文章编号:1674-2974(2023)02-0063-11 DOI:10.16339/j.cnki.hdxbzkb.2023266

基于MeAEG-Net的异常流量检测方法研究

黎文伟[1,2†],岳子乔[1],王涛[3]

[1. 湖南大学 信息科学与工程学院,湖南 长沙 410082;
2. 区块链底层技术及应用湖南省重点实验室(湖南大学),湖南 长沙 410012;
3. 湖南城市学院 信息科学与工程学院,湖南 益阳 413000]

摘 要:异常流量检测现有方法大都是基于有监督的学习,在现实生活中获取并标记异常流量数据样本是极为困难的,存在诸多限制.此外,由于网络异常数据的多样性和复杂性,各种检测方法的自适应性较差,对新出现的异常流量难以判断.针对上述问题,本文设计了一个基于生成对抗网络和记忆增强模块的半监督异常流量检测框架MeAEG-Net(Memory Augment Based on Generative Adversarial Network),通过只训练正常流量样本数据,比较生成器模块输入流量底层特征的重构误差来达到检测异常的目的.在模型中使用生成对抗网络来更好地训练生成器,生成器采用自编码器加解码器的结构来解决自编码器易受噪声影响的问题,并在自编码器子网络中添加记忆增强模块来削弱生成器模块的泛化能力,增大异常流量的重构误差.实验证明,本文提出的方法能在只学习正常流量数据样本的前提下达到很好的异常流量检测效果.

关键词:异常流量检测;生成对抗网络;记忆增强模块;重构误差;半监督学习

中图分类号:TP393 **文献标志码**:A

Research on Abnormal Traffic Detection Method Based on Memory Augment-generative Adversarial Network

LI Wenwei[1,2†], YUE Ziqiao[1], WANG Tao[3]

[1. College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China;
2. Hunan Provincial Key Laboratory of Blockchain Infrastructure and Application (Hunan University), Changsha 410012, China;
3. School of Information Science and Engineering, Hunan City University, Yiyang 413000, China]

Abstract: Most of the existing abnormal traffic methods are based on supervised learning. It is extremely difficult to obtain and mark abnormal traffic data samples in real life, and there are many limitations. In addition, due to the diversity and complexity of abnormal network data, the adaptability of various detection methods is poor, and it is difficult to judge the new abnormal traffic. Based on the above problems, this paper designs a semi-supervised abnormal flow detection framework, MeAEG-Net (Memory Augment Based on Generative Adversarial Network), to detect anomalies by training only normal flow sample data and comparing the reconstruction errors of the underlying

* **收稿日期**:2022-04-12
基金项目:湖南创新型省份建设专项经费项目(2020GK2006,2020GK2007), Special Funds for Construction of Innovative Provinces in Hunan Province of China(2020GK2006, 2020GK2007)
作者简介:黎文伟(1975—),男,湖南沅江人,湖南大学副教授,博士
† 通信联系人,E-mail: liww@hnu.edu.cn

characteristics of input flow of generator module. A generative adversarial network is used in the model to better train the generator. The generator adopts the structure of an autoencoder and decoder to solve the problem that the autoencoder is susceptible to noise. The memory-augmented module is added to the sub-network of the autoencoder to weaken the generalization ability of the generator module and increase the reconstruction error of abnormal traffic. Experimental results show that the method proposed in this paper can achieve a good effect on abnormal traffic detection under the premise of learning only normal traffic data samples. Finally, the future research direction and challenges have been prospected.

Key words: abnormal traffic detection; generative adversarial network; memory augment module; reconstruction errors; semi-supervised learning

网络异常检测的主要目的就是发现流量异常,以保证网络空间的安全.异常流量的检测通常可以被看作流量的分类问题,逻辑回归、决策树、支持向量机(Support Vector Machine,SVM)等都是经常被用来进行异常流量检测的机器学习方法.随着机器性能的提升和数据量的增加,深度学习在异常检测领域也得到了广泛的应用,并取得了不错的成绩[1-3].深度学习快速发展的同时也促进了图像识别等领域的突破[4-5],这些领域的深度学习方法一般也被用于分类和异常检测,因此为人们提供了一些思路,可以结合其他领域的知识进行异常流量检测的探索.异常流量检测可以监控网络状态,判断当前网络是否正常,对确保网络安全具有非常重要的价值[6-7].基于网络异常检测的研究非常多,但是由于网络异常数据的多样性和复杂性,各种检测方法的自适应性较差,对新出现的异常流量难以判断.除此之外,当前的机器学习和深度学习算法都是基于包括正常数据和异常数据在内的所有数据的学习.正常流量样本数据在现实世界中数量多且易于获得,但获取并标记异常流量数据样本却极为困难.因此,通过半监督学习,本文对异常流量检测问题展开研究,主要贡献如下:

设计了一种基于生成对抗网络的异常流量检测模型,在提出的网络模型中,利用生成对抗网络的思想来更好地训练生成器.生成器由编码-解码-编码的网络结构组成,不同于一般的基于自编码器的异常检测方法,该模型通过比较输入数据的底层特征与生成数据的底层特征之间的重构误差进行异常检测,解决了自编码器结构易受噪声影响的问题.在网络结构中利用适当的批量正规化操作和激活函数来提升网络的性能,使用卷积操作替代池化层,避免关键特征的丢失.实验结果表明,该模型在只学习正常流量数据样本分布的前提下,能够取得非常不错的效果,尤其在分布不均衡的数据集中.

在本文提出的基于生成对抗网络的异常流量检测模型中引入了记忆增强模块,记忆增强模块添加在生成器的编码器子模块后.在记忆增强模块中,使用一个记忆矩阵保存训练阶段的底层特征,当输入一个流量数据时,通过记忆矩阵对该数据的底层特征进行加权求和,使该数据的底层特征更贴近训练阶段正常流量数据样本的底层特征,从而导致输入的异常流量数据能够有更大的重构误差,以进一步提高模型的检测性能.实验结果证明,引入了记忆增强模块能够有更好的检测效果.

1 相关工作

在传统的各种异常检测方法中,支持向量机被认为是分类异常行为的最佳机器学习算法之一.在机器学习最早被应用在异常流量检测的文献中,Kim等人[8]提出了一种基于支持向量机的异常流量检测方法,通过KDD'99数据集上的测试,证明了将支持向量机用于异常流量检测是一种有效的方法.贝叶斯网络是一种概率图形模型,是不确定知识表达和推理领域最有效的理论模型之一.Moore等人[9]使用朴素贝叶斯(Naive Bayesian,NB)的方法对网络流量按照应用分类,在使用最简单的朴素贝叶斯分类器上,能够达到65%的准确率.Williams等人[10]使用朴素贝叶斯、C4.5决策树、贝叶斯网络和朴素贝叶斯树等算法演示了基于一致性和基于相关性的特征选择对特征集约简的性能影响,证明了分类算法虽然相似但是对于分类的效果却是显著不同.

