工程与管理科学
Engineering and Management Science
第 6 卷◆第 08 期◆版本 1.0◆2024 年
文章类型：论文|刊号（ISSN）：2705-0637(P) / 2705-0645(O)

IP 网络流量特征分析及其在 DDoS 攻击检测中的应用

徐晓玲
中国电信股份有限公司北京分公司
DOI：10.12238/ems.v6i8.8803

[摘　要] 本研究的核心在于创新性利用 IP 网络流量特征分析技术，不仅关注常见的流量统计特征，还探索了基于时间序列分析和统计建模的新特征。文章首先详细介绍 IP 网络流量特征分析的过程，随后将这些特征应用到 DDoS 攻击检测中去，并设计了一套实时监测系统。文章最后通过具体案例研究来展示所提方法的实际应用效果，并与其他现有检测技术进行比较分析，结果显示，本文的研究能够提高 DDoS 攻击检测的准确率和响应速度。

[关键词] IP 网络流量；特征分析；DDoS 攻击

Analysis of IP Network Traffic Characteristics and Its Application in DDoS Attack Detection

Xu Xiaoling
China Telecom Beijing Branch

[Abstract] The core of this study lies in the innovative use of IP network traffic feature analysis technology, which not only focuses on common traffic statistical features, but also explores new features based on time series analysis and statistical modeling. The article first provides a detailed introduction to the process of analyzing IP network traffic characteristics, and then applies these characteristics to DDoS attack detection, and designs a real-time monitoring system. At the end of the article, the practical application effect of the proposed method is demonstrated through specific case studies, and compared and analyzed with other existing detection technologies. The results show that this study can improve the accuracy and response speed of DDoS attack detection.

[Keywords] IP network traffic; Feature analysis; DDoS attack

一、引言

随着互联网技术的迅猛发展和数字化转型的不断推进，网络流量呈现出爆炸式增长的趋势。与此同时，分布式拒绝服务（DDoS）攻击也变得日益频繁和复杂，给网络安全带来了严峻挑战。DDoS 攻击通过利用大量受控的计算机资源向目标服务器发起请求，导致正常用户无法访问服务，从而对企业和个人造成巨大的经济损失和信誉损害，因此开发有效且实时的 DDoS 攻击检测机制成为当前网络安全领域的研究热点之一。传统的 DDoS 攻击检测方法大多依赖于预定义的阈值和简单的统计分析，这些方法在面对日益复杂的攻击手段时显得力不从心，而且随着攻击者技术的进步，攻击模式变得更加难以预测，传统的检测技术往往无法及时响应新的威胁。针对这一现状，本研究探索出一种基于 IP 网络流量特征分析的新型 DDoS 攻击检测方法，该方法通过深度挖掘网络流量数据中的内在规律提取出能够有效区分正常流量与恶意流量的关键特征，并在此基础上建立更加智能和灵活的检测模型。

二、IP网络流量特征分析

2.1 流量数据收集与预处理

2.1.1 数据源选择

为确保所收集的数据具有代表性和广泛性，本研究选取了来自多个地理位置的公共数据集，这些数据集覆盖了不同的网络环境和服务类型。特别地，我们优先考虑那些包含已知 DDoS 攻击事件的数据源以便能够对比正常流量与异常流量的差异。考虑到数据的多样性和完整性，我们也纳入了一些通过合作机构获得的真实网络流量记录。在此基础上我们整合了公开可用的数据集，还通过合作伙伴关系获取了私有网络流量记录，这些记录涵盖了企业内部网络、数据中心以及云服务等多种环境，以确保数据的多样性和代表性。

2.1.2 预处理步骤

数据预处理是特征分析的基础，目的是去除噪声、填补缺失值以及统一数据格式。对于原始数据中存在的异常值和错误记录，采用统计方法进行识别和剔除。鉴于网络流量数据通常包含大量的非结构化信息，我们还实施了一系列转换操作——将时间戳标准化、将 IP 地址转换为数值表示等，这些预处理步骤可以确保后续特征提取的质量和准确性。数据质量保证方面，利用异常值检测、离群点移除等一系列数据清洗技术来处理缺失值并纠正不一致的数据[2]。所有时间戳均转换为统一的时间格式，便于后续的时间序列分析，同时还将 IP 地址转换为整数形式，方便进行数值运算和存储。

