基于 CWGAN-GP 平衡化的网络恶意流量识别方法

丁要军*，王安宙

(甘肃政法大学网络空间安全学院 兰州 730070)

【摘要】在网络恶意流量识别任务中，存在恶意流量样本数量与正常流量样本比例不平衡问题，从而导致训练出的机器学习模型泛化能力差、识别准确率低。为此，在网络流量图片化的基础上提出一种利用具有梯度惩罚项的条件 Wasserstein 生成对抗网络 (CWGAN-GP) 对少量数据类进行平衡的分类方法。该方法首先借助网络流量图片化方法将原始流量 PCAP 数据按照流为单位进行切分、填充、映射到灰度图片中；然后使用 CWGAN-GP 方法实现数据集的平衡；最后，在公开数据集 USTC-TFC2016 和 CICIDS2017 上使用 CNN 模型对不平衡数据集和平衡后的数据集进行分类测试。实验结果表明，使用 CWGAN-GP 的平衡方法在精确度、召回率、F1 这 3 个指标上均优于随机过采样、SMOTE、GAN 以及 WGAN 平衡方法。

关 键 词 条件 Wasserstein 生成对抗网络； 数据平衡； 流量扩充； 流量识别

中图分类号 TN915.08; TP181 **文献标志码** A doi:10.12178/1001-0548.2022011

Network Malicious Traffic Identification Method Based on CWGAN-GP Category Balancing

DING Yaojun* and WANG Anzhou

(School of Cyberspace Security, Gansu University of Political Science and Law Lanzhou 730070)

Abstract In the network malicious traffic identification task, there is an imbalance between the ratio of the number of malicious traffic samples and the number of normal traffic samples, which leads to poor generalization ability and low recognition accuracy of the trained machine learning model. To solve this problem, this paper proposes a classification method that balances a small number of data classes by using the conditional Wasserstein generative adversarial network (CWGAN-GP) with gradient penalty items based on the visualization of network traffic. This method first uses the network traffic visualization method to segment, fill, and map the original traffic packet capture (PCAP) data into gray-scale images according to the flow as a unit, and then applies the CWGAN-GP method to achieve the balance of the dataset. Finally, in the public dataset USTC-TFC2016 and CICIDS2017, the convolutional neural network (CNN) model is used to classify and test the unbalanced dataset and the balanced dataset. The experimental results show that the balance method using CWGAN-GP is better than the random oversampling, SMOTE, GAN and WGAN balance methods in the three indicators of Precision, Recall, and F1.

Key words CWGAN-GP; data balance; flow expansion; flow identification

在机器学习和数据挖掘领域，恶意流量不平衡是一种普遍存在的现象。目前对恶意流量不平衡数据分类的研究主要涉及两种方法[1]：一是对分类算法进行新的设计或改进；二是在数据级别上进行操作，主要通过过采样、欠采样或混合采样来达到数据集的平衡。现有的过采样方法主要分为传统方法和基于生成对抗网络的方法。大多数传统的过采样方法都基于 SMOTE 技术[2]，由于这些方法更关注局部数据信息，生成的数据不够真实；另一种是基于生成性对抗网络 (generative adversarial networks, GAN)[3]，它可以捕捉数据的真实分布并直接生成合成数据，用于无监督学习。随后，文献 [4] 提出了一种基于条件生成对抗网络 (conditional generative adversarial networks, CGAN) 的模型，该模型在 GAN

收稿日期：2022－01－06；修回日期：2022－02－24
基金项目：甘肃省高等学校产业支撑计划 (2020C-29)
作者简介：丁要军 (1980－)，男，博士，教授，主要从事网络安全及机器学习等方面的研究.
*通信作者：丁要军，E-mail：dingyj80@163.com

