基于 DAE 和 GRU 组合的流量异常检测方法

尹梓诺[1], 马海龙[1], 胡 涛[1]

[1]解放军信息工程大学信息技术研究所, 郑州 中国 450001

摘要 流量异常检测能够有效识别网络流量数据中的攻击行为, 是一种重要的网络安全防护手段。近年来, 深度学习在流量异常检测领域得到了广泛应用, 现有的深度学习模型进行流量异常检测存在两个问题: 一是数据受噪声影响导致检测鲁棒性差、准确率低; 二是数据特征维度高以及模型参数多导致训练和检测速度慢。为了在降低流量数据噪声影响的基础上提高检测速度和准确性, 本文提出了一种基于去噪自编码器(Denoising Auto Encoder, DAE)和门控循环单元(Gated Recurrent Unit, GRU)组合的流量异常检测方法。首先设计了基于 DAE 的流量特征提取算法, 采用小批量梯度下降算法对 DAE 进行训练, 通过最小化含噪声数据的重构向量与原始输入向量间的差异, 有效提取具有较强鲁棒性的流量特征, 降低特征维度。然后设计了基于 GRU 的异常检测算法, 利用提取的低维流量特征数据训练 GRU, 从而构建异常流量分类器, 实现对攻击流量的准确检测。最后在 NSL-KDD、UNSW-NB15、CICIDS2017 数据集上的实验结果表明: 与其他的机器学习、深度学习方法相比, 本文所提方法的检测准确率最大提升了 18.71%。同时, 本文方法可以实现较高的精确率、召回率和检测效率, 同时具有较低的误报率。在面对数据受到噪声破坏时, 具有较强的检测鲁棒性。

关键词 流量异常检测; 深度学习; 去噪自编码器; 门控循环单元
中图法分类号 TP393 DOI 号 10.19363/J.cnki.cn10-1380/tn.2023.03.02

A Traffic Anomaly Detection Method Based on the Combination of DAE and GRU

Yin Zinuo[1], Ma Hailong[1], Hu Tao[1]

[1] Institute of Information Technology, PLA Information Engineering University, Zhengzhou 450000, China

Abstract Traffic anomaly detection can effectively identify attack behaviors in network traffic data, so it is an important means of network security protection. In the recent years, deep learning technology has been widely used in the field of traffic anomaly detection. The existing traffic anomaly detection methods based on deep learning models have two problems: one is poor robustness and low detection accuracy, which results from the data being affected by noise; the other is low efficiency, which is due to high data characteristic dimension and multiple model parameters. In order to improve detec-tion speed and accuracy on the basis of reducing the impact of noise on traffic data, this paper proposes a traffic anomaly detection method based on the combination of Denoising Auto Encoder (DAE) and Gated Recurrent Unit (GRU). Firstly, we design a traffic feature extraction algorithm based on DAE and use the Mini-Batch Gradient Descent (MSGD) algo-rithm to train DAE. By minimizing the difference between the reconstructed vector of noise-contained traffic data and the original input vector, the traffic features with strong robustness are effectively extracted and the dimension of the features is reduced. Then, an anomaly detection algorithm based on GRU is designed. The extracted low-dimensional traffic data is used to train the GRU to construct the abnormal traffic classifier and realize the accurate detection of the attack traffic. Finally, we have carried out anomaly detection experiments on the NSL-KDD, UNSW-NB15, and CICIDS2017 datasets and the experimental results fully show that compared with other machine learning and deep learning methods, the detec-tion accuracy of our proposed method can be improved by 18.71% at most. At the same time, the proposed method can achieve higher precision rate, recall rate and detection efficiency with lower false positive rate. When the traffic data is damaged by noise, it has strong detection robustness.

Key words *traffic anomaly detection; deep learning; denoising autoencoder; gate recurrent unit

1 引言

随着计算机网络发展和应用服务的不断增加, 各种对计算机网络的攻击活动日益猖獗。网络流量异常检测是发现网络攻击的一种重要手段[1]。流量异常检测的本质可视为对正常流量数据和攻击流量数

通讯作者: 马海龙, 博士, 副研究员, Email: longmanclear@163.com。
本课题得到国家重点研发计划(No. 2018YFB0804002)资助。
收稿日期: 2021-10-31; 修改日期: 2022-01-03; 定稿日期: 2023-01-03

据的分类问题。近年来, 机器学习算法被认为是解决分类问题最有效的方法, 因而在网络流量异常检测中得到大量应用。

在流量异常检测过程中, 传统机器学习方法和浅层学习技术面临如下问题: 流量特征提取准确率低、检测性能对特征依赖性高[2]、检测鲁棒性较差等。这些问题会降低对攻击流量的检测性能。Hinton 等人[3]研究提出的深度学习方法在处理大规模和高维数据方面的独特优势得到了许多学者的认可, 且在图像识别、语音识别、自然语言处理等领域都取得了不错的效果, 也为网络流量异常检测提供了新的研究方向。

深度学习构建一种具有多个隐藏层, 每个隐藏层具有多个单元的网络, 不断学习训练样本来更新网络模型参数, 实现输入向量到输出向量的映射, 以完成各种分类任务。研究人员发现与传统的机器学习技术相比, 深度学习算法在流量异常检测方面具有更多优势: 在进行流量特征选择时, 深度玻尔兹曼机(Deep Boltzmann Machine, DBM) 通过从大量无标签高维数据中提取高级特征, 学习数据的概率分布, 进行异常检测, 可以提高特征提取准确率和检测效率[4]; 自编码器(AutoEncoder, AE)获取隐藏层的神经元作为输入层数据的低维深度表示, 实现降维和特征提取功能[5]; 在训练特征数据方面, Staudemeyer 等人[6]使用长短期记忆(Long Short-Term Memory, LSTM)算法对训练数据进行序列学习, 结果表明, LSTM 算法比其他机器学习算法具有一定的优势; GRU 与 LSTM 网络相比, 在一定程度上提高了对序列数据的学习速度和检测精度[7]。基于上述分析, 这些研究成果在一定程度上提升了网络流量异常检测的效果, 表明深度学习方法是一种有效的流量异常检测工具。

基于深度学习的流量异常检测技术仍面临如下问题: (1)流量数据噪声大, 导致检测鲁棒性差, 检测准确率低。(2)数据特征维度高及算法参数多导致检测时效性差。

在本文中, 设计了 DAE 和 GRU 组合 (the Combination of Denoising auto-encoder and Gated recurrent unit, CDG) 模型, 应用于流量异常检测中。首先采用 DAE 对流量数据进行特征提取来降低数据噪声对检测准确率的影响, 然后对提取到的深层特征表示利用 GRU 进行深度学习, 捕获 DAE 难以获取的时序关系, 构建异常流量分类器, 对数据集中各条数据检测识别。

本文的贡献如下:

1. 提出了一种基于 CDG 的流量异常检测方法, 其总体框架分为三部分: 数据预处理、流量异常检测、分类评估, 分别实现对原始流量数据集的预处理、基于 DAE 的流量特征提取及基于 GRU 的异常检测、对模型检测效果的参数评估。

2. 提出一种基于 DAE 的流量特征提取算法, 通过采用小批量梯度下降法对 DAE 快速训练, 最小化被噪声损坏流量数据的重构向量与原始干净输入向量之间的误差, 获得具有较强鲁棒性的特征, 避免权重参数陷入局部最优的问题。

3. 提出一种基于 GRU 异常检测算法。通过自适应矩估计(Adaptive Moment Estimation, Adam)优化算法, 利用提取的深度流量特征对 GRU 进行训练, 获取数据间的时序关系, 利用历史信息来执行当前分类决策, 从而构建异常流量分类器, 提高流量异常检测的速度和准确率。

4. 为验证本文方法和模型的有效性, 使用公开的流量数据集 NSL-KDD、UNSW-NB15 以及 CICIDS2017 进行实验, 利用多种分类评估指标将本文所提方法与几种典型机器学习、深度学习算法进行比较。实验表明: 本文所提方法在流量异常检测的鲁棒性、准确率和速度方面表现出优越性。

本文的其余部分组织如下: 第 2 节讨论流量异常检测的背景。第 3 节描述基于 CDG 的流量异常检测方法的框架和检测流程。第 4 节重点介绍所提方法的关键算法设计。第 5 节阐述实验过程和结果分析。最后, 在第 6 节进行总结。

2 背景介绍

2.1 相关工作

流量异常检测已成为网络安全防御的重要组成部分。一些深度学习方法在异常流量分类任务[8]中表现出良好的性能。越来越多的研究人员正专注于流量异常检测, 并且已经进行了许多相关研究。

Subba[9]等人提出一种基于人工神经网络的流量异常检测模型, 采用前馈和反向传播算法, 并在模型训练中采用不同的优化算法来最小化整个模型的计算开销, 实现流量异常检测的高性能。Xu 等人[10]提出一种结合 GRU 和多层感知机(Multi-Layer Perceptron, MLP)网络的流量异常检测方法来提升对 NSL-KDD 检测准确率, 但并未考虑到当流量数据受到噪声影响时, 模型的检测性能是否还保持较高水平。Nguyen 等人[11]提出一种包含两个卷积层的卷积神经网络(Convolutional Neural Network, CNN), 用于检测 DoS 攻击。Teng 等人[12]提出一种基于决策树

(Decision Tree, DT)和支持向量机(Support Vector Machines, SVM)结合的自适应协同入侵检测方法提升单一 SVM 对 KDD CUP99 数据集的检测准确率至 89.02%。Yisroel 等人[13]提出一种集成 AE 的无监督流量异常检测方法, 实现对网关等设备的攻击流量的实时检测。Shone 等人[2]提出一种流量异常检测模型 S-NDAE, 它堆叠多个非对称自编码器, 并和随机森林(Random Forest, RF)算法结合以达到更好的训练效果。Yu 等人[14]提出一种使用深度机器学习架构的基于会话的网络入侵检测模型, 利用堆栈自编码器(Stacked Autoencoder, SAE)无监督地从流量数据中自动学习特征, 提升检测僵尸网络流量的性能。Alrawashdeh 等人[15]使用受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)构建深度信念网络(Deep Belief Network, DBN), 将提取的特征送入微调后的逻辑回归分类器, 产生最终的检测结果[16]。Amma 等人[17]提出一种结合 SAE 和 GRU 的流量异常检测方法, 用于在减少检测时间的同时, 提升检测准确率, 但其并未考虑到面对数据受到噪声影响的情况下, 模型的检测鲁棒性, 且仅在 NSL-KDD 数据集上进行实验。Roy 等人[18]使用一种改进的循环神经网络(Recurrent Neural Network, RNN)--双向长短期记忆网络来检测物联网中的攻击, 并在 UNSW-NB15 上取得了较高的检测精度。Mirza 等人[19]提出一种基于 LSTM 神经网络的序列自编码器框架, 用于提升无监督情况下网络入侵检测的准确率。为了提升 GRU 模型的异常检测能力, Agarap 等人[7]使用 SVM 替代 GRU 模型输出层中的 softmax 函数, 结果表明所提模型对攻击的检测能力优于传统的 GRU-softmax 模型。