随着计算机性能的增强和数据规模的扩大,深度学习(Deep Learning,DL)开始作为机器学习的一个研究方向且越来越火热,它能够学习数据的内在特征和规律,属于一种复杂的机器学习,在矩阵检测等方面远远超过了当前的相关技术.Javaid等人[11]提出了一种基于深度学习框架的异常流量检测方法,他们基于模糊神经网络设计了一种自主学习系统,并在NSL-KDD基准数据集上进行了性能的比较,能够达到88%的正确率.An等人[12]提出变分自编码器模型,在NSL-KDD基准数据集上进行实验,基于输入数据的重建概率判断是否属于异常流量,取得了比PCA和AE模型更高的检测性能.在使用全连接神经网络构建异常流量检测系统时,在NSL-KDD数据集上的二分类正确率能够达到81.2%.高妮等人[13]提出了一种结合自动编码器和支持向量机的异常流量检测模型(AN-SVM),首先通过自动编码器学习到数据分布的底层特征,降低维度,接着使用支持向量机进行异常流量检测,在减少了模型训练时间的同时提高了检测的正确率.Wang等人[14]提出一种基于卷积神经网络的异常流量检测模型,将网络流量预处理为二维矩阵作为输入,通过模型提取流量的底层特征,实现端到端的异常流量检测.

生成对抗网络是由Goodfellow等人[15]在2014年提出来的一种机器学习架构,被认为是近几年来在复杂数据的分布上进行无监督学习最具有前景的方法之一.生成对抗网络由生成器和判别器两部分组成,生成器负责生成与输入样本数据相似的伪样本,判别器负责判断生成样本和输入样本的真伪,在训练过程中,通过生成器和判别器彼此的对抗,最终达到纳什平衡,达到了生成器生成的数据更接近原始输入样本数据,学习到了输入数据的概率分布,判别器无法判断出真伪的目的.当前将生成对抗网络大部分用于生成样本和数据增强,刘海波等人[16]提出一种基于生成对抗网络和长短期记忆人工神经网络(Long Short-Term Memory,LSTM)的网络结构,针对缺少高持续性威胁攻击样本的情况,使用生成对抗网络生成大量攻击样本,从而改善了传统高持续性威胁攻击检测率低的问题.Yin等人[17]提出了一种基于生成对抗网络的半监督分类方法,在学习部分带标签数据的前提下,在NSL- KDD数据集进行了多分类实验,实验证明了基于生成对抗网络思想的分类模型比原分类模型有更高的检测效果,也证明了生成对抗网络在异常流量检测领域的适用性.

以上的机器学习方法都是在训练正负标签样的前提下,进行异常检测,当今以有监督的学习为指导的异常流量检测占据了学术主流[18-20].采用有监督学习的方式需要大量且各类别分布平衡的已标记训练样本,在现实生活中,正常流量数据样本数量多且易于获得,但是获取并标记异常流量数据样本是极为困难的.而且,在面对未知种类的异常流量时,当前模型的自适应性较差.自编码器和变分自编码器作为一种半监督的深度学习方法,能够通过合理的网络结构设计和预处理,在训练阶段自己学习到正常流量数据分布,通过模型的重构误差或者重建概率进行异常检测,达到不逊于传统模型的检测性能.

2　记忆增强生成对抗网络结构

2.1　基本流程

本文提出了一个基于记忆增强的生成对抗网络学习框架MeAEG-Net来进行异常流量检测.MeAEG-Net总体框架如图1所示,主要包括了基于记忆增强的自编码子网络和重构编码子网络的生成器模块和判别器模块.流量特征矩阵在进入MeAEG-Net时,生成器接收预处理后生成的样本特征矩阵$\boldsymbol{F}$,在生成器中先经过编码器生成特征矩阵$\boldsymbol{F}$的底层特征向量$\boldsymbol{Z}$,底层特征向量$\boldsymbol{Z}$经过记忆增强模块生成特征向量$\boldsymbol{Z},\boldsymbol{Z}$^再经过解码器生成特征矩阵$\boldsymbol{F}'$;把生成的特征矩阵$\boldsymbol{F}'$输入到重构编码器产生底层特征向量$\boldsymbol{Z}'$;对抗网络输入预处理生成的流量特征矩阵$\boldsymbol{F}$和经过生成器生成的特征矩阵$\boldsymbol{F}'$,输出信息反馈促进生成器训练,直至最后生成接近特征矩阵$\boldsymbol{F}$的矩阵,且底层特征向量$\boldsymbol{Z}$与底层特征向量$\boldsymbol{Z}'$尽可能保持最小误差.网络最后的输出结果为异常流量的概率值$\| \boldsymbol{Z}-\boldsymbol{Z}' \|_1$,根据设置的阈值$\theta$和$\| \boldsymbol{Z}-\boldsymbol{Z}' \|_1$进行判别.下面对MeAEG-Net网络结构中的各个部分进行详细的阐述.

2.2　生成器模块

Gong等人[21]提出了一种基于记忆增强的自编码器网络MemAE,通过在自编码器中添加一个记忆增强模块,改进自动编码器.传统的自编码器网络是输入的原始样本数据通过编码器模块得到底层特征向量$\boldsymbol{Z}$,然后底层向量$\boldsymbol{Z}$再通过解码器模块重构为与输入样本数据类似的数据,通过比较输入数据与输出数据的重构误差来进行异常检测.而由于自动编码器的泛化能力,部分异常的输入数据也能够被很好地重建,因此,记忆增强自编码器的目的就是通过对底层向量$\boldsymbol{Z}$的变换来消减这种泛化能力,使训练

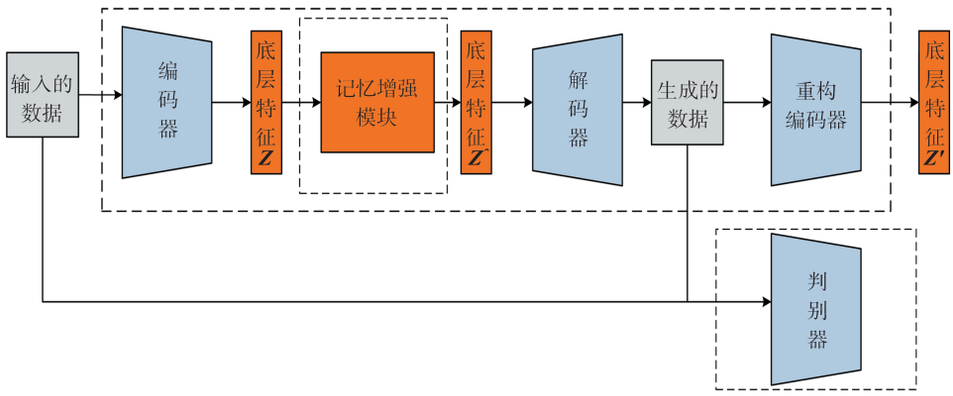

图 1 MeAEG-Net 记忆增强生成对抗网络结构

Fig.1 Memory augment based on generative adversarial network MeAEG-Net

阶段训练的正常样本能够得到贴切的还原,由于底层向量 $\boldsymbol{Z}$ 会变换重构为与正常样本相似的数据,从而使异常数据产生明显的重构误差,达到异常检测的目的.记忆增强模块见图2[21].

