

**硕士学位论文**

**基于NetFlow网络流量异常分析的研究与实现**

**作者姓名：**

**指导教师：**

**学位类别：**

**学科专业：**

**培养单位：**

**年 月**

**A dissertation/thesis submitted to**

**University of Chinese Academy of Sciences**

**in partial fulfillment of the requirement**

**for the degree of**

**Doctor/Master of**

**In**

**By**

**Supervisor:**

**中国科学院大学**

**研究生学位论文原创性声明**

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。承诺除文中已经注明引用的内容外，本论文不包含任何其他个人或集体享有著作权的研究成果，未在以往任何学位申请中全部或部分提交。对本论文所涉及的研究工作做出贡献的其他个人或集体，均已在文中以明确方式标明或致谢。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日 期：

**中国科学院大学**

**学位论文使用授权声明**

本人完全了解并同意遵守中国科学院大学有关收集、保存和使用学位论文的规定，即中国科学院大学有权按照学术研究公开原则和保护知识产权的原则， 保留并向国家指定或中国科学院指定机构送交学位论文的电子版和印刷版文件，且电子版与印刷版内容应完全相同，允许该论文被检索、查阅和借阅，公布本学位论文的全部或部分内容，可以采用扫描、影印、缩印等复制手段以及其他法律许可的方式保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名： 导师签名：

日 期： 日 期：

书脊（此页仅用于制作书脊，不用单独打印放入论文）

黑体小四号，论文题目、作者、中国科学院大学

英文和阿拉伯数字用Times New Roman

|  |
| --- |
| 3cm左右  **论文题目**  **李四**  **中国科学院大学**  3cm左右 |

**摘 要**

**关键词：**，，，，

这里的宏窗体是[单击键入：关键词1]，如果操作失误没有使得宏窗体消失便输入，直接在宏窗体前面输入内容的话，会默认继承前面的文本"关键词："的格式，因为其加粗了，所以其后输入的内容全部默认加粗。正确点选宏窗体输入便不会出现这种问题。

**Abstract**

关键词与摘要间空一行

**Key Words:**

,,,,

注（注的内容在最后的论文里需要删除）：英文摘要与中文摘要的内容应一致，在语法、用词上应准确无误。关键词间用逗号分隔。如无多个关键字请删除多余的关键字。

如有直接粘贴引用pdf文档文本内容到word的话会出现很多换行符，这种情况是因为pdf和word软件是属于两种体系，pdf的文本内容每行直接拷贝出来的话会是一行加换行符号的内容。对于这种情况可以看这篇百度经验[pdf复制到word有空格间隙和换行问题解决办法-百度经验 (baidu.com)](https://jingyan.baidu.com/article/3a2f7c2e646e5867aed61107.html)

**目 录**

[第1章 绪论 1](#_Toc168665238)

[1.1 课题研究背景及意义 1](#_Toc168665239)

[1.2 国内外研究现状 2](#_Toc168665240)

[1.2.1 基于端口号的异常检测 2](#_Toc168665241)

[1.2.2 基于人工规则的异常检测 2](#_Toc168665242)

[1.2.3 基于机器学习的异常检测 3](#_Toc168665243)

[1.2.4 基于深度学习的异常检测 3](#_Toc168665244)

[1.3 本文主要工作 4](#_Toc168665245)

[1.4 本文组织结构 4](#_Toc168665246)

[第2章 相关技术介绍 5](#_Toc168665247)

[2.1 NetFlow技术 5](#_Toc168665248)

[2.1.1 NetFlow技术概述 5](#_Toc168665249)

[2.1.2 NetFlow技术原理 5](#_Toc168665250)

[2.1.3 NetFlow报文格式 5](#_Toc168665251)

[2.1.4 NetFlow技术的应用 5](#_Toc168665252)

[2.2 深度学习概述 5](#_Toc168665253)

[2.3 模型评估方法 5](#_Toc168665254)

[2.4 本章小结 5](#_Toc168665255)

[第3章 AE-GAN网络流量异常发现方法 6](#_Toc168665256)

[3.1 AE-GAN网络流量异常发现模型结构设计 6](#_Toc168665257)

[3.2 实验与结果分析 6](#_Toc168665262)

[3.2.1 实验环境 6](#_Toc168665263)

[3.2.2 NIDS数据集 6](#_Toc168665264)

[3.2.3 实验结果分析 6](#_Toc168665265)

[3.3 科技网流量异常发现 6](#_Toc168665267)

