

基于入侵检测的 SDN 安全运维平台

王明芬¹
WANG Mingfen

摘要

随着软件定义网络（SDN）网络应用的推广，SDN 网络技术也面临许多安全问题，SDN 控制器如果被恶意入侵则会对整个网络造成严重影响。设计一个综合性的可视化安全运维系统，主要包括入侵检测、网络负载均衡、流量监控、安全日志管理等模块。为了更好地保障安全，将整个平台的基础设施、数据库等部分迁移到云端，使得云服务器具有更高的灵活度，从而有效地节省资源，同时，使得用户也可以随时随地访问该平台，从而更好地掌握系统的运营状态。系统融合了人工智能技术和 SDN 网络可编程技术实现了智能化安全运维可视化管理。

关键词

软件定义网络；入侵检测；随机森林；流量监控；可视化运维

doi: 10.3969/j.issn.1672-9528.2024.01.038

0 引言

随着云计算、大数据等技术的发展，对网络需求有了爆炸性的增长。SDN 技术可以通过其集成的逻辑管理机制，使得网络的运行可以被高效地管理和监督，同时也可以通过其统一的南北方向的接口^[1]，使得网络的运行可以被完全监督和管理。从而使得网络可以从传统的静态模式转换到服务器管理的动态模式，从而满足当今云计算、大数据和各种创新应用的需求。但是无论是在南向接口或者是北向应用层接口，SDN 在遭遇到攻击时无法做到有效的防护^[2]。作为下一代网络技术，SDN 尽管有很多优势同时也更面临许多问题。SDN 控制器可以有效地控制整个网络，但一旦发生故障，将会对网络结构产生极大的破坏，因此，它被认为是恶意攻击者最理想的攻击目标^[3]。

本文设计基于云平台的 SDN 入侵检测安全运维系统，集成了入侵检测模型服务器、SDN 控制器、云数据库。其中控制器主要是对 SDN 网络中的网络数据包等数据进行采集，并通过云服务器进行数据存储与备份。入侵检测模型是检测当前网络传输的数据包是否存在异常以及对当前模型进行评估。云数据库首先对 SDN 控制器获取的流量数据、模型检测的结果、Web 界面数据等进行存储和备份，然后将项目部署到服务器上。可视化的网络入侵检测系统为网络管理人员提供更便捷的安全运维平台。

1 系统设计

在 SDN 的网络环境中，通过控制器监控全局拓扑，获取网络中的数据流量。通过北向应用层接口实现流量数据的云端传输^[4]。首先是系统后台登录的安全设计，每个用户登录时浏览器会给一个凭证，当用户退出或者长时间未进行操作时凭证销毁，这样保证了后台系统的安全和可靠。其次是数据的监控和管理，南向接口采用协程的方式获取数据，北向接口以开放接口规范（REST API）获取数据^[5]。然后根据系统需要灵活调用 SDN 南向接口和北向接口，实现数据清洗和端口负载均衡功能。最后是异常数据的检测，将文件转化成二进制文件进行存取，入侵检测模型对异常数据进行响应和预警，保存结果并展示到前端 Web 界面。总体功能设计可大致分为四大部分包括入侵检测、流量监控、负载均衡、安全管理，如图 1 所示。

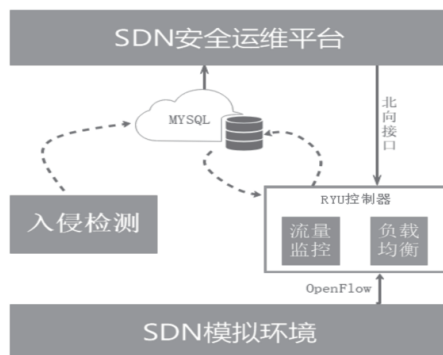


图 1 功能模块简介

1. 福建师范大学协和学院 福建福州 350117

[基金项目] 福建省中青年教育科研项目（JAT200965）；福建师范大学协和学院 2022 年度课程思政示范课程拟培育项目（KCSZ2211）

1.1 流量监控模块

SDN 控制器通过链路层发现协议（LLDP）获取到当前

网络中的链路信息，并根据收集的链路信息来识别网络拓扑结构，对网络设备进行流量监控^[6]。在控制平面，主动监听流表状态的响应报文，当有流量通过时触发监听，然后解析响应报文并采集流量的详细信息存入云数据库。在可视化运维平台可以直接监测流表，特殊情况下支持在平台直接手动配置流表，以维持网络环境的稳定。