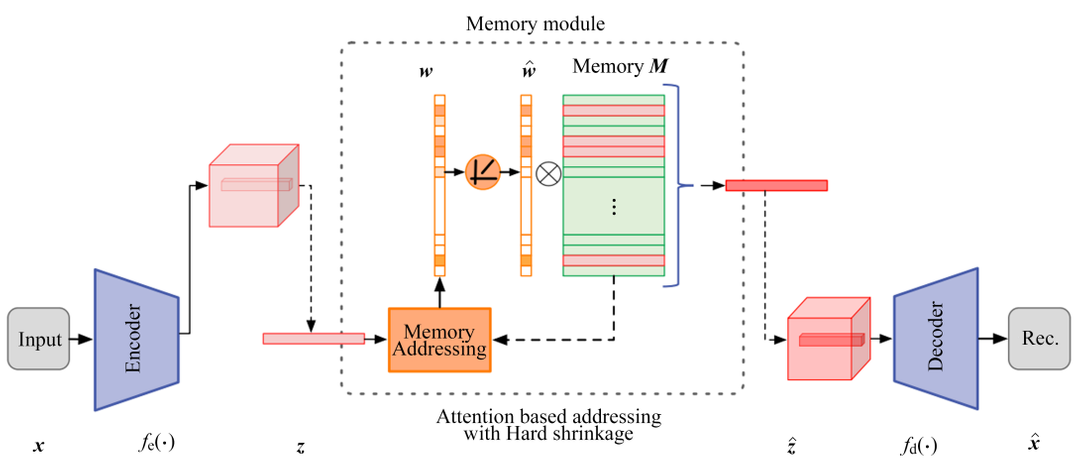

图 2 记忆增强自编码器[21]

Fig.2 Memory augmentated AE[21]

在传统的自编码器网络中,输入数据 X 经过编码器(Encoder)计算得到数据的底层特征向量 $\boldsymbol{Z}$,而底层特征向量 $\boldsymbol{Z}$ 再通过解码器(Decoder)重构为数据 $X^{\wedge}$,见式(1).

$$\begin{aligned} \boldsymbol{Z} &= f_e(X;\ \theta_e) \\ X^{\wedge} &= f_d(Z;\ \theta_d) \end{aligned} \tag{1}$$

在网络中使用记忆增强模块,使得在训练阶段时,输入数据经编码器模块得到底层向量保存成一个记忆矩阵 $\boldsymbol{M}$,$\boldsymbol{M} \in \mathrm{R}^{N\times C}$,其中 N 表示的是存储的底层向量的个数,C 表示的是底层向量的维度,即底层特征向量 $\boldsymbol{Z}$ 的维度.记忆增强模块的输出 $\boldsymbol{Z}^{\wedge}$ 表示为记忆矩阵中 N 个底层向量的加权求和,其中 W 表示的是对应底层向量的权值.见式(2).

$$\boldsymbol{Z}^{\wedge} = W\boldsymbol{M} = \sum_{i=1}^{N} w_i \boldsymbol{m}_i \tag{2}$$

记忆矩阵 $\boldsymbol{M}$ 中每个底层特征向量 $\boldsymbol{m}_i$ 对应的权重 w_i 计算公式见(3).

$$w_i = \frac{\exp(d(\boldsymbol{z}, \boldsymbol{m}_i))}{\sum_{j=1}^{N} \exp(d(\boldsymbol{z}, \boldsymbol{m}_j))} \tag{3}$$

其中 $d(\boldsymbol{z}, \boldsymbol{m}_i)$ 表示的是底层向量 $\boldsymbol{Z}$ 与记忆矩阵每一列底层向量 $\boldsymbol{m}_i$ 的相似度,相似度使用余弦距离(Cosine)计算,见式(4).

$$d(\boldsymbol{z}, \boldsymbol{m}_i) = \frac{\boldsymbol{z}\boldsymbol{m}_i^{\mathrm{T}}}{\|\boldsymbol{z}\|\|\boldsymbol{m}_i\|} \tag{4}$$

在记忆增强模块中,提出了通过使用记忆矩阵中 N 个底层向量的加权求和的方式得到输出的底层向量 $\boldsymbol{Z}^{\wedge}$,而各个权重 w_i 的计算是通过余弦距离得到的,在得到权重 w_i 后,还需要对 w_i 进行约束,因为通过各个底层向量的权重相加也有可能会导致异常被重构,所以,结合激活函数 ReLU 的思想,对 w_i 进行硬阈值约束,即超过阈值的权重被保留,其余的设置为

0,超参数λ表示阈值.见式(5).

$$\hat{w_i} = h(w_i;\ \lambda) = \begin{cases} w_i & \text{如果} w_i > \lambda \\ 0 & \text{其他} \end{cases} \tag{5}$$

为了使记忆增强模块可以计算梯度,结合ReLU函数进行调整,超参数λ一般取[1/N,3/N][21],调整公式见(6).

$$\hat{w_i} = \frac{\max(w_i - \lambda, 0) \times w_i}{|w_i - \lambda|} \tag{6}$$

结合上述过程,基于记忆增强模块的自编码器网络输出数据的计算公式见式(7).

$$\begin{cases} \boldsymbol{Z} = f_e(X;\ \theta_e) \\ \hat{\boldsymbol{Z}} = \hat{W}\boldsymbol{M} = \sum_{i=1}^{N} \hat{w_i}\boldsymbol{m}_i \\ \hat{\boldsymbol{X}} = f_d(\hat{Z};\ \theta_d) \end{cases} \tag{7}$$

生成器模块主要用于提取出输入特征矩阵的一维底层向量和生成特征矩阵,以及提取出生成特征矩阵的一维底层向量,见图3.与传统的使用特征矩阵的重构误差进行异常检测相比,该方法可以有效地减少噪声对模型的影响,且有效地减少了生成器的泛化现象.生成器编码器模块和重构编码器模块中卷积层的每一个卷积核都可以提取特定的特征,不同的卷积核提取不同的特征,输入矩阵经过不同特征的卷积操作就变成了一系列特征矩阵,直观地将这整个操作视为一个单独的处理过程,与此同时,网络结构中不使用池化层,防止关键特征信息的丢失,通过使用批量正规化操作对网络性能进行进一步提升.

