基于联邦全局知识蒸馏的异常网络入侵检测方法

石迎澳[1],李润知[1,2],姬　怡[1]
(1. 郑州大学 网络空间安全学院,河南 郑州 450001;
2. 郑州大学 网络管理中心,河南 郑州 450001)

摘　要:在网络入侵检测领域,联邦学习(FL)作为一种保护数据隐私的分布式处理方法受到了广泛关注。然而,由于参与节点间的数据异构性,传统的联邦学习方法在联合训练过程中往往难以实现高性能。为了解决这一问题,该文提出了一种改进的联邦学习方法,即基于联邦全局知识蒸馏的异常网络入侵检测方法(FLGKD-ANIDS)。该方法在中央服务器中设置了一个缓冲区,用于缓存客户端上传的多轮模型参数。进一步地,这些缓存的参数被用于生成包含多轮全局知识的教师模型参数,指导客户端侧的知识蒸馏过程。这一机制使得客户端能够在注入全局知识特征的情况下训练本地数据。实验在两个公开数据集 UNSW-NB15 和 CIC-IDS2017 上进行,结果显示 FLGKD-ANIDS 在各种数据异构场景下显著提升了模型性能,其性能更接近于集中训练模型的水平。

关键词:入侵检测系统;联邦学习;数据隐私;知识蒸馏;数据异构
中图分类号:TP399　　**文献标识码**:A　　**文章编号**:1673-629X(2025)07-0055-08
doi:10.20165/j.cnki.ISSN1673-629X.2025.0087

Anomaly Network Intrusion Detection Method Based on Federated Learning with Global Knowledge Distillation

SHI Ying-ao[1],LI Run-zhi[1,2],JI Yi[1]
(1. School of Cyber Science and Engineering,Zhengzhou University,Zhengzhou 450001,China;
2. Network Management Center,Zhengzhou University,Zhengzhou 450001,China)

Abstract:In the field of network intrusion detection,Federated Learning (FL) has gained significant attention as a distributed processing method that protects data privacy. However,due to data heterogeneity among participating nodes,traditional FL methods often fail to achieve high performance during joint training. To address this issue,we propose an improved federated learning with global knowledge distillation (FLGKD-ANIDS). It sets up a buffer in the central server to cache multiple rounds of model parameters uploaded by clients. Furthermore,these cached parameters are used to generate teacher model parameters containing multi-round global knowledge,which guide the knowledge distillation process in the client side. This mechanism allows clients to train the local data by injecting global knowledge feature. We conducted experiments on two public available datasets,UNSW-NB15 and CIC-IDS2017. The results show that FLGKD-ANIDS significantly improves model performance across various data heterogeneity scenarios compared to existing federated learning methods,and its performance approaches that of centrally trained models.

Key words:intrusion detection system;federated learning;data privacy;knowledge distillation;data heterogeneity

0　引　言

入侵检测系统(Intrusion Detection System,IDS)作为专门用于保护网络安全的工具,用于监视和分析网络流量数据包及系统日志,以检测潜在的入侵行为[1]。在分布式入侵检测系统中训练机器学习和深度学习模型时面临诸多挑战,尤其是本地数据量不足导致的该节点设备训练出的模型性能较差,以及将数据集中到中心服务器进行训练时可能出现的高通信成本和隐私泄露风险[2-3]。为了解决这些问题,引入联邦学习(Federated Learning,FL)这一分布式机器学习框架,允许多个客户端在不分享原始数据的情况下共同训练全局模型[4]。

然而,客户端之间的数据异构性将导致训练出的全局模型性能降低。具体来说,数据异构性指的是由

收稿日期:2024-11-22　　**修回日期**:2025-03-26
基金项目:河南省高等学校重点科研项目(25B520009)
作者简介:石迎澳(1999-),男,硕士研究生,研究方向为入侵检测、联邦学习;通信作者:李润知(1978-),女,副教授,博士,研究方向为网络安全行为分析、人工智能与医疗数据分析。

于不同客户端遭受的网络攻击类别和攻击数量会存在不同程度的差异,所以客户端之间的数据分布通常是非独立同分布的(non-Independent and Identically Distributed,non-IID)[5-6]。例如,某些客户端可能主要遭受 DDoS 攻击,而其他客户端则更多面对恶意软件或钓鱼攻击。这种 non-IID 特性不仅影响了全局模型的一致性和泛化能力,还可能导致局部模型优化方向的不同,进而影响全局模型的收敛性和性能。这种 non-IID 程度在不同分布式网络环境中是不同的,因此需要科学分析不同异构程度的联邦入侵检测场景对模型训练的影响。由于每个节点上的 IDS 数据并不是从所有客户端的全部数据的全局联合分布中采样的,因此每个客户端模型检测的目标是不同的,所以联邦学习框架下的入侵检测模型也是很难达到集中式训练得到的模型的效果[7-8]。即使每轮通信都从相同的全局模型开始,每个节点的本地模型也会向其局部目标的最差的模型上漂移,因此聚合的全局模型可能不是全局目标的最优[9-10]。

该文通过将知识蒸馏技术引入联邦学习入侵检测中,提出了一种基于联邦学习全局知识蒸馏的异常网络入侵检测方法(FLGKD-ANIDS)。通过在中央服务器上构建一个历史全局模型,并将其作为教师模型传递给客户端,通过在中央服务器缓存多轮历史模型参数构建全局模型参数传递给客户端,通过教师模型指导客户端局部模型学习,提升了局部模型在 non-IID 数据下的性能,以此增强模型在多种异构场景下联邦入侵检测模型的性能。