[3.2.1 科技网NetFlow数据采集与预处理 6](#_Toc168665263)

[3.2.2科技网流量异常发现架构 6](#_Toc168665264)

[3.2.3 实验结果分析 6](#_Toc168665265)

[3.4 本章小结 6](#_Toc168665267)

[第4章异常流量分类算法设计 7](#_Toc168665268)

[4.1 异常分类流程概述 7](#_Toc168665269)

[4.2 基于深度学习的异常流量分类 7](#_Toc168665270)

[4.2.1 异常数据集构造 7](#_Toc168665271)

[4.2.2 数据预处理 7](#_Toc168665272)

[4.2.3 特征选择与提取 7](#_Toc168665273)

[4.2.4 异常检测模型构建 7](#_Toc168665274)

[4.3 实验与结果分析 7](#_Toc168665275)

[4.3.1 实验环境 7](#_Toc168665276)

[4.3.2 数据集 7](#_Toc168665277)

[4.3.3 实验结果与分析 7](#_Toc168665278)

[4.3.4 对比分析 7](#_Toc168665279)

[4.4 本章小结 7](#_Toc168665280)

[第5章 实时异常流量检测原型系统设计与实现 8](#_Toc168665281)

[5.1 实验平台搭建 8](#_Toc168665282)

[5.2 系统框架设计 8](#_Toc168665283)

[5.2.1 系统总体架构 8](#_Toc168665284)

[5.2.2 NetFlow存储模块 8](#_Toc168665285)

[5.2.3 异常发现模块 8](#_Toc168665286)

[5.2.4 异常分类模块 8](#_Toc168665287)

[5.3 系统功能测试 8](#_Toc168665288)

[5.4 本章小结 8](#_Toc168665289)

[第6章 总结与展望 9](#_Toc168665290)

[6.1 工作总结 9](#_Toc168665291)

[6.2 研究展望 9](#_Toc168665292)

[参考文献 11](#_Toc168665293)

[附录 14](#_Toc168665294)

[附录一 16](#_Toc168665295)

[附录二 18](#_Toc168665296)

[致 谢 20](#_Toc168665297)

[作者简历及攻读学位期间发表的学术论文与其他相关学术成果 22](#_Toc168665298)

**图目录**

[图3‑1 点选插入题注 5](#_Toc105403921)

[图3‑2 新建标签 6](#_Toc105403922)

[图3‑3 新建图标签 6](#_Toc105403923)

[图3‑4 设置图标签格式 6](#_Toc105403924)

[图3‑5 7](#_Toc105403925)

[图3‑6 图修改位置 7](#_Toc105403926)

[图3‑7 点选插入新图序 8](#_Toc105403927)

[图3‑8 添加新图序 8](#_Toc105403928)

[图3‑9 删除多余空格 9](#_Toc105403929)

[图3‑10 表修改位置 10](#_Toc105403930)

[图3‑11 点选插入新表序 10](#_Toc105403931)

[图3‑12 添加新表序 11](#_Toc105403932)

[图3‑13 错误序号的二级标题 13](#_Toc105403933)

[图3‑14 重新调整编号 14](#_Toc105403934)

**和上面的附录、附录一和附录二一样，图序和图题会空一个汉字符位（2空格），标准的只需要一个空格，所以需要手动删除一个空格。**

**表目录**

[表3‑1 9](#_Toc105402813)

[表3‑2 11](#_Toc105402814)

[表3‑3 续表表格示例 11](#_Toc105402815)

[表3‑4 表格示例 12](#_Toc105402816)

[表3‑5 表格示例1 12](#_Toc105402817)

[表3‑6 表格示例2 12](#_Toc105402818)

**和上面的附录、附录一和附录二一样，表序和表题会空一个汉字符位（2空格），标准的只需要一个空格，所以需要手动删除一个空格。**

# 绪论

本章首先表明了本课题的研究背景与意义。然后介绍了流量异常检测相关技术的研究发展与现状。最后，归纳了本论文的主要工作内容和论文结构。

## 课题研究背景及意义

根据中国互联网络信息中心在2024年发布的《中国互联网络发展状况统计报告》显示，截至2023年12月，我国网民规模达10.92亿人，较2022年12月新增网民2480万人，互联网普及率达77.5%[1]。未来，随着新技术的不断涌现和互联网用户的持续增长，互联网的发展前景将更加广阔。在中国，伴随着数字经济的蓬勃发展，互联网将广泛应用于金融、教育、医疗等领域，为经济和社会的发展带来更多机遇。