的基础上加入了条件信息生成特定的类别。然而，GAN 与 CGAN 都会受到不稳定训练 (梯度消失) 和模式崩溃的影响。为解决这些问题，WGAN (Wasserstein generative adversarial networks)[5] 随即被提出，该模型使用 EM 距离 (也称为 Wasserstein) 而不是 JS 散度来度量真实数据分布和生成数据分布之间的距离，从理论上解决了梯度消失的问题，可有效缓解模式崩溃问题。但 CGAN 与 WGAN 依然存在着训练只生成质量不高的样本或模型无法收敛。在使用 GAN 模型解决网络恶意数据分类问题中，文献 [6] 提出了 GAN 模型生成具备可执行性和攻击性的恶意网络流样本，但它只针对缓冲区溢出漏洞攻击进行了研究，模型的泛化性有待进一步研究。文献 [7] 提出了 Attack-GAN 模型，用于生成能够敞开入侵检测系统的数据包级别对抗网络流量。文献 [8] 提出将 WGAN-GP 与 CGAN 优势融合的 CWGAN-GP(conditional Wasserstein generative adversarial network-gradient penalty) 方法，该方法不仅生成更真实多样的数据，而且克服了模式崩溃和训练不稳定的问题。

本文将 WGAN-GP 与辅助条件信息相结合，作为一种新的过采样方法，为网络流量图片不平衡数据集中的少数类生成合成样本。

1　CWGAN-GP 原理

与其他 GAN 模型及其衍生版本类似，CWGAN-GP 也是由鉴别器 (discriminator) 和生成器 (generator) 组成，它们以相反的方式训练，通过博弈使得生成的样本与真实数据无法区分。CGAN 模型在原始 GAN 模型上添加辅助条件信息，WGAN-GP 模型以同样的方式扩展到 CWGAN-GP。CWGAN-GP 采用 EM 距离来评估真实样本和模拟样本之间的分布，且加入了条件信息。Wasserstein 距离为：

$$\mathrm{W}(p_{\mathrm{data}},p_{\mathrm{g}})=\inf_{\gamma\in\prod(p_{\mathrm{data}},p_{\mathrm{g}})}\mathrm{E}_{(x,y)\sim\gamma}[\|x-y\|] \tag{1}$$

式中，P_{data}、P_{g} 是真实数据分布和生成数据分布；$\prod(p_{\mathrm{data}},p_{\mathrm{g}})$是所有边缘分布为 p_{data} 和 p_{g} 的联合概率分布。

CWGAN-GP 是通过惩罚鉴别器相对于其输入梯度的范数来替代 WGAN 削减权重，有效克服了 WGAN 模型难以收敛的问题。

为限制生成器生成方向，使生成器能够生成特定类别数据，与 CGAN 类似，CWGAN-GP 模型向生成器和鉴别器添加了附加信息 y，y 可以是类别标签或任何其他类型的辅助信息。本文中条件信息是类别标签。在鉴别器中，将 P_{data}、P_{g}和 y 以联合隐藏表达的形式结合；在生成器中，以相同的形式将条件 y 与生成数据分布 p_{g} 连接。函数形式为：

$$\min_G\max_D V(D,G)=\mathrm{E}_{x\sim p_{\mathrm{data}}(x)}[D(x|y)]-\mathrm{E}_{\tilde{\mathrm{g}}\sim p_{\mathrm{g}}(\mathrm{g})}[D(\tilde{g}|y)]-\lambda\mathrm{E}_{\hat{x}\sim \mathrm{P}_{\hat{x}}}[(\|\nabla_{\hat{x}}D(\hat{x}|y)\|_2-1)^2] \tag{2}$$

其中参数与 WGAN-GP 相同，只是增加了附加条件 y。CWGAN-GP 的优化函数分别为：

$$L(D)=-\mathrm{E}_{x\sim p_{\mathrm{data}}(x)}[D(x|\mathrm{y})]+\mathrm{E}_{\tilde{\mathrm{g}}\sim p_{\mathrm{g}}(\mathrm{g})}[D(\tilde{g}|\mathrm{y})]+\lambda\mathrm{E}_{\hat{x}\sim \mathrm{P}_{\hat{x}}}[(\|\nabla_{\hat{x}}D(\hat{\mathrm{x}}|\mathrm{y})\|_2-1)^2] \tag{3}$$

$$L(G)=-\mathrm{E}_{\tilde{\mathrm{g}}\sim p_{\mathrm{g}}(\mathrm{g})}[D(\tilde{g}|\mathrm{y})] \tag{4}$$