该文的主要贡献如下:

(1)将知识蒸馏方法集成到联邦学习架构中,以提升客户端局部模型的性能。这是通过在服务器端构建历史全局参数模型来实现的。

(2)在客户端的本地知识蒸馏过程中,通过最小化交叉熵损失和蒸馏损失,从硬标签和软标签中学习,从而增强模型的泛化能力和缓解数据异构性导致的性能下降。

(3)实验在 UNSW-NB15 和 CIC-IDS2017 数据集上进行,并且通过在多种异构场景下对不同的联邦学习方法进行比较。

1 相关工作

网络入侵检测系统(Network Intrusion Detection System,NIDS)分为基于签名的和基于异常的两大类,前者适用于快速检测已知攻击但难以应对新型威胁,而后者则通过识别与正常行为的偏差来发现未知攻击,特别是在结合深度学习技术后能更高效地分析大数据并自动提取特征[11]。史嘉宁[12]基于融合模型在网络异常入侵检测方法的应用,在保持分类器高检测率的同时显著减少资源消耗;Imrana Yakubu 等人[13]设计了三种高效可靠的网络入侵检测方法,为研究基于深度学习的 NIDS 开辟了新的方向;王锁成等人[14]进一步改进了基于残差网络的异常流量检测模型,提高了小类别攻击的识别率和检测精度;常利伟等人[15]以卷积神经网络结合多源融合技术,提出了一种层次化网络分析方法,显著提高了攻击识别率,并设计出一种准确有效的网络安全态势感知模型;罗虹富等人[16]提出了一种基于卷积神经网络和双向长短期记忆网络的分层注意力网络入侵检测方法,不仅有效解决了特征提取不足的问题,还通过结合时间序列显著入侵检测模型的整体性能。尽管上述研究在不同方面提升了入侵检测的效果,但它们均属于集中式训练方法,但是并不涉及分布式架构下数据隐私保护和减少通信开销问题的研究。联邦学习作为一种分布式训练架构被引入到 NIDS 中,可以有效解决隐私保护和通信效率的问题[17-18]。

联邦学习是由谷歌提出的一种保护隐私的分布式学习技术,备受关注。作为一种通信高效的框架,它允许客户端使用本地数据训练入侵检测模型,并仅上传模型参数[19]。服务器聚合这些参数以形成全局模型,然后将该模型下发给客户端进行进一步的本地训练迭代,直至通信轮次结束或达到预期的准确率。在整个过程中,服务器与参与训练的各个客户端之间共享的是模型参数而非原始数据,从而降低了通信开销并保护了数据隐私[20]。此外,联邦学习通过共享拥有更大更多元化数据集的客户端知识,有助于缓解数据稀缺性问题,提高最终聚合模型的泛化能力和对未知网络攻击的预测能力。然而,在联邦学习中,如何处理节点间数据异构性导致的模型收敛速度慢和模型精度差的问题是一个关键挑战[18]。

早期联邦学习研究多集中在独立同分布(Independent and Identically Distributed,IID)的数据上,使用传统的 FedAvg 聚合算法在服务器端对客户端模型进行平均以生成全局模型[21]。近年来,越来越多的研究关注于解决不平衡、non-IID 数据的挑战,因为这些数据更贴近现实世界的场景。non-IID 数据带来的主要问题是客户端局部模型参数存在巨大差异导致局部模型被优化到不同的方向[22]。为了解决 non-IID 挑战,研究者提出了多种解决方案,主要包括改进聚合策略和正则化本地训练[23-24]。改进模型聚合策略是指通过聚合其他统计指标给予更有代表性的客户端更多的权重,从而提高模型的整体准确性。然而这些方法并没有完全解决客户端漂移问题,因为它们只是从全局模型层面进行改善。正则化本地模型训练指

通过对本地模型的训练过程施加某种约束或惩罚项以增强其模型的泛化能力,但是在某些联邦场景中无法有效工作。例如,Li 等人[25]提出了 FedProx 方法,在 FedAvg 的基础上增加了一个近端项,使局部模型的发散程度受到全局模型的约束;Karimireddy 等人[26]通过引入控制变量(control variate)来校正客户端漂移问题。

为了缓和数据异构性对联邦入侵检测模型性能的影响,知识蒸馏(Knowledge Distillation,KD)与联邦学习的结合近年来吸引了越来越多的关注。知识蒸馏最早由 Hinton 等人[27]提出,它将一个大型、复杂的神经网络(教师模型)的知识转移到一个小型的、简单的神经网络(学生模型)中,关键步骤是通过将学生网络的软预测与教师网络的软预测对齐。例如,一些工作利用一个代理数据集来进行教师网络和学生网络之间的知识蒸馏。Lin 等人[28]使用代理数据上局部模型的平均 logits 进行聚合,以提高节点模型的性能。Zhu 等人[29]通过共享本地标签数量并通过对生成的潜在特征训练来正则化本地模型。Le 等人[30]提出了 CDKT 方法,利用代理数据集,通过知识蒸馏机制进行跨设备知识转移来提升整体模型的性能。上述研究将知识蒸馏技术应用于联邦学习的网络入侵检测领域,通常需要一个代理数据集作为中介,但由于网络攻击的多样性和某些特殊部门的存在,代理数据集并不总是可用[31-32]。此外,大多数现有的基于联邦学习的网络入侵检测研究仅在客户端间数据同构场景或某种数据异构场景下进行,未能充分证明所提方法在不同数据异构场景中的适用性。

