分类号：
学号：20222108015

密级：公开
单位代码：10759

石河子大学

硕士学位论文

基于集成学习的入侵检测技术研究与系统实现

学位申请人　张思帆
指导教师　邵闻珠
申请学位类别　专业硕士
专业名称　电子信息
研究领域　网络与信息安全
所在学院　信息科学与技术学院

中国·新疆·石河子
2025 年 6 月

分类号：
学号：20222108015

密级：公开
单位代码：10759

石河子大学

硕士学位论文

基于集成学习的入侵检测技术研究与系统实现

学位申请人	张思帆
指导教师	邵闻珠
申请学位类别	专业硕士
专业名称	电子信息
研究领域	网络与信息安全
所在学院	信息科学与技术学院

中国·新疆·石河子
2025 年 6 月

Research and System Implementation of Intrusion Detection Technology Based on Ensemble Learning

A Dissertation Submitted to
Shihezi University
In Partial Fulfillment of the Requirements
for the Degree of
Master of Engineering

By

Zhang Si-fan
（**Network security**）

Dissertation Supervisor:Prof.Shao Wen-zhu

May,2025

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所呈交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发 表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中 作了明确的说明并表示谢意。

研究生签名：张思帆　　时间： 2025 年 5 月 29 日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数 据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：张思帆　　时间： 2024 年 5 月 29 日

导师签名：邵明球　　时间： 2024 年 5 月 12 日

摘要

随着互联网技术的普及，大数据、云计算和区块链等前沿技术的深度应用使网络空间面临严峻安全挑战。网络攻击手段日益复杂化，攻击者从传统端口扫描转向零日漏洞利用，并通过自动化工具大幅降低攻击门槛，导致异常流量检测与防御难度激增。入侵检测系统（IDS）作为网络安全防御的核心环节，其性能优劣直接影响整体防护效果。然而，当前 IDS 技术面临两大关键瓶颈。首先，网络安全数据集中正常与攻击样本数量差异悬殊（通常超 100:1），导致传统机器学习模型训练时偏向多数类别，对少数类别攻击的检测精度不足，且特征提取困难。其次，现有 IDS 多依赖单一算法或简单集成方法，缺乏对复杂攻击场景的适应性，检测流程的时效性难以满足关键基础设施的实时响应需求。为突破这些瓶颈，本文系统研究了集成学习在入侵检测中的创新应用，并设计实现了功能完备的原型系统。以下是核心研究内容与成果：

（1）提出了一种基于语义信息和流量波动阈值融合（Semantic Technology Decision Tree Recursive Feature Elimination）的特征选择方法。该方法通过引入网络攻击的语义特征和动态流量波动特性，优化了传统递归特征消除（RFE）算法。实验表明，该方法在 UNSW-NB15 和 NSL-KDD 数据集上不仅显著降低了特征维度（平均减少 63%），还提升了分类器性能。以 XGBoost 为例，准确率分别达到 92.89%和 93.87%，处理时间较传统方法平均降低 50%，在特征选择效率和模型性能之间实现了良好平衡。

（2）设计了一种层次集成分类模型（Hierarchical Integrated Classification），采用数据重构、特征选择和集成学习策略。利用改进的 SMOTE 处理噪声数据，减少随机波动，并在数据清洗后去除重复、缺失或异常的数据，确保数据质量，以应对数据不平衡问题。同时结合 ST-DT-RFE 方法选择最具代表性的特征子集，消除冗余特征，降低计算复杂度。采用基于攻击感知与高危漏洞率规则的集成方法，整合多个模型所输出的预测成果，在结合零日攻击的扩展数据集 NSL-KDD-2024 上显示，HIC 模型的整体准确率达 96.41%，零日攻击（CVE-2024-0012，CVE-2024-0056，CVE-2024-0034）准确率为 88.5%，85.1%，80.4%，显著优于传统投票和平均集成方法。

（3）开发了一个基于集成学习的入侵检测系统，将上述方法嵌入相应模块。该系统通过多种组件收集网络流量数据和系统终端日志，以可视化形式展示分析结果，并利用 HIC 模型开展入侵检测和攻击分类，同时具备邮件告警功能，能够快速响应安全事件。该系统在检测效能和精准度方面表现优异，可有效提升网络安全防护能力。