在流量监控过程中，平台需要对网络设备进行持续性监控，为此系统开启一个线程持续对交换机发送请求信息。本文主要通过协程对 OpenFlow 南向接口的实时监控与 REST API 北向接口的应用调用实现对交换机状态监控与数据请求，包括请求端口信息和流表信息。

1.2 负载均衡模块

负载均衡的实现通过控制器定义一 OpenFlow 组表（Group Table），组表为 OpenFlow 交换机提供了更加高级的数据包转发特性。当组表类型为 SELECT 时，交换机从该组表所包含的多条路径中选择一条进行转发，再综合使用 OpenvSwitch 的队列保障链路带宽，完成一个基本的二层负载均衡的功能^[7]。

当组表类型为 SELECT 时，系统采用基于用户配置的哈希表选择动作端口。当一个端口被关闭时，交换机会自动将数据包转发到一个预先设定的端口集合，并通过哈希算法进行选择端口，而不是持续将数据包发送给这个关闭的端口，从而有效地减少单点故障对整体网络性能带来的影响。

1.3 入侵检测模型

在并行型集成学习方式中，bagging 方式是其中的一项典型形式，它通过对原始数据集进行随机采样和有放回抽样，生成多个子数据集，然后分别对每个子数据集训练出一个基分类器，再将这些基分类器进行结合，形成一个更加强大的分类器^[8]。

数据集中有的样本在该抽样集中反复存在，但是还有一些样本在数据集中不存在，将这些未出现的样本作为当前模型的验证集来测试模型的泛化程度，作为模型泛化能力的评估的标准，这被称作“包外估计”，令 D_t 表示第 t 个个体学习器对应的采样集， $H^{oob}(x)$ 代表该集成学习器对该样本 x 的包外预测值，公式为：

$$H^{oob}(x) = \arg \max_{y \in \mathcal{Y}} \sum_{t=1}^T \mathbb{I}(h_t(x) = y) \cdot \mathbb{I}(x \notin D_t) \quad (1)$$

根据公式（1），可以估计 bagging 泛化误差的包外影响，公式为：

$$\epsilon^{oob} = \frac{1}{|D|} \sum_{(x,y) \in D} \mathbb{I}(H^{oob}(x) \neq y) \quad (2)$$

随机树林是 bagging 的变种，在 bagging 的基础上引进

了对随机特性的考虑，增加了个体之间的差异性^[9]。随机森林在分类算法上通过集成学习的优化，以重采样技术为核心，将当前样本重新放入原始数据集进行二次抽取，经过多次有放回的抽样得到新的样本训练集对决策树进行训练，重复上面的动作生成多棵决策树。新的数据集由决策树投票进行选择，每棵树的建立依赖于独立抽取的样本。每一棵树的分类能力有限，但是经过随机产生的大量决策树后，最终的随机森林便具有所有测试样本的特性^[10]，流程如图 2 所示。

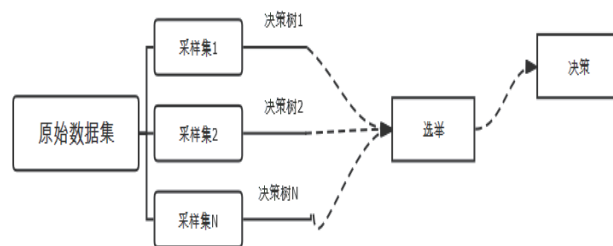


图 2 随机森林模型流程图

1.4 认安全策略定义及管理

当网站受到大量攻击时，针对入侵检测的分类结果，支持用户或管理员选择合适的默认规则启用保护策略，以保障系统的平稳运行。系统设计中存在 4 种不同类型的攻击，包括拒绝服务（DoS）、探测、用户到根（U2R）和远程到本地（R2L）。

初步设计默认规则如下。（1）Dos 攻击，则开启负载均衡来分流异常的大流量，设置合适的最大连接数，超过则断开所有连接并发送警告。（2）Probing 攻击，则减少无用开放端口，实时监测扫描，及时异常警告。（3）R2L 攻击，关闭不必要的远程连接端口，对必须的远程端口做修改避免不必要的连接，定期检测端口状态及时异常警告。（4）U2R 攻击，则定时检查用户表，避免隐藏账户、克隆账户，发现新增管理用户，弹出确认警告。