针对以上问题,该文提出的 FLGKD-ANIDS 方法无需代理数据集,直接利用中央服务器上的历史全局模型作为教师模型,将知识传递给客户端进行本地知识蒸馏。通过最小化交叉熵损失和蒸馏损失策略,提升模型在 non-IID 数据分布下的泛化能力和稳定性。此外,FLGKD-ANIDS 能够在每轮通信中动态更新教师模型,并将其最新的知识传递给所有客户端,使得模型能够更快适应新的攻击模式和变化的数据分布。在 UNSW-NB15 和 CIC-IDS2017 两个公开数据集上进行多组实验,包括多种异构场景下联邦学习方法的对比实验、数据异构性对模型性能影响的实验、不同训练轮次下的模型性能对比实验以及最优缓冲区大小设置的实验,验证 FLGKD-ANIDS 在多种异构场景下的优越性。综上所述,FLGKD-ANIDS 通过引入的无需代理数据集的知识蒸馏机制,提升了局部模型在 non-IID 数据分布下的性能,以此增强在多种异构场景下的联邦入侵检测模型性能。

2　方　法

2.1　FLGKD-ANIDS 框架

图 1 展示了 FLGKD-ANIDS 的模型训练流程,主要由训练集、测试集、数据处理、客户端的本地数据和学生模型、服务器端的全局模型和教师模型,以及最终训练得到的入侵检测模型组成。

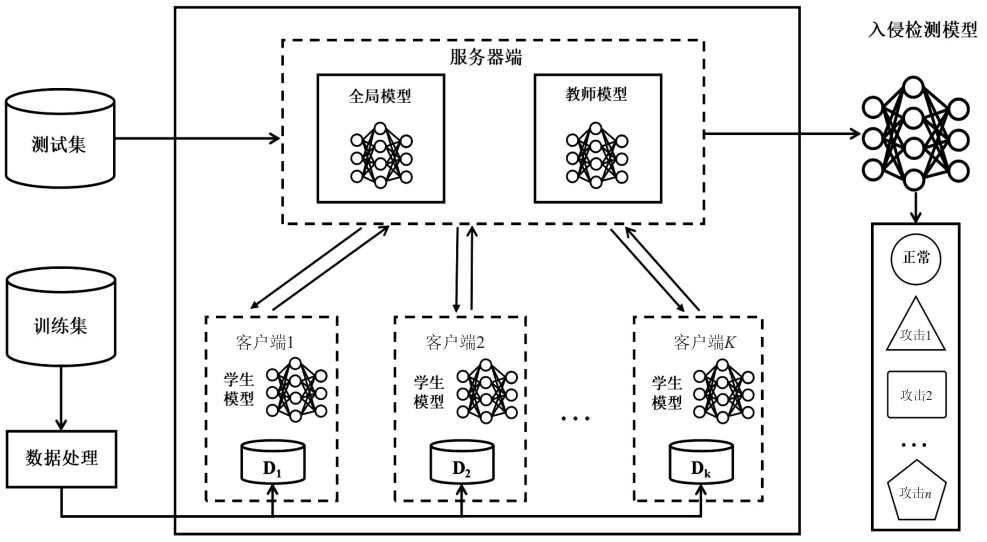

图 1　模型训练流程

具体的训练流程如下:首先,作为训练集的客户端节点的原始入侵检测数据进行数值化、标准化和归一化的数据预处理;然后,被选中的客户端从服务器下载全局模型,进行本地模型训练,并将更新后的模型上传回服务器;随后,服务器聚合这些模型以生成新的全局模型,并将其存储在缓冲区中,累积多轮的历史全局知识生成教师模型,该教师模型再发送回客户端,用于本地知识蒸馏以更新本地模型;当模型收敛或完成预定的通信轮次后,最终的入侵检测模型将在测试集上进行测试,以获得最终的分类结果。

随后,在图 2 中介绍了中心服务器与客户端的详细交互过程,主要目标是通过使用聚合多个通信轮次的全局模型以生成教师模型,将包含多个轮次的历史全局模型的知识从服务器端传递到本地客户端节点的学生模型中。目的是通过服务器中设置缓冲区得到历史全局模型改善模型聚合策略,然后通过知识蒸馏的机制来修正本地模型的训练,从而有效地缓解客户端间数据异构导致联邦入侵检测模型性能不佳的问题。

