基于深度学习与开集识别技术的对抗式DDoS攻击检测技术

Adversarial DDoS Attack Detection Based on Deep Learning and Open Set Recognition Techniques

吴志祥，刘莉丹，高　博（中国联通黑龙江分公司，黑龙江 哈尔滨 150001）
Wu Zhixiang，Liu Lidan，Gao Bo（China Unicom Heilongjiang Branch，Haerbin 150001，China）

摘　要：
网络已成为现代生活不可或缺的一部分，但也面临着诸多的安全风险，特别是分布式拒绝服务（DDoS）攻击。利用人工智能（AI）技术可应对DDoS攻击带来的挑战。基于CNN-Geo和Cycle GAN技术，提出一种包含一个增量学习模块的防御模型，该增量学习模块能够训练未知流量并不断提高模型的防御能力。该模型可以识别偏离学习分布的未知攻击，评估结果表明其准确度超过98.16%，增强了对现实场景中不断演变的DDoS攻击策略的检测和防御能力。

关键词：
DDoS；AI；开放集识别；CNN-Geo；Cycle GAN；增量学习
doi：10.12045/j.issn.1007-3043.2024.08.004
文章编号：1007-3043（2024）08-0018-06
中图分类号：TP181
文献标识码：A
开放科学（资源服务）标识码（OSID）：

Abstract：
The Internet has become an integral part of modern life，but it also faces many security risks，especially Distributed Denial of Service（DDoS）attacks. The use of artificial intelligence (AI) technology can address the challenges posed by DDoS attacks. It proposes a defense model based on CNN-Geo and Cycle GAN techniques，which includes an incremental learning module that is able to train unknown traffic and continuously improve the model's defense capability. This model can identify unknown attacks that deviate from the learning distribution，and the evaluated results show that its accuracy is more than 98.16%，which enhances the ability to detect and defend against the evolving DDoS attack strategies in real scenarios.

Keywords：
DDoS；AI；Open set recognition（OSR）；CNN-Geo；Cycle GAN；Incremental learning

引用格式：吴志祥，刘莉丹，高博. 基于深度学习与开集识别技术的对抗式DDoS攻击检测技术[J]. 邮电设计技术，2024(8)：18-23.

1 概述

近年来网络安全攻击频发，黑客的攻击越来越难以防御，尤其是分布式拒绝服务（DDoS）攻击，被认为是最难以防御的网络攻击手段之一。检测和防御DDoS攻击的挑战在于其模仿合法用户行为的能力，DDoS攻击可通过快速注入大量恶意软件阻碍对服务的访问，使目标受害者大范围的服务中断并可能导致数据中心关闭，后果非常严重[1]。传统的防御机制存在对DDoS攻击误判率高的问题，而且需要大量人力，无法抵御不断演变的攻击[2]。如今人工智能在构建入侵检测系统（IDS）等应用上已取得重大进展，市面上的大部分产品都依靠人工智能技术来自动区分传统流量和攻击流量[3]。然而，这些应用人工智能的IDS面临着巨大的挑战。随着信息活动的迅速扩大，新服务在不断发展，因为此类模型在设计时没有考虑未知攻击，在面对新形式的攻击流量时，缺乏识别和应对未知攻击的能力。最近，这些模型还受到了对抗性攻

收稿日期：2024-07-10

击，有可能扭曲其分类结果。对抗性攻击可以通过添加干扰因素，干扰人工智能模型的辨别能力。基于此，本文旨在利用深度学习领域中的循环广域网（Cycle-GAN）开发一种IDS，用于识别DDoS中的对抗性攻击。该架构采用了深度学习技术，特别是卷积神经网络（CNN-Geo）模型作为IDS的基础。CNN-Geo是一种新颖的防御模型，将几何特征与深度学习技术相结合，可识别和防范未预见的攻击，以提高防御的精度和可靠性。通过集成CNN-Geo模型，架构能够直接检查重建错误，有效地识别网络流量中的异常情况。通过几何度量进一步增强系统根据空间特征对未知样本进行辨别和分类的能力，这种能力有助于主动将未识别的样本转发给工程师进行进一步分析。此外，该架构还包括一个增量学习模块，可随着时间的推移不断提高IDS的性能。