CWGAN-GP 的目标是使 L 最小化[9]，从而实现生成数据与真实数据之间的分布距离更小。与传统的过采样方法相比，CWGAN-GP 直接生成数据，而不只是关注局部信息。

2　基于 CWGAN-GP 的流量图片数据平衡方法

由于深度学习模型对输入数据格式有具体要求，本节介绍数据集的预处理方法，将原始流量数据进行图片化处理，在此基础上对数据扩充平衡。

2.1　网络流量图片化

基于文献 [10] 对数据处理的经验，本节的网络流量图片化主要是使用工具集 USTC-TK2016 将原始流量数据 (PCAP 或 PCAPNG 格式) 处理成 IDX 数据集格式 (更好的迁移到多数模型)。

1) 流量切分：按照流量表示形式将原始 PCAP 文件按照数据流形式切分为多个 PCAP 文件，本数据流是具有相同五元组信息的数据包的时间排序集合。

2) 图片生成：将处理过的文件按照 784 字节进行统一长度处理，即保留文件前 784 字节数据，舍弃文件 785 字节以后的所有信息，如果长度少于 784 字节，则在文件后面补充 0x00；统一长度后的文件按照二进制形式转换为灰度图片，即一个字节对应灰度像素值，如 0x00 对应黑色，0xff 对应白色，输出格式为 PNG。

3) 数据集格式转换：将生成的多类别图像转换为包含图像像素信息和统计信息的 IDX 格式，方便后续数据扩展和分类模型输入。

完成以上处理操作后，每张图片都是 28*28 的灰度图片，这些图片类别之间有良好的区分度，保证了深度学习模型分类有很好的效果。

2.2 流量图片数据平衡

将预处理完成的数据放入 CWGAN-GP 图像生成器中，利用 CWGAN-GP 能稳定地生成多样样本的特点产生新的少数类别流量图片。生成数据是具有真实图像特征且多样性较强的扩充数据，利用这些数据对原始不平衡数据集进行数据扩充，不仅能有效平衡数据集，还能防止像传统上采样技术扩充数据集所造成的数据样本单一、训练模型容易过拟合的问题。CWGAN-GP 网络的鉴别器损失是与生成图片质量高度相关的参数，在生成阶段加入一个判别函数，在鉴别器、生成器损失小于某阈值时输出生成图像，保证生成的图像与原始真实图像的高度相关性。平衡方法如图 1 所示。