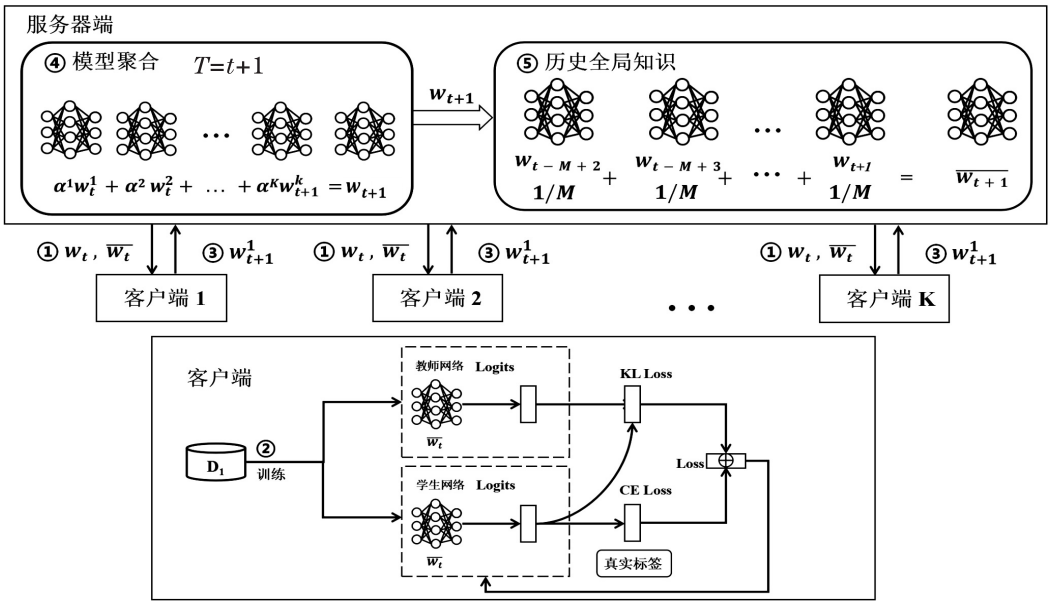

图 2 FLGKD-ANIDS 客户端与服务器的交互过程

2.2 模型聚合和历史全局知识生成

在服务器端执行算法 1,将客户端上传的模型参数进行聚合并生成历史全局知识的教师模型参数,得到每个通信轮次的全局模型。

算法 1 服务器端:模型聚合和历史全局知识生成

输入:客户端总数 K,总通信轮次 T,客户端参与比 C,缓冲区大小 M

输出:完成训练后的 ANIDS 模型

1. 初始化全局 ANIDS 模型 w_1;
2. for 通信轮次 $t = 1,2,\cdots,T$ do
3. $K^t \leftarrow$随机抽取 $C \cdot K$ 个客户端
4. for 每个被选中的客户端 $k \in K^t$ 同时 do
5. 向客户端 k 发送全局模型 w_t 和历史全局模型 $\overline{w_t}$;
6. 从客户端接收 w_t^k;
7. end for
8. 通过公式 1 得到全局模型 w_{t+1};
9. 将 w_{t+1} 存储在缓冲区并通过公式 2 获取教师模型 $\overline{w_{t+1}}$;
10. end for

具体如下:服务器得到参与方上传的本地模型后通过公式 1 聚合得到第 $t+1$ 轮的全局模型 w_{t+1},其中 $\alpha_k = \dfrac{n_k}{\sum_{k=1}^{K} n_k}$,$n_k = |D_k|$,即客户端 k 用于训练的样本数量。

$$w_{t+1} = \sum_{k=1}^{K} \alpha_k w_{t+1}^k \tag{1}$$

通过在服务器设置一个缓冲区,用于存放 M 个轮次的最新全局模型。在第 $t+1$ 轮中,服务器从 K 个客户端上传的本地模型生成全局模型 w_{t+1},然后将其保存在缓冲区中用以生成历史全局知识的教师模型 $\overline{w_{t+1}}$,然后发送到客户端通过本地模型更新以实现知识的传递。具体是通过公式 2 获得教师模型 $\overline{w_{t+1}}$。

$$\overline{w_{t+1}} = \frac{1}{M} \sum_{t=t-M+2}^{t+1} w_t \tag{2}$$

2.3 本地知识蒸馏

在被选中的客户端上执行算法 2,这是将多轮历史全局模型的知识传递到本地模型的关键步骤,即通过增加知识蒸馏的损失值以修正本地模型的训练。具体如下:假设客户端 k 接收到第 t 轮服务器下发的全局模型 w_t 作为本地学生模型,以及历史全局知识 $\overline{w_t}$ 作为教师模型,客户端使用本地私有的入侵检测数据集 D_k,通过知识蒸馏得到一个总体的损失函数,从而更新本地模型得到下一轮的本地模型 w_{t+1}^k。

算法 2 客户端:本地知识蒸馏更新模型

输入:本地迭代轮次数 E,私有 IDS 数据集 D_k,客户端样本量 n_k,初始化本地模型 w_1^k,学习率 η,批大小 B,全局模型 w_t,历史全局模型 $\overline{w_t}$

输出:更新后的本地模型 w_{t+1}^k

1. 从服务器端接收 w_t 和 $\overline{w_t}$;
2. 将 w_t 设置为本地学生模型;
3. 通过公式 7 从 $\overline{w_t}$ 中蒸馏知识得到 w_{t+1}^k;
4. for 本地迭代 $e = 1,2,\cdots,E$ do

5. for 每个 Batchsize　$b \in B$　do

6. $w \leftarrow w - \eta \cdot L(w,b)$,更新本地模型参数;

7.　　end for

8. end for

9. return w_{t+1}^k ,发送到服务器端

在知识蒸馏过程中,本地模型训练过程的损失函数中包含多个损失项,具体如下:假设 x 是模型的输入,目标 y 的维度是 d。$p(w,x)$ 是由模型 w 输出的 logits 向量 $z(w,x)$ 通过式 3 得到的软标签,即每个类别输出的概率。正如 Hinton 等人[27]所提出的,使用温度标度 τ 来软化预测概率以得到更好的蒸馏效果。

$$p(w,x) = \frac{\exp(z(w,x)/\tau)}{\sum_d \exp(z(w,x)/\tau)} \tag{3}$$

之后,$p_s(w_t, x)$ 是通过学生模型 w_t 得到的学生软标签,$p_t(\overline{w_t}, x)$ 是通过教师模型 $\overline{w_t}$ 得到的教师软标签。