关键词：入侵检测；攻击感知；特征选择；集成学习；零日攻击

Abstract

With the popularization of Internet technology, the in-depth application of cutting-edge technologies such as big data, cloud computing and blockchain makes cyberspace face severe security challenges. The methods of network attacks are becoming increasingly complex, with attackers shifting from traditional port scanning to zero day vulnerability exploitation and significantly reducing attack thresholds through automated tools, resulting in a significant increase in the difficulty of detecting and defending against abnormal traffic. As the core component of network security defense, the performance of intrusion detection systems (IDS) directly affects the overall effectiveness of protection. However, current IDS technology faces two key bottlenecks. Firstly, there is a significant difference in the number of normal and attack samples in network security datasets (usually exceeding 100:1), which leads to traditional machine learning models being biased towards the majority of categories during training, resulting in insufficient detection accuracy for attacks in minority categories, and difficulty in feature extraction. Secondly, existing IDS mostly relies on a single algorithm or simple integration methods, lacking adaptability to complex attack scenarios, and the timeliness of the detection process is difficult to meet the real-time response requirements of critical infrastructure. To overcome these bottlenecks, this article systematically studied the innovative application of ensemble learning in intrusion detection and designed and implemented a fully functional prototype system. The following are the core research contents and achievements:

(1)A novel feature selection method based on the integration of semantic information and traffic fluctuation thresholds, termed Semantic Technology Decision Tree Recursive Feature Elimination(ST-DT-RFE), is proposed. By incorporating semantic features of cyber attacks and dynamic traffic fluctuation characteristics, this method optimizes the traditional Recursive Feature Elimination(RFE)algorithm. Experimental results on the UNSW-NB15 and NSL-KDD datasets demonstrate that this approach significantly reduces feature dimensionality(by an average of 63%)while enhancing classifier performance. For instance, when using XGBoost, accuracy rates of 92.89%and 93.87%were achieved on the respective datasets. Processing time was reduced by an average of 50%compared to traditional methods, achieving a favorable balance between feature selection efficiency and model performance.

(2)A hierarchical integrated classification model is designed, employing strategies of data reconstruction, feature selection, and model integration. The model utilizes an improved SMOTE technique to handle noisy data, mitigate random fluctuations, and ensure data quality by removing redundant, missing, or anomalous data after cleaning. This addresses the issue of data imbalance. Additionally, the ST-DT-RFE

method is employed to select the most representative feature subsets, eliminating redundant features and reducing computational complexity. The model integrates predictions from multiple models using an attack-aware and high-risk vulnerability rate-based ensemble approach. On the extended NSL-KDD-2024 dataset incorporating zero-day attacks, the HIC model achieved an overall accuracy of 96.41%, with accuracy rates of 88.5%, 85.1%, and 80.4%for zero-day vulnerabilities(CVE-2024-0012, CVE-2024-0056, CVE-2024-0034), significantly outperforming traditional voting and averaging ensemble methods.

(3)An intrusion detection system based on ensemble learning is developed, integrating the aforementioned methods into corresponding modules. The system collects network traffic data and system terminal logs through multiple components, visualizes analysis results, and performs intrusion detection and attack classification using the HIC model. It also includes email alert functionality to enable rapid response to security incidents. The system performs excellently in terms of detection efficiency and accuracy, effectively enhancing network security protection capabilities.

Key words: Intrusion Detection;Attack Awareness;Feature Selection;Ensemble Learning;Zero-day Attack

目录

第 1 章 绪论

在互联网持续发展的背景下，信息系统正面临日益严峻和复杂的安全挑战。入侵检测作为一种具有前瞻性和主动性的安全防御策略，在精确识别和有效分类各类攻击行为方面，无疑扮演着至关重要的角色。然而，在实际的入侵检测过程中，攻击样本与正常样本数量的显著差异，为检测和分类工作带来了诸多难题。以经典的 CIC-IDS-2018 数据集为例，其中正常样本与攻击样本的数量比例严重失衡，这种极度不均衡的样本分布状况，显著增加了从海量数据中精确提取有效特征的难度，并使得仅依赖单一模型完成高效准确的检测与分类任务变得极为困难，几乎成为一项难以克服的艰巨挑战。因此，当前的学术研究和技术探索主要集中在两个核心方向。其一，深入研究如何运用更为先进、高效的数据特征提取方法，通过这些方法来尽可能地缓解甚至彻底消除样本不均衡问题对入侵检测工作所产生的负面影响。其二，众多学者也在积极致力于探索兼具高度可靠性和卓越稳定性的集成学习技术。期望借助这些先进的融合方法，在复杂的网络环境中，以更高的精确度检测出各种异常行为，并对其进行科学、准确的分类。