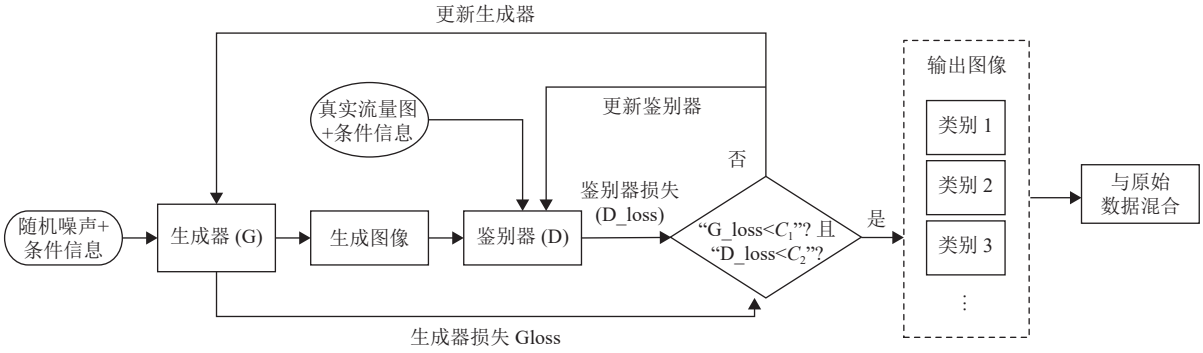


图 1 基于 CWGAN-GP 数据平衡方法

1) 将需要扩充的少数类真实流量图制作成 IDX 数据集格式，输入进 CWGAN-GP 模型中进行训练。

2) 在生成器损失小于 C_1 且鉴别器损失小于 C_2 时，将生成器骗过鉴别器的图像按类别输出。本文的 C_1、C_2 绝对值大小分别为 1.0 和 0.2，在大量训练情况下记录生成器、鉴别器损失大小变化，如图 2 所示，C_1、C_2 是在考虑时间效率下相对收敛的损失值大小。

3) 将生成的流量图片与原始流量图片进行合，完成原始数据集的平衡。

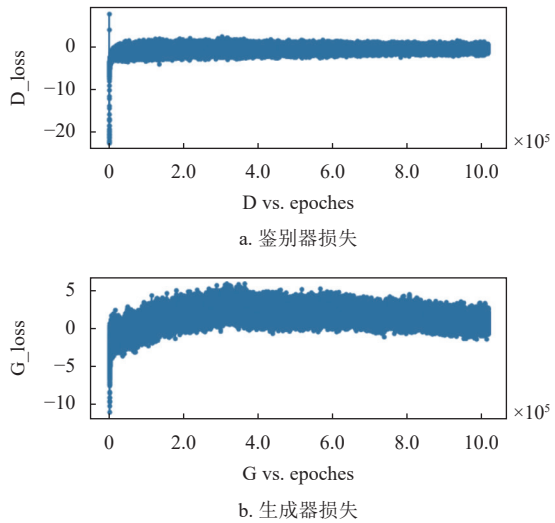


图 2 生成器、鉴别器损失变化

平衡完成的数据是和原始流量图片高度相关且具有多样性，如图 3 所示，可以看出利用此方法可以生成人眼均可明确分辨的相关图像，且具有一定的多样性。生成数据相关程度在实验结果中得到验证。

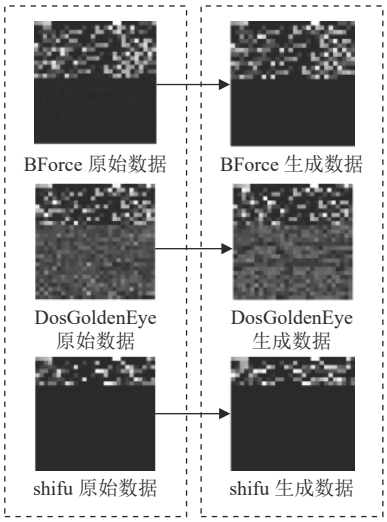


图 3 生成流量图与真实流量图对比

3 实验结果分析

3.1 不平衡数据集

实验使用的是公共数据集 USTC-TFC2016 和 CIC-IDS2017[11] 中的部分数据，数据集由原始 PCAP

文件组成。从 USTC-TFC2016 中选择了 9 类应用程序，其中包括 6 类正常应用流量和 3 类异常应用流量；从 CIC-IDS2017 选择 10 类应用程序，9 类异常应用流量和 1 类正常流量，数据集均有较大的不平衡。使用不同方法使数据集实现平衡，将数据样本随机划分成 90% 的训练集和 10% 的测试集。数据集平衡前后分布如表 1 和表 2 所示。为保证实验有效性，数据集的测试集部分均为真实数据，生成数据只对训练集部分进行平衡。

