CHAPTER 25

A Generative Adversarial Network (GAN) Solution for Synthetically Generated Botnet Attack Data Samples

Nikolaos Peppes, Theodoros Alexakis, Emmanouil Daskalakis, Evgenia Adamopoulou, and Konstantinos Demestichas

Introduction

The widespread adoption of digital services in people's daily lives has resulted in an increased demand for cybersecurity. With the proliferation of new software and hardware, detecting known botnets or other types of

N. Peppes (✉) • T. Alexakis • E. Daskalakis • E. Adamopoulou
Institute of Communication and Computer Systems, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece
e-mail: npeppes@cn.ntua.gr; talexakis@cn.ntua.gr; edaskalakis@cn.ntua.gr; eadam@cn.ntua.gr

K. Demestichas
Department of Agricultural Economics and Rural Development, Agricultural University of Athens, Athens, Greece
e-mail: cdemest@aua.gr

I. Gkotsis et al. (eds.), *Paradigms on Technology Development for Security Practitioners*, Security Informatics and Law Enforcement, https://doi.org/10.1007/978-3-031-62083-6_25

attacks has become a daunting task for cybersecurity professionals. Botnets as one type of cyberattack can have disastrous consequences [1, 2], as they allow attackers to remotely control infected machines, since they have the potential to impact numerous devices in parallel, particularly within IoT networks, due to a large number of devices interconnected.

Cybersecurity incidents are predominantly addressed reactively, subsequently to the occurrence of an attack, necessitating the engagement of cybersecurity professionals to respond and mitigate resultant damage. To combat these infections, cybersecurity experts are developing proactive systems that utilize machine-learning and deep learning (ML & DL) technologies. Consequently, the primary dataset for cybersecurity analysis predominantly comprises historical attack data. This essentially implies that nearly all cybersecurity systems are developed based on historical attack patterns, rendering them susceptible to emerging variants. Nonetheless, many organizations refrain from sharing their attack data, resulting in a scarcity of such information, consequently hindering the effective training of ML or DL models and the development of such systems.

The current study proposes a methodology for generating botnet-type data in a tabular format. This methodology employs an 8-layer generative adversarial network (GAN) model [3] to evaluate its effectiveness in generating synthetic data with high precision while minimizing computational expenses The generated samples will be assessed using a wide range of graphical data quality indicators, such cumulative sums, absolute log mean and STD diagrams, correlation matrices, and heatmaps.

The remainder of this study contains Section "Related Works" that investigates related research on botnet attack generation techniques, Section "BNGAN: A Proposed Solution For Addressing the Data Issue" that provides a more in-depth explanation of the GAN model design methodology, whereas Section "Data Generation Results Evaluation" focuses on evaluating the synthetic dataset's significance and Section "Conclusion and Future Work" examines the revealed discoveries.

Related Works

The escalating damage inflicted on computer systems by botnet attacks has underscored the imperative need to delve deeper into detection methods. Consequently, a plethora of studies in this domain can be found in the existing literature. Within the context of our study's core elements outlined earlier, the works discussed in this section primarily

concentrate on two key aspects: the generation and classification of botnet attack datasets.

Yin et al. [4] concentrated on augmenting botnet detection. Their research introduced a GAN designed to generate nearly lifelike botnet attack samples, enhancing the training of machine-learning classifiers. The Bot-GAN consistently supplied "synthetic" data to the discriminator, which classified these samples using a softmax function. This approach resulted in improved accuracy and precision when compared to pretrained models utilizing the original imbalanced dataset. Pursuing a similar route to mitigate the challenges posed by imbalanced datasets, Song et al. [5] introduced the GAN-efficient lifelong learning algorithm (ELLA) solution. Their methodology demonstrated that dataset expansion through a GAN architecture not only boosted the performance of traditional ML solutions for botnet identification but also enhanced the lifelong learning approach of the ELLA algorithm.

Tram Truong-Huu and his team [6] investigated the application of GANs in network anomaly detection. They employed multiple datasets to assess GANs' performance in comparison with other network anomaly detection methods. Their experiments revealed significant improvements over existing deep learning techniques, indicating promise in detecting unknown anomalous behavior and zero-day attacks focusing on botnet traffic.

Zhong et al. [7] introduced MalFox, a solution designed to demonstrate the limitations of existing black box detectors. MalFox employs a convolutional GAN and adopts a confrontational strategy to create perturbation paths. These paths incorporate up to three methods (Obfusmal, Stealmal, and Hollowmal) to generate adversarial malware examples. Their results showed promising performance, with an accuracy of approximately 99%, while the detection rate of the generated samples was at a lower percentage, around 45%.