2.3　判别器模块

判别器模块判断输入的样本数据为真实矩阵还是生成器生成的矩阵,本章提出的MeAEG-Net判别器使用卷积神经网络,最后使用Sigmod函数输出矩阵为真的概率值,见图4.

判别器本质上类似分类器,目的在于判断输入的流量特征矩阵是生成的特征矩阵$\boldsymbol{F}'$还是原始矩阵$\boldsymbol{F}$.

2.4　损失函数设计

生成器的损失函数L_G由四部分组成,第一部分是生成网络中$\boldsymbol{F}$和$\boldsymbol{F}'$的重构损失L_{con},用于减少经过预处理的流量特征矩阵$\boldsymbol{F}$和经过生成器生成的重构特征矩阵$\boldsymbol{F}'$在像素层面上的差距.L_{con}见式(8).

$$L_{con} = \left\| \boldsymbol{F} - \boldsymbol{F}' \right\|_2 \tag{8}$$

第二部分在判别器中,针对在矩阵特征方面的优化损失L_{adv},用于减少流量特征矩阵$\boldsymbol{F}$和在生成器中生成的流量特征矩阵$\boldsymbol{F}'$的差距,L_{adv}见式(9).

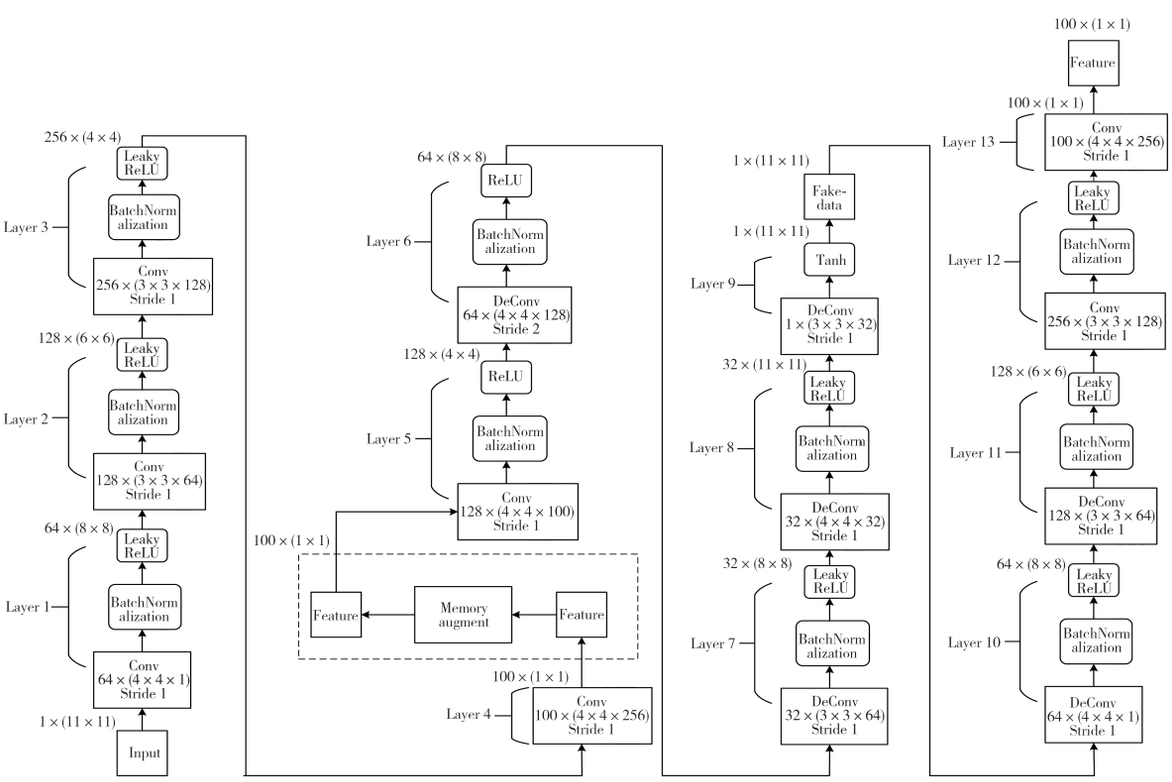

图3　生成器网络结构

Fig.3　Generator network architecture

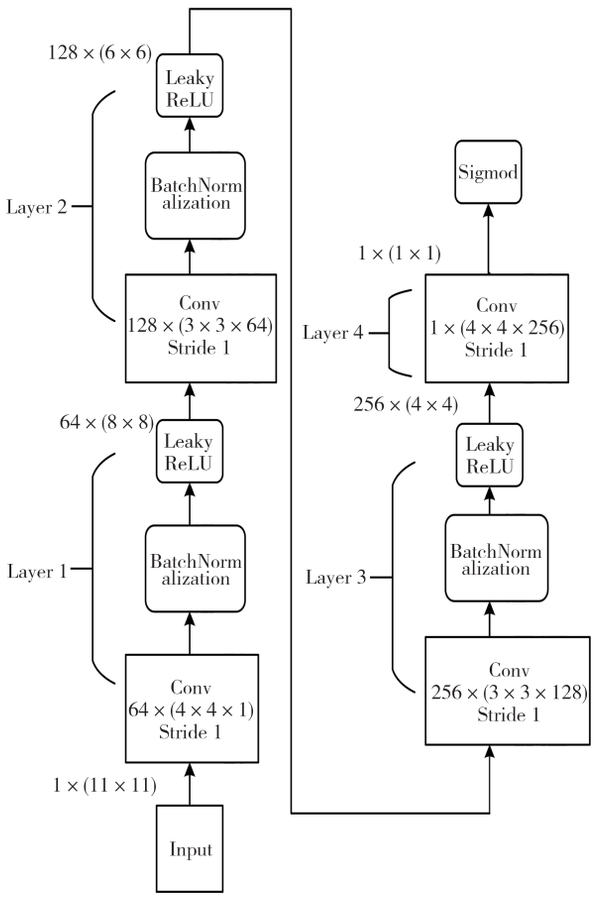

图4 判别器网络结构

Fig.4 Discriminant network architecture

$$L_{adv} = \| f_{adv}(\boldsymbol{F}) - f_{adv}(\boldsymbol{F}') \|_2 \tag{9}$$

第三部分中针对重构特征矩阵$\boldsymbol{F}'$进行编码得到的底层向量$\boldsymbol{Z}'$,对于正常流量数据而言,希望得到的底层特征向量$\boldsymbol{Z}'$能够与矩阵$\boldsymbol{F}$编码得到的底层向量$\boldsymbol{Z}$尽可能相似,因此引入一个潜在向量之间的误差优化损失函数L_{enc},见式(10).

$$L_{enc} = \| \boldsymbol{Z} - \boldsymbol{Z}' \|_2 \tag{10}$$

第四部分是对于记忆增强模块中权重矩阵的稀疏,期望的权重矩阵的熵越小越好,损失函数为L_{att},见式(11).

$$L_{att} = E(W) \tag{11}$$

对整个生成网络模型使用损失函数L_G进行训练,见式(12).

$$L_G = \alpha L_{con} + \beta L_{adv} + \delta L_{enc} + \sigma L_{att} \tag{12}$$

其中α、β、δ和σ是调节分配各损失的权重,需要在实验中进行调整.