表 1　USTC-TFC2016 数据集平衡前后分布

编号	应用类别	平衡前		平衡后	
		数量	占比/%	数量	占比/%
1	Ftp(正常)	51 689	41.6%	13 000	11%
2	MySql(正常)	11 900	9.5%	13 000	11%
3	SMB(正常)	13 800	11.1%	13 000	11%
4	Wordofwarcraft(正常)	14 185	11.4%	13 000	11%
5	Miuref	14 000	11.2%	13 000	11%
6	Weibo(正常)	13 000	10.4%	13 000	11%
7	Shifu	2 000	1.6%	13 000	11%
8	Cridex	1 700	1.3%	13 000	11%
9	Neris	1 800	1.4%	13 000	11%
	总计	124 074	100%	117 000	100%

表 2　CIC-IDS2017 数据集平衡前后分布

编号	应用类别	平衡前		平衡后	
		数量	占比/%	数量	占比/%
1	normal(正常)	10 003	7.2%	10 000	10%
2	Web-BForce	1 229	0.9%	10 000	10%
3	DDOS	40 932	29.4%	10 000	10%
4	DOSGoldenEye	6 817	4.9%	10 000	10%
5	DOSHulk	12 697	9.1%	10 000	10%
6	FTP-Patator	3 594	2.6%	10 000	10%
7	PortScan	54 000	38.8%	10 000	10%
8	slowhttptest	3 796	2.7%	10 000	10%
9	SSH-Patator	2 681	1.9%	10 000	10%
10	DoSSlowloris	3 505	2.5%	10 000	10%
	总计	136 360	100%	100 000	100%

3.2　模型参数设置

对于相同的超参数，设置均保持相同；对于不同的超参数，单独设置。GAN、WGAN 和 CWGAN-GP 均为 4 层神经网络，噪声空间的维数设置为 100，batch_size 设置为 64。WGAN 和 CWGAN-GP 将 α、$\beta 1$、和 $\beta 2$ 分别设置为 0.002、0.5 和 0.999；clip_value 均为 0.01；critic 设置为 5，即当生成器训练 1 个 batch 时，判别器要接着训练 5 次；梯度惩罚系数 λ 在 CWGAN-GP 中设置为 10，训练批次均为在满足 2.2 节条件下且生成数据数量足够时停止；SMOTE 方法的 k_neighbors 为 5。

本文使用的分类模型是经典 LeNet-5[12] 的 CNN 结构，CNN 模型使用交叉熵损失，batch_size 设置为 50，训练轮次均为 2 000。

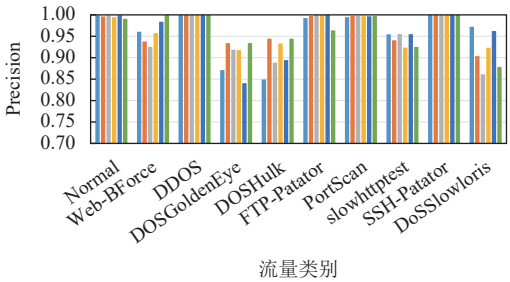
3.3　评价标准

本文使用评价网络流量分类器的性能指标有精确度 (precision)、召回率 (recall) 和 F1-score。

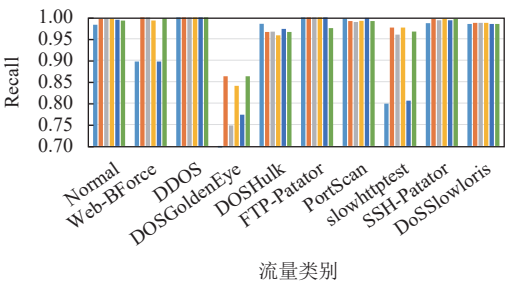
3.4　实验结果及分析

实验在一台配置了 Intel(R)Core(TM) i7-7700HQ CPU @2.80 GHz 处理器、16 GB 内存、GPU(GeForce GTX1050)、Win10 系统的笔记本电脑上运行。使用 TensorFlow 1.15.0+Keras 2.3.1 深度学习平台实现分类操作，基于 GAN 及其衍生版本的数据生成方法使用 torch1.3.1+GPU 实现。