2 相关技术

2.1 DDoS攻击原理

DDoS攻击是一种网络攻击形式，可通过大量的网络攻击来耗尽目标计算机的网络或系统资源，目的是阻碍向授权用户提供服务[4]，其原理如图1所示。常见的DDoS攻击类型有Ping Flood、Smurf Attack、SYN Flooding和DNS放大等[5]。由于DDoS攻击的分布式特性，识别恶意流量成为一项挑战。

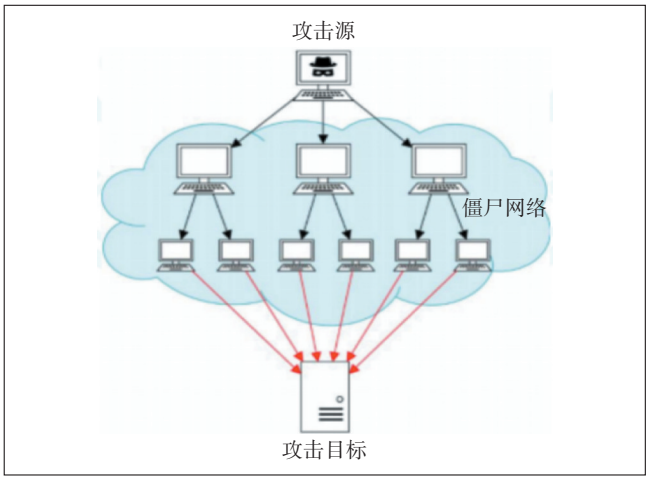

图1　DDoS攻击原理

2.2 流量过滤和IP阻断

流量过滤和IP阻断是缓解DDoS攻击的基本技术。根据特定标准有选择地允许或阻止网络流量，有助于保护网络和系统免受恶意流量的攻击[6]。这种技术在缓解DDoS方面特别有用，它允许组织识别并阻止与DDoS相关的已知恶意IP地址。通过将这些IP列入黑名单，企业可以有效地在网络边界阻止恶意流量，减少对网络的影响[6]。实施流量过滤和IP阻断的方法有访问控制列表（ACL）、入侵检测/防御系统（IDS/IPS）和专用的DDoS缓解服务[7]。

2.3 卷积神经网络（CNN）

卷积神经网络（CNN）是一种高效的深度学习模型，它以动物的视觉皮层为模型，被设计用于自主学习并从输入数据中提取相关特征，这使CNN在分析网络流量和检测潜在攻击时非常有效[8]。训练CNN时需向其提供标注数据，从而使网络获得在输入流量样本和各自标签之间建立关联的能力。CNN在IDS中的功效很大程度上取决于是否存在广泛而多样的训练数据。

2.4 开放集识别（OSR）

识别开放集模式的任务比识别闭集模式的任务复杂得多，主要是因为需要有效管理未识别的模式[9]，封闭集分类与开放集识别比较如图2所示。OSR架构集成了重构和分布技术，这种方法旨在确定超球的分布，提高检测的有效性。该理论涉及通过类似于概率密度函数来估计空间分布。如果最近获得的样本偏离了可接受的范围，就会被归类为未知。

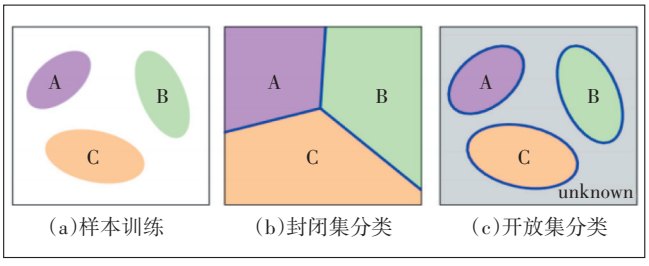

图2　封闭集分类与开放集识别比较

2.5 未知DDoS攻击检测

可以通过极值理论识别未知的DDoS攻击。极值理论[10]采用高斯混合模型[11]（GMM）及其相关方法对输入分布进行估计，通过将深度学习框架与GMM分布阈值和双向长短期记忆模型[10]（BI-LSTM）进行二元分类，利用BI-LSTM模型获得的属性值作为无法识别实体的线索。

