基于联邦学习的大数据隐私保护与协同网络安全检测模型研究

李琛

中国科技出版传媒股份有限公司,北京 100717

摘要：为应对大数据环境下隐私泄露与跨域协同检测难题，本文构建了一种基于联邦学习的网络安全检测模型。通过边缘节点本地训练与中央聚合器参数融合，实现多源数据的隐私保护建模。模型融合差分隐私与对抗防御机制，在保障数据安全的同时提升了检测鲁棒性。实验以 UNSW-NB15 和 CIC-IDS2017 数据集为基础，结果显示该模型在准确率、F1 值、误报率等指标上优于传统方法，且在通信开销与攻击防护方面具备明显优势。研究成果为多方参与场景下的安全协同检测提供了有效技术路径。

关键词：联邦学习；隐私保护；协同检测；大数据安全；网络攻击识别

引言

信息技术的发展推动大数据成为网络空间的重要资产，金融、电信、医疗等领域数据结构复杂、更新频繁，使得网络攻击手段愈加多样与隐蔽。传统依赖中心化方式的安全检测体系难以满足当前高频攻击与大规模数据处理的需求，逐渐暴露出响应滞后、泛化能力弱、隐私风险高等问题。隐私法规日益严苛的背景下集中收集原始数据用于训练的方式已难以适应“数据本地化”的合规要求。由于数据受限于技术、制度与法律障碍，跨域共享困难，数据孤岛现象严重，严重制约了网络安全协同防御体系的构建。在此背景下，如何不暴露原始数据的前提下，实现多机构之间的联合建模与高效检测，成为当前大数据安全治理中亟需解决的关键问题。

1 理论基础与关键技术分析

1.1 联邦学习基本原理与框架

联邦学习是一种分布式机器学习方法，其核心思想是在不集中原始数据的前提下，通过本地训练和模型参数聚合，实现多参与方协同构建全局模型的目标[1]。根据数据分布的不同，联邦学习可分为横向联邦学习、纵向联邦学习和联邦迁移学习。横向联邦学习适用于参与方数据特征维度相似但样本不重合的场景，如多个银行的用户行为分析；纵向联邦学习则适用于特征空间互补但样本部分重合的情况，典型如金融机构与电商平台联合建模；而联邦迁移学习用于参与方之间数据分布和特征空间差异较大但仍希望共享知识的场景。在联邦学习中，参数聚合策略是其关键环节，常用方法包括 FedAvg、FedOpt 和 FedProx。FedAvg 通过简单平均本地更新参数实现模型同步，适用于数据分布相对均衡的环境；FedOpt 则通过优化器调整全局模型的收敛路径，提高训练效率；FedProx 在损失函数中引入正则项，以缓解数据异构带来的不一致性问题。与传统集中式机器学习相比，联邦学习不仅减少了数据传输，提高了训练的隐私安全性，还能更好地适应边缘计算环境下多设备协同的现实需求，展现出更强的应用适配能力与制度合规性[2]。

1.2 大数据环境下隐私保护机制

大数据安全建模过程中如何在保障数据价值最大化的同时实现隐私保护，是模型设计的重要挑战。差分隐私是一种基于数学保障的数据匿名化技术，其核心在于通过在输出中引入随机噪声，使得攻击者难以通过结果反推出任一用户的具体信息[3]。ε-差分隐私控制了这种扰动程度，使得模型训练过程在保证有效性的同时兼顾隐私强度。同态加密是一种支持加密状态下执行运算的加密机制，在联邦学习中可用于对模型梯度进行加密，使得协调服务器在不知明文的前提下完成聚合操作，从而防止模型参数泄露。安全多方计算技术通过在多个参与方之间进行密文运算，确保各方输入数据在整个计算过程中始终保持不可见，特别适用于高度敏感的数据协同建模任务。另一个重要支撑是可信执行环境（Trusted Execution Environment, TEE），该机制通过在硬件层面隔离出受保护的执行区，确保模型训练过程中的代码与数据不会被外部系统窃取或篡改，极大提升了模型训练阶段的可信性。上述隐私保护机制各具优势，在联邦学习框架中往往需要组合使用，以平衡性能开销与安全保障之间的权衡关系[4]。