其中第一个 Loss 项是通过学生模型 w_t 输出的软标签 p_s 和真实标签 y 通过交叉熵损失函数公式 4 得到的 Loss_s。

$$\text{Loss}_s = \text{CE}(p_s(w_t,x), y_{\text{TrueLabel}}) \tag{4}$$

另外一个 Loss 项是通过教师模型 $\overline{w_t}$ 输出的软标签 p_t 和学生模型 w_t 输出的软标签 p_t 通过 KL(Kullback-Leibler)散度公式 5 计算得到的 Loss_t,用于将历史全局知识传递到本地模型。

$$\text{Loss}_t = \text{KL}(p_s(w_t, x) \parallel p_t(\overline{w_t}, x)) \tag{5}$$

最后是本地模型的总损失函数 Loss,表示为:

$$\text{Loss} = \text{Loss}_s + \alpha \text{Loss}_t \tag{6}$$

其中,α 是 Loss_t 的蒸馏系数,和文献[22]相同,$\alpha = 0.005$。

客户端 k 使用公式 7 通过反向传播来更新本地模型,获得第 $t+1$ 轮的本地模型 w_{t+1}^k。

$$w_{t+1}^k = w_t^k - \eta \frac{\partial \text{Loss}}{\partial w_t^k} \tag{7}$$

3　实　验

3.1　数据集和实验环境设置

3.1.1　数据集

实验在两个公开的网络入侵检测数据集(UNSW-NB15 和 CIC-IDS2017)上进行。UNSW-NB15 数据集包含 9 种攻击类型以及正常类型、49 个特征和 2 540 044条数据。这 9 种攻击类型分别是 Fuzzers、Analysis、Backdoors、DoS、Exploits、Generic、Reconnaissance、Shellcode 和 Worms。CIC-IDS2017 数据集包含 14 种攻击类型和正常类型、78 个特征和 2 830 743 条数据。这 14 种攻击类型分别是 Dos Hulk、PortScan、DDoS、DoS GoldenEye、FTP-Patator、SSH-Patator、DoS slowloris、DoS slowhtttest、Bot、Web Attack-Brute Force、Web Attack-XSS、Infiltration、Web Attack-Sql Injeetion、Heartbleed。

3.1.2　数据的异构分布

为了模拟真实的分布式入侵检测环境,该文通过 Dirichlet 分布 Dir(α)对客户端数据进行划分,以创建不同的异构场景。α 表示集中参数,α 越小,数据的异构程度越高[33]。设置 $\alpha = 100$ 可以达到近似同构的数据分布。通过调整 α 值来控制数据的异构程度,并设置了五种场景:α 为 0.05、0.1、1、10 和 100。

3.1.3　实验参数设置

在 UNSW-NB15 数据集上选用 CNN_LSTM 模型,该模型包括两层卷积层,一层 LSTM 层,两层全连接层。在 CIC-IDS2017 数据集上选择了一维卷积神经网络,该模型包括三层卷积层,三层全连接层,并使用 ReLU 激活函数。所有客户端的模型以及教师模型都使用相同的参数进行初始化。联邦学习的参数设置:本地训练轮数 $E = 10$,通信轮数在 UNSW-NB15、CIC-IDS2017 上分别为 $T = 100$ 和 $T = 50$,客户端总数 $K = 20$,客户端存活率 $C = 0.4$。对于本地模型训练,Batchsize = 128,学习率在 UNSW-NB15、CIC-IDS2017 上分别为 0.000 1、0.001,优化器使用 Adam。

3.1.4　基线算法

为了全面评估 FLGKD-ANIDS 的有效性和优越性,该文将其与以下最新的入侵检测算法进行了详细比较。FedAvg-ANIDS[21]:这是一种经典的联邦学习算法,直接对参与节点上传的异常网络入侵检测模型进行平均。FedProx-ANIDS[25]:在 FedAvg 的基础上,FedProx 引入了一个正则化项来优化本地模型的更新。CDKT-ANIDS[30]:这是一种基于知识蒸馏和代理数据集的跨设备知识传输方法。Centralized-ANIDS[34]:集中式训练入侵检测模型的方法。

3.1.5　评价标准

由于数据异构性导致的模型性能存在较大的差异,为了更直观地进行实验对比,本实验将最具代表性的准确率 Acc 作为评价指标,记录联邦入侵检测场景中每一轮异常检测模型在测试集上的准确率,通过计算平均准确率 Acc_{Avg} 和最优准确率 Acc_{Best},分别代表不同联邦学习方法得到的异常检测模型的整体性能和最优性能。

$$S = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{8}$$

其中,S 表示准确率 Acc,T_p 和 F_p 分别为正确预测正分类的个数和错误预测正分类的个数,T_n 和 F_n 分别为正确预测负分类的个数和错误预测负分类的个数。

$$S_{\text{Avg}} = \frac{1}{T}(S_1 + S_2 + \cdots + S_T) \tag{9}$$

$$S_{\text{Best}} = \max\{S_1, S_2, \cdots, S_T\} \tag{10}$$

其中，S_{Avg} 和 S_{Best} 分别表示平均准确率 Acc_{Avg} 和最优准确率 Acc_{Best}，T 为通信轮次，S_T 表示第 T 轮通信得到的入侵检测模型的准确率。