使用损失函数L_D对异常流量检测模型中的判别器模块进行训练,定义y为真实的标签值0或1,定义Adv为判别器,则L_D见式(13).

$$L_D = -y\log(\mathrm{Adv}(F')) - (1-y)\log(1-\mathrm{Adv}(F)) \tag{13}$$

2.5 模型分析

模型在输入阶段不使用随机噪声产生矩阵,而是直接将样本特征矩阵作为输入,相比于使用随机噪声作为输入的方式,该方式能够更好地提取出原始矩阵的关键特征信息,使得网络训练阶段能够学习到关键底层特征,从而提高了判别性能.在Berman等人[20]的网络结构中,卷积核都采用了4×4的大小,采用其结构则编码器模块只有三个卷积层,本章增加了卷积层的层数,更有利于训练阶段进行特征提取.

通过在编码器和判别器前三个卷积层中缩小卷积核的大小,能够在具有相同感知视野的条件下,提升了网络的深度,在一定程度上提升了神经网络效果的同时可以有效减少计算参数量.在卷积层中取消了池化层的使用,避免了池化层的下采样操作造成关键特征的丢失.通过在浅层Tanh函数和深层ReLU函数的结合使用,一定程度上提升了网络性能.

模型在损失函数方面,相对于一般的网络模型来说,生成器模块的损失函数由三部分构成.除了判别器的相反值和原始矩阵和生成矩阵的差异度之外,还增加了一项生成器模型中编码器生成的底层向量与重构编码器生成的底层向量之间的L2正则化.由于是均方误差,如果误差大于1,那么平方后,相比L1正则化而言,误差就会被放大很多,如果不是为了进行特征选择,一般使用L2正则化模型效果更好.因此模型会对样例更敏感.如果样例是一个异常值,模型会调整最小化异常值的情况,以牺牲其他更一般样例为代价,因为相比单个异常样例,那些一般的样例会得到更小的损失误差.

文章的网络中采用了自编码器+重构编码器+判别器的网络结构,通过对卷积核大小和步长的控制,对网络层数进行控制,让模型的复杂度尽量减小,计算复杂度缩小的同时提高模型性能,同时在网络中使用批量归一化及激活函数等操作提高网络的表达能力.

记忆增强模块是为了消减生成器的泛化能力,在记忆增强模块中,采用记忆矩阵的方法,记忆矩阵里存储的都是在训练阶段学习到的正常数据样本的

底层向量,鼓励输入的数据经编码器得到的底层向量通过记忆模块能够表示正常数据得到的底层向量,这样重构出来的数据与正常数据十分相似,当输入的样本数据为异常数据时,采用此方法得到的输出会与输入产生更大的重构误差,由此增强了检测能力.

3 实验结果与分析

3.1 数据集和实验设置

本章实验中采用在入侵检测领域经常使用的NSL-KDD数据集,数据集解决了KDD'99数据集中存在的一些固有问题,不包含冗余数据,而且训练集和测试集的记录数量非常合理,使得该数据在入侵检测领域得到广泛的使用,可以帮助研究人员用来比较不同的入侵检测方法.使用NSL-KDD Train$^+$中的全部正常流量数据作为训练集,NSL-KDD Test$^+$和NSL-KDD Test$^-$[21]作为测试集.首先使用Weka工具进行CorrelationAttributeEval分析得出属性与类别的相关性排名,去除相关性为0的num_outbound_cmds这一属性,接着对数据集中的离散值和连续值进行相应的预处理.

本文使用独热编码对离散值进行处理.当使用独热编码时,针对协议类型tcp、udp、icmp,在相同的场景下,分别赋值为(1,0,0)、(0,1,0)和(0,0,1).采用该种方式,模型判断之间的相似度是一致的,将目标主机的网络服务类型特征service和链接正常或错误的状态flag也通过独热编码处理.

在NSL-KDD连续值的预处理中,经常使用归一化方法.本文中,先将取值范围大的特征,如src_bytes的取值为[0~ 1 379 963 888],dst_bytes的取值为[0~1 309 937 401]等进行数值的标准化,数值的标准化见式(14).

$$X'_{ij}=\frac{X_{ij}-E(X_j)}{\mathrm{Stad}_j} \tag{14}$$

式中,$E(X_j)$表示特征j的平均值,Stad_j表示平均绝对误差,相对于标准差对于孤立点具有更好的鲁棒性,见式(15).

$$\mathrm{Stad}_j=\frac{1}{n}\left(\left|X_{1j}-E(X_j)\right|+\left|X_{2j}-E(X_j)\right|+\cdots+\left|X_{nj}-E(X_j)\right|\right) \tag{15}$$

再将标准化的每一个数值归一化到[0,1]的范围上,归一化见式(16).

$$X'_{ij}=\frac{X_{ij}-\min(X_j)}{\max(X_j)-X_{ij}} \tag{16}$$

使用NSL-KDD数据集的其他40个特征经过上述流量预处理方法得到1×121维的流量向量,再使用Python的reshape()函数,转化为11×11的流量特征矩阵.

模型的检测性能由AUC值、正确率和F_1-Measure三个指标来衡量[19-20],正确率是二分类和入侵检测中最直观的评价指标,AUC指标代表着ROC曲线下面的面积,ROC曲线是假阳性率(False Positive Rate,FPR)和真阳性率(True Positive Rate,TPR)为横纵坐标的曲线,和阈值的选择无关,能更好地表示模型的性能.F_1-Measure被广泛使用于评价分类结果的质量,是精确率(precision)和召回率(recall)综合衡量的一个指标.

此方案通过深度学习框架PyTorch实现,采用GPU加速,Python版本为3.7.0,pytorch1.2.0,Weka版本为3.8.4,cudn版本为cudnn-10.0-windows10-x64-v7.4.2.24.