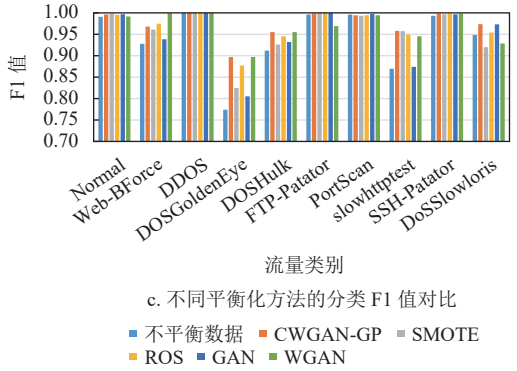
使用 CNN 模型对 5 种平衡方法和原始数据进行实验测试。只对训练集进行处理，测试集全部为原始数据。随机选取 10% 的样本集作为测试集，并对剩余 90% 的数据集进行平衡化作为训练集。为消除随机划分数据集对结果的影响，将实验重复 5 次取平均值作为最终结果，结果如图 4、图 5 所示。



a. 不同平衡化方法的分类 Precision 对比

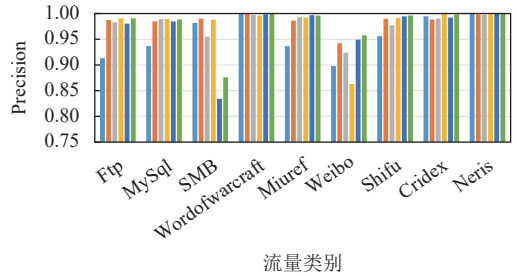


b. 不同平衡化方法的分类 Recall 对比

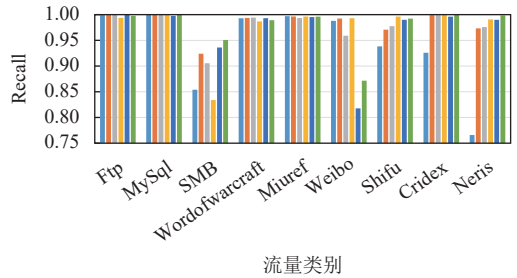


c. 不同平衡化方法的分类 F1 值对比

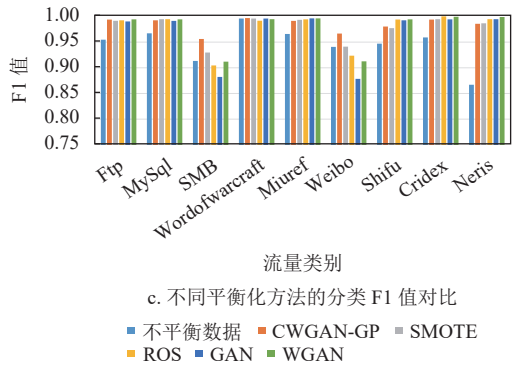
图 4 CIC-IDS2017 分类结果对比



a. 不同平衡化方法的分类 Precision 对比



b. 不同平衡化方法的分类 Recall 对比



c. 不同平衡化方法的分类 F1 值对比

图 5 USTC-TFC2016 分类结果对比

经过对比得到在相同训练参数下，使用 CWGAN-GP 方法进行流量图片平衡后的数据集相较于其他分类方法识别效果提升明显，且在多数类别上识别均值高于使用其他平衡方法。CWGAN-GP 方法的最终效果优于其他方法，F1 值在 IDS2017 上相较于原始数据提高近 3%，较于 SMOTE 方法提高近 2%；在 TFC2016 上较原始提高近 4%，较 SMOTE 提高近 1.3%；在两个数据集上均有识别率较低的流量类别，这些流量类别具有一定隐蔽性，不易被识别出来，但本方法对识别率有较大提高，验证了本方法具有一定的鲁棒性，且不易出现模式崩溃和收敛困难的问题。由于相较于 WGAN-GP 添加了辅助信息，在生成效率上也有很大提高。均值结果统计如表 3、表 4 所示。