在本章的论述中，首先针对入侵检测这一领域的研究背景、所具备的重要意义以及国内外在此方面的研究现状，进行了全面且精炼的概括性阐述。随后，以详尽且有条理的方式，深入剖析了本文所着重研究的具体内容以及精心规划的研究路线。最后，对整篇文章的结构布局进行了清晰明了的说明，旨在为读者呈现一个系统、完整且逻辑严谨的研究框架与论述脉络。

1.1 研究背景与意义

在数字化时代，互联网的迅猛发展深刻地改变了社会的各个层面。然而，网络空间安全面临的挑战日益复杂且严峻。零日攻击由于厂商尚未发现或尚未发布补丁，成为黑客攻击的关键。一旦被恶意利用，攻击者可能在毫无预警的情况下突破系统防线，对系统的完整性和保密性构成威胁。挖矿恶意软件悄无声息地潜入用户设备，消耗电力资源、降低设备性能、加速硬件老化，甚至可能导致系统崩溃。蠕虫病毒通过自我复制的方式传播，像传染病一样在网络中蔓延，占用大量带宽资源，导致网络瘫痪。这些网络攻击的频率和猖獗程度日益增加，持续对网络空间安全构成威胁。《瑞星：2024 年中国网络安全报告》描绘了当前网络威胁的严峻态势。2024 年，尽管瑞星“星核”平台截获的病毒、恶意网址、手机病毒样本数量有所下降，但网络安全事件仍然频繁发生，波及医疗、政府、金融等关键领域，给经济和社会带来了沉重的损失和震动。在恶意软件与网址方面，

虽然病毒样本及感染次数有所下降，但木马病毒的占比仍然最大；勒索软件样本和感染次数减少，而挖矿病毒样本却有所增长。全球恶意网址总量虽有所下降，但美国的恶意 URL 数量最多，中国位居第三，香港的恶意网址数量则居首位。在移动安全领域，手机病毒样本数量有所减少，但以信息窃取类病毒为主，其危害不容忽视。报告还列出了手机病毒和漏洞的前五名，其中 Google Android 系统的漏洞被提及，成为移动安全的重大隐患。在企业安全方面，重大事件频发，例如 Change Healthcare 遭受勒索软件攻击，暴露出企业网络安全的漏洞，凸显了网络安全对企业运营和声誉的重要性。漏洞问题一直是网络安全的顽疾，微软 Office 漏洞经常被攻击者利用。同时，多个高级持续性威胁（APT）攻击组织活跃，对全球信息安全构成了前所未有的挑战。

随着上述网络攻击风险日益加剧，如同幽灵般的 APT 攻击组织在全球各领域游荡，随时可能发起致命一击，各大公司和政府部门深刻意识到网络安全防护的紧迫性。为了抵御外部威胁，部署了多种网络安全设备。常见的网络安全设备诸如综合性入侵感知系统，以及依据多样过滤精细程度运作的情境感知防火墙[1]。这些设备能识别并报警多数攻击模式，但多为单机运行，面对海量网络流量数据，易因不堪重负而宕机，导致防护失效[2]。随着互联网的发展，仅依靠边界安全设备难以绝对防范入侵。一旦攻击者突破边界防御进入内网实施攻击，内网中的漏洞防御极为关键。未来安全研究聚焦于达成外部防御与内部检测的双重机制，同时对每台主机实施近实时的异常监测，并最大限度地减少对主机性能的干扰。

近年来，机器学习研究不断深入，其方法广泛应用于生活的诸多领域[3]。在入侵检测研究中，也有助于构建更加智能、高效和可靠的安全防御体系。国内企业以腾讯的大禹网络安全防护系统为例，该系统深度整合了人工智能与机器学习算法。通过对网络流量的实时监测与海量数据的深度分析，大禹系统能够迅速捕捉到异常流量波动。其独特的多维度威胁识别模块，可精准筛选出隐藏在正常流量中的恶意攻击线索，运用智能关联分析技术，有效识别出 APT 攻击的各个阶段，成功还原完整的攻击链条，进而对 APT 攻击展开精确检测与强力抵御，最大程度保障企业核心信息资产的安全，避免数据泄露风险[4]。