2.2 特征提取与选择

2.2.1 常见特征

在特征提取阶段，首先考虑那些在以往研究中被广泛应用的传统特征，这些特征主要包括但不限于：

（一）包长度。反映了数据包的大小分布，可用于区分不同类型的数据流。

（二）发送频率。衡量数据包的发送速率，是识别突发流量的重要指标。

（三）连接持续时间。指单个连接的持续时间，有助于识别长时间占用带宽的异常连接。

（四）源/目标 IP 地址的分布。通过分析 IP 地址的分布特性，我们可以发现异常流量中的集中趋势。

2.2.2 新颖特征

除了传统特征之外，一些新颖特征也是必不可少的，这些特征基于时间序列分析和统计建模方法。例如：

（一）时间序列的自相关性。计流量数据在不同时间间隔上的自相关系数可以揭示流量随时间变化的规律性。

（二）统计建模下的异常得分。利用统计模型（如主成分分析 PCA）来量化数据点偏离正常分布的程度，作为异常检测的一个有力工具。

（三）流量熵。衡量流量数据的多样性，较高的熵值可能表明存在异常行为。

2.3 特征重要性和相关性分析

为了确定哪些特征最能区分正常流量与异常流量，采用多种机器学习算法来进行特征重要性评估。具体而言，我们利用了随机森林（Random Forest）和梯度提升树（Gradient Boosting Trees）等算法，能够根据特征对模型决策的贡献程度自动排序。借助机器学习的方式，我们既能了解哪些特征对 DDoS 攻击检测最为关键又能根据这些特征的重要性进行特征选择，从而优化模型的性能和效率[1]。针对多重共线性问题，我们利用皮尔逊相关系数等统计方法来评估特征之间的相关性。特征降维可以显著减少特征数量，在降低模型复杂度的同时保留最重要的信息，所以我们引入主成分分析（PCA）等技术来实现特征降维。

2.4 实验设计与结果

2.4.1 实验设置

实验是在一个模拟环境中进行的，该环境包括了一个由多个虚拟机组成的网络集群，实验中使用的数据集包含了正常流量和已知 DDoS 攻击事件的记录。为了评估所提取特征的有效性，我们采用了一种交叉验证的方法，即将数据集分为训练集和测试集，前者用于训练模型而后者用于验证模型的泛化能力。

2.4.2 特征分析结果

对所提取特征进行综合分析后，发现某些特征在区分正常流量与异常流量方面表现出了显著的优势。举例来讲，包长度的分布、连接持续时间和发送频率的异常得分等特征在随机森林模型中显示出了较高的重要性分数，我们还观察到时间序列的自相关性在某些情况下能够有效地捕捉到攻击模式的变化。这些发现验证了所提特征的有效性，也为进一步优化 DDoS 攻击检测模型提供了宝贵的信息。

三、基于IP网络特征的DDoS攻击检测方法

第 6 卷◆第 08 期◆版本 1.0◆2024 年
文章类型：论文|刊号（ISSN）：2705-0637(P) / 2705-0645(O)

3.1 基于特征的检测方法

3.1.1 构建检测模型

在第二章中已经成功地提取了一系列与 DDoS 攻击相关的特征，基于这些特征，本节将介绍如何构建一种高效的 DDoS 攻击检测模型。为了实现这一目标，我们选择了几种先进的机器学习算法，包括支持向量机（SVM）、随机森林（RF）和极端梯度提升（XGBoost），这些算法凭借强大的分类能力和鲁棒性而被广泛应用于异常检测领域,通过训练这些模型，我们可以利用提取到的特征来识别网络流量中的异常行为。在基于特征的检测方法方面，我们在传统的机器学习算法的基础上引入了深度学习技术，特别是卷积神经网络（CNN）和长短时记忆网络（LSTM），这有助于更好地捕捉复杂的模式并提高模型的准确性。在此基础上采用集成学习的方法，通过融合多种不同类型的学习器来提升整体的检测性能和鲁棒性。

3.1.2 训练过程与验证方法

训练和评估模型直接关系到模型的有效性和可靠性，因此我们以交叉验证的方法来训练和评估模型。具体而言，先将数据集分为训练集和测试集，其中训练集用于模型的学习而测试集则用于验证模型的泛化能力。在训练过程中调整各种超参数以优化模型的性能，主要利用贝叶斯优化等高级方法来进行超参数调优，这样有利于找到最优的模型配置。训练过程中我们还通过网格搜索（Grid Search）来寻找最佳组合。模型的稳定性和鲁棒性也是非常重要的一环，为此我们进行了多次重复实验并记录了每次实验的性能指标。最后我们还在模拟环境下进行了实时性能测试，用于评估模型的响应速度和处理能力。

