网络空间安全学院（密码学院）

2023-2024 学年度第 2 学期

Y00309 网络安全态势感知 课程论文

论文题目：一种新的两阶段深度学习网络入侵检测模型:LSTM-AE

学生学号及姓名：**20213006839** 甄五四

20213006525 梁浩哲

学生所在班级：信息安全理科实验班

论文提交时间：**2024** 年 **6** 月 **25** 日

授课教师：

教师评阅：

成绩：_____

教师签名：_____

摘　　要

机器学习和深度学习技术被广泛用于评估入侵检测系统(IDS)，这些系统能够快速、自动地识别和分类对网络和主机的网络攻击。然而，当破坏性攻击变得越来越广泛时，就会出现更多挑战，需要全面应对。许多入侵检测数据集是公开访问的，以便供网络安全研究者进一步分析。然而，之前的研究并没有详细检查所提出的模型在各种可公开访问的数据集上的性能。由于攻击的动态性及其快速变化的攻击技术，必须定期更新和对可公开访问的入侵数据集做基准测试。本文将深度神经网络(DNN)和卷积神经网络(CNN)，开发一种灵活有效的 IDS 的深度学习模型进行研究，这些 IDS 能够检测并将它们与所提出的检测网络攻击的模型进行比较。网络行为的不断发展和攻击的快速增长，需要 IDS 的发展以及通过静态和动态方法对长期产生的大量数据集进行评估。这种研究使得识别未来网络攻击的最有效算法成为可能。

我们提出了一种新的两阶段深度学习技术，使用混合长短期记忆(LSTM)和自编码器(AE)来检测入侵检测攻击。使用 CICIDS2017 和 CSE-CICDIS2018 数据集确定 LSTM-AE 的最佳网络参数。实验结果表明，该混合模型效果良好，适用于现代场景下的攻击检测。

关键词：卷积神经网络；深度神经网络；网络入侵检测；深度学习；两阶段模型；LSTM-AE

Abstract

Machine learning and deep learning techniques are widely used to evaluate intrusion detection systems (IDS) capable of rapidly and automatically recognizing and classifying cyber-attacks on networks and hosts. However, when destructive attacks are becoming more extensive, more challenges develop,needing a comprehensive response. Numerous intrusion detection datasets are publicly accessible for further analysis by the cybersecurity research community. However, no previous research has examined the performance of the proposed model on a variety of publicly accessible datasets in detail. Due to the dynamic nature of the attack and its rapidly changing attack techniques, the publicly accessible intrusion datasets must be updated and benchmarked regularly. The deep neural network (DNN) and convolutional neural network (CNN) are examined in this article as types of deep learning models for developing a flexible and effective IDS capable of detecting and comparing them with the proposed model in detecting cyberattacks. The constant development of network behavior and the fast growth of attacks need the development of IDS and the evaluation of many datasets produced over time through static and dynamic methods. This kind of research enables the identification of the most efficient algorithm for identifying future cyber-attacks.We proposed a novel two-stage deep learning technique hybridizing Long-Short Term Memory (LSTM) and Auto-Encoders (AE) for detecting attacks. The CICIDS2017 and CSE-CICDIS2018 datasets are used to determine the optimum network parameters for the proposed LSTM-AE. The experimental results show that the proposed hybrid model works well and is applicable for detecting attacks in modern scenarios.

Keywords: Convolutional neural networkd；eep neural networks；network intrusion detection；deep learning；two-stage model；LSTM-AE

目　　录

1 引言

信息和通信技术(ICT)系统和网络管理各种敏感的用户数据，这些数据容易受到内部和外部攻击者的攻击。这些攻击可能是手动的，也可能是自动的，并不断改进其能力，导致未被发现的数据泄露。随着计算机网络在各个领域的应用越来越广泛，网络安全变得越来越重要。许多企业通过使用防火墙、反垃圾邮件方法和反病毒软件等传统安全技术来保护自己免受网络攻击。不幸的是，这些技术无法识别新的或复杂的威胁[1]。因此，网络入侵检测系统(NIDS)现在被用作监控网络流量和识别入侵事件的第二防线[2]。NIDS 是一种非常有效的防御技术，能够缓解复杂的攻击和威胁[3]。

入侵检测一直是网络安全研究的一个重要领域，因为识别对受保护的内部网络的异常访问至关重要[4]，[5]。NIDS 是由网络设备通过交换机、路由器、TAP (network terminal access points)等网络设备镜像采集而成。这些设备作为监控工具来监控网络违规和策略违规行为[6]。许多公司将 NIDS 与防火墙和应用程序防火墙结合使用，以保护同一网络和系统上的 web 服务器。最近，先进的网络攻击通过使用编码和混淆等不规则模式来绕过安全措施。为了解决这些问题，我们使用基于人工智能的入侵检测系统来识别传统基于签名的入侵检测系统无法检测到的变体攻击。

传统上，入侵检测严重依赖于传统的方法，如加解密方法、协议控制、防火墙和杀毒软件模型。虽然这些方法在识别有限的攻击方面是成功的，但它们在检测大量攻击方面存在困难，并导致高误报率。具体来说，黑客发起了大量的拒绝服务(DoS)攻击，这些攻击很难用传统的方法识别和防御。目前的大多数研究已经将重点转向将机器学习(ML)技术用于入侵检测。与传统技术相比，它们倾向于提高识别率，同时降低与管理大规模攻击相关的开销。支持向量机(Support V vector Machines, svm)能够根据训练数据的属性识别测试数据集中的目标攻击，并且具有较高的内存效率[7]。基于支持向量机的入侵检测模型采用超平面和核函数来识别攻击类。K-nearest neighbor(KNN)设计提供了一种简单、快速、高效的解决方案[8]。它将输入分类为搜索空间中的样本，并使用单个或多个特征向量来确定哪些样本与正常或攻击类最相似。为了最小化与检测攻击相关的错误率，朴素贝叶斯方法使用对特征、偏差和方差有独立假设的概率模型[9]。随机森林