为此，结合机器学习技术设计出一种具备准确的检测能力和快速的响应能力的入侵检测系统来使网络攻击得到快速响应和阻止变得尤为重要。入侵检测系统的设计可以帮助及时发现和防范恶意攻击，保护数据的机密性和完整性，避免因网络入侵造成的信息泄露、系统瘫痪和财产损失[5]。它能够通过监测网络流量和用户行为，准确识别出异常活动和不寻常的模式，快速响应和应对潜在的威胁。通过引入多种技术手段和算法，入侵检测系统能够大大提高网络安全性，并为组织提供一个可靠的防线。同时，还可以减少误报率并提高网络效率，帮助组织节省资源和时间。因此，入侵检测系统的设计对于确保网络安全、保护敏感数据和维护业务连续性至关重要。

1.2 研究现状

在当前国际与国内的研究领域，入侵检测技术的研究重点涵盖了特征选择、机器学习以及集成学习等多个方面。特征选择的核心目的在于从初始特征集合中筛选出最具代表性和相关性的特征子集，通过剔除冗余或不相关的特征，不仅能够降低数据维度，减少计算资源的消耗，同时亦能提升模型的精确度与泛化性能。在入侵检测的应用场景中，精确的特征选择对于模型高效识别各种攻击行为具有至关重要的作用。机器学习技术凭借其固有的优势，能够自适应地学习新的攻击特征，但该方法对数据规模和计算能力提出了较高的要求；集成学习则通过巧妙地融合多个学习器（这些学习器可以是多种不同的机器学习模型）来实现，期望能够整合多个模型的优势；基于深度学习的方法则依赖于神经网络实现特征的自动提取，显著提高了检测的准确性，尽管需要大量的资源进行训练。

1.2.1 流量特征优化研究现状

入侵检测问题在学术领域被广泛视为一个兼具分类与特征优化双重挑战的复杂课题。其中，流量特征优化作为特征优化的关键组成部分，其重要性不言而喻。流量特征优化的核心目的在于显著提升特征集的信息表达能力，能够高效地剔除冗余流量特征以及不相关流量特征，从而有效降低计算复杂度。这一过程不仅优化了特征集的质量，还显著提升了入侵检测系统的整体性能，使其在面对复杂多变的网络攻击时能够更加精准地识别和响应[6]。在特征优化研究领域，相关探索与成果呈现蓬勃发展的趋势，已形成了较为系统的理论与实践体系，涵盖了特征选择、特征变换、特征构建以及特征搜索等多种方法[7]。在具体研究实践中，Roy 等学者[8]运用遗传算法驱动的包裹式策略作为特征搜索手段，把逻辑回归模型的预测精度与特征子集的维度作为适应度函数的衡量指标，以此来评判特征，进而筛选出规模最小且相关性极高的特征组合。谢等学者[9]以无监督学习为理论基础，构建了一种结合谱聚类的无监督特征选择模型。在此过程中，运用谱聚类算法对特征按照相似度进行分类，同时考量特征的区分度与独立性，将其乘积作为衡量特征重要性的标准，选取具有代表性的特征。Wei 等学者[10]创新性地引入了参考向量的概念，通过动态调整参考向量的方向和大小，来灵活地引导最优特征选择的方向和范围。这种改进方法能够更精准地适应不同数据集的特征分布特性，有效提升了特征选择的准确性和效率。Alazzam 等学者[11]以鸽子启发优化器为蓝本，开发出新型二进制连续鸽子启发优化器。其创新点在于，运用余弦相似度定义鸽子飞行速度，搭建从鸽子优化到特征选择优化问题的映射桥梁。这一优化器将鸽子飞行特性与特征选择优化目标有机结合，为高维数据特征降维开辟新路径。SaiSindhuTheja 等学者[12]乌鸦搜索算法与相