综上所述, 研究人员提出的不同模型在分类器优化和流量异常检测技术方面都取得了良好的进展, 并且更多地聚焦于提高模型的检测性能。但在检测鲁棒性以及时效性方面的关注有所欠缺。本文基于深度学习方法, 从流量数据特征提取和检测两方面考虑, 提出了 CDG 模型进行流量异常检测, 利用 DAE 提取出鲁棒性强的低维特征, 并利用 GRU 网络学习流量数据的时序信息。将基于 CDG 的流量异常检测方法应用于网络流量数据集 NSL-KDD、UNSW-NB15 和 CICIDS2017, 来检验其检测未知攻击和新型网络攻击的能力, 使用多种评估指标, 包括准确率、精确率、召回率、误报率、F1 值、时间等来评估所提出方法的性能。

2.2 数据集介绍

在数据集方面, 本文利用公开流量数据集 NSL-KDD、UNSW-NB15 和 CICIDS2017 进行流量异常检测, 其特征情况如下。

NSL-KDD 数据集包含 41 个特征, 可分为 4 类: (1)TCP 连接基本特征: 即每条连接的基本属性, 如持续时间、协议类型等, 可用于 DoS 攻击检测; (2)TCP 连接内容特征: 数据包负载中的内容特征, 如登陆失败次数等, 用于检测 U2R 和 R2L 两类嵌入在数据负载里的攻击; (3)基于时间的网络流量统计特征: 当前连接与其短时间前后的连接记录之间存在的某些相关联系统计特征, 如过去两秒内与当前连接具有相同目标主机的连接数等, 由于网络攻击的强关联性, 这类特征可在一定程度反映入侵行为; (4)基于主机的网络流量统计特征: 当前连接之前 100 个连接记录中与当前连接具有相同目标主机的统计信息, 如前 100 个连接中与当前连接具有相同目标主机的连接数等。这类特征可用于检测 probe 攻击这类慢速扫描攻击。

UNSW-NB15 数据集包含 42 个特征, 同样可分为 5 大类: (1)基本特征: 包含连接记录的总持续时间等 14 个特征; (2) 内容特征: 包含源 TCP 序列号、目的 TCP 序列号等 8 个特征; (3)时间特征: 包含记录的开始时间、持续时间等 7 个特征; (4)流量特征: 包含交换协议特征; (5)额外原始特征: 包含源 ip、端口和目的 ip、端口是否一致, 100 个连接中包含相同服务和源地址的连接数, 以及其他的相关统计特征共 12 个。

CICIDS2017 数据集包含 78 个特征, 包含连接基本特征、流量上行、下行统计特征、流量时间统计特征以及各关键标志位值等特征。三个数据集的具体特征如表 1 所示。

2.3 相关理论

2.3.1 AE

AE 由编码器和解码器组成, 包含输入层、隐藏层和输出层。编码过程将原始输入数据压缩得到新的特征表示, 解码过程将此特征表示解码重构, 生成与原始数据接近的重构数据。通过不断减小输入输出的差异来训练模型。经过无监督训练, AE 仅能尝试将输入复制到输出, 不能确保获得有用的特征。

为保证 AE 可以学到有用特征, Vicent 等人[20]提出 DAE。DAE[21]是 AE 的扩展[22]。通过引入一个损坏过程, 如对数据引入噪声等, 得到受损数据作为输入, 使得网络在学习过程中去除噪声等损坏因素, 重构原始未被损坏数据。通过对 DAE 进行训练, 可以从原始高维特征中学习到特征的低维深度表示, 能更好地获取鲁棒性较强的低维特征。

表 1 NSL-KDD、UNSW-NB15 和 CICIDS2017 数据集的特征
Table 1 Features of NSL-KDD, UNSW-NB15 and CICIDS2017 datasets

数据集	数据特征
NSL-KDD	TCP 连接基本特征: (1)duration,(2)protocol_type,(3)service,(4)flag,(5)src_bytes,(6)dst_bytes,(7)land,(8)wrong_fragment,(9)urgent TCP 连接的内容特征: (10)hot,(11)num_failed_logins,(12)logged_in,(13)num_compromised,(14)root_shell,(15)su_ attempted,(16)num_root, (17)num_file_creations, (18)num_shells, (19)num_access_files,(20) num_outbound_cmds,(21)is_hot_login, (22)is_guest_login 基于时间的网络流量统计特征: (23)count,(24)srv_count,(25)serror_rate,(26)srv_serror_rate,(27)rerror_rate,(28)srv_rerror_rate,(29)same_srv_rate,(30)diff_srv_rate,(31)srv_diff_host_rate 基于主机的网络流量统计特征: (32)dst_host_count,(33)dst_host_srv_count,(34)dst_host_same_srv_rate,(35)dst_host_diff_srv_rate,(36)dst_host_same_src_port_rate,(37)dst_host_srv_diff_host_rate,(38)dst_host_serror_rate,(39)dst_host_srv_serror_rate,(40) dst_host_rerror_rate,(41)dst_host_srv_rerror_rate.
UNSW-NB15	基本特征: (1)state,(2)dur,(3)sbytes,(4)dbytes,(5)sttl,(6)dttl,(7)sloss,(8)dloss,(9)service,(10)sload,(11)dload,(12)spkts,(13)dpkts(14)rate 内容特征: (15)swin,(16)dwin,(17)stcpb,(18)dtcpb,(19)smeansz(20)dmeansz,(21)trans_depth, (22)res_bdy_len 时间特征: (23)sjit,(24)djit,(25)sintpkt,(26)dintpkt,(27)tcprtt,(28)synack,(29)ackdat 流量特征: (30) proto 额外原始特征: (31)is_sm_ips_ports,(32)ct_state_ttl,(33)ct_flw_http_mthd,(34)is_ftp_login,(35)ct_ftp_cmd,(36)ct_srv_src,(37)ct_srv_dst,(38)ct_dst_ltm,(39)ct_src_ltm,(40)ct_src_dport_ltm,(41)ct_dst_sport_ltm,(42)ct_dst_src_ltm
CICIDS2017	(1)Destination Port (2)Flow Duration, (3)Total Fwd Packets, (4)Total Backward Packets, (5)Total Length of Fwd Packets, (6)Total Length of Bwd Packets, (7)Fwd Packet Length Max, (8)Fwd Packet Length Min, (9)Fwd Packet Length Mean, (10)Fwd Packet Length Std, (11)Bwd Packet Length Max, (12)Bwd Packet Length Min, (13)Bwd Packet Length Mean, (14)Bwd Packet Length Std, (15)Flow Bytes/s, (16)Flow Packets/s, (17)Flow IAT Mean, (18)Flow IAT Std, (19)Flow IAT Max, (20)Flow IAT Min, (21)Fwd IAT Total, (22)Fwd IAT Mean, (23)Fwd IAT Std, (24)Fwd IAT Max, (25)Fwd IAT Min, (26)Bwd IAT Total, (27)Bwd IAT Mean, (28)Bwd IAT Std, (29)Bwd IAT Max, (30)Bwd IAT Min, (31)Fwd PSH Flags, (32)Bwd PSH Flags, (33)Fwd URG Flags, (34)Bwd URG Flags, (35)Fwd Header Length, (36)Bwd Header Length, (37)Fwd Packets/s, (38)Bwd Packets/s, (39)Min Packet Length, (40)Max Packet Length, (41)Packet Length Mean, (42)Packet Length Std, (43)Packet Length Variance, (44)FIN Flag Count, (45)SYN Flag Count, (46)RST Flag Count, (47)PSH Flag Count, (48)ACK Flag Count, (49)URG Flag Count, (50)CWE Flag Count, (51)ECE Flag Count, (52)Down/Up Ratio, (53)Average Packet Size, (54)Avg Fwd Segment Size, (55)Avg Bwd Segment Size, (56)Fwd Header Length, (57)Fwd Avg Bytes/Bulk, (58)Fwd Avg Packets/Bulk, (59)Fwd Avg Bulk Rate, (60)Bwd Avg Bytes/Bulk, (61)Bwd Avg Packets/Bulk, (62)Bwd Avg Bulk Rate, (63)Subflow Fwd Packets, (64)Subflow Fwd Bytes, (65)Subflow Bwd Packets, (66)Subflow Bwd Bytes, (67)Init_Win_bytes_forward, (68)Init_Win_bytes_backward, (69)act_data_pkt_fwd, (70)min_seg_size_forward , (71)Active Mean, (72)Active Std, (73)Active Max, (74)Active Min, (75)Idle Mean, (76)Idle Std, (77)Idle Max, (78)Idle Min

2.3.2 RNN

传统的深度神经网络模型中数据流是单向的, 即从输入层经过隐藏层到输出层, 而忽略了样本间的时序关系, 导致流量异常检测过程中可能会丢失一些信息。因此, 产生了用于学习序列数据的神经网络 RNN。其本质是: 上一个时刻的网络状态信息将会作用于下一个时刻的网络状态。输入层信息单向流向隐藏层, 隐藏层节点的自连接和互联结构可以实现信息的充分交换, 有效解决时序依赖关系。流量数据间也存在前后关联关系, 因此可将 RNN 应用于流量异常检测中。尽管可以训练 RNN 可以捕获序列数据的时间依赖关系, 但是由于“梯度消失”问题, 难以捕获序列中的长期依赖关系 [23]。为了克服这一问题, 人们设计了特殊类型的 RNN, 最有效的是 LSTM 和 GRU。