方法[10]使用综合特征选择技术和内在指标对攻击类分类的特征进行排序。其他方法，如 k-means 聚类和逻辑回归，利用聚类和回归来确定攻击类型。然而，传统的机器学习算法在识别具有高度集成特征的攻击方面存在显著缺陷[6]。此外，这些算法在处理噪声和多维交通数据时表现不佳。已经开发了结合两种或多种 IDS 技术的混合机器学习方法，但由于模型复杂性大，它们往往效率较低。

深度学习(DL)算法最近在入侵检测方面取得了进展[11]。基于非线性结构化深度学习系统(如深度神经网络(DNN)[12]、卷积神经网络(CNN)[13]、循环神经网络(RNN)[14]和长短期记忆(LSTM)[15])的入侵检测模型显示出增强的学习行为和提高的入侵检测准确率。此外，近年来硬件设计已经升级，以适应 DL 模型广泛的安全性增强。CNN 对多层感知器进行正则化，因此使用全链接网络确定攻击类别。由于这个问题的严重性，确保网络系统的安全势在必行。使用工具来协助管理和安全操作是至关重要的。此外，这些工具必须自动化，以简化异常事件的检测和对策的实施，以减轻敌对代理人的影响。在本研究中，我们提出了一种采用数据预处理和模型训练的两阶段检测系统。该模型将 LSTM 与 AE 相结合，以提高检测和准确性，如图 1 所示。

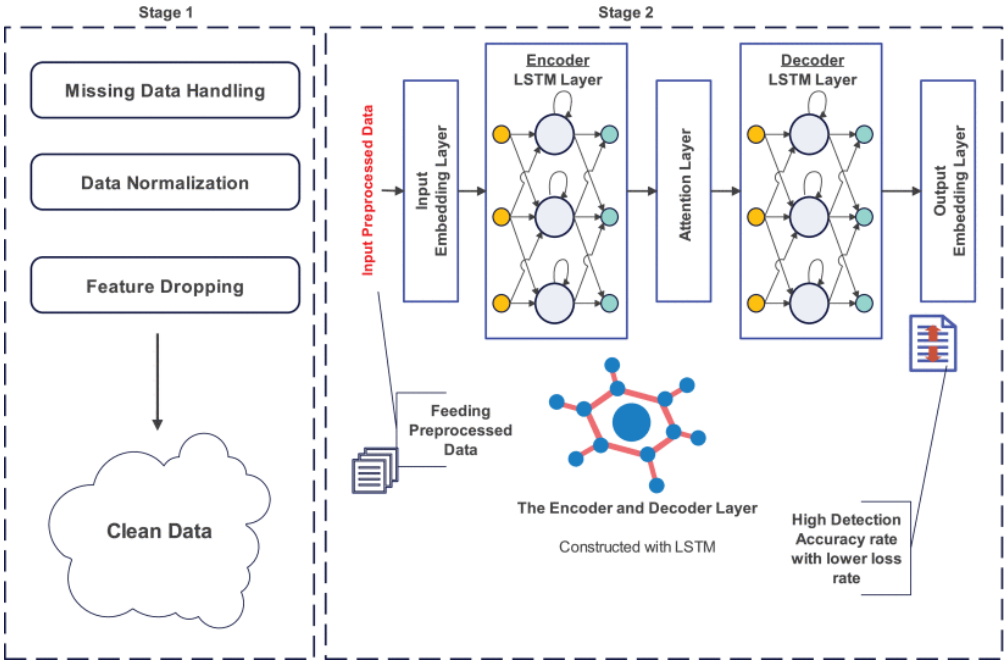

图 1 提出 LSTM-AE 模型框架

在本文中，我们的主要研究目标是提出一种鲁棒的 IDS，能够高效地处理大量复杂的原始网络数据，并提供有效的检测和更高的性能结果。

本文的主要贡献可以总结如下:

1) 我们通过混合 LSTM 和 AE(称为 LSTM-AE)提出了一种基于两阶段深度学习的 IDS，其中数据经过过滤以减少过拟合和欠拟合。
2) LSTM-AE 可以有效地平衡高度不平衡数据集的降维和特征保留。因此，用两个数据集对所提出的模型进行了测试。
3) LSTM-AE 比其他流行的入侵检测模型具有更高的检测性能。

为了分析 LSTM-AE 模型在 IDS 中的性能，可以使用两个常见的误差函数 MSE(均方误差)和 MAE(平均绝对误差)来比较模型的性能。MSE 测量预测值和实际值之间的平方差的平均值，而 MAE 测量绝对差的平均值。一般来说，MSE 对大误差更敏感，可以用来对较大的误差更严厉地惩罚模型，而 MAE 对异常值更健壮，可以提供更好的平均误差估计。为了比较使用 MSE 和 MAE 的 LSTM-AE 模型的性能，我们可以在网络流量数据集上训练模型，并使用这两个误差函数在测试数据集上评估其性能。在两个误差函数方面表现更好的模型被认为是更好的模型。

2 相关工作

自从计算机体系结构发明以来，人们就开始研究与 NIDS 和 HIDS 相关的安全问题。自 2017 年以来，已经发表了许多使用各种深度学习方法进行入侵检测的研究。下面总结了基于模型、特征、数据集和性能度量的深度学习入侵检测的相关研究。Liu 等[16]认为，与其他 IDS 分类器相比，基于卷积神经网络(convolutional neural networks, cnn)的入侵检测模型具有最好的检测率和准确率。近年来，基于传统机器学习的几种技术、模型和方法已经被开发和提出，以解决网络入侵检测问题。本节介绍机器学习和深度学习技术在网络入侵检测和网络入侵防御领域的应用现状。Alzahrani 和 Hong[17]建议在 IDS 中使用基于签名的人工神经网络来识别 DDoS 攻击。运行测试后发现，与基于签名的方法和使用人工神经网络的方法相比，联合策略的准确率更高，达到 99.98%。Kim 等人[18]开发了 AI-IDS，并使用 DNN 模型使用 CICIDS 2017 的实时数据和公共 HTTP 数据集来评估深度学习模型的性能，准确率为 91.69%。当 CNNLSTM 应用于同一数据集时，他们可以获得 98.07%的准确率。同时，论文[18]强调了目前 IDS 检测的不足，并指出其对安全的关键性，需要改进。此外，该研究建议使用 CNN-LSTM 模型在高性能计算环境