3.2 实验结果和分析

3.2.1 整体性能比较

为了验证文中方法的有效性，在表 1 中展示了 FLGKD-ANIDS 分别与另外两个基线方法以及集中式网络入侵检测方法在两个数据集和不同异构环境中得到的异常网络入侵检测模型在测试集上的准确率。结果显示，FLGKD-ANIDS 在两个数据集和不同级别的数据异构场景中都表现出了最佳性能，能够更加接近集中式网络入侵检测模型的性能。具体来说，在 $\alpha = 0.05$ 的高数据异构场景下，尽管 FLGKD-ANIDS 在 UNSW-NB15 数据集上的最优模型性能与最好的基线模型相同，但整体表现比最好的基线模型高出 2.21 百分点；而在 CIC-IDS2017 数据集上，尽管最优模型与最好的基线模型相差不大，但 FLGKD-ANIDS 方法整体表现比最好的基线模型高出 26.51 百分点，显示出明显的优势。在 $\alpha = 0.1$、1.0、10、100 的不同数据异构程度下，FLGKD-ANIDS 在 UNSW-NB15 数据集上与最优的基线方法相比，整体表现分别高出 2.86 百分点、1.19 百分点、0.6 百分点和 1.02 百分点；在 CIC-IDS2017 数据集上，与最优的基线方法相比，整体上分别高出 4.94 百分点、0.36 百分点、0.56 百分点和 0.02 百分点。虽然 FLGKD-ANIDS 与最好的基线方法在最优模型上的表现相差不大，但 FLGKD-ANIDS 得到的最优模型性能始终最好。此外，从整体上看，FLGKD-ANIDS 相对于最好的基线方法有着明显的提升，尤其是在数据异构性最高的情况下，提升尤为明显。

表 1 FLGKD-ANIDS 与基线方法在不同异构场景下的测试准确率 %

数据集	Dir(α)	FedAvg-ANIDS		FedProx-ANIDS		CDKT-ANIDS		FLGKD-ANIDS		Centralized-ANIDS	
		Acc_{Avg}	Acc_{Best}	Acc_{Avg}	Acc_{Best}	Acc_{Avg}	Acc_{Best}	Acc_{Avg}	Acc_{Best}	Acc_{Avg}	Acc_{Best}
UNSW-NB15	α=0.05	62.74	78.21	58.34	80.19	59.45	78.21	64.95	80.19	–	–
	α=0.1	67.21	78.21	61.59	77.22	66.76	79.09	70.07	79.20	–	–
	α=1.0	77.48	85.14	74.33	81.18	77.57	83.16	78.67	85.14	–	–
	α=10	79.09	85.15	79.17	84.17	78.66	84.16	79.77	85.15	–	–
	α=100	79.67	85.15	79.33	84.20	79.60	85.15	80.69	85.15	81.30	85.15
CIC-IDS2017	α=0.05	42.48	86.09	38.86	81.79	46.26	86.08	68.99	86.09	–	–
	α=0.1	75.93	90.43	72.78	91.30	79.25	90.64	80.87	91.30	–	–
	α=1.0	91.60	96.39	92.44	96.52	92.16	96.32	92.80	96.52	–	–
	α=10	92.71	96.52	92.93	97.30	92.97	97.39	93.49	97.39	–	–
	α=100	92.99	97.39	94.01	97.39	93.94	97.39	94.03	97.39	96.40	99.13

3.2.2 数据异构性的影响

图 3 和图 4 分别展示了数据异构程度对两个数据集上 ANIDS 模型性能的影响。随着数据异构程度的降低，联邦框架下的 ANIDS 模型性能逐渐提高。此外，值得注意的是，文中方法在所有数据异构场景下获得的 ANIDS 模型在测试集上的准确率均为最高。

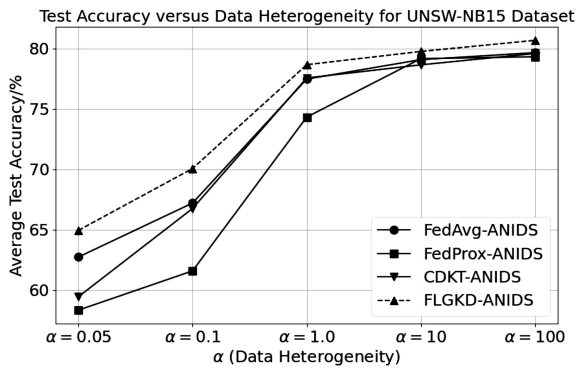

图 3 UNSW-NB15 数据集上数据异构性对模型性能的影响

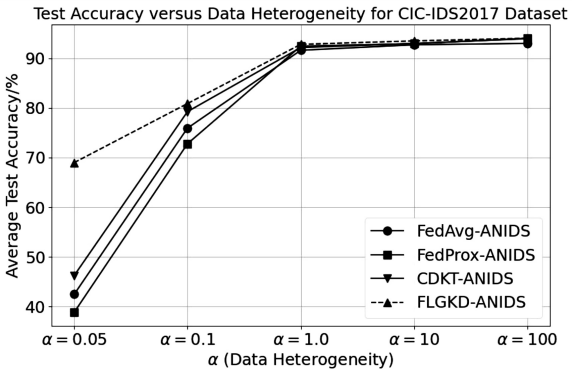

图 4 CIC-IDS2017 数据集上数据异构性对模型性能的影响

3.2.3 不同通信轮次的准确性比较

为了更细致地分析 FLGKD-ANIDS 的有效性，表 2 展示了两个数据集在数据异构程度 $\alpha = 0.05$ 和缓冲区大小为 3 的场景下，文中方法与基线方法在不同轮次