随着用户规模的增加和网络技术的发展，网络安全问题也迎来了新的挑战。根据国家计算机网络应急技术处理协调中心发布的《2020年中国互联网网络安全报告》显示，我国境内有5541万个IP地址受到了恶意程序攻击，占到了全国IP地址总数的14.2%，勒索病毒手段不断提升，联网数据库泄露风险问题突出，云平台网络安全事件居高不下[2]。网络安全是网络持续健康发展的重要课题，由于网络创建的初期，对网络安全方面欠考虑，设计的网络协议存在安全方面的缺陷，自身的不完善导致缺失了天然的保护屏障。如今，网络面向世界的开放性和全球信息化的大力推进，人们从网络获取信息的途径也越来越简单，人们可以随时随地的自由的访问网络信息和资源，其中不乏很多不法分子采用非法手段窃取机密信息，恶意破坏网络环境，或者有目的的攻击提供网络服务的系统，由此带来的经济损失不可估计，而且攻击手段越来复杂，越来越隐蔽，目前的检测手段很难追踪到攻击者，这导致攻击者抱着侥幸心理，肆意妄为，这更加重了安全问题的严重性。安全漏洞的不断增多导致安全事件频发，因此，相关网络安全措施的实施刻不容缓。

在众多流量监控和分析技术中，NetFlow技术因其能够详细记录网络流量信息而备受关注。NetFlow是一种由Cisco开发的网络协议，它能够捕获并记录通过网络设备的数据包信息，为网络流量分析、流量分类和异常检测提供了强大的支持[3]。通过分析NetFlow数据，可以有效感知网络中流量的整体情况，识别网络中的各种应用流量，并及时发现异常流量行为，从而提高网络的整体安全性。因此，基于NetFlow技术的异常检测研究具有重要的理论价值和实际应用意义。

## 国内外研究现状

针对恶意的网络攻击行为，网络流量检测技术应运而生，其通过各种手段来解析网络流量数据并识别出具有攻击行为的流量，有效的保障网络安全，维护网络环境，一直以来都被广泛的研究。网络流量异常检测的发展大致可以分为基于端口号、人工规则、机器学习和深度学习四个阶段[4]。

### 基于端口号的异常检测

基于端口号的网络流量分类方法是最早期的传统分类方法。该方法基于识别得到的流量端口号与互联网数字分配机构所发布的知名端口号对比，即可识别出流量数据所属的不同应用类型。比如：FTP文件传输服务的端口号是21号端口，80号端口是HTTP超文本传输服务的使用端口，POP3服务则使用110号端口等。此方法的实现原理非常简单，算法具有时间复杂度低、检测速度快等优点，但随着互联网的迅速发展，此方法受到的限制也越来越多，弊端也逐渐显现出来。比如某些情况下，服务器的端口号并非是固定不变的，而是采用动态分配端口的方式。后来加密流量的出现，也使得从流量数据包中提取出端口号的难度大大增加。Sen[5]等人对缺省端口号的Kazaa、P2P等协议进行了基于端口的检测实验， 该协议的多数流量数据均未使用IANA发布的端口号，实验结果表明，检测的准 确率仅为30%。Moore等人[6]通过实验结果验证了使用IANA官方公布的端口号 进行流量分类，准确率也不够高，准确率在70%以下。Madhukar等人[7]则通过实验验证了基于端口号的检测方法准确率较低的结论，在他们的实验中，有一半左右的流量都无法通过端口号得到正确的识别。

### 基于人工规则的异常检测

基于人工规则的异常流量检测通过分析网络流量的特定参数，如包大小、包延迟和协议类型等，来检测异常流量。该算法通过收集一定时间范围内的流量数据，计算各参数的统计特征，将新的流量数据与已有统计特征进行比较，如果偏差超过一定阈值，则认为是异常流量。例如，Zhu[8]等人构建了基于交叉熵的混合网络流量和容量模型，将网络异常流量中的特征参数设为一系列指标的阈值，当流量通过时，如果指标超过阈值，则判定为异常。He[9]等人提出了一种DDoS攻击防御方案SDCC，结合带宽检测和数据流检测，通过基于置信度滤波（CBF）方法计算数据包的CBF得分，判断低于阈值的数据包为攻击数据包。

基于人工规则的流量检测方法主要是通过端口、负载等匹配规则进行流量识别，其性能取决于人工规则设计的好坏，并且随着计算机网络的不断发展和网络加密技术的出现，这类检测方法的局限性越来越明显，难以适应新型网络环境和满足攻击流量检测的需求。