3 技术框架及算法

3.1 未知检测框架

本文所提出的未知检测框架即CNN-Geo（见图3），该框架将强大的CNN[12]架构与几何度量计算模块相结合以应对未知DDoS攻击挑战。CNN-Geo利用了

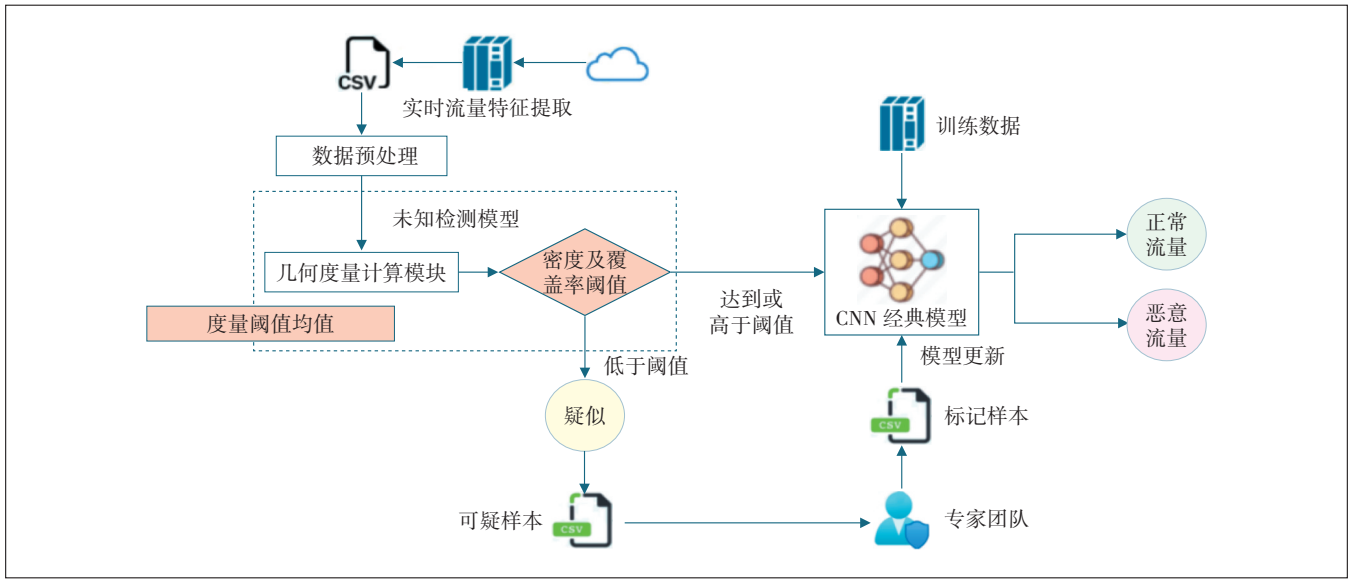

图3　未知检测框架结构

CNN模型捕捉输入数据中空间和时间模式的能力，从而实现有效的流量分类。利用几何度量计算模块可增强模型[13]检测未知样本的能力，可通过识别偏离既定标准的数据点来检测未知样本。这种方法允许CNN-Geo建立一个度量阈值，区分已知分布内的样本和分布外的异常值。在分类过程中，通过筛选满足阈值条件的元素，CNN-Geo可以有针对性地进行分析优化，并优先处理置信度较高的样本。采用稀疏分类交叉熵[14]损失函数，使优化过程更直接。选择的编码方法，减轻了标签之间的线性依赖性问题。总的来说，CNN-Geo的理念集中于利用CNN的模式识别能力，结合几何度量分析，提高模型对交通流量进行准确分类和检测未知样本的能力。