表 3 CIC-IDS2017 分类结果均值统计

数据集类型	Precision	Recall	F1
原始数据	0.959 4	0.932 9	0.940 8
CWGAN-GP平衡	0.965 2	0.977 7	0.974 0
SMOTE平衡	0.954 5	0.964 1	0.957 7
ROS平衡	0.964 7	0.974 0	0.968 9
GAN平衡	0.963 2	0.942 0	0.951 4
WGAN平衡	0.963 5	0.973 4	0.967 9

表 4 USTC-TFC2016 分类结果均值统计

数据集类型	Precision	Recall	F1
原始数据	0.957 1	0.940 2	0.945 6
CWGAN-GP平衡	0.985 3	0.983 3	0.984 1
SMOTE平衡	0.979 3	0.978 2	0.978 6
ROS平衡	0.978 9	0.976 8	0.976 6
GAN平衡	0.970 0	0.968 5	0.968 3
WGAN平衡	0.978 1	0.977 4	0.977 3

4 结 束 语

本文提出了利用流量图片化结合 CWGAN-GP 来处理恶意流量识别领域的不平衡问题，此方法通过学习原始数据的真实分布来生成新的数据。

在两个不同的恶意流量不平衡数据集上，通过使用深度学习分类算法 CNN 对基于 CWGAN-GP 方法进行了评估。实验结果表明，CWGAN-GP 在所有指标下均优于其他过采样方法，但需要更多的时间进行训练。在未来的研究中，将对 CWGAN-GP 进行更加深入的理论研究，加速其训练和收敛过程；并探索更加合理的生成数据评价指标。

参 考 文 献

[1] SUH S, LEE H, LUKOWICZ P, et al. CEGAN: Classification enhancement generative adversarial networks for unraveling data imbalance problems[J]. Neural Networks, 2021, 133: 69-86.

[2] SUN J, LANG J, FUJITA H, et al. Imbalanced enterprise

credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates[J]. Information Sciences, 2018, 425: 76-91.

[3] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 2014 Advances in Neural Information Processing Systems. New York: Curran Associates, 2014: 2672-2680.

[4] DOUZAS G, BACAO F. Effective data generation for imbalanced learning using conditional generative adversarial networks[J]. Expert Systems with Applications, 2018, 91: 464-471.

[5] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of wasserstein gans[C]//Proceedings of the 2017 Advances in Neural Information Processing Systems. New York: Curran Associates, 2017: 5767-5777.

[6] 潘一鸣, 林家骏. 基于生成对抗网络的恶意网络流生成及验证[J]. 华东理工大学学报 (自然科学版), 2019, 45(2): 344-350.
PAN Y M, LIN J J. Generation and verification of malicious network flow based on generative adversarial networks[J]. China Journal of East University of Science and Technology (Natural Science), 2019, 45(2): 344-350.

[7] CHENG Q, ZHOU S, SHEN Y, et al. Packet-Level adversarial network traffic crafting using sequence generative adversarial networks[EB/OL]. [2021-11-30]. https://arxiv.org/abs/2103.04794.

[8] MING Z A, TONG L, RUI Z, et al. Conditional Wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification[J]. Information Sciences, 2020, 512: 1009-1023.

[9] WANG P, LI S, YE F, et al. PacketCGAN: Exploratory study of class imbalance for encrypted traffic classification using CGAN[C]//IEEE International Conference on Communications (ICC). [S.l.]: IEEE, 2020: 1-7.

[10] 王伟. 基于深度学习的网络流量分类及异常检测方法研究[D]. 合肥: 中国科学技术大学, 2018.
WANG W. Deep learning for network traffic classification and anomaly detection[D]. Hefei: University of Science and Technology of China, 2018.

[11] Intrusion detection evaluation dataset (CICIDS2017). [EB/OL]. [2020-11-05]. https://www.unb.ca/cic.

[12] LECUN Y, BOTTOU L. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

编　辑　税　红