### 基于机器学习的异常检测

随着互联网的快速发展，网络应用层出不穷，数据流量的频率和业务量呈指数级增长，现代网络环境具备大规模、动态化和复杂化的特点，这使得传统静态匹配算法在网络恶意流量检测上的效果不再满足需求。近年来，机器学习被应用到流量检测技术中，通过人工设计特征建立模型，对网络攻击进行检测，贝叶斯分类、决策树和支持向量机等机器学习方法都被应用到了网络流量检测中并且取得了不错的效果[10]。

Moore等人[11]将遗传算法用于网络流量特征提取，并结合机器学习算法进行流量识别，实验证明该方法对流量异常检测有较高的准确率。Hasan等人[12]通过对多种主流机器学习算法进行分析，比较它们在网络异常流量检测中的准确率。Medagliani等人[13]提出了一种基于自动编码器和受限深度玻尔兹曼机的方法，从未标记的数据集中提取特征并降低维数，在KDD99数据集上实现了92.12%的检测准确率。Kong等人[14]设计了一种入侵检测系统（IDS），解决了有监督学习算法中的局部优化问题，在KDDCUP99数据集上表现出良好性能。

### 基于深度学习的异常检测

深度学习是机器学习的分支，凭借其自身的多层神经网络结构，能更好的表示的特征，在自然语言处理和计算机视觉领域都取得了突破性的成果。在流量检测领域的研究中，深度学习也开始提供新的解决思路和应用方案，网络流量包含了大量的信息，可以很好的利用深度学习进行特征提取，检测网络攻击流量[15]。

Liu等人[16]通过引入注意力机制的卷积神经网络（CNN），将CNN提取的特征输入到LSTM中，解决了记忆丢失和梯度分散的问题。Zhang等人[17]基于遗传改进算法和深度信念网络（DBN）提出了一种自适应异常检测模型，通过遗传算法迭代优化神经元数量，提高检测效率。Maimo等人[18]针对5G网络中的入侵检测问题，提出了一种面向5G的异常检测系统模型架构，能够根据5G用户设备采集的网络流量实时自适应检测异常流量，并优化资源消耗。Elsayed等人[19]将自编码器与LSTM集成到无监督学习中，通过重构误差作为阈值判断流量类型，进一步结合单类支持向量机（OC-SVM）提高检测性能。Ma等人[21]利用KDD99数据集应用深度神经网络检测入侵行为。Niyaz等人[22]研究了基于深度信念网络的NSL-KDD数据集入侵检测方法。这两种方法的共同点是它们的模型均从人工设计的流量特征中进行学习。黎佳玥等[23]提出一种结合了中长短记忆网络和卷积神经网络的预测模型，使用CICIDS2017数据集进行训练和性能测试，总体准确率达到了99.57%。王伟博士[24]提出了一种使用表征学习方法进行恶意流量分类的方法，实验精度达到了98%以上，接近了实际应用的水平。