3.2 训练流程和参数设置

在训练过程中,首先固定生成器网络参数,将预处理后生成的流量特征矩阵$\boldsymbol{F}$和经过生成器生成的特征矩阵$\boldsymbol{F}'$输入判别器,进行监督训练,利用正向传播计算输出结果,通过对输出结果与实际结果的求偏差判断是否超过容许范围,否则进行反向传播以调整对抗网络参数,使其可以更好地区分流量特征矩阵$\boldsymbol{F}$和生成的矩阵$\boldsymbol{F}'$.然后固定对抗网络参数,训练生成器模块,在训练过程中利用正向传播计算输出结果,通过对输出结果与实际结果的求偏差判断是否超过容许范围,否则进行反向传播,使生成器生成的特征矩阵$\boldsymbol{F}'$更加贴近输入之前的特征矩阵$\boldsymbol{F}$,最终对抗网络无法做出正确判断,且生成器生成的特征矩阵$\boldsymbol{F}'$经过重构编码器得到的底层向量$\boldsymbol{Z}'$更加贴近特征矩阵$\boldsymbol{F}$经过重构编码器得到的底层向量$\boldsymbol{Z}$.然后计算各层中的误差,通过梯度下降算法更新各层权值,随着训练次数的增加,最终生成训练好的网络模型.

在测试阶段,直接将需要检测的流量特征矩阵$\boldsymbol{F}$输入网络模型中,经过一系列传播计算,使模型同

样的首先进行底层特征的提取操作,生成器中编码器模块的编码得到一个底层向量$\boldsymbol{Z}$,再经过解码器生成相似的特征矩阵$\boldsymbol{F}'$,通过后续的传播计算,特征矩阵$\boldsymbol{F}'$经过重构编码器得到一个底层向量$\boldsymbol{Z}'$,输出结果为异常流量的概率值$\| \boldsymbol{Z}-\boldsymbol{Z}' \|_1$,最后根据设置的阈值$\theta$和$\| \boldsymbol{Z}-\boldsymbol{Z}' \|_1$进行判别.见图5.

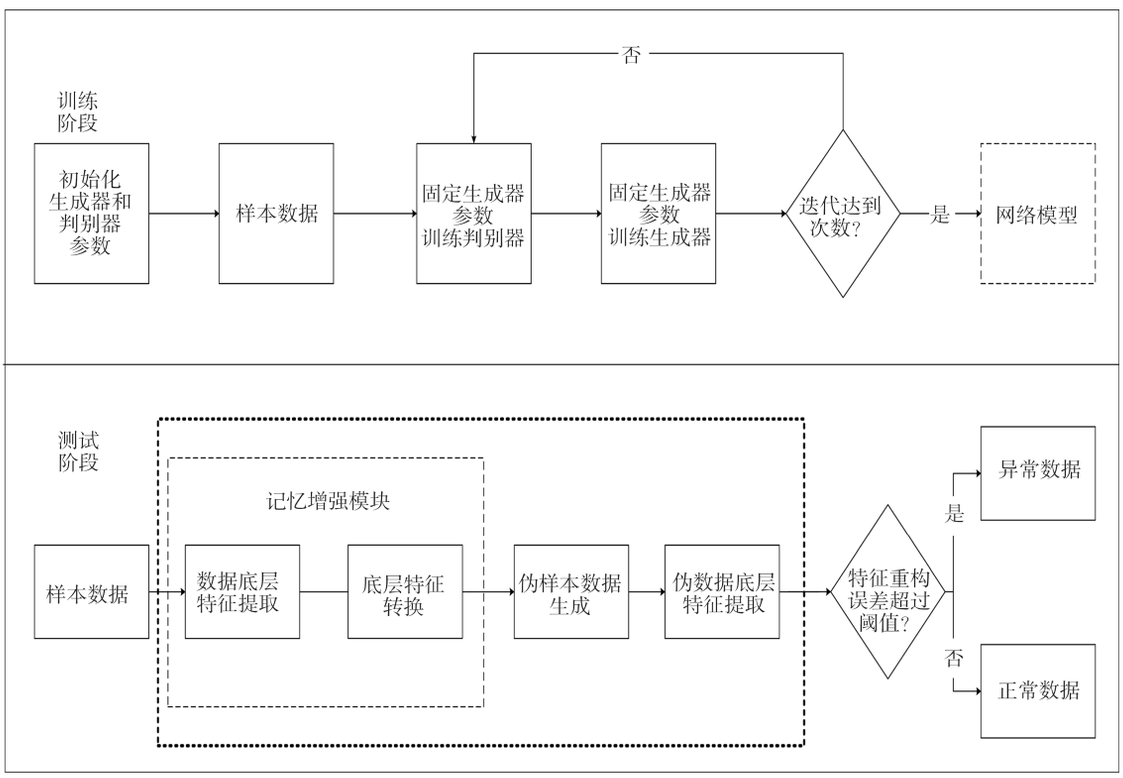

图5 模型训练和测试流量图

Fig.5 Model training and test flow charts

本模型采用自适应矩估计(Adam)优化器来训练网络.基础学习率(Base_lr)为0.000 2,Adam参数中一阶矩估计的指数衰减率(beta1)为0.5、二阶矩估计的指数衰减率(beta2)为0.999.网络模型中的卷积层使用高斯初始化,初始权值在均值为0、方差为0.02的高斯分布中采样.生成器模块损失函数中的四个权重参数分别为1、50、1、0.01,底层向量通道数为100,记忆增强模块的记忆矩阵储存底层向量个数N为10 000,硬阈值设置为0.000 3,根据GPU显存的大小,训练阶段每一批次(batch_size)的大小设置为128,网络模型设置为当训练25个epoch时结束训练.

3.3 实验结果

3.3.1 与有监督机器学习方法的比较

本文使用Weka机器学习和挖掘工具,在NSL-KDD数据集上,研究二分类(Normal,Anomaly)下,随机树、朴素贝叶斯、支持向量机、多层感知机、决策树(J48)等机器学习算法的检测准确率和AUC值和F_1-Measure,以及本章使用的MeAEG-Net模型检测准确率、AUC值和F_1-Measure,并进行性能比较.其中,传统机器学习方法在NSL-KDD Train+全部的数据上进行训练,而MeAEG-Net只使用数据集中的正常样本.使用NSL-KDD Test+和NSL-KDD Test-21作为测试集.

