhD in 2015 in Computer and Communication Engineering from the National Cheng Kung University. His interests are Cyber-Security, Wireless Network, Group Communication and Reliable Transmission.

Mr. Cheng-En Hsieh

He received his B.S. degree from the Department of Communication Engineering, National Central University, Taiwan in 2019. He got the M.S. degree in National Cheng Kung University in Taiwan. His research focuses on network communication and cyber security.

Ms. Wei-Min Lin

She received her B.S. degree from the Department of Electrical Engineering, Yuan Ze University, Taiwan in 2020. She is acquiring the master's degree in Department of Electrical Engineering / Institute of Computer and Communication Engineering, National Cheng Kung University in Taiwan.

Prof. Jung-Shian Li

He is a full Professor in the Department of Electrical Engineering, National Cheng Kung University, Taiwan. He graduated from the National Taiwan University, Taiwan, with B.S. in 1990 and M.S. degrees in 1992 in Electrical Engineering. He obtained his PhD in 1999 in Computer Science from the Technical University of Berlin, Germany. He teaches communication courses and his research interests include wired and wireless network protocol design, network security, and network management. He is the director of Taiwan Information Security Center @ National Cheng Kung University. He serves on the editorial boards of the International Journal of Communication Systems.

Prof. Chu-Fen Li

She is an Associate Professor in the Department of Finance at the National Formosa University, Taiwan. She received her PhD in information management, finance and banking from the Europa-Universität Viadrina Frankfurt, Germany. Her current research interests include intelligence finance, e-commerce security, financial technology, IoT security management, as well as financial institutions and markets. Her papers have been published in several international refereed journals such as European Journal of Operational Research, Journal of System and Software, International Journal of Information and Management Sciences, Asia Journal of Management and Humanity Sciences, and others.