3.2 CNN分类器

本文提出了一个分类器模型[15]，其输入为一个9×9的矩阵，该矩阵代表了网络流量。该模型的输出包括2个预测级别，即良性分类和攻击分类。框架由多层卷积组成，每层卷积都经过批量归一化层、剔除层和全连接层。CNN在识别模式和提取特征方面的效果在及时识别和缓解DDoS攻击方面发挥着至关重要的作用。

3.3 未知检测模块

开发未知检测模块的目的是解决网络安全中检测不明网络攻击的问题。本文中的模块分批对数据集进行分割，可以评估上述批次数据集与初始数据集之间的相似度，无需同时处理整个数据集，这提高了模块的处理速度和效率。为了评估异常值，本文根据平均值构建了一个阈值，具体如式(1)和式(2)所示。

$$D_{\text{threshold}} = \frac{2}{(N-1)N}\sum_{i=1}^{N}\sum_{j\neq i}\text{Density}\left(\text{bath}_i,\text{bath}_j\right) \quad (1)$$

$$C_{\text{threshold}} = \frac{2}{(N-1)N}\sum_{i=1}^{N}\sum_{j\neq i}\text{Density}\left(\text{bath}_i,\text{bath}_j\right) \quad (2)$$

其中，N为批次数。阈值在拟议模块中具有重要意义，因为它有助于区分已识别和未识别的攻击。数据的度量密度和覆盖结果低于阈值水平的情况将被归类为异常值(见图4)。这一过程可以检测和隔离陌生的网络入侵。

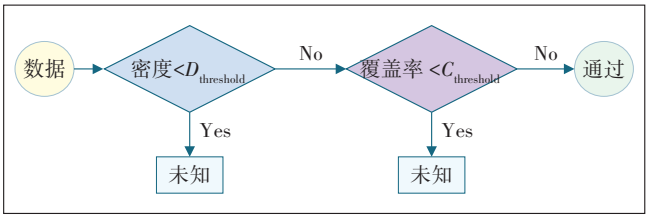

图4　利用几何度量阈值的未知检测策略

未知检测模块进行密度指标的计算，并遵守算法1的程序(见图5)。覆盖率计算的算法如图6所示。

3.4 增量学习模块[16]

本文的模型包含一个增量学习模块，该模块能够检测出来源不明的样本。一旦检测到无法识别的流量，就会通知通信专家对数据进行分类，以便后续完善模型。此外，在训练过程中还会调整模型的学习率，以防出现严重遗漏以前学习过的信息的情况。

输入：D_R，真实样本数据集；D_U，未知样本数据集；Ng，真实样本的数量；N_u，未知样本的数量；k，用于密度计算的近邻数量；Count，计数，包含每个未知样本的真实样本邻近球体数组
输出：密度值

```
Real sample ← r ∈ D_R
Unknown sample ←u ∈ D_U
Define distance between the unknown sample m and real sample n:
    d_mn = distance(u_m, r_n)
Define k^th nearest neighbour distances for real sample
    NNDn=nearest_neighbour_distances(r_n, k)
Count=array[N_u] of zeros
for i in range N_u:
 for j in range N_R:
  if d_ij<NNDi
   Count[i]=Count[i]+1
end for
end for
Density = mean (Count)
return Density
```

图5　算法1：密度计算

输入：D_R，真实样本数据集；D_U，未知样本数据集；N_R，真实样本数；N_u，未知样本数；k，计算覆盖率时使用的近邻数；Count，包含至少一个未知样本的真实样本邻近球体数组
输出：覆盖值

```
Real sample ←r ∈ D_R
Unknown sample ← u∈ D_U
Define distance between the unknown sample m and real sample n:
    d_mn= distance(u_m, r_n)
Define kth nearest neighbour distances for real sample n:
    NND_n= nearest_neighbour_distances(r_n, k)
Count = array[N_R] of zeros
for i in range N_R:
 for j in range N_u:
  if d_ji < NND_i
   Count[i] =1
  break
end for
end for
Coverage = mean (Count)
return Coverage
```