对基学习策略相结合，提出了一种相对乌鸦搜索算法用于特征优化。该算法通过生成相对解并评估其适应度来选择特征，借助相对基学习策略改进传统乌鸦搜索算法，使其更契合特征选择优化需求。在算法运行中，生成一系列相对解模拟乌鸦间相互学习与竞争，不断探索特征空间，同时引入适应度评估体系量化评价每个相对解的质量，依据评估结果迭代更新，逐步逼近最优特征组合，为特征选择提供新思路。Panigrahi 等学者[13]将混合朴素贝叶斯决策表与多目标进化特征选择方法相融合，构建出一种创新的特征选择框架。该框架在特征解空间中进行智能探索，借助混合朴素贝叶斯决策表的强大分类能力和概率推理优势，为多目标进化特征选择方法提供指导，使其能够在复杂的特征空间中更高效地搜索最优特征组合。Karuppiah 等学者[14]基于模糊粗糙集的特征优化方法，来封装具有不确定性的相关数据。具体而言，模糊粗糙集通过计算数据对象的隶属度，评估特征在区分不同类别中的有效性和相关性，从而筛选出对分类任务最有价值的特征子集。这种方法在处理包含噪声和不确定性的数据时表现出色。Zhao 等学者[15]选择 Sentinel-1 和 Sentinel-2 遥感数据集提取初始特征，随后使用差分进化特征选择算法筛选出十个特征子集，最后利用主成分分析对这些特征子集进行降维处理，以提高数据的可处理性和模型的效率。Mohammadi 等学者[16]基于线性相关系数的过滤器和基于墨鱼算法的包装器进行特征选择。利用过滤器对初始特征进行排序，然后将排名靠前的特征作为包装器的输入。结合两级筛选机制的方法有效降低了特征维度，还显著提升了墨鱼算法的性能。Qu 等学者[17]设计了一种单分支自监督视觉表示学习方法，使用渐进式级联降维模块，进行适应特征选择和激活，以获取最具代表性的特征。该方法的创新性在于将自监督学习与特征优化相结合，通过自适应模块动态调整特征选择策略，能够更好地适应不同数据集的特征分布。同时，渐进式级联降维模块不仅有效降低了特征维度，还保留了特征的核心信息。

1.2.2 入侵检测研究现状

Belavagi 等学者[18]在研究中对比了逻辑回归、随机森林等模型，结果表明在入侵检测领域，随机森林相较于其他模型更具成效。不过，他们并未对该模型展开有效的优化。Du 等学者[19]在其研究中，将 LSTM（长短期记忆网络）应用于系统日志数据的入侵检测工作。考虑到系统日志处于持续变化的状态，他们还给出了在线更新模型以适应变化的方法。LSTM 网络在处理序列数据方面具有独特优势，能够有效捕捉数据中的时间依赖性，适用于分析具有时间序列特性的系统日志数据。通过在线更新模型，该方法能够及时适应系统日志的变化，保持入侵检测的准确性和时效性。Taher 等学者[20]在研究中把有监督学习模型与特征过滤进行结合，尝试找出能更精准识别攻击流量的组合方式。经研究发现，在入侵检测中，包装特征选择的神经网络在性能表现上优于传统机器学习算法。但遗憾之处在于，他们没有就其他优秀的有监督学习模型，从多个维度展开对比。

Paulo 等学者[21]运用随机森林模型做入侵检测与攻击分类，该模型对复杂网络流量数据适应性和稳定性良好，经参数调优，检测准确率高，降低误报率，检测速度快且能应对大数据计算压力。但对新型攻击检测能力不足，解释性欠缺。未来可结合其他算法提升性能，探索增强解释性的方法。田志宏等学者[22]成功构建了一套基于 URL 解析的网络攻击检测系统。该系统以前沿的分布式深度学习技术为支撑，具备强大的边缘设备部署能力，能够有效应对网络边缘环境下的复杂攻击场景。同时，系统通过整合多个并行深度学习模型，实现了对网络攻击检测模型的高效更新与优化，显著提升了检测的准确性和时效性。MShafiq 等学者[23]提出一套框架模型，用于网络恶意流量的识该框架所采用的过滤模型（CoCRISPAUC）能够筛选出有用特征，去除会干扰攻击分类的无用特征。Chen 等学者[24]提出了一种入侵检测策略，该策略结合了神经网络无监督学习算法和白名单过滤技术。通过白名单机制筛选出异常通信行为，同时运用神经网络无监督离线样本进行训练和学习。即使在信息不全的情况下，也能通过对规则库的优化，提升异常通信检测的效率和准确性。这种结合方法不仅能够有效应对复杂多变的网络环境，还能在数据不完整或不充分的情况下，保持较高的检测性能。Ezzarii 等学者[25]提出了一种基于表观遗传算法的入侵检测分类器，用于对网络攻击进行分类。实验表明，该方法的检测率优于普通遗传算法分类器，且在速度上也更快。徐汝志等学者[26]为攻克注入式数据篡改检测难题，构建优化聚类模型。该模型整合多源检测指标进行聚类分析，有效提升了自身稳定性。Choi 等学者[27]开发出一套基于自动编码器的无监督攻击分类系统。该系统借助训练数据中的异常比例来设定损失边界值，能够达到 91.70%的检测率。Laetitia 等学者[28]鉴于标记数据集成本高昂，且可用的标记数据集稀缺，提出基于自动编码器的无监督学习模型，能从可视化网络图中精准识别内部威胁异常事件。通过 CIC-IDS-2017 数据集仿真实验，验证其优势。