中进行有效载荷级深度学习。Zhang 等[19]使用蒙特卡罗树搜索方法(Monte Carlo tree search method, MCTS)生成跨站脚本攻击的对抗性实例。此外，作者利用 GAN 框架来增强入侵检测模型识别对抗性攻击的能力。在实验阶段，CICIDS-2017 数据集被用于创建新的跨站攻击。从数据集中提取 XSS 攻击流量和常规流量实例。GAN 检测模型识别跨站攻击及其对抗实例的准确率达到 99.9%以上。Srinivas 和 Manivannan[20]提出了一种深度学习方法，用于检测和阻止医疗物联网网络上基于 dos 的 Hello flood 攻击。这种攻击通过发送大量 Hello 数据包来降低网络速度，并使用深度信念网络(DBN)模型进行了验证。在[20]中，BAU-ROA 被用于帮助 DBN 模型更有效地运行并提供更好的结果。Rider Optimization Method 是一种计算参数较少的基本优化算法，而 BAUU - ROA 是为了进一步提高其性能而创建的一种元启发式算法。测试发现，BAU-ROA 方法优于其他优化算法，因为它能够增强 DBN 的整体性能。Ujjan 等人[21]利用深度学习模型解决了物联网网络 SDN 检测早期阶段网络安全中采用的基于采样的技术。在他们的研究[21]中，SAE 系统包括一个减少后续层的编码器和一个增加后续层的解码器，就像使用了对称系统一样。为了探索深度学习对入侵检测的影响，使用自适应轮询和 sFlow 技术使用了 SAE 模型，并对结果进行了评估。作为实验的结果，sFlow(两个样本的 CPU 使用研究)和 Adaptive polling(准确率研究)分别取得了成功的发现[21]。为了识别 DDoS 攻击，Priyadarshini 和 Barik[22]在文中描述的 LSTM 深度学习模型在云计算和雾计算环境的 SDN 控制层中实现。以特定时间间隔记录的网络数据包最适合 LSTM 训练，因为它们保留了前一个数据包对当前数据包的影响的信息。LSTM 深度学习模型被确定由三个隐藏层组成，其中有 128 个隐藏单元，作为测试的结果是足够的。当 LSTM 模型应用于 ISCX 2012 和 IDS CTU-13 僵尸网络数据集(合并时称为 Hogzilla)时，达到了这种精度水平。模拟 DDoS 攻击是通过使用开源工具完成的。他们的模型在测试数据集上达到了 98.88%的准确率。Krishnan 等[23]提出了深度自编码器(Deep Autoencoder, NDAE)深度学习模型和随机森林(Random Forest, RF)浅层机器学习方法来对抗 SDN 安全威胁。根据[23]，在 NDAE 中，自编码器不像传统的自编码器那样具有编码器-解码器结构。为了解决浅层机器学习分类器的缺点，研究人员倾向于使用深度学习模型。我们之所以选择 NDAE[23]，是因为它的准确率更高，并且占用的 CPU 和训练时间更少。通过使用 NSL-KDD 和

CICIDS2017 数据集，对 NSL-KDD 和 CICIDS2017 数据集中采用的 DDoS 攻击检测模型进行了评估。为了测试用于识别 DDoS 攻击的模型的有效性，使用了 NSL-KDD 和 CICIDS2017 数据集。使用 NDAE 混合模型进行评估，发现该模型适合用于入侵检测系统，准确率分别为 99.60%和 99.24%。Kanna 和 Santhi[24]提出了优化 CNN (OCNN)和分层多尺度 LSTM (HMLSTM)，并对公开可用的 IDS 数据集进行了评估;NSLKDD、ISCX-IDS 和 UNSWNB15。不需要单独的特征来组合该技术。使用 OCNN-HMLSTM 模型进行评估，发现该模型适合用于入侵检测系统，准确率分别为 90.67%、95.333%和 96.334%。Hussain 和 Hnamte[25]提出了一种用于检测 SDN 环境攻击的 DNN 模型。该模型使用 KDD-CUP99、NSL-KDD 和 UNSW-NB15 数据集进行训练，准确率分别达到 99.61%、98.12%和 81.70%。该研究没有提出任何预处理或数据清理。Mighan 和 Kahani[26]提出了堆叠自动编码器与支持向量机(SAE-SVM)模型的杂交方法来检测异常。该模型使用 ISCX-2012 和 CIC-IDS2017 数据集进行训练，准确率分别达到 95.98%和 99.49%。Lu 和 Tian[27]提出了一种将 SAE 和 Attention-BiLSTM 相结合的改进 LSTM，用于高效的通信入侵检测。该模型采用 UNSW-NB15 数据集进行训练，检测准确率达到 99.41%。Binbusayyis 和 Vaiyapuri[28]提出了一种对抗性深度学习方法(CNN、GAN、LSTM 和 MLP)， GAN 框架在识别 DDoS 攻击方面表现良好。此次评估是使用名为 CICDDoS 2019 的公共数据集进行的，因为该数据集包含最新类型的 DDoS 攻击。Bae 和 Joe[33]提出了一种基于 LSTM-AE 的时间序列 UAV 异常检测模型。该模型没有针对入侵检测系统进行专门的测试，主要关注的是学习时间，没有提到检测准确率。因此，他们提出的 IDS 模型的适用性与检测精度无关，而只与可用性有关。Lindemann 等人[34]研究了 LSTM 网络用于异常检测，重点介绍了 2015 年至 2020 年间提出 LSTM、基于编码器-解码器和混合方法的 12 个来源。这些资料比较了 LSTM 网络与其他 ML 和 DL 模型的性能，但他们没有特别关注 IDS。Musleh 等[35]提出了基于层叠自编码器 LSTM 的自动发电控制(AGC)来维持电网的稳定和运行。尽管这些技术是无模型的，不需要精确的系统模型，但仍然有许多障碍需要克服，例如可能难以拟合的高维观测值。此外，训练期间使用的受限攻击场景对这些检测算法识别零日攻击的能力提出了质疑。Mushtaq 等[36]提出了一种混合 Auto-Encoder 和 LSTM 的两阶段 IDS。使用 NSL-KDD 数据集进行训练，该模型对攻击的分类准确率达到