时的模型准确率。结果显示,FLGKD-ANIDS 在多数轮次中的模型表现优于基线方法,并且其性能更接近集中式场景中的异常网络入侵检测模型。具体而言,在 UNSW-NB15 数据集上,除了在第 10 轮时三种方法的模型准确率均为 14.85% 外,随着轮次增加,FLGKD-ANIDS 在第 20 轮、第 50 轮和第 100 轮时的性能最佳,分别比最好的基线方法高出 16.61 百分点、1.39 百分点和 0.69 百分点,特别是在第 20 轮次,提升显著。在 CIC-IDS2017 数据集上,FLGKD-ANIDS 在第 10 轮、第 20 轮、第 30 轮和第 50 轮时的表现均为最佳,分别比最好的基线方法高出 55.66 百分点、4.25 百分点、0.13 百分点和 29.57 百分点,特别是在第 10 轮和第 50 轮时,性能提升显著。综上所述,FLGKD-ANIDS 在高数据异构场景下的异常网络入侵检测模型性能优于基线方法,并且更加接近于集中式场景下的表现。

表 2　FLGKD-ANIDS 与基线方法在不同通信轮次下准确率的对比(Buffersize=3, α =0.05)　%

方法	UNSW-NB15 TestAcc at different communication round				CIC-IDS2017 TestAcc at different communication round			
	10	20	50	100	10	20	30	50
FedAvg-ANIDS	14.85	36.63	64.35	76.23	11.30	16.52	80.74	25.22
FedProx-ANIDS	14.85	41.58	65.34	76.23	13.91	13.91	26.08	30.43
CDKT-ANIDS	12.34	55.67	68.90	76.54	11.30	12.17	78.26	53.04
FLGKD-ANIDS	14.85	72.28	70.29	77.23	69.57	20.87	80.87	82.61
Centralized-ANIDS	81.18	81.18	80.20	80.20	97.39	96.52	96.78	96.91

3.2.4　缓冲区大小的影响

为了探究 FedGKD-ANIDS 的最佳缓冲区大小设置,该文比较了不同大小的缓冲区在不同数据异构场景下对模型性能的影响,如表 3 所示。由于缓冲区仅在服务器端进行存储和计算,不会增加通信成本。实验结果表明,在 UNSW-NB15 数据集上,当 α =0.05 和 0.1 时,缓冲区大小为 7 时模型的整体性能最佳,缓冲区大小为 3 次优;而在 α =1.0、10、100 时,缓冲区大小为 3 时模型的整体性能最佳。在 CIC-IDS2017 数据集上,无论数据异构程度如何,缓冲区大小为 3 时模型的整体性能始终最优。综上所述,缓冲区大小为 3 在两个数据集及不同的数据异构场景下均表现出最佳的整体性能。

表 3　缓冲区大小对 FLGKD-ANIDS 平均测试准确性 Acc_{Avg} 的影响　%

数据集	Buffersize	α = 0.05	α = 0.1	α = 1.0	α = 10	α = 100
UNSW-NB15	1	64.03	65.07	77.48	80.10	80.22
	3	64.75	68.50	78.67	80.33	80.69
	5	64.44	66.59	77.97	79.79	79.77
	7	64.95	70.07	77.40	79.13	79.18
CIC-IDS2017	1	45.98	77.70	91.60	92.71	92.99
	3	70.64	81.79	92.90	93.49	94.03
	5	68.99	80.86	92.80	93.32	93.70
	7	70.14	81.47	92.88	93.39	93.90

4　结束语

针对联邦学习中客户端数据异构性导致的模型性能下降问题,该文提出了一种结合知识蒸馏技术的方法 FLGKD-ANIDS。该方法通过在服务器中设置缓冲区来存储多轮历史全局模型参数,并计算包含多轮全局知识的教师模型,进而通过知识蒸馏将这些知识传递给客户端模型,以此提升模型性能。为了验证 FLGKD-ANIDS 的有效性,该文在多个联邦异构场景和集中式场景下与基线模型进行对比。实验结果表明,在不同的数据异构场景下,FLGKD-ANIDS 的性能优于其他联邦学习方法。通过分析不同训练轮次下模型性能的表现,进一步证实了该方法的优越性。最后,通过缓冲区大小的实验,确定了最佳缓冲区设置。

未来的研究将进一步优化联邦学习中本地模型的构建,探索不同网络入侵检测模型在联邦学习框架下的表现及其对整体模型性能的影响。

参考文献:

[1] 王振东,张　林,李大海. 基于机器学习的物联网入侵检测系统综述[J]. 计算机工程与应用,2021,57(4):18-27.

[2] AGRAWAL S,SARKAR S,AOUEDI O,et al. Federated learning for intrusion detection system:concepts,challenges and future directions[J]. Computer Communications,2022,195:346-361.