图6　算法2：覆盖率计算

3.5 对抗性检测框架

为了制定一个理解性防御框架，该研究对CNN-Geo防御系统进行了实验，评估其在抵御更有害的攻击时的效果。图7详细说明了CNN-Geo所采用的对抗性防御框架的架构与未知攻击检测框架，它保留了CNN-Geo模型作为骨干。

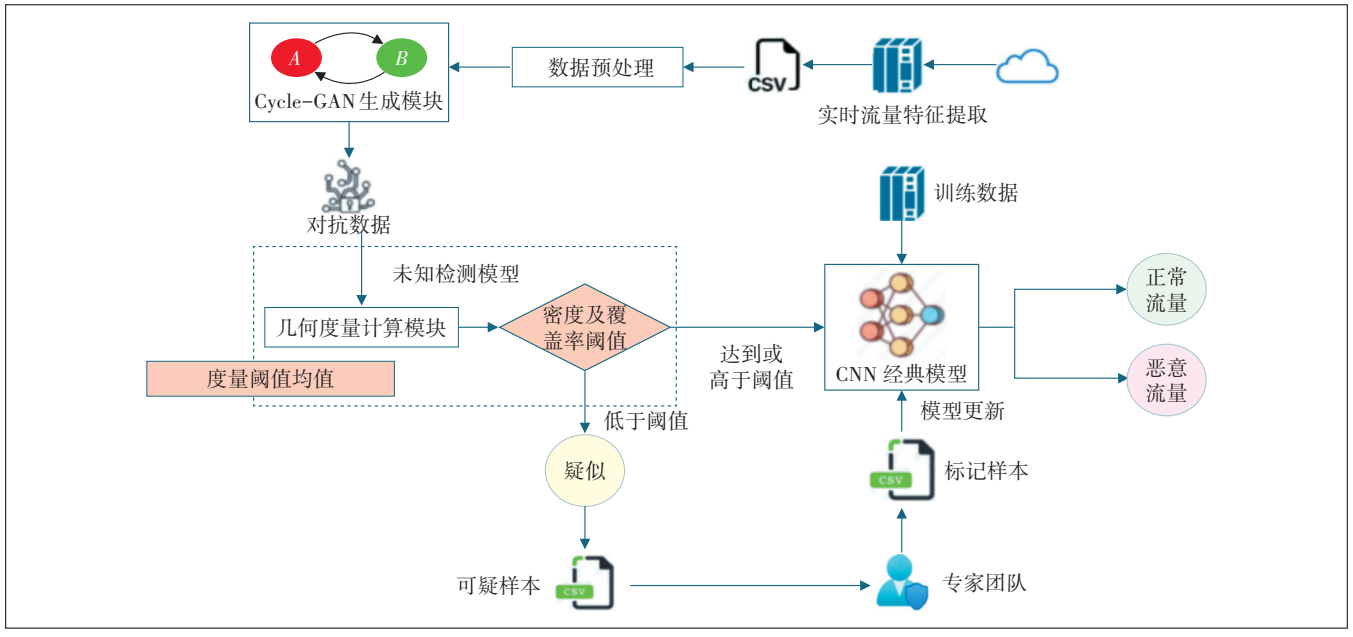

图7　对抗检测框架结构

3.6 CycleGAN生成架构

CycleGAN生成模型如图8所示，它可在没有配对实例的情况下进行训练。本模型有一对生成器G_A和G_B，前者负责生成与“攻击”相关的数据，表示为Φ_A，后者则生成“良性域”的数据，表示为Φ_B。生成器根据输入数据执行数据转换，特别是来自不同域的数据。