89%。Mahmoud 等人[37]提出了一种基于 AE-LSTM 的模型来检测物联网环境中的异常。该模型使用 NSL-KDD 数据集进行训练，攻击分类准确率达到 98.88%。这项研究没有包括任何数据预处理来提高训练成绩。Altunay 和 Albayrak[38]提出了一种基于 CNN+ lstm 的混合型 IDS。采用 UNSW-NB15 数据集对该模型进行训练，二元分类准确率为 93.21%，多类分类准确率为 92.9%。提出的系统被强调用于工业物联网网络。尽管检测准确率很高，但训练、验证和测试的损失非常高。没有提出提高性能的建议。Issa 和 Albayrak[39]提出了将 CNN 和 LSTM 杂交的 DDoS 检测方法。使用 NSL-KDD 数据集对模型进行训练，准确率达到 99.20%。数据集陈旧，可能无法反映现代攻击场景。Wu 等[29]提出了一种基于鲁棒变压器的入侵检测系统(Robust Transformer-based Intrusion Detection System, RTIDS)，模型使用 CICIDS2017 和 CIC-DDoS2019 数据集进行训练，准确率分别达到 99.35%和 98.58%。该研究还在训练模型之前进行了数据清理。Wang 等[30]提出了一种基于流形和决策边界的声发射检测系统(MANDA)来检测异常。使用 NSL-KDD 和 CICIDS-2017 数据集对模型进行训练，准确率达到 98.41%，假阳性率为 5%。该模型表现良好，但在训练前没有对数据进行预处理。Umair 等[31]提出了一种用于检测网络入侵的混合多层深度学习模型。该模型使用 KDDCUP99 和 NSL-KDD 数据集进行训练，准确率达到 99%。特征选择也在训练阶段之前进行。Ravi 等[32]提出了基于循环深度学习的攻击检测和分类方法。特征选择阶段采用基于核的主成分分析(KPCA)，训练阶段采用 KDD-Cup-1999、UNSWNB15、WSN-DS 和 CICIDS-2017 数据集。该模型使用 WSN-DS 数据集训练后的检测准确率为 98%，而使用其他数据集训练后的检测准确率为 99%。只有少数关于入侵检测的研究表明，深度学习成功地优于传统技术。用于网络入侵检测的无监督特征学习技术和算法包括深度信念网络(dbn)、dnn、受限玻尔兹曼机(rbm)和自动编码器(AE)等。

3 系统框架

我们提出的方法框架解决了传统方法的一些缺陷，例如当数据量增加或变量之间的相关性变得复杂时，从头开始、过拟合和短期记忆问题。该框架引入了一种结合了自编码器（AE）和长短期记忆（LSTM）网络的两阶段模型，并结合数据预处理阶段，能够有效地区分攻击类别与正常流量。

该模型的架构如图 2 所示。该架构结合了自编码器和 LSTM 网络，旨在通过

两阶段的模型来改进网络入侵检测系统（NIDS）的性能。

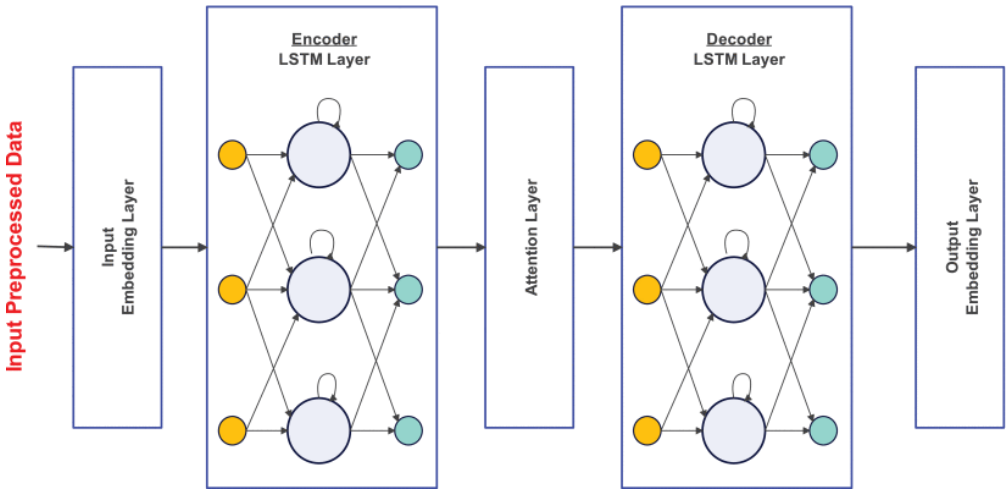

图 2 LSTM-AE 模型体系结构

4 方法

4.1 算法设计

1.数据预处理阶段：首先对原始网络流量数据进行清洗和归一化处理，以提高数据质量并减少噪声。这一步骤确保输入数据的一致性，并提高后续模型训练的有效性。

2. 自编码器（AE）阶段：自编码器用于对输入数据进行编码和解码。编码器将高维的输入数据压缩成低维的瓶颈特征表示（即编码向量），而解码器则试图从这些瓶颈特征中重建输入数据。通过最小化重建误差，自编码器能够学习输入数据的主要特征和结构。