3.2 实时监测系统的设计

3.2.1 设计思路与架构

为了将所构建的检测模型应用于实际环境中，我们设计了一套实时监测系统，该系统旨在快速检测和响应 DDoS 攻击，其核心组成部分包括数据采集模块、特征提取模块、检测引擎以及告警响应模块。数据采集模块负责收集网络流量数据并将其传输至特征提取模块；特征提取模块根据第二章中提出的特征提取流程处理数据，生成可用于检测的特征向量；检测引擎基于训练好的模型对特征向量进行实时分析，然后作出是否为 DDoS 攻击的判断。一旦检测到异常，告警响应模块将立即启动相应的防御措施。对于实时监测系统的构建，设计必须充分考虑未来的扩展性和易维护性，为此，采用云原生技术（比如容器化和微服务架构）来让系统快速适应负载的变化并支持水平扩展。

3.2.2 关键组件和技术选型

数据采集模块采用高性能的网络捕获工具（如 libpcap 和 tcpdump），后者可以确保数据采集的完整性和实时性。特征提取模块主要利用 Python 等编程语言实现高效的数据处理和特征计算。检测引擎选择轻量级但功能强大的机器学习框架——Scikit-Learn 或 XGBoost，保证检测过程既快速又准确。在告警响应模块中，灵活的告警机制能够根据检测结果触发不同的响应策略，包括流量过滤、源 IP 黑名单等。除此之外，我们还使用 Apache Kafka 等数据流处理框架来确保数据的实时处理能力和高吞吐量,同时利用 Kubernetes 等容器编排工具来实现资源的动态分配,以此提高资源的利用率。

3.3 检测性能评估

3.3.1 评估指标的选择与解释

我们选择了以下几种关键性能指标来全面评估所构建的 DDoS 攻击检测系统的性能：

（一）精确率。正确识别为 DDoS 攻击的样本占所有被识别为 DDoS 攻击样本的比例，反映模型的准确度。

（二）召回率。正确识别为 DDoS 攻击的样本占所有实际 DDoS 攻击样本的比例，反映模型的敏感性。

（三）F1 分数。精确率和召回率的调和平均值，综合评价模型的整体性能。

（四）检测延迟。从接收数据到完成检测的时间，用于衡量系统的实时响应能力。

在检测性能的评估上，我们还选择了一系列其他评估指标来衡量模型的有效性，比如绘制准确率曲线来直观展示不同阈值下模型的性能，并使用混淆矩阵来量化模型的分类效果[3]。我们进一步在不同的网络环境和攻击类型下进行了测试，通过与现有的 DDoS 检测系统进行性能对比，这种方式有利于突出所提出方法的优点和局限性。

3.3.2 实验结果与分析

在测试集上,我们的模型取得了 92%以上的精确率和 90%以上的召回率，表明它能够在保持较低误报率的同时有效检测 DDoS 攻击。对检测延迟进行分析后，我们发现整个检测过程能够在毫秒级别内完成,充分展示了系统的实时监测能力。