## 本文主要工作

## 本文组织结构

# 相关技术介绍

## NetFlow技术

### NetFlow技术概述

### NetFlow技术原理

### NetFlow报文格式

### NetFlow技术的应用

## 深度学习概述

## 模型评估方法

## 本章小结

# 基于无监督学习的网络流量异常检测方法

## 网络流量异常检测流程概述

## 基于无监督学习的网络流量异常检测

### NetFlow数据采集

### 数据预处理

### 算法流程

## 实验与结果分析

### 实验环境

### 数据集

### 流量异常检测结果分析

### 科技网流量异常检测

## 本章小结

# 基于深度学习的异常流量分类方法

## 异常检测流程概述

## 基于深度学习的异常流量分类

### 异常数据集构造

### 数据预处理

### 特征选择与提取

### 异常检测模型构建

## 实验与结果分析

### 实验环境

### 数据集

### 实验结果与分析

### 对比分析

## 本章小结

# 科技网实时异常流量检测系统设计与实现

## 实验平台搭建

硬件环境

软件环境

## 系统框架设计

### 系统总体架构

### NetFlow存储模块

### 异常检测模块

### 异常流量分析模块

## 系统功能测试

## 本章小结

# 总结与展望

## 工作总结

## 研究展望

# 参考文献

1. CNNIC．第49次《中国互联网络发展状况统计报告》[EB/OL]. (2024-03-22)[2024-06-02]. https://www.cnnic.cn/n4/2024/0322/c88-10964.html
2. CNCERT/CC．2020年中国互联网网络安全报告[EB/OL]. (2021-07-21)[2024-06-02]. https://www.cert.org.cn/publish/main/46/2021/20210721130944504525772/20210721130944504525772\_.html
3. D. R. Kerr and B. L. Bruins, “Network Flow Switching and Flow Data Export,” 2001.
4. 蹇诗婕, 卢志刚, 杜丹等. 网络入侵检测技术综述[J]. 信息安全学报, 2020, 第5卷(4):96-122.
5. Sen S, Spatscheck O, Wang D. Accurate, scalable in-network identification of p2p traffic using application signatures[C]//Proceedings of the 13th international conference on World Wide Web. 2004: 512-521.
6. Moore A W, Papagiannaki K. Toward the accurate identification of network applications[C]//International workshop on passive and active network measurement. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 41-54.
7. Madhukar A, Williamson C. A longitudinal study of P2P traffic classification[C]//14th IEEE international symposium on modeling, analysis, and simulation. IEEE, 2006: 179-188.
8. Zhu Z, Wang L, Xia L, et al. Flow performance analysis of software defined networking based on cross entropy[C]//2015 8th International Congress on Image and Signal Processing (CISP). IEEE, 2015: 1556-1560.
9. He H, Hu Y, Zheng L H. Efficient DDoS attack detection and prevention scheme based on SDN in cloud environment[J]. Journal on Communications, 2018, 39(4): 139-151.
10. 王荣明. 基于机器学习的入侵检测算法设计与实现[D].电子科技大学,2021.DOI:10.27005/d.cnki.gdzku.2021.003257.
11. Moore A W, Zuev D. Internet traffic classification using bayesian analysis techniques[C]//Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems. 2005: 50-60.
12. Hasan M, Islam M M, Zarif M I I, et al. Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches[J]. Internet of Things, 2019, 7: 100059.
13. Medagliani P, Leguay J, Duda A, et al. Internet of things applications-from research and innovation to market deployment[J]. 2014.
14. Kong L, Huang G, Wu K. Identification of abnormal network traffic using support vector machine[C]//2017 18th international conference on parallel and distributed computing, applications and technologies (PDCAT). IEEE, 2017: 288-292.
15. Aldweesh A, Derhab A, Emam A Z. Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues[J]. Knowledge-Based Systems, 2020, 189: 105124.
16. Liu Y, Garg S, Nie J, et al. Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach[J]. IEEE Internet of Things Journal, 2020, 8(8): 6348-6358.
17. Zhang Y, Li P, Wang X. Intrusion detection for IoT based on improved genetic algorithm and deep belief network[J]. IEEE Access, 2019, 7: 31711-31722.
18. Maimó L F, Gómez Á L P, Clemente F J G, et al. A self-adaptive deep learning-based system for anomaly detection in 5G networks[J]. Ieee Access, 2018, 6: 7700-7712.
19. Elsayed M S, Le-Khac N A, Dev S, et al. Detecting abnormal traffic in large-scale networks[C]//2020 International Symposium on Networks, Computers and Communications (ISNCC). IEEE, 2020: 1-7.
20. Said Elsayed M, Le-Khac N A, Dev S, et al. Network anomaly detection using LSTM based autoencoder[C]//Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks. 2020: 37-45.
21. Ma T, Wang F, Cheng J, et al. A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks[J]. Sensors, 2016, 16(10): 1701.
22. Javaid A, Niyaz Q, Sun W, 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). 2016: 21-26.
23. 黎佳玥, 赵波, 李想等. 基于深度学习的网络流量异常预测方法[J]. 计算机工程与应用, 2020, 第56卷(6):39-50.
24. 王伟.基于深度学习的网络流量分类及异常检测方法研究[D].中国科学技术大学,2018.

# 附录

如果存在多个附录则删除该页，保留“附录一”，“附录二”等，其中“附录”二字（多个附录则为“附录一”等）与题名间空一个汉字符位。

# 附录一

如没有附录一删除该页

# 附录二

如没有附录二删除该页

# 致 谢

年月

# 作者简历及攻读学位期间发表的学术论文与其他相关学术成果

**作者简历：**

年月——年月，在获得学士学位。

年月——年月，在获得硕士学位。

年月——年月，在攻读博士/硕士学位。

工作经历：

年月——年月，在任。

**已发表（或正式接受）的学术论文：**（书写格式同参考文献）

**申请或已获得的专利：**（书写格式同参考文献，无专利时此项不必列出）

**参加的研究项目及获奖情况：**