在得到入侵检测结果后，将结果存入数据库并在平台上显示。用户通过查看入侵检测结果，可以选择启用固定的默认规则、自定义规则等方式启用防御策略。

1.5 运维可视化

SDN 安全运维平台整体环境、服务平台、数据库都搭建在腾讯云上。部署成本较低，云服务部署不需要专门的运维人员，节省了时间成本。云服务器具有极高的安全性和可靠性，能够实现快速、灵活的计算资源配置，大大降低了软硬件采购的成本。可视化的前后端交互模块是通过 Flask 框架进行部署，它是用 Python 编写的微框架。使用 Flask 框架，可以创建一个基于机器学习的 Web 应用，它可以将前端网页

与后端模型数据相结合,从而实现入侵检测模型与页面之间的有效交互,如图3所示。

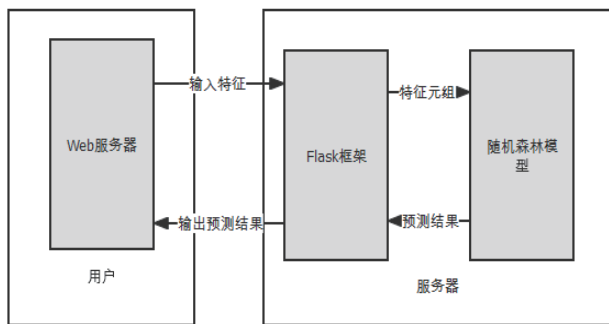


图3 可视化部署

可视化技术设计是以springboot框架为核心的多种框架整合。首先使用spring resttemplate依赖实现后端访问远程服务器,然后通过SqlSession实现对云数据库的读写,并通过JSON(JavaScript object notation)的轻量级的数据交换格式实现前后端数据的交互,最后通过Layui框架实现界面的设计。具体设计流程见图4。

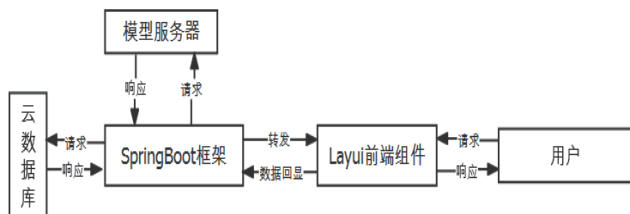


图4 可视化技术设计流程

2 测试与分析

2.1 数据集

NSL-KDD相较于KDD99数据集,它在冗余数据上做了新的优化,可以作为有效的数据集来检测模型的能力。当前网络的NSL-KDD数据集包括4个子数据集^[11]:KDDTrain+、KDDTrain+_20Precent、KDDTest+、KDDTest+21。数据集中存在4种不同类型的攻击,包括拒绝服务(DoS)、探测、用户到根(U2R)和远程到本地(R2L),如表1所示。

表1 数据集异常类型

异常类型	描述
PROBING	这是一个找目标过程,进行监视扫描
R2L,REMOTE-TO-LOCAL	这是一个远程->本地,是一个入侵的过程
U2R,USER-TO-ROOT	本地提权
DOS	资源消耗

2.2 模型性能评估与分析

精确率(precision)是针对预测结果而言的,其含义是在被所有预测为正的样本中实际为正样本的概率,召回率(recall)是针对原样本而言的,其含义是在实际为正的样本中被预测为正样本的概率。本文希望精确率和召回率都很高,但实际上这两个指标是矛盾的,因此,不可能同时达到precision与recall的最佳状态。为了获取最佳的结果,在设定阈值时,应该考虑各种不同的目标。例如,为了获取更高的精确率,应该降低一些召回率;为了获取更佳的召回率,应该降低一些精确率。而在这两者的平衡点 F_1 值同时考虑精确率和召回率,则是最平衡的方案^[12]。通过调整 F_1 参数,既可以提升准确性,又可以降低召回风险,从而实现双赢。

经过对比分析,可以发现,随机森林在测试集上的准确率明显高于决策树,其召回率也更高, F_1 值也更大。随机森林在测试集上的性能指标有较明显的优势,如图5所示。

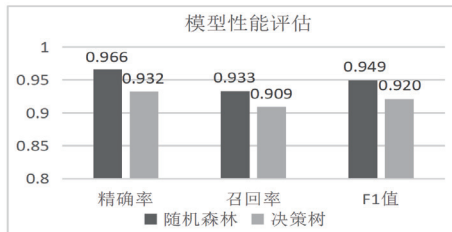


图5 模型性能评估

2.3 可视化入侵检测平台

入侵检测模块需要对系统设备进行周期性更新,本文的设备信息主要是对系统服务器信息做轮询的,定期获取信息并刷新页面信息。记录被攻击详情、启用规则、系统事件等重要事件的记录。入侵检测结果如图6所示,用于记录一周内受到异常攻击的总量、类型及其占比。