LSTM 最初由 Hochreiter 和 Schmidhuber 提

出[24]。与传统的 RNN 相比, LSTM 精心设计了循环体结构, 可以有效地解决信息的长期依赖性, 避免梯度消失或爆炸。LSTM 具有复杂的内部结构和大量参数, 在实际应用中, 训练收敛速度较慢。

GRU 是由 Cho 等人[25]提出, 是 LSTM 的一种变体。GRU 网络简化了 LSTM 存储单元, 并使用两个门(重置门、更新门)来实现离散时间序列长期数据的选择性存储。和 LSTM 相比, GRU 在结构上较为简单, 且参数较少, 可以更快地训练, 在检测性能和收敛性方面具有很大的优势。

3 基于 CDG 的流量异常检测方法

3.1 总体框架

目前流量异常检测中主要存在流量数据噪声多, 以及实时性差的问题, 我们考虑到 DAE 可以对流量数据进行自动特征提取, 既保留数据的整体特征, 又从数据中迭代提取出更复杂可靠的特征, 降低噪声对流量异常检测的影响, 减少特征维度, 因而将其应用于流量异常检测中。对于流量数据, 根据不同的网络协议, 可以将多个网络流量字节组合成一个网络数据包, 再将多个数据包组合成一个数据流, 这些网络流量的字节、数据包和数据流与自然语言处理中的字符、句子和段落非常相似。对流量数据进行异常检测类似于将自然语言中的一个段落划分为正样本和负样本[26]。并且许多攻击过程是一个持续的过程, 其特征模式常分散在多条数据的特征中, 而非单独出现在某条数据特征中, DAE 只能学习数据特征间的线性或非线性关系, 忽略了数据序列之间的时间依赖关系。近年来研究人员发现 GRU 神经网络在学习自然语言的时序特征时具有良好的性能。因此, 本文基于类似思想, 利用GRU学习多个流量特征向量间的时间关系, 提高流量异常检测的准确率, 同时相比其他RNN变体, GRU还具有结构简单, 参数更少, 训练低维深度特征数据速度较快, 耗时更少的优点。基于此, 提出了基于 CDG 的流量异常检测方法。其总体框架如图 1 所示, 主要由数据预处理模块、流量异常检测模块和分类评估模块组成。

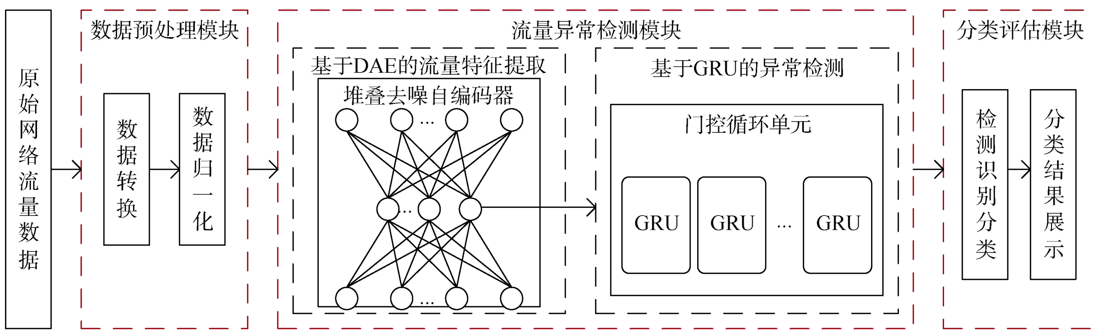


图 1 基于 CDG 流量异常检测方法框架

Figure 1 Framework Structure of Traffic Anomaly Detection Method Based on CDG

数据预处理模块: 对原始数据集预处理, 主要包括数据转换、数据归一化等操作, 使数据集转化成可供模型使用的标准输入格式, 满足输入数据的要求。训练集和测试集分别用于模型训练和测试。

流量异常检测模块: 流量异常检测模块是基于 CDG 的流量异常检测方法的重点。其处理过程包括基于 DAE 的流量特征提取和基于 GRU 的异常检测。基于 DAE 的流量特征提取算法利用 DAE 对数据进行特征提取, 获取有效且可靠特征, 降低特征维度。基于 GRU 的异常检测算法利用 GRU 对从原始流量数据中提取出的深度特征数据进行分类, 通过捕获时间前后流量数据间的关联关系, 将学习到的历史信息存储记录保留到当前时刻, 来确定网络的最终输出分类结果。使用预处理后的流量数据训练集对该模块进行训练, 训练结束后保存模型。

分类评估模块: 利用流量数据测试集对流量异常检测模块进行测试, 得到二分类和多分类的混淆矩阵, 根据混淆矩阵, 计算准确率、精确率、召回率、误报率、F1-值、AUC 值等性能参数, 并计算模型的训练时间和测试时间, 利用这些性能参数展示分类结果, 完成模型性能评估。

3.2 检测流程

本文所提出的流量异常检测方法检测流程如图 2 所示。首先使用 DAE 学习流量数据的特征。然后将DAE学习的特征输入到GRU算法进行训练, 得到训练好的 CDG 模型。最终, 测试数据被送入 CDG 模型中完成流量异常检测。基于 CDG 的流量异常检测方法的具体步骤如下:

步骤 1: 预处理流量数据集, 主要包括数据转换和数据归一化处理。

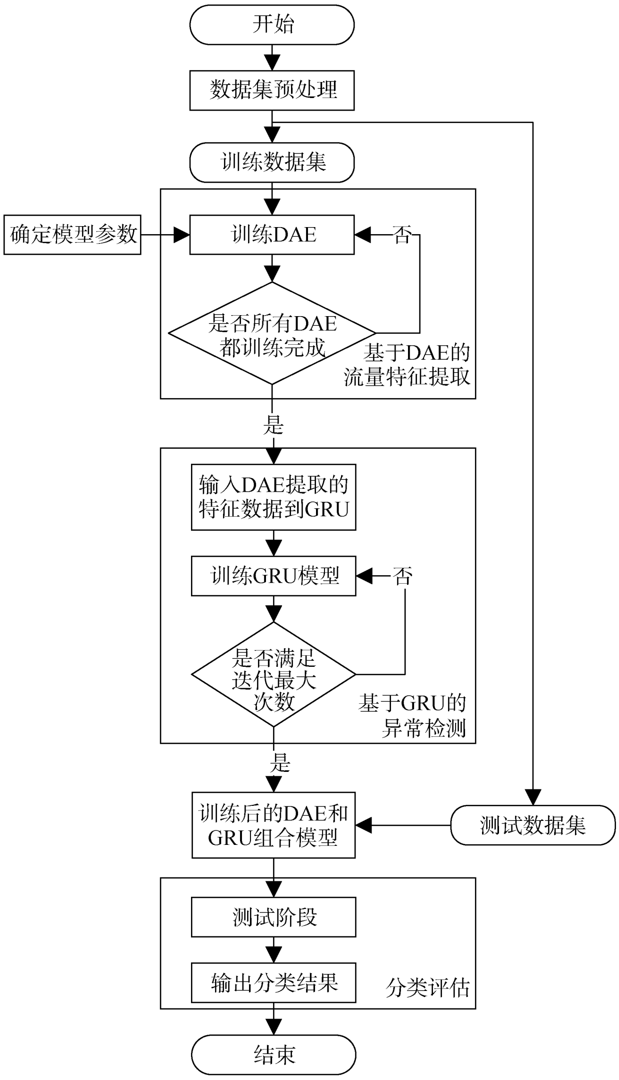


图 2 检测流程

Figure 2 Detection Precedures

步骤 2: 初始化模型参数, 确定网络模型的结构。

步骤 3: 对 DAE 进行无监督训练, 将输入输出的重构误差不断减小。

步骤 4: 重复步骤 3 直到 DAE 训练完成。

步骤 5: 利用 GRU 算法学习 DAE 提取的特征, 在学习过程中不断优化模型权值和阈值, 直到达到指定的训练时间。

步骤 6: 模型测试。输入测试数据集到训练后的 CDG 模型中, 获取测试数据集的分类结果。

4 关键算法设计

针对本文所提方法中各模块的关键算法, 在本节中进行详细描述。

4.1 数据集预处理

原始流量数据集中个别数据会包含缺失值或无效值, 如某些样本的个别特征值为“nan”或“infinity”, 即为无效样本数据, 这类样本无法用于模型训练, 因此需要进行数据清洗, 删除无效样本。

流量数据集中包含字符型数据, 如 NSL-KDD 数据集中的 3 个非数字特征(‘protocol_type’, ‘service’, ‘flag’), 需要对其进行属性映射, 将字符型数据转化成数值型。本文采用 scikit-learn 库中 LabelEncoder 函数对非数值性数据进行标签编码, 使其转为数值型数据。

为了消除数据集中特征数据由于量纲不同导致的巨大差异, 需要对数据进行归一化处理。本文使用 Min-Max 归一化方法将各特征值归一化到[0,1]区间中, 其公式如式(1)所示:

$$x' = \frac{x - Min_i}{Max_i - Min_i} \tag{1}$$

其中, x 是第 i 个属性的某一个样本值, Min_i 是第 i 个属性列的最小值, Max_i 是第 i 个属性列的最大值, x' 是 x 归一化后的值。数据集预处理过程如算法 1 所示。

算法 1.数据集预处理

输入: 原始数据集(NSL-KDD、UNSW-NB15、CICIDS2017)

输出: 预处理后的数据集(M*N 维, M 为预处理后的样本数, N 为预处理后各数据集的特征数)