3. 长短期记忆（LSTM）阶段：LSTM 是一种改进的递归神经网络（RNN），能够有效处理序列数据中的长依赖关系。LSTM 使用输入门、遗忘门和输出门来控制信息的流动，从而避免了传统 RNN 中的梯度消失问题。在该架构中，LSTM 网络用于进一步处理从自编码器获得的瓶颈特征，捕捉数据中的时间依赖性和模式。

4. 瓶颈层：瓶颈层连接编码器和解码器，是一个低维特征表示层。其主要目的是通过压缩输入数据来保留关键信息，同时丢弃冗余信息。

5. 重建与预测：最后，解码器对瓶颈特征进行解码，重建出原始输入数据，并计算重建误差。模型还可以利用这些特征进行网络攻击的预测，通过最小化预测误差来调整和优化模型。

通过结合自编码器和 LSTM，提出的 LSTM-AE 模型能够在处理大规模和复杂网络流量数据时，既能捕捉短期和长期的依赖关系，又能有效区分正常流量和异常流量，从而提高网络入侵检测的准确性和鲁棒性

4.2 算法性能分析

算法的性能通过实验结果和多种评估指标来分析，包括准确率、精确率、召回率、F1 分数和误报率。实验使用 CICIDS2017 和 CSE-CICIDS2018 数据集来验证提出模型的有效性。结果表明，提出的 LSTM-AE 模型在各种网络攻击场景下具有良好的检测能力，显著提高了异常检测的准确性和鲁棒性。

4.3 算法复杂度分析

算法的复杂度主要体现在训练和推理过程中：

1. 自编码器的复杂度：自编码器的计算复杂度取决于其层数和每层的节点数。假设 AE 有 L 层，每层有 N 个节点，则其时间复杂度大致为$O(LN^2)$。

2. LSTM 的复杂度：LSTM 的计算复杂度也与其层数和每层的节点数有关。假设 LSTM 有 M 层，每层有 K 个节点，则其时间复杂度大致为$O(MK^2)$。

3. 整体复杂度：整体模型的复杂度为自编码器和 LSTM 复杂度的组合，即$O(LN^2 + MK^2)$。数据预处理阶段的复杂度相对较低，主要包括数据清洗和归一化处理。通过合理设计和优化，可以在保证检测性能的同时，尽量降低计算复杂度和资源消耗。

5 实验

为了验证 LSTM-AE 神经网络在异常检测中的质量，我们还构建了两个额外的模型——简单的 DNN 和 CNN 模型，并与 LSTM-AE 进行比较。DNN 神经网络是一个 4 层的全连接网络，输出层为 softmax 层。输入层和输出层的神经元数量均为 81。三个隐藏层的神经元数量分别为 128、256 和 128。CNN 卷积神经网络有两层卷积层，接着是一个隐藏层，然后是一个 softmax 层作为输出层。输入层和输出层的神经元数量均为 81，隐藏层的神经元数量为 256。为了分析和验证所提出的用于 NIDS（网络入侵检测系统）的 LSTM-AE 模型，以及 DNN 和 CNN 模型，使用了两个

公开可访问的数据集。这两个数据集是 CICIDS-2017 和 CSE-CICIDS-2018 基准数据集。

5.1 数据采集

CICIDS-2017 数据集解决了用于入侵检测评估的实时网络流量数据集不足的问题，包含最新的测试数据，特别是类不平衡信息。相比其他 IDS 数据集，CICIDS-2017 有超过 80 个特征，数据维度更优。数据集包括 8 个会话的流量监控记录，涵盖正常和 14 种攻击类型，数据分布不均。CSE-CICIDS-2018 数据集由通信安全机构和加拿大网络安全研究所合作创建，包含七种攻击场景，包括暴力破解、僵尸网络和 DDoS 等。攻击网络由 50 台计算机组成，受害者组包括 420 台 PC 和 30 台服务器。数据集按天分组，共八天，包含超过 80 个网络流量特征，并以 CSV 格式提供。这两个数据集均开源，可在 https://www.unb.ca/cic/datasets 下载。其中基于删除空值和重复值的方法对 CICIDS-2017 清洗前后分布如图 3、图 4：