The significance of GANs for data augmentation, especially in the cybersecurity realm, was underscored also by Habibi et al.'s Conditional Tabular GAN (CTGAN) model [8]. They experimented with various CTGAN versions and parameters to identify the most effective one. The outcomes demonstrated CTGAN's ability to preserve the structure of both continuous and discrete data. This provided a solution for ML classifiers or detectors, addressing dataset imbalances and training these algorithms for novel threats, given that GAN-generated data are novel and unseen.

Lingam et al. [9] conducted a study on imbalanced data concerning bot identification. Their objective was to tackle the issue of imbalanced data for ML classifiers by employing a GAN with a gated recurrent unit (GRU). This enabled them to generate synthetic data closely resembling real data, effectively balancing benign user and bot classes. Results indicated that their approach outperformed ML methods trained solely on the original Twitter dataset, achieving an average accuracy of approximately 91% with the GAN-generated dataset.

BNGAN: A Proposed Solution for Addressing the Data Issue

Generative adversarial networks (GANs) [1] utilize an architecture that generates new data based on input data and random noise. GANs consist of two components: the generator and discriminator. The generator uses random noise to create realistic data, while the discriminator classifies input samples as either real or fake. Both components are optimized based on the discriminator's ability to accurately classify real and fake data.

Hence, there is significant importance in conducting experiments involving various GAN architectures and adjusting their hyperparameters to discover the most suitable model tailored to a particular dataset and objective. Approaches like hyperparameter optimization and architectural exploration serve as valuable tools in pinpointing the ideal GAN structure and hyperparameters tailored to a specific task.

This study aims to evaluate the effectiveness of a proposed 8-layer GAN architecture called BNGAN in generating synthetic data that accurately represent malicious cyber-attacks, specifically botnet attacks [3]. To accomplish this, the study evaluates the performance of the proposed 8-layer GAN model [10] for both the generator and discriminator, using the CTU-13 dataset [11] from the Stratosphere IPS. This dataset includes captures of diverse malware samples and normal traffic, with 32 million packets. The training dataset has 216,352 records, with 140,849 marked as "0" for malware and 75,503 labeled as "1" for legitimate. The evaluation dataset has 88,258 records without any labels.

The study utilizes the BNGAN model architecture, which is designed to generate 1D synthetic data from the input dataset. The model was implemented using Tensorflow 2.0 and Keras API. The proposed BNGAN architecture utilizes the sequential API to stack the different layers of the

deep neural network. The generator model is built using the sequential API and consists of an input layer for accepting appropriately scaled, randomly generated noise with the intended size. This input is then processed through six subsequent hidden layers utilizing the "ReLU" activation function, ultimately leading to an output layer. This output layer employs a "linear" activation function, aligning its dimension with that of the preprocessed dataset.

The discriminator, by itself, takes the form of a sequential model, composed also of eight dense layers. In the initial seven layers, the "ReLU" activation function is utilized, while the last layer employs the "sigmoid" function to classify input samples as either authentic (genuine) or counterfeit (malware). To bolster the model's precision, a 20% dropout rate is applied to both the visible (input) layer and the six concealed layers within the discriminator model. The ultimate choice of this dropout rate was reached through a series of iterative experiments, considering its influence on preventing overfitting while ensuring the model's capability to capture pertinent data patterns.

After detailing the generator and discriminator models, the proposed BNGAN model is characterized as a sequential model that integrates these components in an adversarial manner. Figure 25.1 illustrates how the BNGAN model uses (preprocessed) botnet data samples to generate synthetic, tabular data.

Data Generation Result Evaluation

In the previous chapter, the generator and discriminator models were established, combining them to form the comprehensive BNGAN model. Subsequently, the training process was initiated to facilitate the generation of datasets mirroring the originals. The training process encompassed a total of 1000 epochs, with each epoch involving batch training of a predefined size for both the generator and discriminator networks. In this process, the discriminator received as input a predetermined batch of data samples from the original dataset as well as the generated output data sample from the generator. For each (data) batch, the discriminator computed the loss for both the genuine and the generated data. The losses computed (by the discriminator) served to refine the predictions made by the discriminator model, subsequently enabling the computation of generator losses and gradients via backpropagation techniques. In this iterative process, the generator persists in enhancing the quality of the