1)取出数据集中的所有数据

2)对数据集中的数值进行判断, 删除包含缺失值或无效值的样本

3)对特征进行判断

4) 对于 NSL-KDD 数据集, 如果特征是 protocol_type、service 或 flag, 则取出其所有特征值

5)创建标签编码器, 对三个特征的特征值进行编码,得到 NSL-KDD 编码后的数据集

6)对于 UNSW-NB15 数据集, 如果特征是 proto、service 或 state, 则取出其所有特征值

7) 创建标签编码器, 对三个特征的特征值进行编码,得到 UNSW-NB15 编码后的数据集

8)对编码后的数据集, 计算每个特征的最小值和最大值

9)对数据集中所有数据, 利用公式(1)计算三个数据集分别标准化后的数据集

4.2 基于 DAE 的流量特征提取算法

处理大规模高维数据和降低流量数据噪声影响是流量异常检测过程的第一个重要问题。传统的 PCA 等线性降维方法无法学习数据中的非线性信息,

从而导致后续流量异常检测准确率不高。而 DAE 能够表征线性变换和非线性变换, 学习到流量特征的深层表示, 提高特征提取的鲁棒性, 降低特征维度, 以减少后续流量异常检测时间复杂性。因此提出基于 DAE 的特征提取算法, 其训练过程如图 3 所示。

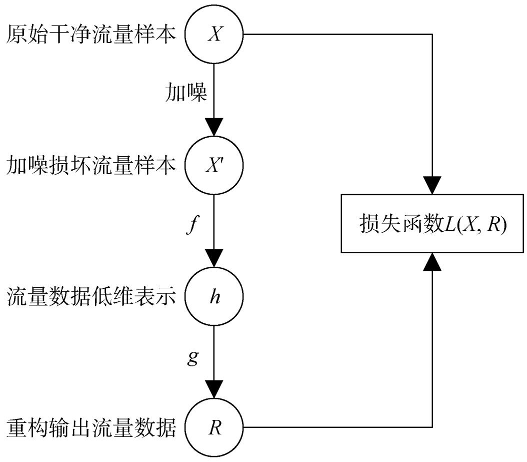


图 3　DAE 训练过程
Figure 3　Training DAE process

设 X 是流量数据集训练样本中的一个样本, 对其加噪以获取损坏流量样本 X'。通过编码器对 X'压缩变换, 构造隐藏层的低维神经元 h, 即流量数据的低维深层表示。解码器 g 将 h 解码重构得到输出数据 R。最小化原始未加噪声的流量数据 X 与经过解码重构的输出流量数据 R 之间的误差来调整和优化网络参数。其重构误差的表达式如式(2)所示:

$$L(x, g(f(x'))) = \|x - g(f(x'))\|^2 \tag{2}$$

编码器将输入层映射到隐藏层公式如式(3)所示:

$$h = f(WX' + b) \tag{3}$$

解码器将隐藏层映射到输出层公式如式(4)所示:

$$R = g(\hat{W}h + \hat{b}) \tag{4}$$

其中 W 和 $\hat{W}$ 为权重矩阵, b 和 $\hat{b}$ 为偏置值, 且 $\hat{W} = W^T$

其损失函数 $L(W, b)$, 其公式如式(5)所示:

$$L(W,b) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\|x_i - g(f(x_i'))\|^2\right) + \frac{\lambda}{2}\sum_{l=1}^{m_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(W_{ji}^l)^2 \tag{5}$$

其中, 第一项为输入输出的均方误差, n 是输入层神经元的个数, x_i 表示输入的第 i 个流量样本, f 是编码器, g 是解码器, $g(f(x'))$ 是对原始流量样本的重构输出。为防止训练过程出现过拟合, 在输入输出均方误差基础上, 引入第二项正则项, 又称权重衰减项, λ 是权重衰减系数, 用于减少权重幅值, W_{ji}^l 是第 l 层第 i 个神经元与第 $l+1$ 层第 $j+1$ 个神经元之间的连接权重, m_l 是网络层数, s_l 是第 l 层神经个数。

基于 DAE 的流量特征提取算法使用流量数据训练集对 DAE 进行无监督训练, 利用小批量梯度下降算法更新 DAE 网络的权重参数, 求解出目标函数的最优解, 以实现从高维数据中得到低维可靠的深层特征表示, 降低数据维度。算法训练过程如算法 2 所示。

算法 2. 基于 DAE 的流量特征提取算法

输入: 预处理后的流量数据 X, 加入一定比例噪声的流量数据 X', 每个小批量的样本数 *batchsize*, 迭代次数 l
输出: 流量数据的低维表示 Y
1)初始化参数权重偏置
2)FOR i in l do
3)FOR j in *batchsize* do
4)计算每个小批量中第 j 个样本对应隐藏层向量的第 k'个元素和重构向量的第 k 个元素
5)END FOR
6)利用上一循环计算出的元素, 通过参数更新规则最小化式(5)计算要更新的参数
7)END FOR
8)训练结束, 得到最终网络参数, 利用其计算, 取得流量数据的低维深层表示 Y

算法复杂度分析: 设小批量样本大小为 *batchsize*, 迭代次数为 l, 隐藏层总数为 N, 输入输出层中神经元数分别为 I 和 O, 隐藏层中神经元数为 h_j, j 为第 j 个隐藏层。对于具有单个隐藏层的 DAE 模型, 应在训练前阶段执行 $I * h$ 和 $h * O$ 矩阵乘法操作。因此, 计算样本的时间复杂度为 $o(I * h + h * O)$ 。具有多个隐藏层的 DAE 算法, 由于使用小批量梯度下降方法, 其整体时间复杂度为 $o(bachisize * l * 2(I * h_1 + \sum_{j=1}^{N-1} h_j * h_{j+1} + h_N * O))$ 。

4.3　基于 GRU 的异常检测算法

流量数据结构化信息是影响流量异常检测的重要特征, 这些特征在基于传统深度学习的流量异常检测中取得一定的效果。而流量数据的另一个重要特征是时序关系, 某些攻击行为会表现为流量特征在一段时间内的持续变化。传统的神经网络(如 CNN)输出只与当前的输入数据有关, 难以捕获流量数据间的时序关系。GRU 可以捕获时间前后数据间的关系, 是处理序列数据的有效工具, 具有更强的信息选择能力和序列学习能力, 适合解决检测长期依赖性攻击的问题(如对于 NSL-KDD 中的 probe 攻击, 可

利用GRU学习其基于主机的网络流量统计特征和基于时间的网络流量统计特征来捕获其时间前后的关联关系进行检测)。因此本文将具有判别能力的GRU作为分类器，设计了基于GRU的异常检测方法，进行流量异常检测，从而输出数据的分类结果，其前向计算单元如图4所示。

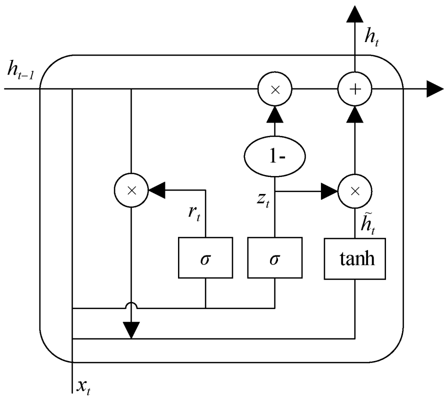


图 4　GRU 前向计算单元

Figure 4　GRU forward calculation unit

其更新门、重置门和存储单元的状态方程如公式(6)-(9)所示:

$$z_t = \sigma(W_z \cdot x_t + U_z \cdot h_{t-1} + b_z) \tag{6}$$

$$r_t = \sigma(W_r \cdot x_t + U_r \cdot h_{t-1} + b_r) \tag{7}$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot x_t + r_t * (U_{\tilde{h}} \cdot h_{t-1}) + b_{\tilde{h}}) \tag{8}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{9}$$

其中时间步 t 的输入 x_t 是DAE输出深度特征向量序列，h_t 是时间步 t 的隐层状态，σ 是sigmoid函数，r_t 是重置门，用于控制丢弃前一时刻状态信息的程度，重置门的值越小，说明丢弃的越多。z_t 表示更新门。更新门用于控制前一时刻的流量状态信息被带入到当前状态中的程度，更新门的值越大，说明前一时刻的状态信息带入越多。也就是说，当流量数据流入GRU单元时，更新门 z_t 和重置门 r_t 可以控制是否可以通过以及可以通过多少上一时间步的流量数据状态信息。

基于GRU的异常检测方法通过学习DAE提取的低维深度流量特征，实现异常样本的快速准确检测。为实现GRU网络反向传播过程中梯度下降的快速收敛性，采用Adam算法作为优化算法最小化损失，根据计算出的损失不断调整权重和偏差，算法训练过程具体如算法3所示:

算法 3. 基于 GRU 的异常检测方法

输入: 特征提取后的流量数据 X, 迭代次数 l

输出: 输出预测结果 Y

续表

算法 3. 基于 GRU 的异常检测方法

1)FOR i in l do

2) 计算输出层的输入

3) 计算输出层的输出

4) 得到最终输出后，计算单个训练样本在某个时间步的损失

5) 计算单个训练样本在所有时间步的损失

6) 采用 adam 优化方法最小化损失函数，更新参数

7) 利用公式(6)计算更新门 z_t

8) 利用公式(7)计算重置门 r_t

9) 利用公式(8)-(9)计算隐藏层状态 h_t

10)END FOR

11)根据训练得到的网络参数，保存 GRU 模型

12)利用训练好的 GRU 模型计算测试数据的预测结果 Y

算法复杂度分析：设输入神经元维度为 x，输出神经元维度为 h，算法的时间复杂度为 $o(h*(h+x)+h)$。

5　实验设计和结果分析

5.1　实验环境和超参数设置

为了验证本文方法的检测性能，我们不仅对早期的流量数据集NSL-KDD进行了异常检测，还对近年来新兴的流量数据集UNSW-NB15和CICIDS2017进行了异常检测，充分验证了本文方法对未知攻击和新型网络攻击的检测能力。所有实验均在Windows 10 PC Intel(R) Core (TM) i7-8700 CPU @ 3.20GHz 3.19GHz, 32.0GB RAM环境下进行，使用Python 3.5实现本文算法。

参数的选择对模型的学习能力和检测效果有很大的影响。本文基于NSL-KDD、UNSW-NB15、CICIDS2017数据集进行实验。三个数据集的实验中，NSL-KDD特征数量和UNSW-NB15特征数量分别为41和42个，在NSL-KDD数据集上确定模型的最优参数，并将这些参数应用于NSL-KDD和UNSW-NB15数据集。对于NSL-KDD数据集，CDG模型的输入层包含41个神经元；对于UNSW-NB15，输入层包含42个神经元。DAE包含三个隐藏层，隐藏层各层节点数为[34, 24, 14]，其输入层与隐含层、隐含层与输出层均采用全连接方式。对于CICIDS2017数据集，CDG模型的输入层包含78个神经元，DAE隐藏层各层节点数为[64, 48, 32]。将DAE提取后的深度特征数据作为GRU的输入，GRU包含128个单元，