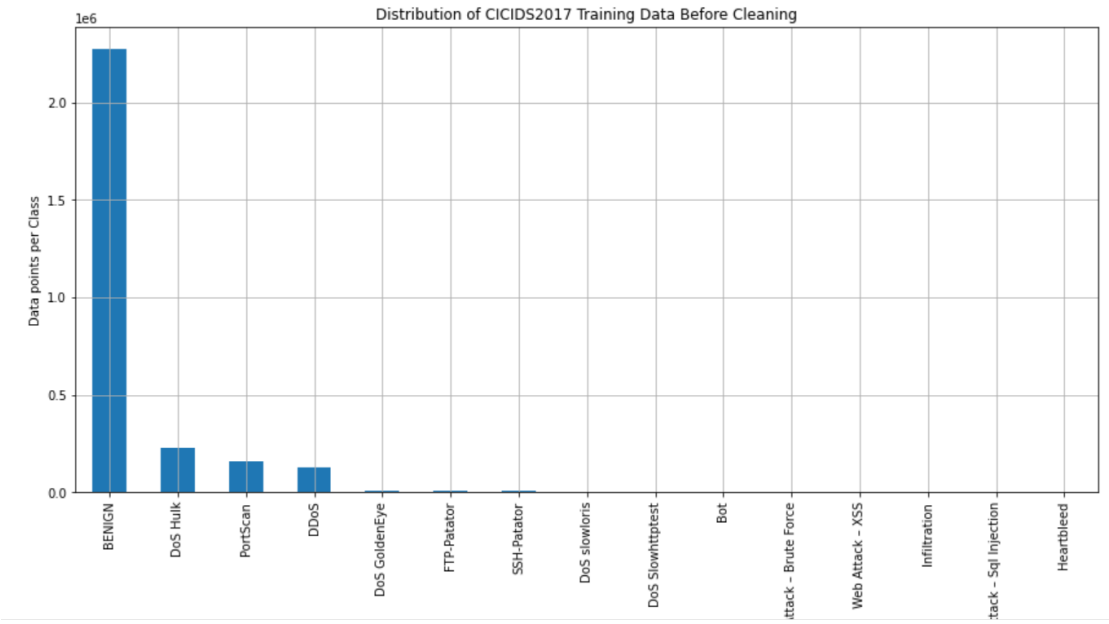

图 3　CICIDS-2017 数据集清洗前

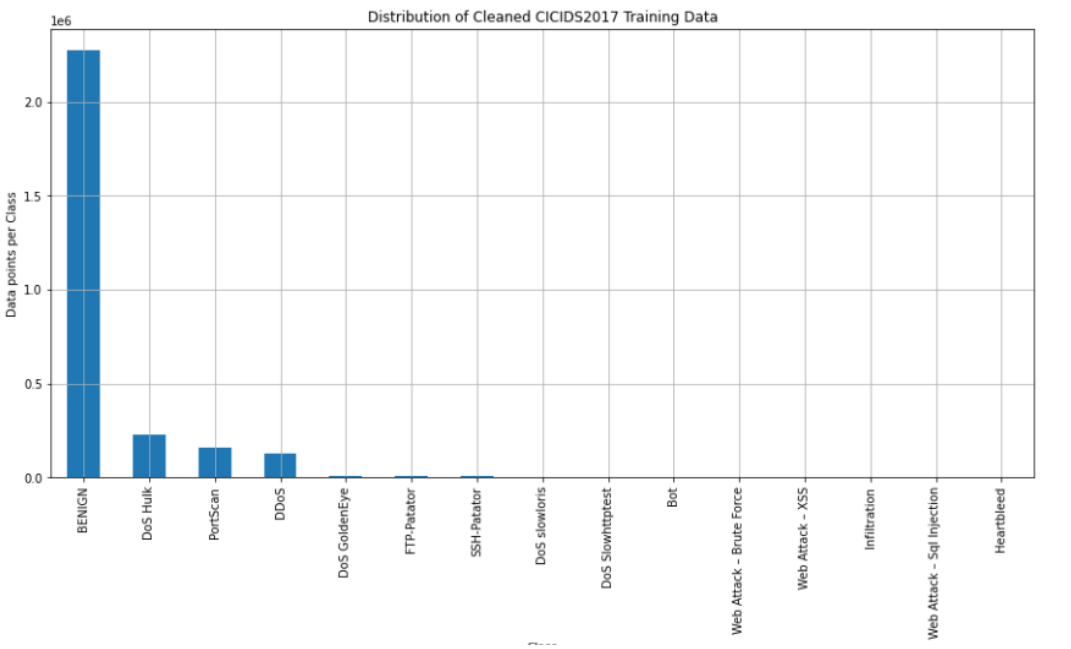

图 4　CICIDS-2017 数据集清洗后

5.2 实验设置及环境

实验是在配置 NVIDIA GeForce RTX 4090 服务器上使用 Python 3.8.10 和 DL 应用程序 Keras3 和 TensorFlow4 的开发库进行的，CUDA 版本如图 5。

```
root@autodl-container-e3534eb58d-2c312bd3:~/autodl-tmp/LSTM-AE# nvidia-smi
Mon Jun 24 14:02:52 2024
+-----------------------------------------------------------------------------------------+
| NVIDIA-SMI 550.54.14              Driver Version: 550.54.14      CUDA Version: 12.4     |
|-----------------------------------------+------------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id          Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |           Memory-Usage | GPU-Util  Compute M. |
|                                         |                        |               MIG M. |
|=========================================+========================+======================|
|   0  NVIDIA GeForce RTX 4090        On  |   00000000:38:00.0 Off |                  Off |
| 30%   32C    P8             22W /  450W |       3MiB /  24564MiB |      0%      Default |
|                                         |                        |                  N/A |
+-----------------------------------------+------------------------+----------------------+

+-----------------------------------------------------------------------------------------+
| Processes:                                                                              |
|  GPU   GI   CI        PID   Type   Process name                              GPU Memory |
|        ID   ID                                                               Usage      |
|=========================================================================================|
|  No running processes found                                                             |
+-----------------------------------------------------------------------------------------+
```

图 5　CUDA 环境

5.3 评价指标

为了评估 LSTM-AE 模型的性能，使用了与大多数过去 NIDS 研究中相同的评估标准。特别是，通过对相关文献的研究，发现准确率、精确率、召回率、F1 分数和误报率（FAR）指标经常被用来评估大多数入侵检测系统的效果。

以下公式用于计算这些指标：

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\text{F1-Score} = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

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

TP、TN、FP 和 FN 分别代表真正例（True Positive）、真负例（True Negative）、假正例（False Positive）和假负例（False Negative）。最后，F1 分数是召回率和准确率的调和平均值，适当地表明了系统的性能。

5.4 优化和分层

当学习率从 0.001 降低到 0.0001 时，所有评估指标的准确性都有所提高。然而，如果将学习率降低到 0.0001 以下，所有评估指标都变得无效。损失和准确度测量表明，降低学习率会对 NIDS 模型的性能产生负面影响。为了优化，我们使用 Adam，一种自适应学习率方法，它为不同参数计算单独的学习率。Adam 通过估计梯度的一阶和二阶动量，为神经网络的每个权重调整学习率。我们比较了建议的 LSTM-AE 模型与具有最佳设置的 CNN 模型以及在原始 CICIDS2017 和 CSE-CICIDS2018 数据集上经过最佳设置的 DNN 模型的性能，如图 6、7 所示。在确保误报率不增加的情况下，LSTM-AE 模型显著提高了模型检测异常实例的能力。