1.2.3 集成学习研究现状

集成学习技术的关键目的是把多个模型的优势结合起来，从而提高恶意攻击检测率。Li 等学者[29]提出一种集成学习方法，用于攻击高效分类。该方法运用权重投票机理对新增样本进行归类，在投票进程中依据各分类器的失误比率实时调控其权重，进而提升分类的精准度。Dickson 等学者[30]提出内核集成技术，可自动筛选多核，确定集成模型参数。经实践验证，该技术在复杂数据环境下展现出卓越性能。刘金平等学者[31]提出一种基于核极限学习机的攻击分类模型。借助装袋法融入稀疏随机特征训练学习机，选择性集成策略融合多模型，该策略基于边缘距离最小化原则。然而，由于实现复杂，需反复计算样本相似度，限制了其在大规模网络系统中的应用。Zhou 等学者[32]提出一款基于曲线下面积自适应增强算法的集成模型，通过粒子群优化算法融合多个 AdaBoost 分类器。但该策略提升了模型间耦合度，致使在其他领域的迁移应用受阻。Saikat Das[33]在

研究中融合自然语言处理和集成学习用于攻击分类，实现了较高检测率。Kumar 等学者[34]融合决策树、朴素贝叶斯以及随机森林算法，构建出一种可识别医疗物联网所遇网络攻击的多级检测框架。该框架通过整合多个模型来提升入侵检测性能，但空间复杂度显著增加。Tsogbaatar 等学者[35]针对数据异构和不均衡难题，提出了一种基于软件定义网络（SDN）的动态攻击检测集成学习框架。该框架具备高检测率和强可靠性。Zhang 等学者[36]提出一种攻击识别方法，通过堆叠策略集成模型。然而，堆叠集成法耦合度高，子分类器不稳定会直接影响入侵检测效果。

从上述内容可以看出，研究人员在网络入侵检测的分类与特征提取方向已开展诸多研究，但现有技术在检测稳定性与分类准确率上仍存在提升空间。因此，本文将实施有效的数据标准化与特征选择操作，以获取网络攻击的强数据特征，并设计集成学习模型，构建出稳定的网络入侵检测系统。数据标准化可消除特征间的量纲差异，提高模型收敛速度与精度；特征选择能去除冗余特征，降低数据维度与计算复杂度，增强模型泛化能力。集成学习模型则结合多个基学习器的优势，提升系统稳定性和鲁棒性，使其在复杂网络攻击下可靠运行。

1.3 研究内容

本论文深入分析了现有的网络入侵检测方法，揭示了当前技术在数据预处理、特征选择及检测分类等关键环节的不足之处。在此基础上，以网络流量的特性为研究出发点，系统地研究入侵加检测技术并构建实时系统，旨在有效提升入侵检测的准确性和稳定性。研究内容主要包括以下几个方面：

第一，在特征选择阶段，为优化特征权重计算，使模型更关注对分类结果影响显著的特征，以递归特征消除（RFE）为核心方法开展研究。该研究通过精准计算和优化网络攻击数据的特征权重达成这一目标。运用 RFE 手段，逐步去除对模型贡献较小的特征，进而确定最优特征子集。针对该子集，运用决策树、KNN、随机森林、XGBoost 以及其他特征选择方法，对分类性能与准确性进行综合评估。

第二，基于 LightGBM、CatBoost、XGBoost 和 RF 等机器学习算法，开展攻击感知分类模型的集成策略研究。兼顾特征选择与攻击分类，融合数据重构、降维及集成等策略应对复杂挑战。数据预处理时，改进 SMOTE 处理噪声数据，降低波动干扰，清洗数据保证质量。融合网络流量特征与梯度递归消除评估特征重要性，筛选关键特征，降低计算复杂度，提升准确性。模型构建采用攻击感知与高危漏洞率规则集成法，融合多模型预测结果。依据相关信息和规则为各模型赋权，加权平均预测结果，全面捕捉攻击特征，提高分类准确率。

第三，在网络入侵检测系统的设计与实现中，开展对网络流量特征分析的研究。在