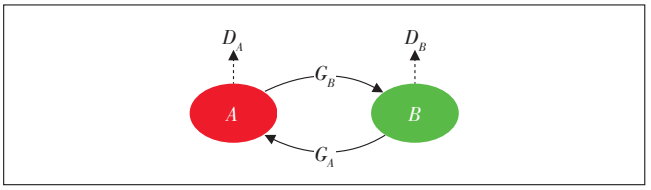

图8　CycleGAN生成

从变量Φ_A中获得的信息传输到系统G_A，而系统G_B则接收来自变量Φ_B的输入数据。G_A所采用的方法是创建与目标分布Φ_B中分布流量非常相似的对抗数据，每个生成器都与一个鉴别器竞争。初始判别器D_A从生成器G_A中提取真实攻击数据和对抗性信息，在真假实例之间做出二元分类决策。第2个鉴别器D_B的任务是辨别真正的良性流量和由G_B生成的对抗性流量。

与传统的GAN架构类似，CycleGAN采用了对抗性零和框架来训练判别器和生成器组件。生成器的目标是增强其欺骗鉴别器的能力，而鉴别器则尽力检测合成的伪造数据。训练过程旨在达到一种平衡状态。此外，生成器还需进行正则化，以生成与源领域相对应的合成数据，而不仅仅为源领域生成对抗数据。要完成上述任务需将生成的数据输入到相应的生成器中，并在结果与主要生成器之间建立相关性。循环的特点是2个生成器之间的信息传输，循环一致性的过程如图9所示。

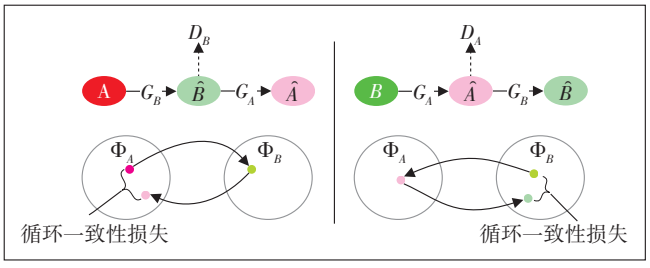

图9　周期一致性映射

此外，架构设计包含身份映射（见图10）。预计从目标域获取的输入数据将产生类似的结果。不过，这种特殊设计的实施是自由裁量的，它能增强输入数据的匹配过程。

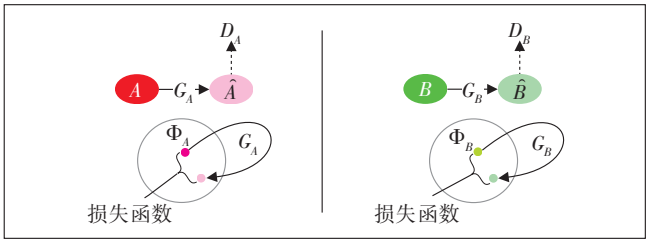

图10　身份映射

3.7 CycleGAN生成的损失函数

“周期一致性映射”和“身份映射”这2个映射函数与对抗损失具有很强的相关性。与映射函数相关的损失函数为$G_A:\Phi_B\rightarrow\Phi_A$，其判别式$D_A$如式（3）所示。

$$L_{\text{adv}}\left(G_A,\ D_A\right)=E_{A\in\Phi_A}\left[\log\left(D_A(A)\right)\right]+E_{B\in\Phi_B}\left[1-\log\left(D_A\left(G_A(B)\right)\right)\right] \tag{3}$$

其中，A是攻击训练集Φ_A中的攻击流量，B是良性训练集Φ_B中的良性流量。

3.8 周期一致性损失

数据转换过程必须具备将B恢复到初始状态的能力，每个无害数据需恢复到其初始状态，如式（4）所示。

$$\forall B\in\Phi_B,\ B\rightarrow G_A(B)\rightarrow G_B\left[G_A(B)\right]\approx B \tag{4}$$

同样，对于来自Φ_A的每个A，G_A和G_B也应满足以下要求：

$$\forall A\in\Phi_A,\ A\rightarrow G_B(A)\rightarrow G_A\left[G_B(A)\right]\approx A \tag{5}$$