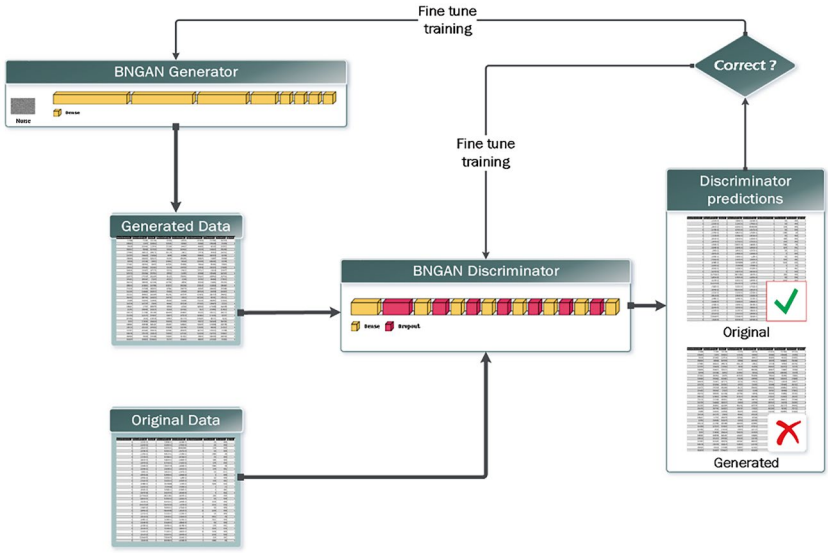

Fig. 25.1 BNGAN Model Implementation

synthetically generated data samples by constantly adjusting its weights based on these gradients.

Visual representations, such as diagrams, prove to be an effective way of assessing and illustrating the similarity between datasets (synthetically) generated by a GAN model and real data. These similarity estimation methods offer valuable insights into the fidelity and precision of the generated dataset, aiding researchers in pinpointing areas where enhancements to the GAN model might be necessary to produce synthetic data that closely mirrors real data. Furthermore, the GAN model's performance in creating synthetic data that closely resembles the real data can be determined. The choice of diagram types is contingent upon the nature of the analyzed data and the particularly considered objectives of the research. The current study includes the following diagrams to evaluate the generated data: correlation matrices with heatmaps, highlighting clusters illustrating distinctions between the real and generated datasets, cumulative sum (cumsum) diagrams for visualizing the accumulation over time of the original and the generated data and STD diagrams to compare the (similarity) scores between the original and generated datasets from the GAN

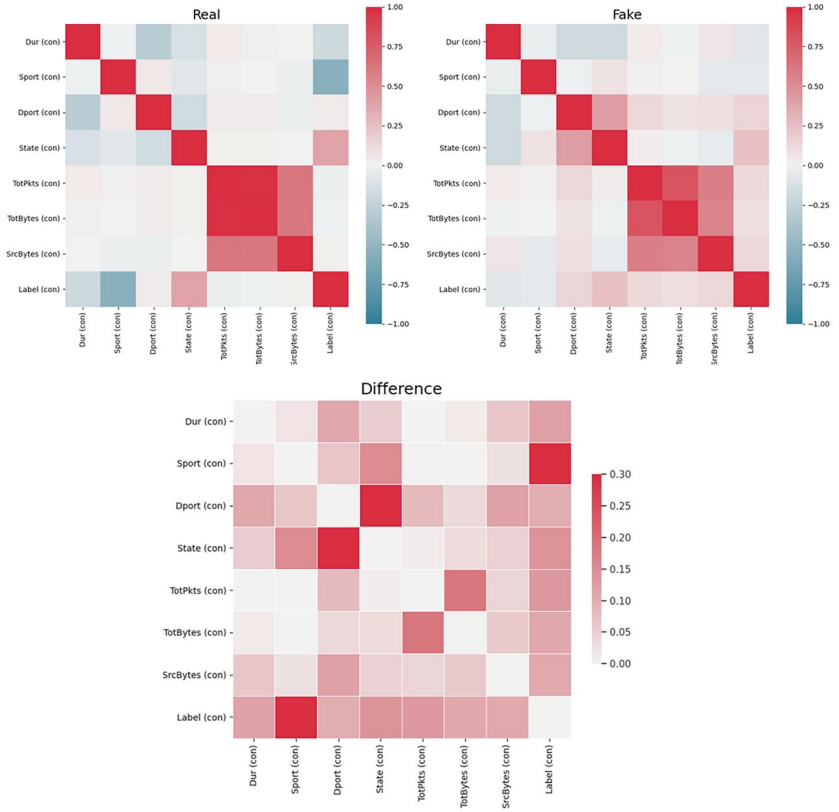

Fig. 25.2 Correlation Matrices with Heatmap

model. Figures 25.2, 25.3, and 25.4 visualize the comparison results between the original and the generated data from the GAN model.