图6和图7以及表1和表2表示在二分类任务

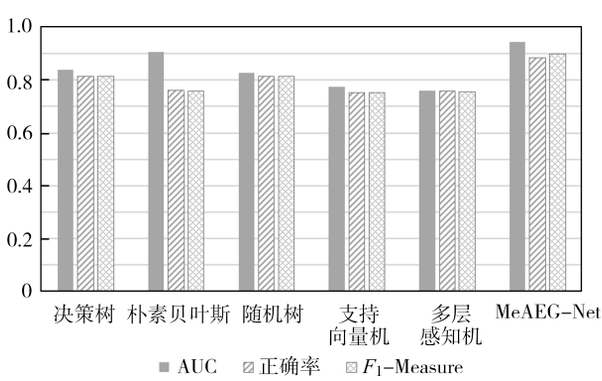

图6 MeAEG-Net与传统机器学习算法的比较(NSL-KDD Test+)

Fig.6 Comparison between MeAEG-Net and the traditional machine learning methods(NSL-KDD Test+)

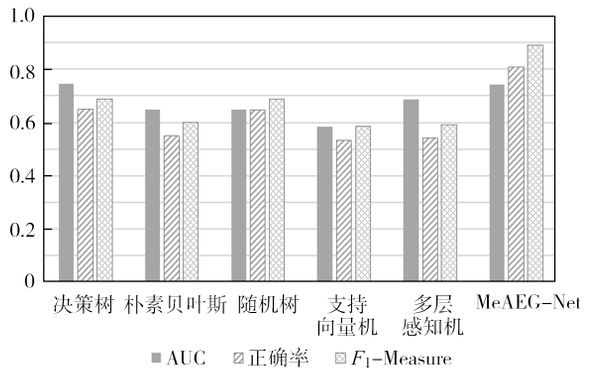

图 7　MeAEG-Net与传统机器学习算法的比较(NSL-KDD Test-21)

Fig.7　Comparison between MeAEG-Net and the traditional machine learning methods(NSL-KDD Test-21)

表 1　与传统机器学习的性能比较(NSL-KDD Test+)

Tab. 1　Performance comparison with traditional machine learning(NSL-KDD Test+)

方法	AUC	正确率/%	F_1-Measure
决策树	0.840	81.50	0.815
朴素贝叶斯	0.908	76.10	0.759
随机树	0.827	81.40	0.814
支持向量机	0.775	75.40	0.752
多层感知机	0.759	75.80	0.757
MeAEG-Net	0.946	88.50	0.901

表 2　与传统机器学习的性能比较(NSL-KDD Test-21)

Tab. 2　Performance comparison with traditional machine learning(NSL-KDD Test-21)

方法	AUC	正确率/%	F_1-Measure
决策树	0.746	64.90	0.689
朴素贝叶斯	0.650	54.90	0.599
随机树	0.649	64.80	0.687
支持向量机	0.585	53.30	0.585
多层感知机	0.687	54.10	0.592
MeAEG-Net	0.744	80.90	0.890

下,使用本文提出的MeAEG-Net与传统的机器学习算法在测试集上的检测性能.从数据中可以看出本文提出的MeAEG-Net在只学习正常流量数据样本的前提下,相对于传统的机器学习,在学习全部数据样本能够取得更好的检测效果.

从以上数据可以得出,本文提出的模型在NSL-KDD Test+数据集上,与比较的传统机器学习方法在AUC值上提高了0.038到0.187,正确率提高了7%到12.7%,F_1-Measure提高了0.086到0.149.这证明了模型能够在仅仅使用训练集中正常流量数据的前提下,达到比传统的机器学习算法在有监督学习的情形下更好的检测性能.

因为NSL-KDD Test-21测试集中存在数据分布不均衡的现象,所以一般的机器学习算法检测性能较差.从表2和图7的数据可以得出,本文提出的模型在AUC值上与设计的决策树相差不大,但相对于其他方法提高了0.057到0.159,正确率提高了16%到27.6%,F_1-Measure提高了0.199到0.305.这证明了模型能够在仅仅使用训练集中正常流量数据的前提下,在不均衡的数据集中能够达到比传统机器学习算法更优异的检测性能.

3.3.2　与半监督深度学习方法的比较

本文使用的方法,是基于正常样本的训练,达到能检测出异常样本的目的.因此,选取An Jinwon等人[12]的变分自编码器(VAE)与本文采用同样网络结构的自编码模型(CAE),编码-解码-编码模型(CAEE)以及基于生成对抗网络的CAEE模型AEG-Net作为本模型的比较对象.

图8和图9以及表3和表4表示在二分类任务下,使用本文提出的MeAEG-Net与对照模型在测试集上的检测性能.从数据中可以看出本文提出的MeAEG-Net在只学习正常流量数据样本的前提下相对于对照模型能够取得更好的检测效果.

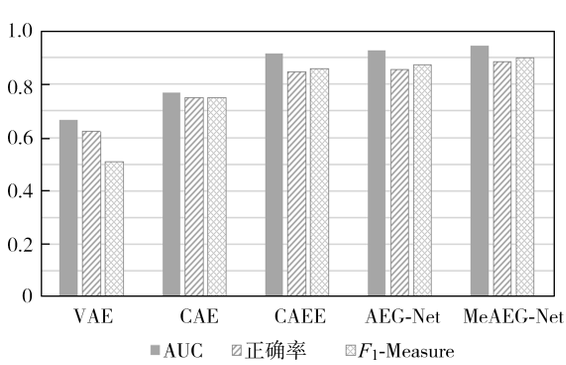

图 8　MeAEG-Net与半监督深度学习方法的比较(NSL- KDD Test+)

Fig.8　Comparison between MeAEG-Net and semi-supervised deep learning methods(NSL-KDD Test+)

从以上数据可以得出,本文提出的模型在NSL-KDD Test+数据集上,与比较的深度学习方法在AUC值上提高了0.019到0.279,正确率提高了2.8%到26.2%,F_1-Measure提高了0.027到0.392.这证明了模型与同样采用基于半监督学习的深度学习算法AE与AEE等相比,在NSL-KDD Test+数据集上具有更好的检测性能.