本文使用循环一致性损耗来实现上述特征，其计算如式（6）所示。

$$L_{\text{cyc}}\left(G_A,G_B\right)=E_{A\in\Phi_A}\left\{\left\|G_A\left[G_B(A)\right]-A\right\|_1\right\}+E_{B\in\Phi_B}\left\{\left\|G_B\left[G_A(B)\right]-B\right\|_1\right\} \tag{6}$$

3.9 Identity丧失

该架构具有身份映射功能。预计源于目标域的输入数据会产生类似的结果，具体来说，$G_A(A)\approx A$和$G_B(B)\approx B$。利用同一性损失可以方便地实现式（7）。

$$L_{\text{ide}}\left(G_A,G_B\right)=E_{A\in\Phi_A}\left[\left\|G_A(A)-A\right\|_1\right]+E_{B\in\Phi_B}\left[\left\|G_B(B)-B\right\|_1\right] \tag{7}$$

3.10 信号发生器损耗函数

生成器生成与原始源数据相似的数据，鉴别器区分翻译和真实的样本。生成器的目标是减少损失函数，而鉴别器的目标是最大化损失函数。式（8）为建议框架中的生成器损失函数。

$$\min_{G_A,G_B}L\left(G_A,G_B,D_A,D_B\right)=L_{\text{ADV}}\left(G_A,D_A\right)+L_{\text{ADV}}\left(G_B,D_B\right)+\lambda\times L_{\text{cyc}}\left(G_A,G_B\right)+\mu\times L_{\text{ide}}\left(G_A,G_B\right) \tag{8}$$

其中，λ和μ分别是控制循环一致性损失和特征损失的相对权重参数。

鉴别器的目标是模拟基于机器学习的检测，上述输入可作为一种手段，促使生成器生成对抗性流量，生成的流量必须不被目标系统基于ML的检测器检测到。算法3描述了用于生成流量的CycleGAN框架的功能（见图11）。

4 效果验证

通过实验对研究进行了验证，验证使用的数据集

输入：G_A，攻击域生成器；D_A，攻击域判别器；G_B，良性域生成器；D_B，良性域判别器；Φ_A，攻击域；Φ_B，良性域
输出：训练后的CycleGAN生成架构系统

1. for n epochs do
2. Attack sample batch $\leftarrow \{A^{(i)}\}_{i=1}^{m} \in \Phi_A$
3. Benign sample batch $\leftarrow \{B^{(i)}\}_{i=1}^{m} = 1 \in \Phi_B$
4. Generate m sample of $G_B(A) \rightarrow \hat{A}$ and $G_A(B) \rightarrow B$
5. Generate m sample of $G_B(G_B(A))$and $G_B(G_A(B))$
6. Generate m sample of $G_A(A)$and $G_B(B)$
7. Update the discriminator D_A and D_B according to the adversarial loss function using Equation（2）
8. $\max_{DA} L_{adv}(G_A, D_A)$
9. $\max_{DB} L_{adv}(G_B, D_B)$
10. Update the Generator G_A and G_B according to the total CycleGAN loss function using Equation（7）
11. $\min_{GA,DA}(G_A, G_B, D_A, D_B)$
12. end for

图11　算法3：CycleGAN生成架构生成系统的训练

为CICIDS2017-Tuesday和CICDDoS2019等，验证结果显示了CNN-Geo在检测未知攻击方面的有效性。该模型有效降低了未知攻击带来的风险，评估结果表明其鉴别未知DDoS攻击的准确度超过98.16%。利用CNN的强大功能，CNN-Geo能够提取和分析网络流量中的复杂模式，从而准确检测和及时应对潜在威胁。

5 总结

本文探讨了对抗性攻击的概念及其与Cycle GAN的关联。通过采用先进的检测技术和增量学习，CNN-Geo能够不断提高其防御能力，且面对复杂的对抗性攻击，CNN-Geo仍能保持较高的准确性。通过结合CNN和几何分析的优势，CNN-Geo提供了一个有效的解决方案来应对未知和对抗性攻击带来的日益严峻的挑战。它为网络安全提供了一种前景广阔的方法，可确保在线服务的不间断运行并保护企业免受潜在的DDoS攻击破坏性影响。

参考文献：

[1] kaspersky. Internet Security：What is it，how can you protect yourself online? [EB/OL]. [2024-01-23]. https://www. kaspersky. com/resource-center/definitions/what-is-Internet-security.