[3] POPOOLA S I,ANDE R,ADEBISI B,et al. Federated deep learning for zero-day botnet attack detection in IoT edge devices[J]. IEEE Internet of Things Journal,2021,9:3930-

3944.
[4] 刘艺璇,陈 红,刘宇涵,等. 联邦学习中的隐私保护技术[J]. 软件学报,2022,33(3):1057-1092.
[5] MOTHUKURI V,PARIZI R M,POURIYEH S,et al. A survey on security and privacy of federated learning[J]. Future Generation Computer Systems,2021,115:619-640.
[6] QIN Y,KONDO M. Federated learning-based network intrusion detection with a feature selection approach[C]//2021 international conference on electrical, communication, and computer engineering(ICECCE). Kuala Lumpur:IEEE,2021:1-6.
[7] ZHU H,XU J,LIU S,et al. Federated learning on non-IID data:a survey[J]. Neurocomputing,2021,465:371-390.
[8] CAMPOS E M,SAURA P F,GONZ'ALEZ-VIDAL A,et al. Evaluating federated learning for intrusion detection in internet of things: review and challenges[J]. Computer Networks,2022,203:108661.
[9] LI L,FAN Y,TSE M,et al. A review of applications in federated learning[J]. Computers & Industrial Engineering,2020,149:106854.
[10] ZHANG J,GUO S,QU Z,et al. Adaptive federated learning on non-iid data with resource constraint[J]. IEEE Transactions on Computers,2021,7:1655-1667.
[11] AHMAD Z,KHAN A S,CHEAH W S,et al. Network intrusion detection system:a systematic study of machine learning and deep learning approaches[J]. Transactions on Emerging Telecommunications Technologies,2021,32(1):e4150.
[12] 史嘉宁. 基于融合模型的网络异常入侵检测方法应用[D]. 呼和浩特:内蒙古财经大学,2024.
[13] YAKUBU I. 基于异常的网络入侵检测的深度学习方法[D]. 成都:电子科技大学,2023.
[14] 王锁成,陈世平. 一种基于残差网络改进的异常流量入侵检测模型[J]. 小型微型计算机系统,2023,44(12):2757-2764.
[15] 常利伟,刘秀娟,钱宇华,等. 基于卷积神经网络多源融合的网络安全态势感知模型[J]. 计算机科学,2023,50(5):382-389.
[16] 罗虹富,王 恒,马自强. 基于 CNN 和 BiLSTM 的分层注意力网络入侵检测方法[J]. 计算机技术与发展,2024,34(11):95-100.
[17] LI B,WU Y,SONG J,et al. DeepFed:federated deep learning for intrusion detection in industrial cyber - physical systems[J]. IEEE Transactions on Industrial Informatics,2020,17(8):5615-5624.
[18] YANG Q,LIU Y,CHEN T,et al. Federated machine learning:concept and applications[J]. ACM Transactions on Intelligent Systems and Technology,2019,10(2):1-19.
[19] CRIADO M F,CASADO F E,IGLESIAS R,et al. Non-iid data and continual learning processes in federated learning:a long road ahead[J]. Information Fusion,2022,88:263-280.
[20] DIAO E,DING J,TAROKH V. HeteroFL:computation and communication efficient federated learning for heterogeneous clients[J]. arXiv:2010.01264,2010.
[21] MCMAHAN B,MOORE E,RAMAGE D,et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial intelligence and statistics. Fort Lauderdale:PMLR,2017:1273-1282.
[22] CHIARO D,PREZIOSO E,IANNI M,et al. FL-enhance:a federated learning framework for balancing non-IID data with augmented and shared compressed samples[J]. Information Fusion,2023,98:101836.
[23] ZHANG L,SHEN L,DING L,et al. Fine-tuning global model via data-free knowledge distillation for non-iid federated learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans:IEEE/CVF,2022:10174-10183.
[24] YAO D,PAN W,DAI Y,et al. Fed gkd:towards heterogeneous federated learning via global knowledge distillation[J]. IEEE Transactions on Computers,2023,73(1):3-17.
[25] LI T,SAHU A K,ZAHEER M,et al. Federated optimization in heterogeneous networks[C]//Proceedings of machine learning research. Austin:PMLR,2020:429-450.
[26] KARIMIREDDY S P,KALE S,MOHRI M,et al. SCAFFOLD:stochastic controlled averaging for on-device federated learning[C]//Proceedings of the 37th international conference on machine learning. Vienna:PMLR,2020:5132-5143.
[27] HINTON G,VINYALS O,DEAN J. Distilling the knowledge in a neural network[J]. arXiv:1503.02531,2015.
[28] LIN T,KONG L,STICH S U,et al. Ensemble distillation for robust model fusion in federated learning[C]//Proceedings of the 34th conference on neural information processing systems. Vancouver:PMLR,2020:2351-2363.
[29] ZHU Z,HONG J,ZHOU J. Data-free knowledge distillation for heterogeneous federated learning[C]//Proceedings of the 38th international conference on machine learning. Vienna:PMLR, 2021:12878-12889.
[30] LE H Q,NGUYEN M N H,PANDEY S R,et al. CDKT-FL:cross-device knowledge transfer using proxy dataset in federated learning[J]. Engineering Applications of Artificial Intelligence,2024,133:108093.
[31] YOO J,CHO M,KIM T,et al. Knowledge extraction with no observable data[C]//Advances in neural information processing systems. Vancouver:PMLR,2019:32.
[32] LOPES R G,FENU S,STARNER T. Data-free knowledge distillation for deep neural networks[J]. arXiv:1710.07535,2017.
[33] HSU T M H,QI H,BROWN M. Measuring the effects of non-identical data distribution for federated visual classification[J]. arXiv:1909.06335,2019.
[34] 李可欣. 基于联邦学习的网络入侵检测方法研究[D]. 大连:辽宁师范大学,2022.