之后连接一个全连接层作为输出层, 将 GRU 输出的向量转换成标签向量的维度, 输出层包含 1 个单元, 用于区分网络的正常连接和攻击类型。学习速度和迭代次数由实际经验决定。采用小批量梯度下降法和 adam 算法对模型进行训练。经过多次模型超参数的调整, 使学习效率最大化、学习效果最佳的超参数配置如表 2 所示。

表 2 超参数配置

Table 2 Hyper-parameter configuration

超参数	值
Batch Size	16
Epoch of AE	10
AE 层数	3
Number of units for GRU	128
Epoch of GRU	10
Activation	RELU
Learning rate	0.001

5.2 数据集

(1) NSL-KDD: 该数据集是对 KDD Cup99 数据集的改进。所有的数据都来自一个模拟的美国空军局域网。实验的训练数据为 7 周网络流量; 测试数据为 2 周网络流量。测试数据和训练数据具有不同的概率分布, 测试集中包含了一些训练集中没有出现的攻击类型, 使得流量异常检测更加真实。与原始的 KDD Cup99 数据集相比, NSL-KDD 数据集删除了其中的冗余数据和重复记录, 更适用于异常检测。其详细描述见表 3。

(2) UNSW-NB15: 该数据集由澳大利亚安全实验室在 2015 年收集的网络流量数据生成, 是一个综合性的网络攻击流量数据集, 包括训练数据集和测试数据集。它由正常数据和 9 种攻击流量数据构成。其详细描述见表 4。

表 3 NSL-KDD 数据集的详细描述

Table 3 Details of the NSL-KDD Dataset

数据类型	描述	Train	Test
正常数据	正常连接记录	67343	9711
攻击数据	DoS: 攻击者的目的是破坏网络资源	45927	5741
	Probe: 攻击者通过探测或扫描网络或系统来获取其配置的详细统计信息	11656	1106
	U2R: 攻击者获得主机的 root 或超级用户访问权限	52	67
	R2L: 来自远程计算机的非法访问	995	2199
总计		125973	22544

表 4 UNSW-NB15 数据集的详细描述

Table 4 Details of the UNSW-NB15 Dataset

数据类型	描述	Train	Test
正常数据	正常连接记录	5600	37000
攻击数据	Analysis:一种通过端口扫描和 web 应用渗透的攻击类型	2000	677
	Backdoor:一种绕过安全机制进入计算机程序或系统的攻击	1764	583
	DoS:一种消耗大量网络资源, 导致网络无法提供正常服务的攻击	12264	4089
	Exploit:一种通过触发系统或软件中的漏洞来控制目标系统的攻击	33532	11132
	Fuzzers:一种通过输入大量随机生成的数据来发现安全漏洞而使程序或者网络暂停服务的攻击	18184	6062
	Genetic:一种对抗所有分组密码的攻击技术	40000	18871
	Reconnaissance:一种收集所有信息以逃避安全控制的攻击	10491	3496
	Shellcode: 一种通过发送利用特定漏洞的代码来控制目标计算机的攻击	1133	378
	Worms: 一种通过不断复制自身来传播的恶意计算机病毒	130	44
总计		175300	82332

(3) CICIDS2017: 该数据集由加拿大网络安全研究所于 2017 年在真实环境下一周收集到的网络数据集, 包含最新的常见攻击。数据集包含 8 个文件, 共 3119345 条样本, 78 个特征, 其中包含 15 个类别标签(1 个正常标签和 14 个攻击标签)。由于数据集中包含一些标签缺失样本和特征缺失样本, 对其删除后数据集的详细描述见表 5。

5.3 实验评估指标

本文利用准确率(Accuracy)、精确率(Precision)、召回率(Recall)、假阳率(False Positive Rate, FPR)、 F1-Score、AUC(Area Under Curve, ROC 曲线下的面积)值对各模型的分类能力进行评估。这些值基于表 6 混淆矩阵得到。真正例(True Positive, TP): 真实类别和预测类别均为攻击的样本数; 假负例(False Negative, FN): 真实类别为攻击, 预测类别为正常的样本数; 假正例(False Positive, FP): 真实类别为正常, 预测类别为攻击的样本数; 真负例(True Negative, TN): 真实类别和预测类别均为正常的样本数。

表 5　CICIDS2017 数据集的训练和测试连接记录
Table 5　Training and testing connection records of CICIDS2017

数据类型	描述	连接记录的数量
正常数据	BENIGN	2359087
攻击数据	DoS Hulk	231072
	PortScan	158930
	DDoS	41835
	DoS GoldenEye	10293
	FTP-Patator	7938
	SSH-Patator	5897
	DoS Slowloris	5796
	DoS Slowhttptest	5499
	Bot	1966
	Web Attack-Brute Force	1507
	Web Attack-XSS	652
	Infiltration	36
	Web Attack-SQL Injection	21
	Heartbleed	11
总计		2830540

评估指标的定义如下:

准确率(Accuracy): 整个测试数据集中被正确识别的连接记录的比例, 其定义如式(10)所示:

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

精确率(Precision): 预测为攻击的样本中预测正确的比例, 其定义如式(11)所示:

$$Precision=\frac{TP}{TP+FP} \tag{11}$$

召回率(Recall): 又称真正率(True positive rate, TPR), 表示所有攻击样本中, 被正确预测的比例, 其定义如式(12)所示:

$$Recall=\frac{TP}{TP+FN} \tag{12}$$

F1-Score: 精度和召回率的谐波平均值。如果 F1-Score 较高, 则检测模型更好(F1-Score∈[0,1])。F1-Score 的定义如式(13)所示:

$$F1-score=\frac{2*precision*recall}{precision+recall} \tag{13}$$

假阳率(False Positive Rate, FPR), 又称误报率, 即所有正常样本中, 被预测为攻击样本的比例, 其定义如式(14)所示:

$$FPR=\frac{FP}{FP+TN} \tag{14}$$

ROC(Receiver Operating Characteristics) 曲线: ROC 根据 y 轴 TPR 与 FPR 之间的权衡绘制。AUC 为 ROC 曲线下面积。如果 AUC 较高, 机器学习模型就会更好。AUC 定义如式(15)所示:

$$AUC=\int_0^1 \frac{TP}{TP+FD}\, d\frac{FP}{TN+FP} \tag{15}$$

表 6　混淆矩阵
Table 6　Confusion Matrix

实际	预测	
	攻击	正常
攻击	*TP*	*FN*
正常	*FP*	*TN*

5.4　实验过程及结果分析

在网络流量异常检测领域, 任何检测方法最终都将应用于不同的实际场景, 但由于应用环境、数据大小和网络环境不同, 没有唯一的标准来衡量算法的优越性。因此将基于 CDG 的流量异常检测方法应用于 NSL-KDD、UNSW-NB15 和 CICIDS2017 数据集进行实验分析和验证。

5.4.1　模型鲁棒性测试

当数据被加入噪声, 受到破坏时, 可能会影响入侵检测方法的检测准确率。在实际工程应用中, 如在对网络流量数据进行获取、解析的过程中, 难免会引入大量噪声, 或由于传输错误、人为操作错误等原因也会引入噪声。流量数据中的噪声具有未知性, 真实环境下的噪声往往是多源复合而成的, 高斯噪声被研究人员认为是对真实噪声的最好模拟, 是一种测试模型鲁棒性的典型方法[27-29]。因此, 为验证本文方法的检测鲁棒性, 对实验数据经过预处理后加入高斯白噪声, 它服从正态分布, 均值为 0, 方差为 1。加入随机噪声后, 数据维度保持原状。当被测数据属性特征破坏率(即噪声因子)为 0.1,0.2,0.3,0.4,0.5 时, 对比 AE、DAE、CDG 模型对 NSL-KDD 数据集的检测准确率, 实验结果如图 5 所示。在实验过程中由于 AE 和 DAE 不具有分类特性,将 AE 与 softmax 分类器结合成为自编码网络, 将 DAE 与 softmax 分类器结合成为去噪自编码器网络。在后文的对比实验中, 也使用同样的 AE 和 DAE 网络进行实验。

由图 5 可知当被测数据加噪、特征遭到破坏时，基于 AE 的异常检测方法检测准确率明显降低，且随着破坏率增加，检测准确率不断下降；而基于 DAE 的异常检测方法检测准确率保持稳定，在 91%左右浮动，增大被测属性特征破坏率也不会导致检测准确率下降，表明 DAE 具有较强鲁棒性；基于 CDG 的流量异常检测方法保持同样的检测鲁棒性，且检测准确率最高，在 98%左右浮动，表明本文方法在取得良好检测鲁棒性的同时具有最高的检测准确率。

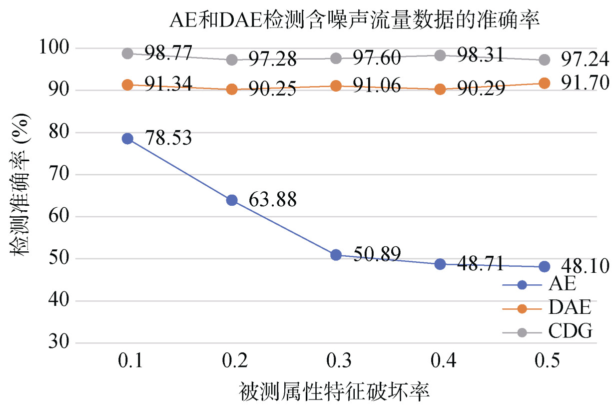


图 5 模型的鲁棒性测试结果

Figure 5 Model robustness test results

5.4.2 基于 NSL-KDD 数据集的二分类实验

该实验为基于 NSL-KDD 数据集的二分类实验，将所有的攻击类型纳入异常类，正常数据归为正常类。实验过程与 3.2 节中的流程基本一致，首先对数据集中所有数据按照 4.1 节中的预处理过程进行数据预处理，然后对数据进行异常检测。由于相关工作中的一些现有流量异常检测方法主要是针对特定攻击类型(如文献[11]仅针对 DoS 攻击、文献[14]仅针对僵尸网络流量)，或针对特定场景(如文献[18]针对物联网中的攻击进行检测等)，无法直接与本文方法进行对比。典型的传统机器学习模型和深度神经网络对使用场景、研究环境、研究问题的更具普适性。因此，为验证本文方法的有效性，我们将本文方法与典型机器学习算法 RF、DT 和常见深度学习算法[30-31]AE、DAE、极限学习机(Extreme Learning Machine, ELM)、深度神经网络(Deep Neural Networks, DNN)、LSTM、GRU 和组合模型 DAE-LSTM 及相关工作中现有较为流行的且检测性能较好的两种模型 GRU-MLP[10]和 DBN[15] 对 NSL-KDD 数据集二分类的检测性能进行比较。