[2] DE NEIRA A B，KANTARCI KANTARCI，NOGUEIRA M. Distributed denial of service attack prediction：Challenges，open issues and opportunities[J]. Computer Networks，2023，222：109553.

[3] LAZENBY S. DDoS attacks in the financial industry：how to protect your infrastructure and payments [EB/OL]. [2024-01-23]. https://www.inetco.com/blog/ddos-attacks-in-the-financial-industry/.

[4] Imperva. DDoS in the time of COVID-19 [EB/OL]. [2024-01-23]. https://www. imperva. com/resources/resource-library/reports/ddos-in-the-time-of-covid-19/.

[5] IRWIN L. DDoS attacks soar as organisations struggle with effects of COVID-19 [BE/OL]. [2024-01-23]. https://www. itgovernance. eu/blog/en/ddos-attacks-soar-as-organisations-struggle-with-effects-of-covid-19.

[6] Norton. What is a DDoS Attack? [EB/OL]. [2024-01-23]. https://community.norton.com/es/node/1315971.

[7] Cloudflare. DDoS threat report for 2023 Q1 [EB/OL]. [2024-01-23]. https://xhh.club/e/dWd8VDD3yXf38dcmbLF7/.

[8] RAHMAN O，QURAISHI M A G，LUNG C H. DDoS attacks detection and mitigation in SDN using machine learning [C]//2019 IEEE World Congress on Services（SERVICES）. Piscataway：IEEE，2019：184-189.

[9] MAHJABIN T，XIAO Y，SUN G，et al. A survey of distributed denial-of-service attack，prevention，and mitigation techniques [J]. International Journal of Distributed Sensor Networks，2017，13（12）：1550147717741463.

[10] NISHANT R，KENNEDY M，CORBETT J. Artificial intelligence for sustainability：challenges，opportunities，and a research agenda [J]. International Journal of Information Management，2020，53：102104.

[11] GAURAV A，GUPTA B B，ALHALABI W，et al. A comprehensive survey on DDoS attacks on various intelligent systems and it′s defense techniques [J]. International Journal of Intelligent Systems，2022，37（12）：11407-11431.

[12] WANG B F，LANG B，XIAO N，et al. AspIOC：aspect-enhanced deep neural network for actionable indicator of compromise recognition [C]//Information Security. Cham：Springer，2022：411-421.

[13] LYU L J，TU Y，ZHANG Y J. Deep learning assisted key recovery attack for round-reduced simeck32/64 [C]//Information Security. Cham：Springer，2022：443-463.

[14] SUN L L，LI H，YU S W，et al. HeHe：balancing the privacy and efficiency in training CNNs over the semi-honest cloud [C]//Information Security. Cham：Springer，2022：422-442.

[15] CHENG J R，YIN J P，LIU Y，et al. DDoS attack detection using IP address feature interaction [C]//2009 International Conference on Intelligent Networking and Collaborative Systems. Piscataway：IEEE，2009：113-118.

[16] WANG C G，ZHENG J，LI X Y. Research on DDoS attacks detection based on RDF-SVM [C]//2017 10th International Conference on Intelligent Computation Technology and Automation（ICICTA）. Piscataway：IEEE，2017：161-165.

作者简介：
吴志祥，毕业于吉林大学，高级工程师，硕士，主要从事网络安全技术研究、网络安全技术培训和网络安全事件应急处置工作；刘莉丹，毕业于哈尔滨工业大学，高级工程师，硕士，主要从事DCN网络维护工作；高博，毕业于哈尔滨理工大学，高级工程师，硕士，主要从事互联网、大数据、云计算、人工智能（机器学习）相关创新研发工作。