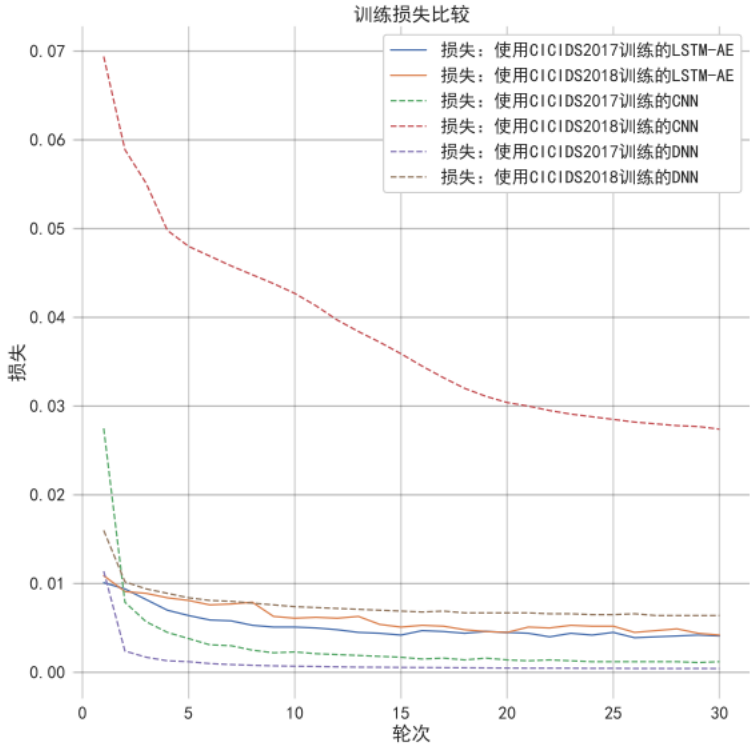

图 6 训练损失比较

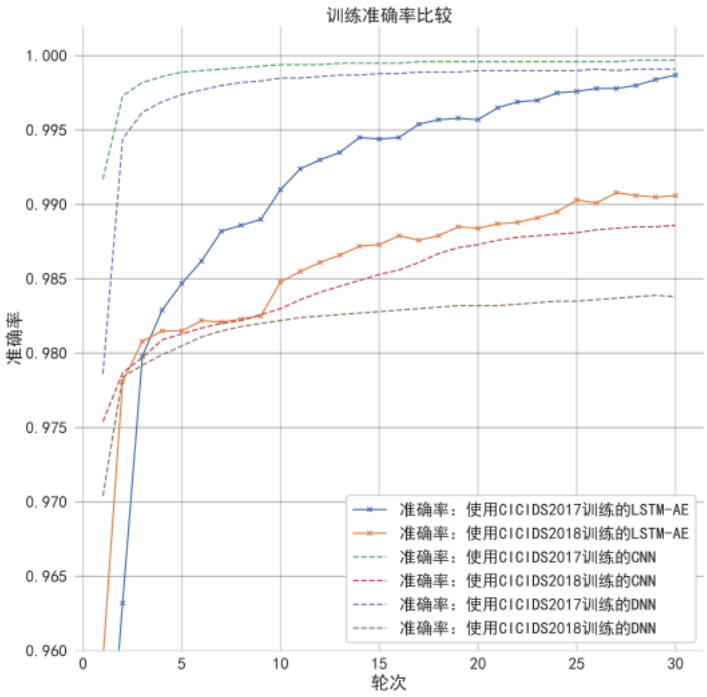

图 7 训练准确率比较

表 1 显示了三种深度学习模型的性能比较：DNN、CNN 和提出的 LSTM-AE 在两个不同的数据集 CICIDS-2017 和 CSE-CICIDS-2018 上。该表有两个主要部分，

每个数据集一个部分，每个部分包括模型在准确性、损失、召回率、精确度和 F 度量方面的分数，以及训练和推理时间。在每一列中，都会显示模型的分数，其中行表示评估指标，列表示三个模型。每个指标都使用百分比或时间值进行评估。在每个指标中，粗体值表示三个模型中的最佳性能。例如，在 CICIDS-2017 部分中，LSTM-AE 模型在准确度、损耗、召回率、精度和 F 测量方面具有最佳性能。但是，CNN 和 DNN 模型具有更好的训练和推理时间。同样，在 CSE-CICIDS-2018 部分中，LSTM-AE 模型在准确率、召回率、精度和 F 测量方面优于 CNN 和 DNN 模型，但 CNN 和 DNN 模型具有更好的训练和推理时间。

CICIDS-2017			
Score	**CNN**	**DNN**	**LSTM-AE**
Accuracy	99.97%	99.93%	**99.99%**
Loss	0.0009	0.0003	0.0005
Recall	99.97%	99.93%	**99.99%**
Precision	99.97%	99.93%	**99.99%**
F-Measure	99.97%	99.93%	**99.99%**
Training Time (epoch)	~73s	~55s	**~184s**
Inference Time	24.65s	28.61s	**53.66s**
CSE-CICIDS-2018			
Score	**CNN**	**DNN**	**LSTM-AE**
Accuracy	98.98%	98.55%	**99.10%**
Loss	0.0245	0.0107	0.0040
Recall	98.97%	98.55%	**99.10%**
Precision	98.94%	98.41%	**99.07%**
F-Measure	98.86%	98.31%	**99.02%**
Training Time (epoch)	~190s	~130s	**~462s**
Inference Time	62.55s	73.14s	**128.24s**

表 1 DNN、CNN 和提出的 LSTM-AE 性能比较

表 1 进一步展示了深度学习模型的训练时间可能会受多种因素影响，例如模型的复杂性、数据集的大小以及所采用的优化技术。正如图 8 所示，对于较小的数据集和简单的模型，当我们的提出模型在 GPU 上进行训练时，训练时间可能仅需几秒钟。为了优化训练时间，可以采用多种技术，如小批量训练、提前停止和迁移学习。此外，还可以使用预训练模型作为进一步训练的起点，从而减少训练所需的时间。