Based on the results depicted previously, the cumulative sum diagrams reveal notable insights regarding the similarity between real and generated datasets for eight variables. Five of these variables (Dur, TotPkts, TotBytes, SrcBytes, and Label) exhibit a consistent, steadily increasing similarity score in both datasets, suggesting a continuous pattern. In contrast, the remaining three variables (Sport, Dport, and State) display a fluctuating pattern with abrupt spikes and drops in the similarity score, indicating deviations in the synthetic dataset. These fluctuations suggest certain data

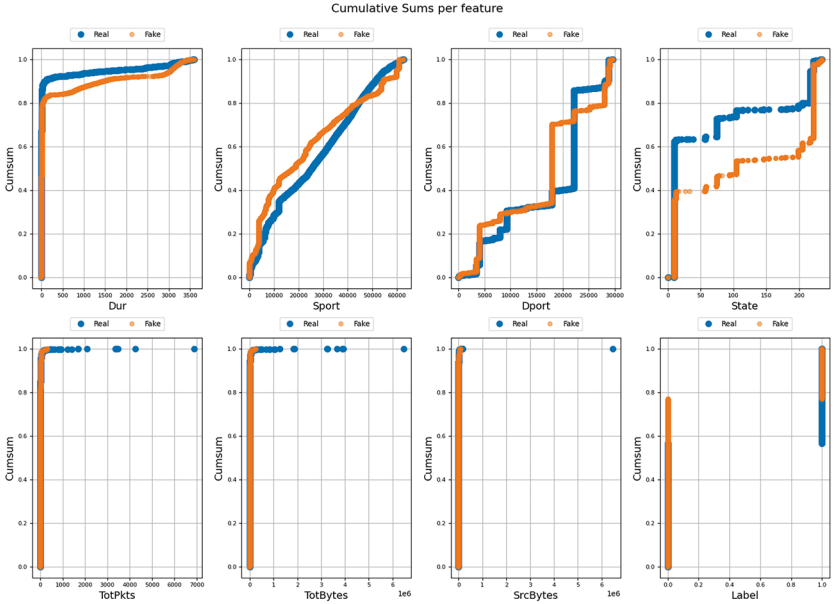

Fig. 25.3 Cumulative Sum Diagrams

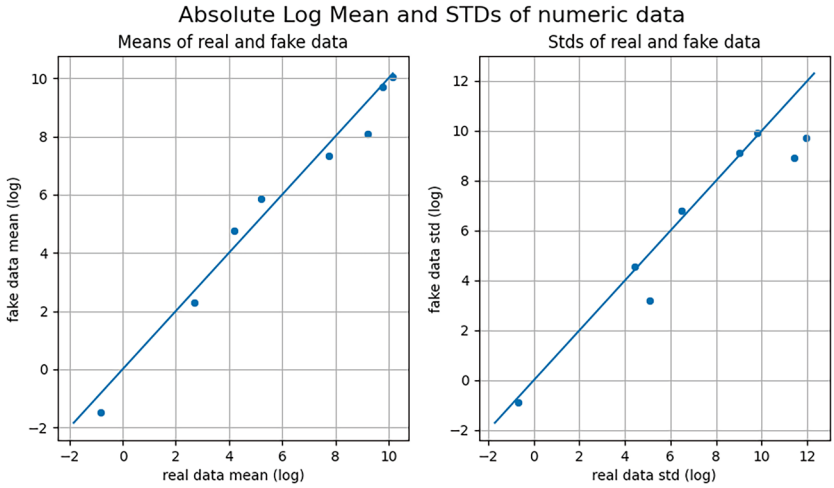

Fig. 25.4 STD Diagrams

points significantly diverge from the overall pattern, contributing to lower overall similarity scores for these variables. Furthermore, the cumulative sum diagrams suggest that the GAN model might require more training epochs to produce a synthetic dataset closely in order to resemble the real one. Moving to the correlation matrix diagrams, the real dataset illustrates a strong positive correlation among its variables. However, in the generated synthetic dataset, the positive correlations are weaker, and no significant negative correlations emerge. Additionally, a strong positive correlation arises between the various features in the "Difference" section, signifying that as the epochs progress, the generated data faithfully replicates patterns and characteristics from the real dataset in a realistic manner. Finally, examining the absolute mean and standard deviation diagrams reveals that the synthetic dataset contains higher values for certain features compared to the real dataset. This variance may suggest that the generated data for some features could not precisely mirror the real data, at least initially. However, as the number of training epochs increases, the synthetic dataset progressively aligns more closely with the real dataset provided.