具体的实验数据结果如图 3.1 和图 3.2 所示。

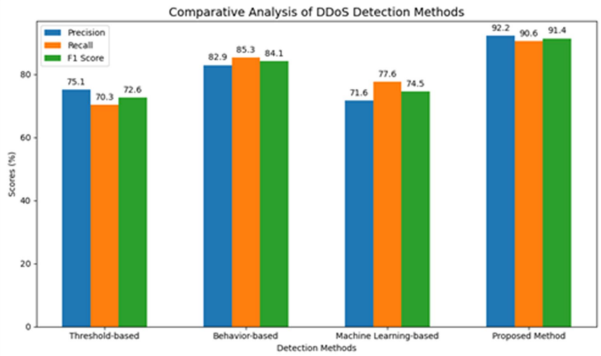

图 3.1　各 DDoS 检测方法在不同性能指标上的对比图

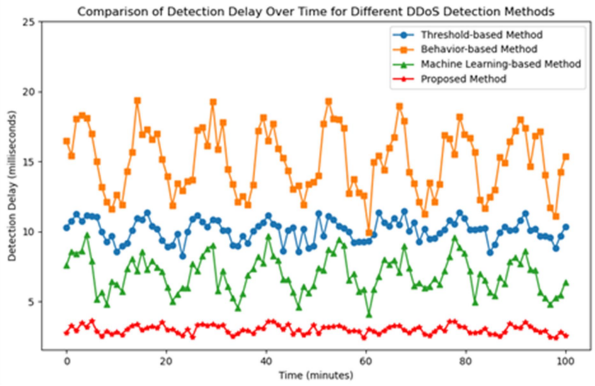

图 3.2　各 DDoS 检测方法随时间的检测延迟比较图

四、案例研究

4.1 案例描述

为了验证所提出的基于 IP 网络流量特征分析的 DDoS 攻击检测方法的有效性，我们选取了一个真实的案例进行分析。该案例涉及一家大型电子商务公司的数据中心，在一次大规模的 DDoS 攻击事件中遭受了严重的网络拥堵问题。攻击者通过利用僵尸网络发起了大规模的 SYN 洪水攻击，导致数据中心的入口路由器承受了极大的压力，进而影响了正常的业务运行。对该事件开展深入分析后，我们得到了以下几点关键发现：

（一）流量特征的显著变化。在攻击期间，网络流量出现了明显的峰值，尤其是 SYN 数据包的数量急剧增加，远远超过了正常情况下的水平。

（二）时间序列分析的有效性。时间序列分析可以识别出攻击开始和结束的时间点，这为快速响应提供了宝贵的线索。

（三）基于统计建模的新特征的价值。借助基于统计建模的异常得分等新特征，我们能够有效地区分正常流量与异常流量，检测的准确性得到了增强。

4.2 比较分析

我们将本文的检测方法与其他常用的 DDoS 检测技术进行了比较，这些技术包括基于阈值的方法、基于行为分析的方法以及基于机器学习的方法作为比较基准。不同的 DDoS 检测方法有着各自的优缺点和适用场景。基于阈值的方法的优点在于简单易实现，适用于流量模式较为稳定的环境[4]，缺点是容易受到动态网络环境的影响，误报率较高。基于行为分析的方法可以捕捉到长期的行为模式，因而适用于检测慢速攻击，然而这种方法需要大量的历史数据来建立行为基线，对资源消耗较大。基于机器学习的方法擅长处理复杂的数据分布，具有极强的适应性，但是训练过程可能较为复杂且需要精心设计特征工程。本文所提方法结合了多种特征提取技术和机器学习算法，既具备高准确性又能快速响应，适用于复杂多变的网络环境。

比较结果显示，基于 IP 网络流量特征分析的方法在诸多方面表现出色。准确性方面，相比基于阈值的方法，本文所提方法能够更准确地区分正常流量与攻击流量，特别是在复杂网络环境中。相较于基于行为分析的方法，我们的方法能够更快地适应新的攻击模式，这是因为我们不仅考虑了常见的流量特征，还引入了基于时间序列分析的新特征。响应速度同样排名第一，本文的方法在保持较高准确率的同时实现了更快的响应速度，这对于实时检测至关重要。

五、结束语

本研究成功地探索了一种基于 IP 网络流量特征分析的 DDoS 攻击检测方法，实验结果表明所提方法能够准确识别 DDoS 攻击，显著提升了检测的准确性和响应速度。未来，我们计划进一步优化特征提取过程，探索更多基于深度学习的技术来增强模型的鲁棒性。在此基础上我们还将致力于将此方法应用于更广泛的网络环境以应对日益复杂的网络安全挑战。

[参考文献]

[1]李全乐. 网络流量特征分析方法的研究与实现[D]. 北京邮电大学，2022. DOI：10.26969/d.cnki.gbydu.2022.003536.

[2]骆志成. 基于流量特征分析的网络攻击溯源技术研究[D]. 南京理工大学，2021. DOI：10.27241/d.cnki.gnjgu.2021.001643.

[3]朱凌. 基于深度学习的通信网络 DDoS 攻击在线检测方法[J]. 网络安全技术与应用，2024，(07)：17-19.

[4]孙佳奇，谭小波，郭浩然，等. 基于动态阈值的可变速率 DDoS 攻击检测方法[J]. 沈阳工程学院学报（自然科学版），2024，20（01）：48-54+61. DOI：10.13888/j.cnki.jsie（ns）.2024.01.008.