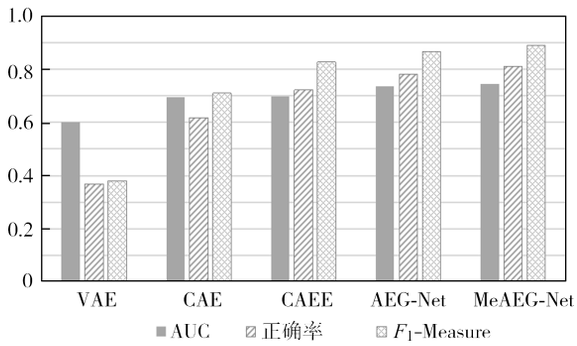

图9 MeAEG-Net与半监督深度学习方法的比较(NSL-KDD Test-21)

Fig.9 Comparison between MeAEG-Net and semi-supervised deep learning methods(NSL-KDD Test-21)

表3 与半监督深度学习方法的性能比较(NSL-KDD Test+)

Tab. 3 Performance comparison with semi-supervised deep learning methods(NSL-KDD Test+)

方法	AUC	正确率/%	F_1-Measure
VAE	0.667	62.30	0.509
CAE	0.770	75.00	0.751
CAEE	0.917	84.80	0.859
AEG-Net	0.927	85.70	0.874
MeAEG-Net	0.946	88.50	0.901

表4 与半监督深度学习方法的性能比较(NSL-KDD Test-21)

Tab. 4 Performance comparison with semi-supervised deep learning methods(NSL-KDD Test-21)

方法	AUC	正确率/%	F_1-Measure
VAE	0.602	36.90	0.380
CAE	0.693	61.80	0.709
CAEE	0.697	72.10	0.826
AEG-Net	0.735	78.00	0.864
MeAEG-Net	0.744	80.90	0.890

从以上数据可以得出,本文提出的模型在NSL-KDD Test-21数据集上,与比较的深度学习方法在AUC值上提高了0.009到0.142,正确率提高了2.9%到44%,F_1-Measure提高了0.026到0.51.这证明了模型与同样采用基于半监督学习的深度学习算法AE与AEE等相比,在NSL-KDD Test-21数据集上具有更好的检测性能.

实验结果表明,在相同的实验条件下,本文提出的MeAEG-Net以KDD Train+正样本数据为训练集,以KDD Test+和KDD Test-21为测试集,在仅仅训练KDD Train+训练集中的正常样本数据的前提下,在鉴别异常样本数据的性能方面优于基于传统机器学习的分类模型,同时优于同样只使用正常流量样本做异常检测的基于卷积自编码器CAE和VAE模型以及本文使用的基于卷积神经网络的编码-解码-编码结构CAEE与AEG-Net,提升了对异常流量的检测能力.

4 结束语

与目前主流的异常流量检测模型不同,本文提出的模型基于生成对抗网络思想,通过生成对抗的方式更好地训练生成器模型,并在生成器网络中引入记忆增强模块以此来提高异常流量数据的检测能力.模型在训练阶段学习正常流量数据样本的数据分布,通过比较输入流量数据的重构误差进行异常检测,减少了标记异常流量数据的工作.与此同时,本文还引入了记忆增强模块,通过消减生成器模块的泛化能力,从而加大异常流量的重构误差,有效地提高了模型的性能.在现实生活中,根据网络环境,网络管理员可以手动调整阈值,从而可以根据网络需求在灵敏度和特异性之间进行权衡,与其他现有方法相比,这是一个巨大的优势.下一步,将基于本模型深入研究真实网络流量数据下的检测性能.

参考文献

[1] VINAYAKUMAR R, ALAZAB M, SOMAN K P, et al. Deep learning approach for intelligent intrusion detection system[J]. IEEE Access, 2019:41525-41550.

[2] DONG B, WANG X. Comparison deep learning method to traditional methods using for network intrusion detection[C]//2016 8th IEEE International Conference on Communication Software and Networks. Beijing, China: IEEE, 2016:581-585.

[3] RAJKUMAR N, D'SOUZA A, ALEX S, et al. Long short-term memory-based recurrent neural network approach for intrusion detection [C]//Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB). 2018:837-846.

[4] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. Medical Image Analysis, 2017, 42:60-88.

[5] AKCAY S, ATAPOUR-ABARGHOUEI A, BRECKON T P. GANomaly: semi-supervised anomaly detection via adversarial training[J]. arXiv e-prints, 2018, arXiv:1805.06725.

[6] ZHANG Y Z, XIAO J, YUN X C, et al. DDoS attacks detection and control mechanisms[J]. Journal of Software, 2012, 23(8): 2058-2072.

[7] 吴春琼. 基于特征选择的网络入侵检测模型[J]. 计算机仿真,2012,29(6):136-139.
WU C Q. Network intrusion detection model based on feature selection[J]. Computer Simulation, 2012, 29(6): 136-139. (in Chinese)

[8] KIM D S, PARK J S. Network-based intrusion detection with support vector machines[C]// International Conference on Information Networking. Berlin, Heidelberg: Springer, 2003: 747 - 756.

[9] MOORE A W, ZUEV D. Internet traffic classification using Bayesian analysis techniques[J]. ACM SIGMETRICS Performance Evaluation Review, 2005, 33(1): 50-60.

[10] WILLIAMS N, ZANDER S, ARMITAGE G. A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification[J]. Comput Commun Rev, 2006, 36: 5-16.

[11] JAVAID A, NIYAZ Q, SUN W Q, et al. A deep learning approach for network intrusion detection system[C]//Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS). New York, USA: ACM, 2015: 21-26.

[12] AN J W, CHO S Z. Variational autoencoder based anomaly detection using reconstruction probability[R]. Special Lecture on IE, 2015.

[13] 高妮,高岭,贺毅岳,等. 基于自编码网络特征降维的轻量级入侵检测模型[J]. 电子学报,2017,45(3):730-739.
GAO N, GAO L, HE Y Y, et al. A lightweight intrusion detection model based on autoencoder network with feature reduction[J]. Acta Electronica Sinica, 2017, 45(3): 730-739. (in Chinese)

[14] WANG W, ZHU M, WANG J L, et al. End-to-end encrypted traffic classification with one-dimensional convolution neural networks[C]//2017 IEEE International Conference on Intelligence and Security Informatics. Beijing, China: IEEE, 2017: 43-48.

[15] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, QC, Canada: MIT Press, 2014: 2672-2680.

[16] 刘海波,武天博,沈晶,等. 基于GAN-LSTM的APT攻击检测[J]. 计算机科学,2020,47(1):281-286.
LIU H B, WU T B, SHEN J, et al. Advanced persistent threat detection based on generative adversarial networks and long short-term memory[J]. Computer Science, 2020, 47(1): 281-286. (in Chinese)

[17] YIN C L, ZHU Y F, LIU S L, et al. An enhancing framework for botnet detection using generative adversarial networks[C]//2018 International Conference on Artificial Intelligence and Big Data (ICAIBD). Chengdu, China: IEEE, 2018: 228-234.

[18] BUCZAK A L, GUVEN E. A survey of data mining and machine learning methods for cyber security intrusion detection[J]. IEEE Communications Surveys & Tutorials, 2016, 18(2): 1153-1176.

[19] 刘海燕,丛菲. 基于深度学习的网络入侵异常检测综述[J]. 信息系统工程,2020(9):50-51.

[20] BERMAN D S, BUCZAK A L, CHAVIS J S, et al. A survey of deep learning methods for cyber security[J]. Information. 2019, 10(4): 122.

[21] GONG D, LIU L Q, LE V, et al. Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection[J]. arXiv preprint, 2019, arXiv: 1904.02639.