Conclusion and Future Work

As digital tools continue to evolve and become more prevalent, the need for effective cybersecurity measures has become increasingly critical. The primary objective of this study is to outline a comprehensive methodology for generating synthetic data for botnet attacks using generative adversarial networks (BNGAN). The generation process utilizes an open-source dataset, the CTU-13 dataset, provided by Stratosphere IPS, which is a collection of network traffic captures that have been widely used in the field of cybersecurity research. This tabular format data is used as input for the suggested BNGAN architecture [11]. The BNGAN model generates over 200,000 new botnet data samples that closely resemble the original data. Subsequently, the generated botnet data samples are evaluated using a wide range of graphical data quality indicators, including cumulative sums, absolute log mean and STD diagrams and correlation matrices with heatmaps, to assess the quality of the generated data. Overall, this proposed methodology provides a promising approach to improving botnet attack detection and prevention. The future prospect of this research involves expanding data categories and domains into various fields, encompassing diverse data formats and addressing a broader range of

cyberthreats. Furthermore, an important avenue of exploration is the integration of lifelong learning techniques, both for data generation and the zero-day detection and classification of such attacks.

References

1. Shinan, K., Alsubhi, K., Alzahrani, A., & Ashraf, M. U. (2021). Machine learning-based Botnet detection in software-defined network: A systematic review. *Symmetry, 13*. https://doi.org/10.3390/sym13050866
2. Check point check point research reports a 38% increase in 2022 global cyberattacks Available online: https://blog.checkpoint.com/2023/01/05/38-increase-in-2022-global-cyberattacks/. Accessed on 22 July 2023.
3. Randhawa, R. H., Aslam, N., Alauthman, M., Rafiq, H., & Comeau, F. (2021). Security hardening of Botnet detectors using generative adversarial networks. *IEEE Access, 9*, 78276–78292. https://doi.org/10.1109/ACCESS.2021.3083421
4. Yin, C., Zhu, Y., Liu, S., Fei, J., & Zhang, H. (2018). An Enhancing framework for botnet detection using generative adversarial networks. In *Proceedings of the 2018 international conference on artificial intelligence and Big Data (ICAIBD)* (pp. 228–234).
5. Song, C., Wushouer, M., & Tuerho, G. (2022). Botnet detection based on generative adversarial network and efficient lifelong learning algorithm. In *Proceedings of the 2022 international conference on Big Data, information and computer network (BDICN)* (pp. 48–54).
6. Truong-Huu, T., Dheenadhayalan, N., Pratim Kundu, P., Ramnath, V., Liao, J., Teo, S. G., & Praveen Kadiyala, S. (2020). An empirical study on unsupervised network anomaly detection using generative adversarial networks. In *Proceedings of the proceedings of the 1st ACM workshop on security and privacy on artificial intelligence* (pp. 20–29). Association for Computing Machinery: New York, NY, USA.
7. Zhong, F., Cheng, X., Yu, D., Gong, B., Song, S., & Yu, J. (2023). MalFox: Camouflaged adversarial malware example generation based on Conv-GANs against Black-Box detectors. *IEEE Transactions on Computers*, 1–14. https://doi.org/10.1109/TC.2023.3236901
8. Habibi, O., Chemmakha, M., & Lazaar, M. (2023). Imbalanced tabular data modelization using CTGAN and machine learning to improve IoT Botnet attacks detection. *Engineering Applications of Artificial Intelligence, 118*, 105669. https://doi.org/10.1016/j.engappai.2022.105669
9. Lingam, G., Yasaswini, B., Jagadamba, P. V. S. L., & Kolliboyana, N. (2022). An improved Bot identification with imbalanced data using GG-XGBoost. In

Proceedings of the 2022 2nd international conference on intelligent technologies (CONIT) (pp. 1–6).

10. Peppes, N., Alexakis, T., Adamopoulou, E., & Demestichas, K. (2023). The effectiveness of zero-day attacks data samples generated via GANs on deep learning classifiers. *Sensors, 23*. https://doi.org/10.3390/s23020900
11. García, S., Grill, M., Stiborek, J., & Zunino, A. (2014). An empirical comparison of Botnet detection methods. *Computers & Security, 45*, 100–123. https://doi.org/10.1016/j.cose.2014.05.011