为避免单一测试集进行检测的偶然性和随机性，本文首先采用十折交叉验证法对各类模型的检测性能进行比较。随机分割整个 NSL-KDD 数据集(包含训练集和测试集)为 10 等份，每次选取其中 9 份用于训练，剩余 1 份用于检测，最终检测结果取 10 次试验平均值，实验结果如表 7 所示。基于 AE 的异常检测方法检测准确率和精确率最差，不超过 90%，误报率最高；与 AE 相比，基于 DAE 的流量异常检测方法检测性能均有提升。ELM 和 DBN 模型的检测检测准确率、精确率、召回率、F1-score 以及 AUC 值相比 AE 和 DAE 模型明显提升，但其误报率最差。基于 RF 和基于 DT 的异常检测方法以及基于 DNN、基于 LSTM、基于 GRU 的异常检测方法的误报率都小于 7%，其误报率相比前几个模型明显下降。与组合模型 DAE-LSTM、文献[10]的 GRU-MLP 和单一的机器学习、深度学习模型进行流量异常检测相比，本文方

表 7 基于 NSL-KDD 数据集二分类性能测试检测结果

Table 7 NSL-KDD binary test results

模型	accuracy	precision	recall	FPR	F1-score	AUC
RF	93.55%	94.36%	95.08%	5.62%	94.72%	93.27%
DT	93.15%	98.06%	93.27%	6.75%	95.60%	94.85%
AE	86.56%	82.42%	94.20%	21.68%	87.91%	86.26%
DAE	91.27%	91.88%	91.26%	8.70%	91.56%	91.27%
ELM	94.04%	98.31%	93.18%	41.66%	95.68%	81.33%
DNN	96.00%	96.21%	95.69%	3.70%	95.95%	96.00%
DBN	93.39%	98.67%	92.97%	49.88%	95.73%	71.01%
LSTM	96.20%	94.02%	96.95%	6.76%	96.43%	96.09%
GRU	96.94%	96.40%	95.73%	3.36%	95.76%	96.90%
DAE-LSTM	96.92%	96.96%	96.79%	4.80%	96.87%	96.92%
GRU-MLP	94.73%	96.95%	94.05%	3.13%	95.03%	94.70%
CDG	98.56%	98.72%	96.85%	1.36%	97.63%	97.60%

法的检测准确率、精确率、召回率、F1-score 以及 AUC 值都最高, 误报率最低, 验证了 CDG 模型相比其他模型具有更好的综合检测性能。

原始 NSL-KDD 数据集独立地划分了训练集和测试集, 包含正常数据和 4 种攻击大类, 这 4 种攻击大类又可分为 39 种攻击类型, 其中有 22 种攻击类型同时分布于训练集和测试集中, 其余 17 种攻击类型仅出现在测试集中。因此, 为测试各类方法对未知异常的检测能力, 本文利用预处理后的原始训练集进行模型训练, 训练完成后使用预处理后的原始测试集进行测试。将测试结果同上文提及的几种检测方法进行比较。实验结果如表 8 所示, 基于 GRU-MLP 的流量异常检测方法具有 93.08%的检测准确率, 本文方法的准确率相比其提升了 1.88%。基于 CDG 的流量异常检测方法在各个性能指标上相比其他方法都表现出更好的性能。由于原始测试集中包含一些训练集中没有的攻击, 数据分布存在一定差异, 所以各类方法的检测准确率相比上一实验都有所下降, 但本文仍以 94.83%的检测准确率、98.14%的精确率、94.41%的召回率、96.24%的 F1-score 以及 95.78%的 AUC 值以及 2.83%的误报率进行二分类, 这也表明本文方法对于未知的攻击类型也具有很好的检测能力。

表 8 基于 NSL-KDD 数据集的未知异常二分类检测结果

Table 8 Unknown anomaly binary classification test result based on NSL-KDD

模型	accuracy	precision	recall	FPR	F1-score	AUC
RF	84.24%	86.11%	84.32%	14.35%	85.21%	86.37%
DT	76.21%	95.58%	61.06%	35.27%	74.51%	80.49%
AE	83.60%	77.20%	87.88%	19.65%	82.19%	84.12%
DAE	84.65%	76.01%	94.06%	22.47%	84.08%	85.80%
ELM	74.49%	86.67%	65.23%	8.82%	74.44%	76.59%
DNN	74.02%	58.91%	93.05%	38.71%	73.03%	78.67%
DBN	71.60%	89.72%	56.61%	7.19%	69.42%	74.94%
LSTM	92.52%	88.44%	95.90%	10.39%	92.10%	92.25%
GRU	92.75%	95.72%	87.06%	2.95%	91.19%	92.06%
DAE-LSTM	91.01%	95.67%	93.50%	3.06%	94.00%	94.89%
GRU-MLP	93.08%	92.64%	91.46%	5.49%	92.05%	92.98%
CDG	94.83%	98.14%	94.41%	2.83%	96.24%	95.78%

5.4.3 基于 UNSW-NB15 数据集的二分类实验

该实验为基于 UNSW-NB15 数据集的二分类实验。实验过程与 3.2 节中的流程基本一致, 首先对数据集中所有数据按照 4.1 节中的预处理过程进行数据预处理, 然后对数据进行异常检测。

按照 UNSW-NB15 数据集中的原始分配情况利用预处理后的数据进行模型的训练和测试。将上文提及的几种检测方法同本文方法的检测结果进行对比, 实验结果如表 9 所示。在大多数情况下, CDG 模型实现了比其他模型更好的检测性能, 以 95.09%的

表 9 基于 UNSW-NB15 数据集二分类检测结果

Table 9 UNSW-NB15 binary test results

模型	accuracy	precision	Recall	FPR	F1-score	AUC
RF	82.30%	63.55%	79.01%	24.32%	70.44%	72.33%
DT	76.38%	72.16%	95.12%	12.72%	82.43%	84.85%
AE	76.88%	79.66%	85.77%	38.91%	82.60%	73.43%
DAE	82.88%	85.70%	87.91%	26.04%	86.79%	80.93%
ELM	76.51%	72.37%	92.75%	13.20%	81.30%	81.40%
DNN	73.42%	72.02%	84.59%	31.54%	77.80%	66.39%
DBN	86.98%	75.94%	87.06%	25.78%	75.04%	76.01%
LSTM	92.71%	93.21%	95.56%	12.36%	94.37%	91.59%
GRU	93.78%	94.18%	94.56%	10.37%	94.37%	92.09%
DAE-LSTM	93.41%	96.50%	94.61%	9.27%	95.55%	93.14%
GRU-MLP	93.98%	96.73%	94.84%	20.67%	95.78%	87.09%
CDG	95.09%	98.89%	95.91%	7.53%	97.37%	96.09%

检测准确率、98.89%精确率、95.91%召回率、97.37%的 F1-score 以及 96.09%的 AUC 值以及 7.53%的误报率进行检测，在各个性能指标上也都表现出最好效果。检测准确率相比 RF 算法提升了 15.49%，相比 DT 算法提升了 18.71%，相比其他深度学习方法算法中检测准确率最高的 GRU-MLP 模型提升了 1.18%。这充分表明本文方法对于新型网络攻击也具有很好的检测能力。

5.4.4 基于 CICIDS2017 数据集的二分类实验

本文使用 CICIDS2017 数据集对我们的模型和方法在检测新型网络攻击方面进行了进一步验证。实验过程与在 NSL-KDD 数据集一致，实验结果如表 10 所示。由实验结果可知，本文的方法的准确率比 DT 提高了 1.18%，比 DAE 提高了 17.14%，比 GRU 提高了 2.59%，效果明显提升。在其他检测性能上，本文方法也表现出最好的性能，具有 99.12%精确度，97.80%的召回率，0.85%的检测误报率，98.46%的 F1-score 和 98.47%的 AUC 值，综合性能较好。

5.4.5 基于 NSL-KDD 数据集的多分类实验

当数据集中的攻击样本细分为不同的攻击类型时，此实验为多类分类实验。NSL-KDD 数据集中的攻击类型分为 Dos、Probe、U2R、R2L 四个类别，因此本实验为基于 NSL-KDD 数据集的多分类实验。由于在以上实验中，除本文方法外，相比其他检测方法，基于 GRU 的流量异常检测方法检测准确率、综合性能最好，且本文方法在分类方面也使用了基于 GRU 的异常检测方法，因此仅比较本文方法与基于 GRU 的流量异常检测方法在多分类方面的检测性能。实验过程与 5.4.2 节中的流程基本一致，首先对数据集中所有数据按照 4.1 节中的预处理过程进行数据预处理，然后将对数据进行训练。检测结果如图 6-10 所示。图 6 为本文方法对测试集进行五分类得到的混淆矩阵，可明显看出，本文方法在区分正常和攻击数据方面取得了不错的效果(14851 条测试数据中仅 45 条误报、51 条漏报)。