图6 入侵检测结果

流表监控是维持网络稳定的重要功能,包括监控错包、发送包的数量等一系列的操作。在网络发生异常的时候,流量监控可以快速定位网络的病因。由于在流量监控过程中,平台需要对网络设备进行持续性监控,为此系统开启一个线程持续对交换机发送请求信息,在向交换机发送请求信息后,交换机会回送报文即向控制器发送端口和流表的信息。此时系统对数据进行解析并存储到云数据库中用于平台显示,如图7所示。



图 7 流表监控

平台的日志管理多维度显示系统安全性能，管理者能快速掌握图形化实时信息，有助于其对业务运行情况进行判断和风险预估。可视化平台作为系统运行状况评估标准，其中主要包括网络性能指标检测，如网络带宽利用率、抖动率、丢包率和时延等关键指标的统计，方便用户及时了解网络运行状况，如图 8 所示。



图 8 网络性能可视化

3 结语

在 SDN 网络可以实现接口可编程的基础上，通过流量监控模块进行获取数据并记录，对网络数据进行异常检测。搭建可视化运维平台利用北向接口与控制器通信，实现运维功能与 SDN 控制器的分离，让运维更加直观明了，摆脱命令行界面用户体验感差的缺点。

SDN 安全运维平台具备入侵检测、流量监控、安全管理、日志管理等功能，兼顾网络的内部稳定与外部安全问题，是一个全方面多功能的 SDN 管控平台。通过将机器学习算法应用于入侵检测，本文的系统可以有效地检测出大量、复杂的网络流量，并发现新的、未知的攻击行为。SDN 安全运维平台整体环境搭建在云上，云服务器提供安全可靠的弹性计算服务，可以通过公网访问本系统，方便用户随时查看系统运行情况，实现资源的共享。

参考文献：

- [1] 张朝昆, 崔勇, 唐蜀祯, 等. 软件定义网络 (SDN) 研究进展 [J]. 软件学报, 2015, 26(1): 62-81.
- [2] 王赞, 于尧, 赵雨佳, 等. 家庭物联网中基于 SDN 的入侵

检测防御机制 [J]. 控制工程, 2021, 28(5): 1027-1032.

- [3] 张立群, 林海涛, 邹文明, 等. 基于 Open-Flow 的软件定义网络流规则冲突检测系统 [J]. 计算机应用, 2022, 42(2): 6-13.

- [4] 赵云, 李莉, 沈苏彬, 等. 支持可定制 QoS 服务的 SDN 北向接口设计与实现 [J]. 计算机技术与发展, 2016, 26(11): 6-11.

- [5] 王洋, 汤光明, 王硕, 等. 基于 API 调用管理的 SDN 应用层 DDoS 攻击防御机制 [J]. 网络与信息安全学报, 2022, 8(2): 73-87.

- [6] 董帅, 张安琳, 黄道颖, 等. 基于 Open-Flow 的 SDN 中链路层拓扑发现的优化 [J]. 火力与指挥控制, 2020(8): 45-48.

- [7] 张一凡, 韩卫占. 基于 SDN 网络大象流负载均衡算法研究 [J]. 计算机测量与控制, 2023, 31(1): 257-263.
- [8] 付炜, 杨洋. 基于卷积神经网络和随机森林的音频分类方法 [J]. 计算机应用, 2018, 38(A2): 5-9.
- [9] 江轲, 张宏进. 一种改进随机森林算法及其在入侵检测中的应用 [J]. 电子设计工程, 2021, 29(22): 85-88+92.
- [10] 周杨, 王春林, 郭锐. 基于随机森林算法的数据中心运维异常告警方法 [J]. 现代电子技术, 2023, 46(8): 143-148.
- [11] 张小云, 康晓霞. 基于决策树算法的网络入侵检测系统设计与评估 [J]. 信息技术, 2023, 47(2): 117-122.
- [12] 王博, 华庆一, 舒新峰. 基于云平台日志的故障检测和复杂构件系统即时可靠性度量研究 [J]. 计算机科学, 2022, 49(12): 11-18.

【作者简介】

王明芬 (1981—), 女, 副教授, 研究方向: 网络通信、机器学习等。

(收稿日期: 2023-09-23)