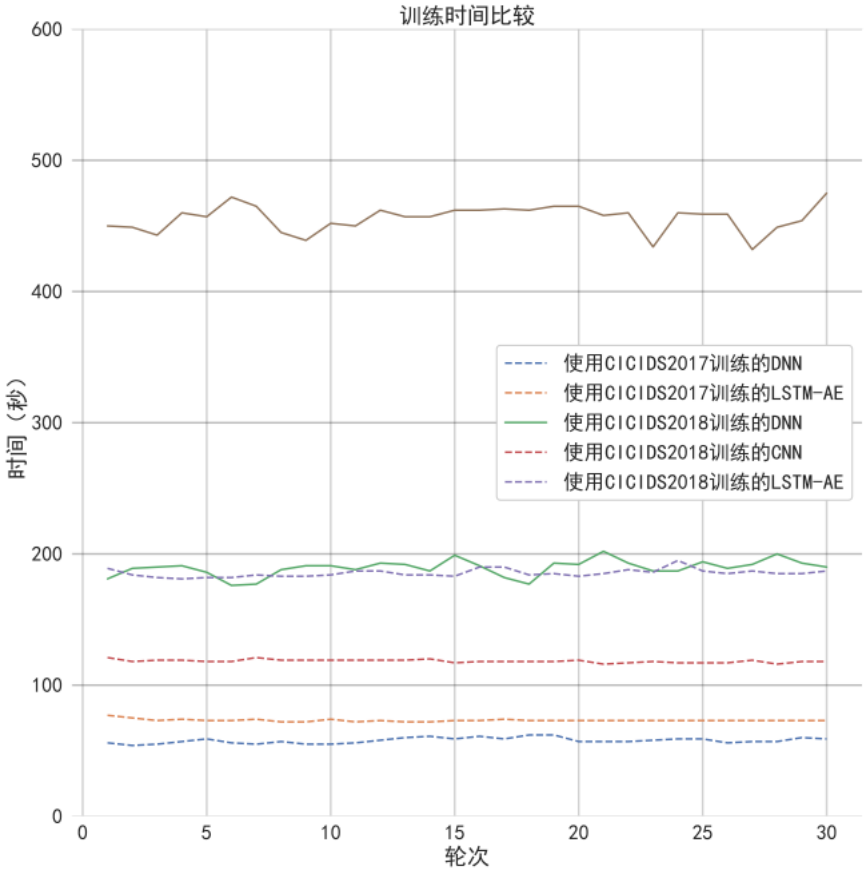

图 8 训练模型所花费的时间

6 总结及展望

本文提出了一种新型的两阶段入侵检测系统，采用高效框架对网络活动进行分析。该系统采用了分布式深度学习模型，包括 DNN 模型和 CNN 模型，并与提出的 LSTM-AE 模型进行了全面比较。我们使用了两个数据集进行模型的训练和测试：CICIDS2017 和 CSE-CICIDS2018。根据我们的了解，该系统能够以分布式方式检测恶意活动，并利用所提出的混合模型来更准确地识别攻击。该模型在农业、医学、语言翻译等多个深度学习领域都有应用，并且由于数据清洗，训练过程中的损失率得到了极大的降低。LSTM-AE 模型在入侵检测领域具有广阔的前景，有潜力进一步提升其性能。例如，研究人员可以尝试不同的架构，如堆叠或双向 LSTM-AE 模型，以评估它们是否能够取得更好的结果。此外，还可以探索注意力机制的应用，帮助模型专注于关键特征并忽略无关特征，从而提高性能。另一个潜在的方向是使用迁移学习技术，通过在大型数据集上预训练模型，并在较小的数据集上微调，来提高准确性并减少训练时间。此外，采用集成方法，如组合多

个 LSTM-AE 模型，有助于减轻过拟合问题，并提高模型对新数据的泛化能力。随着网络安全领域不断出现新型攻击，LSTM-AE 模型可能需要不断调整以保持有效性。研究人员可以收集新的数据集，并对模型进行重新训练，以确保其能够持续检测最新的威胁。总之，LSTM-AE 模型在入侵检测方面具有巨大的潜力，为未来的发展和改进提供了许多机会。

参考文献

[1]G. De Carvalho Bertoli et al., "An End-to-End Framework for Machine Learning-Based Network Intrusion Detection System," in IEEE Access, vol. 9, pp. 106790-106805, 2021, doi: 10.1109/ACCESS.2021.3101188.

[2]B. Mukherjee, L. T. Heberlein and K. N. Levitt, "Network intrusion detection," in IEEE Network, vol. 8, no. 3, pp. 26-41, May-June 1994, doi: 10.1109/65.283931.

[3]F. A. Khan, A. Gumaei, A. Derhab and A. Hussain, "A Novel Two-Stage Deep Learning Model for Efficient Network Intrusion Detection," in IEEE Access, vol. 7, pp. 30373-30385, 2019, doi: 10.1109/ACCESS.2019.2899721.

[4]S. A. Althubiti, E. M. Jones and K. Roy, "LSTM for Anomaly-Based Network Intrusion Detection," 2018 28th International Telecommunication Networks and Applications Conference (ITNAC), Sydney, NSW, Australia, 2018, pp. 1-3, doi: 10.1109/ATNAC.2018.8615300.

[5]X. Zhang, Y. Zhou, S. Pei, J. Zhuge and J. Chen, "Adversarial Examples Detection for XSS Attacks Based on Generative Adversarial Networks," in IEEE Access, vol. 8, pp. 10989-10996, 2020, doi: 10.1109/ACCESS.2020.2965184.

附：我们对论文所提出的模型进行了复现，实验所需代码地址 ZhenWusi/LSTM-AE: 网络安全态势感知课程设计 (github.com)**，由于数据集较大需自行下载，可在** https://www.unb.ca/cic/datasets **下载，未上传到仓库。**