表 10 基于 CICIDS2017 数据集二分类检测结果

Table 10 CICIDS2017 binary test results

模型	accuracy	precision	recall	FPR	F1-score	AUC
RF	97.54%	97.60%	97.62%	3.67%	97.61%	97.34%
DT	97.32%	97.61%	97.63%	3.26%	97.62%	96.81%
AE	81.08%	83.38%	87.95%	31.13%	85.61%	78.41%
DAE	84.06%	85.72%	90.10%	26.67%	87.87%	81.72%
ELM	94.93%	93.86%	95.72%	12.36%	94.78%	93.12%
DNN	92.67%	96.19%	92.20%	6.48%	94.15%	92.86%
DBN	97.68%	99.61%	86.93%	13.65%	92.74%	94.76%
LSTM	94.79%	92.60%	94.64%	2.18%	94.91%	94.79%
GRU	96.98%	98.64%	93.23%	1.28%	96.86%	95.97%
DAE+LSTM	96.23%	98.64%	93.74%	1.28%	96.13%	96.22%
GRU-MLP	95.98%	96.14%	95.79%	3.84%	95.97%	95.98%
CDG	98.47%	99.12%	97.80%	0.85%	98.46%	98.47%

预测值

实际值	Normal	DoS	Probe	U2R	R2L
Normal	6420	45	0	0	0
DoS	0	4796	37	0	0
Probe	0	2	1596	0	0
U2R	0	0	0	89	43
R2L	51	40	74	2	1656

图 6 CDG 模型对 NSL-KDD 测试集五分类的混淆矩阵

Figure 6 CDG model's confusion matrix for the multi classification of NSL-KDD test set

图 7 为本文方法同基于 GRU 的异常检测方法对 NSL-KDD 测试集多分类的检测准确率对比图，可以发现本文方法对正常数据的检测准确率为 99.30%，对 DoS 的检测准确率为 99.21%，对 Probe 的检测准确率为 98.31%，对 U2R 的检测准确率为 94.40%，对 R2L 的检测准确率为 94.92%。在检测准确率方面对流量数据多分类取得了不错的效果，且都比基于 GRU 的异常检测方法准确率高。而两种方法在 U2R 类型的检测准确率都低于其他攻击类型的检测准确率，这是由于数据集中 U2R 类型数据数量较少，在训练集中仅有 52 条，且此类数据并不像其他攻击类

型表现出明显频繁的序列模式，仅表现在嵌入在数据包的数据负载里面(类似R2L或正常数据)，因此模型训练不充分，无法准确学习其特征，导致将 U2R 攻击类型数据准确检出的数量较少，针对这类攻击的准确检测能力不强。

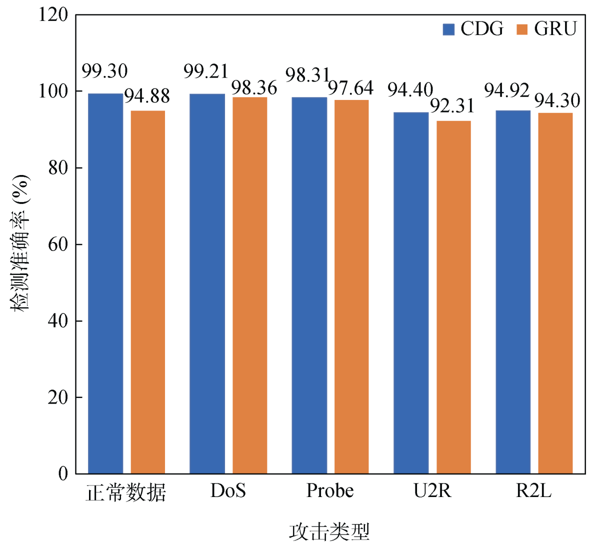


图 7　CDG 模型与 GRU 模型对 NSL-KDD 测试集多分类的检测准确率比较

Figure 7　Comparison of accuracy of CDG and GRU for NSL-KDD classification

图 8 为本文方法同基于 GRU 的异常检测方法对 NSL 测试集五分类的检测精确度对比图。可以直观发现本文方法对正常数据的检测精确度为 99.21%，对 DoS 检测精确度为 98.20%，对 Probe 的检测精确度为 93.50%，对 U2R 的检测精度为 97.80%，对 R2L 的检测精确度为 97.50%。可以看出本文方法的检测精确度均优于基于 GRU 的异常检测方法。对于 U2R 攻击类型，其训练数据较少，训练不充分，导致类别模糊，基于 GRU 的异常检测方法在流量异常检测过程中，将一些其他类别数据检测为 U2R 类别。并且在测试集中 U2R 类别的测试数据数量较少。根据精确率的定义可知，U2R 的精确率是检测为 U2R 的所有数据中正确检测为 U2R 的比例，因此其检测精确率极低。

图 9 为本文方法同基于 GRU 的异常检测方法对 NSL-KDD 测试集多分类的检测召回率对比图。可以直观发现本文方法对正常数据、DoS、Probe 和 R2L 攻击的检测召回率均高于 90%，但对 U2R 的检测召回率为 67.42%，这是由于U2R类型的训练数据较少，攻击特征类似 R2L 攻击，且 U2L 作为一种攻击大类，又包含若干攻击小类，其某些攻击小类并未包含在训练集中，导致模型在训练过程中不能十分准确地学习其特征，因此在检测时将 U2R 类型的部分数据检测为其他类别，检测召回率不高。

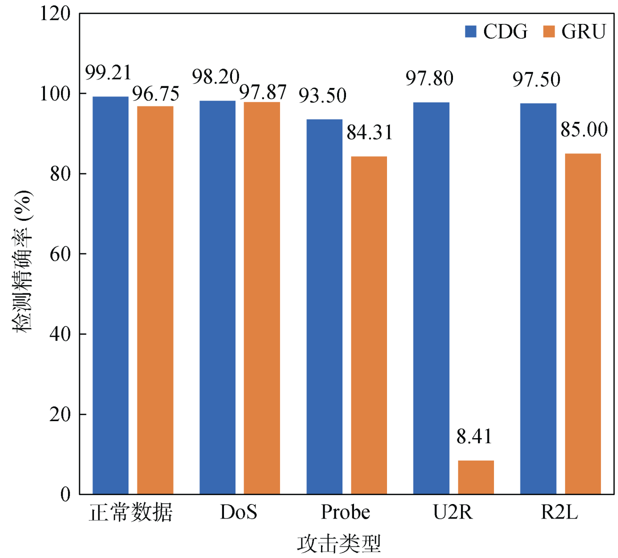


图 8　CDG 模型与 GRU 模型对 NSL-KDD 测试集多分类的检测精确率比较

Figure 8　Comparison of precision of CDG and GRU for NSL-KDD classification

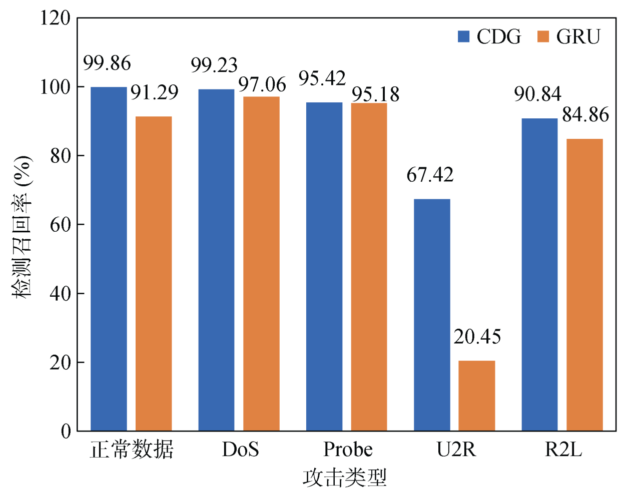


图 9　CDG 模型与 GRU 模型对 NSL-KDD 测试集多分类的检测召回率比较

Figure 9　Comparison of recall of CDG and GRU for NSL-KDD classification

图 10 为本文方法同基于 GRU 的异常检测方法对 NSL 测试集五分类的检测误报率对比图。可以直观发现本文方法对正常数据的检测误报率为 0.29%，对 DoS 检测误报率为 0.66%，对 Probe 的检测误报率为 0.87%，对 U2R 的检测误报率为 0.02%，对 R2L 的检测误报率为 0.01%。可以看出本文方法的检测误报率均低于基于 GRU 的异常检测方法。验证了本文方法具有较低的误报率。

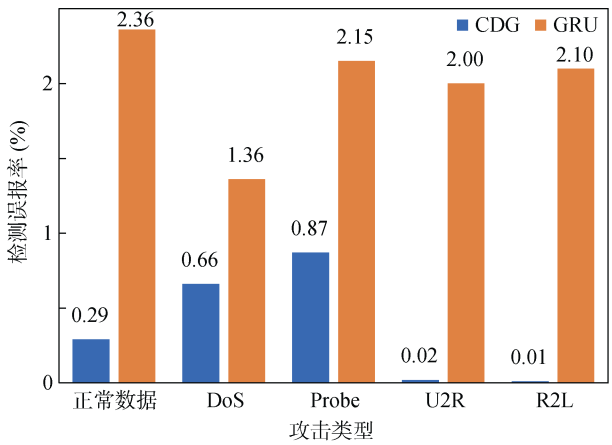


图 10　CDG 模型与 GRU 模型对 NSL-KDD 测试集多分类的检测误报率比较

Figure 10　Comparison of FPR of CDG and GRU for NSL-KDD classification

综合以上分析，本文方法在准确率、精确率、召回率、误报率四个评价指标上，相对基于 GRU 的异常检测方法均获得了不错的提升，准确率最高可达 99.30%，误报率均不超过 1%。但在区分不同攻击类型、面对样本不平衡方面还有进一步提升空间。

5.4.6　模型运行时间分析

将 CDG 模型应用于流量异常检测中，除了确保该模型能够实现更好的检测性能外，还需考虑该模型的时间性能。因此在实验过程中，对比本文方法同其他方法在三个数据集上进行训练和测试的时间指标。对于训练时间，本文将各个深度学习模型从训练开始到其在验证集上的损失函数趋于稳定，在验证集上达到最好效果时的时间，作为其训练时间进行对比。对于 AE 和 DAE 模型，当每次训练最大数据量 batchsize 为 128，训练轮数 epoch 为 10 时，其 loss 趋于稳定；同样，对于 ELM，其 batchsize 为 64，epoch 为 20；对于 DNN，其 batchsize 为 64, epoch 为 100；对于 DBN，其 batchsize 为 128，epoch 为 50；对于 LSTM 和 GRU，其 batchsize 为 128， epoch 为 20；对于 GRU-MLP 和 DAE-LSTM，其 batchsize 为 32，epoch 为 20。不同模型的训练时间和测试时间分别如表 11 和表 12 所示。

从表 11 中可以发现，针对三个数据集, AE、DAE 和 ELM 模型的训练时间最短, CDG 模型相比这三种模型需要更长的训练时间，但是相差不大。并且与其他机器学习和深度学习方法相比，本文方法的训练时间大大减少，更好地满足实时性的要求。

各个模型在测试集上的检测时间如表 12 所示，

表 11　不同模型的训练时间

Table 11　Training time of different methods

算法	时间(s)		
	NSL-KDD	UNSW-NB15	CICIDS2017
RF	357.2	427.3	570.8
DT	360.0	456.4	564.2
AE	38.4	50.9	250.3
DAE	45.2	50.4	246.7
ELM	58.2	79.8	323.1
DNN	220.8	335.7	553.4
DBN	259.7	369.5	576.2
LSTM	188.1	231.8	381.4
GRU	166.7	185.0	346.1
DAE-LSTM	139.4	175.1	336.3
GRU-MLP	170.2	290.3	365.7
CDG	124.4	158.1	305.7

表 12　不同方法的测试时间

Table 12　Testing time of different methods

算法	时间(ms)		
	NSL-KDD	UNSW-NB15	CICIDS2017
RF	6.28	9.74	11.86
DT	6.43	10.32	10.68
AE	0.51	3.95	3.87
DAE	0.32	3.95	3.65
ELM	1.57	7.31	6.17
DNN	7.29	11.24	12.21
DBN	8.34	12.58	13.26
LSTM	5.53	9.25	9.36
GRU	5.69	9.00	8.30
DAE-LSTM	2.47	7.24	7.58
GRU-MLP	6.37	11.78	10.20
CDG	2.01	6.67	6.25

从表 12 可以发现，CDG 模型的检测时间虽然比 AE 和 DAE 长，但是相差不大，在可接受范围内，说明本文方法能够实时进行流量异常检测。

6　结论

为了提高网络流量异常检测的综合性能，本文提出基于 CDG 的流量异常检测方法。相较于传统的深度学习异常检测方法，基于 CDG 的流量异常检测方法能够降低数据特征维度，减小数据噪声，增加模型的鲁棒性，缩短训练时间，提高检测准确率。本文使用 NSL-KDD、UNSW-NB15 和 CICIDS2017 数据集对所提出方法进行实验验证其性能，结果表明在二分类和多分类情况下，本文方法的综合检测性

能都优于典型机器学习和深度学习方法, 能够有效地检测未知攻击, 更好地满足系统的安全需求。下一步将研究如何提高模型在多分类中对 R2L 和 U2R 流量攻击数据的检测性能, 提高算法的适应性。

参考文献

[1] Boahen E K, Elvire Bouya-Moko B, Wang C D. Network Anomaly Detection in a Controlled Environment Based on an Enhanced PSOGSARFC[J]. *Computers & Security*, 2021, 104: 102225.

[2] Shone N, Ngoc T N, Dinh Phai V, et al. A Deep Learning Approach to Network Intrusion Detection[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018, 2(1): 41-50.

[3] Hinton G E, Osindero S, Teh Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. *Neural Computation*, 2006, 18(7): 1527-1554.

[4] Gao N, Gao L, Gao Q L, et al. An Intrusion Detection Model Based on Deep Belief Networks[C]. *2014 Second International Conference on Advanced Cloud and Big Data*, 2015: 247-252.

[5] Sadaf K, Sultana J. Intrusion Detection Based on Autoencoder and Isolation Forest in Fog Computing[J]. *IEEE Access*, 2020, 8: 167059-167068.

[6] Althubiti S, Nick W, Mason J, et al. Applying Long Short-Term Memory Recurrent Neural Network for Intrusion Detection[C]. *SoutheastCon*, 2018: 1-5.

[7] Agarap A F M. A Neural Network Architecture Combining Gated Recurrent Unit (GRU) and Support Vector Machine (SVM) for Intrusion Detection in Network Traffic Data[C]. *The 2018 10th International Conference on Machine Learning and Computing*, 2018: 26-30.

[8] Wang Y, Jiang Y M, Lan J L. FCNN: An Efficient Intrusion Detection Method Based on Raw Network Traffic[J]. *Security and Communication Networks*, 2021, 2021: 1-13.

[9] Subba B, Biswas S, Karmakar S. A Neural Network Based System for Intrusion Detection and Attack Classification[C]. *2016 Twenty Second National Conference on Communication*, 2016: 1-6.

[10] Xu C Y, Shen J Z, Du X, et al. An Intrusion Detection System Using a Deep Neural Network with Gated Recurrent Units[J]. *IEEE Access*, 2018, 6: 48697-48707.

[11] Nguyen S N, Nguyen V Q, Choi J, et al. Design and Implementation of Intrusion Detection System Using Convolutional Neural Network for DoS Detection[C]. *The 2nd International Conference on Machine Learning and Soft Computing*, 2018: 34-38.

[12] Teng S H, Wu N Q, Zhu H B, et al. SVM-DT-Based Adaptive and Collaborative Intrusion Detection[J]. *IEEE/CAA Journal of Automatica Sinica*, 2017, 5(1): 108-118.

[13] Mirsky Y, Doitshman T, Elovici Y, et al. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection[EB/OL]. 2018: arXiv: 1802.09089. https://arxiv.org/abs/1802.09089.

[14] Yu Y, Long J, Cai Z P. Session-Based Network Intrusion Detection Using a Deep Learning Architecture[M]. Modeling Decisions for Artificial Intelligence. Cham: Springer International Publishing, 2017: 144-155.

[15] Alrawashdeh K, Purdy C. Toward an Online Anomaly Intrusion Detection System Based on Deep Learning[C]. *2016 15th IEEE International Conference on Machine Learning and Applications*, 2017: 195-200.

[16] Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks[J]. *Science*, 2006, 313(5786): 504-507.

[17] Bhuvaneswari Amma N G, Selvakumar S, Leela Velusamy R. SAGRU: A Stacked Autoencoder-Based Gated Recurrent Unit Approach to Intrusion Detection[M]. Advances in Intelligent Systems and Computing. Singapore: Springer Singapore, 2020: 41-50.

[18] Roy B, Cheung H. A Deep Learning Approach for Intrusion Detection in Internet of Things Using Bi-Directional Long Short-Term Memory Recurrent Neural Network[C]. *2018 28th International Telecommunication Networks and Applications Conference*, 2019: 1-6.

[19] Mirza A H, Cosan S. Computer Network Intrusion Detection Using Sequential LSTM Neural Networks Autoencoders[C]. *2018 26th Signal Processing and Communications Applications Conference*, 2018: 1-4.

[20] Vincent P, Larochelle H, Bengio Y, et al. Extracting and Composing Robust Features with Denoising Autoencoders[C]. *The 25th international conference on Machine learning*, 2008: 1096-1103.

[21] Kachuee M, Darabi S, Moatamed B, et al. Dynamic Feature Acquisition Using Denoising Autoencoders[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(8): 2252-2262.

[22] Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1798-1828.

[23] Chawla A, Lee B, Fallon S, et al. Host Based Intrusion Detection System with Combined CNN/RNN Model[M]. ECML PKDD 2018 Workshops. Cham: Springer International Publishing, 2019: 149-158.

[24] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.

[25] Cho K, Merrienboer B V, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [EB/OL]. 2014, ArXiv: 1406.1078v3.

[26] M. Farhadloo, Statistical models for aspect-level sentiment analysis[D]. Ph.D.thesis, University of California, Merced, USA (2015). URL http://www.escholarship.org/uc/item/2ks913br.

[27] Dong S Q, Zhang B. Network Traffic Anomaly Detection Method Based on Deep Features Learning[J]. *Journal of Electronics & Information Technology*, 2020, 42(3): 695-703.
(董书琴, 张斌. 基于深度特征学习的网络流量异常检测方法[J]. *电子与信息学报*, 2020, 42(3): 695-703.)

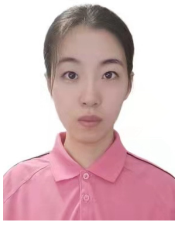
[28] Zhang G L, Wang X D, Li R, et al. Network Intrusion Detection Method Based on Stacked Denoising Sparse Autoencoder and Extreme Learning Machine[C]. *2020 2nd International Conference on Information Technology and Computer Application*, 2021: 194-199.

[29] Zhang H P, Wu C Q, Gao S, et al. An Effective Deep Learning Based Scheme for Network Intrusion Detection[C]. *2018 24th International Conference on Pattern Recognition*, 2018: 682-687.

[30] Wang Z D, Liu Y D, He D J, et al. Intrusion Detection Methods Based on Integrated Deep Learning Model[J]. *Computers & Secu-*

rity, 2021, 103: 102177.

[31] Yin S L, Zhang X L, Zuo L Y. Intrusion Detection System for Dual Route Deep Capsule Network[J/OL]. *Journal of Computer Research and Development*, 2021: 1-11. (2021-04-01). https://kns.cnki.net/kcms/detail/11.1777.tp.20210331.1727.004.html.
(尹晟霖, 张兴兰, 左利宇. 双重路由深层胶囊网络的入侵检测系统[J/OL]. *计算机研究与发展*, 2021: 1-11. (2021-04-01). https://kns.cnki.net/kcms/detail/11.1777.tp.20210331.1727.004.html.)



尹梓诺 于 2019 年在哈尔滨工业大学信息安全专业获得学士学位。现在信息工程大学学校网络空间安全专业攻读硕士学位。研究领域为网络安全、入侵检测。研究兴趣包括: 网络入侵检测、大数据处理等。Email: yinzinuo1997@163.com

马海龙 于 2011 年在信息工程大学通信与信息系统专业获得博士学位。现任信息工程大学信息技术研究所副研究员。研究领域为网络安全、路由工程。研究兴趣包括: 创新网络体系、网络安全管控等。Email: longmanclear@163.com

胡涛 于 2021 年在信息工程大学网络空间安全专业获得博士学位。现任信息工程大学信息技术研究所助理研究员。研究领域为网络安全、新型网络体系结构。研究兴趣包括: 新型网络体系结构、软件定义网络等。Email: hutaondsc@163.com