1.3 网络安全检测技术发展概述

网络安全威胁的复杂化驱动着检测技术的不断演进。从 DoS 攻击、端口扫描到钓鱼攻击、恶意流量注入，攻击手段层出不穷，对应的检测需求也呈现出动态、实时、多维的特征[5]。传统基于规则匹配的方法由于依赖已有攻击特征库，对新型变种攻击难以识别，因此逐渐被数据驱动的机器学习方法所替代。近年来，深度学习模型因其在特征自动提取、复杂模式识别方面的显著优势，被广泛引入网络安全检测中。卷积神经网络（DNN）在识别静态流量模式方面表现出色，而循环神经网络（LSTM）在处理流量时间序列数据方面具有优势，可有效捕捉攻击行为的时序特征。但深度模型训练所依赖的大规模标签数据难以集中获取，训练代价高、对抗攻击敏感等问题仍需重视。在模型部署方面，当前主流检测系统多采用中心化结构，部署在云端或核心服务器上，具有较高资源依赖性；而部分系统尝试将模型嵌入到本地或边缘设备中以提升响应效率与实时性。更为先进的方式是采用云边协同架构，通过将模型训练、更新与推理任务在云端与本地间灵活分配，兼顾了计算效率、通信开销与数据保护需求，这也为联邦学习等协同建模机制的落地应用创造了良好条件。

2 模型架构设计与系统实现

2.1 联邦协同安全检测系统架构

为实现跨组织、多节点之间的网络安全威胁联合检测，并在保障数据隐私的前提下进行协同训练，本研究构建了一个包含边缘节点、本地模型与中央聚合器三层结构的联邦安全检测系统。在系统架构中，各边缘节点负责本地数据的预处理与模型更新，而中央聚合器仅对加密参数进行汇总与迭代优化，从不接触原始数据，确保数据不出本地。数据分布方面，由于来源于不同行业或机构，存在显著的非独立同分布特征（Non-IID），对训练稳定性与收敛速度提出更高要求，因此在本地预处理过程中采用标准归一化与平衡抽样技术，以缓解数据偏差带来的影响。在通信机制方面，系统通过端到端的 SSL 加密协议实现传输安全，结合模型轮次压缩与量化技术，降低模型同步时带宽压力。此外，设计中还采用异步接入策略支持节点灵活加入与退出，提升系统弹性。

2.2 模型训练与优化流程

在实际训练过程中，每个边缘节点先基于本地数据划分样本，按比例设置训练集与验证集，通过本地初始化构建轻量化神经网络模型。节点完成若干轮本地训练后，提取模型梯度或权重变化信息，并加密后上传至中央聚合器。聚合器采用加权平均或正则优化策略对参数进行融合，再将更新后的模型参数分发至各节点继续下一轮迭代，形成“本地训练—集中聚合—模型分发”的闭环流程。针对同步策略在大规模部署中容易引发阻塞的问题，系统支持同步与异步模式并行运行。其中同步模式便于统一收敛控制，适用于设备性能相近的场景；而异步模式更适配边缘节点资源差异大、网络质量波动频繁的实际应用需求。为了量化不同策略在实际部署中的性能差异，本文设计对比实验并整理如下。

表 1 不同模型更新策略的性能对比

更新方式	模型收敛轮次	平均通信时延（ms）	检测准确率（%）	网络鲁棒性评分
同步模式	25	430	92.6	中
异步模式	28	210	91.2	高
混合调度	26	280	93.1	高

由表 1 可见，在通信效率与模型精度之间，混合调度策略表现出较好的平衡特性。虽然异步模式在收敛速度上略低于同步模式，但在通信延迟与网络适应性方面具有明显优势，尤其适用于多区域部署场景。

2.3 隐私保护与安全机制融合

为了保障模型训练过程中信息的安全性与完整性，本系统引入多种隐私保护与攻击防御机制。差分隐私策略通过在参数上传前引入可控噪声，有效抑制逆向推理攻击对个体样本的重建可能性，ε -值根据任务敏感性与训练轮次数进行动态调整。在模型传输过程中，加入随机扰动机制提升模型不可识别性，进一步增强系统抗推理能力。针对联邦学习中可能出现的恶意节点上传伪造参数、污染聚合模型等问题，系统内置对抗检测模块，采用基于行为特征的异常识别方法，结合聚合过程中的梯度轨迹偏离分析，及时识别可疑节点并隔离处理。此外，聚合端采用轻量级验证机制核验上传参数的一致性与合法性，增强整体系统的鲁棒性与防攻击能力。通过多层次、全流程的隐私与安全融合机制，保障了大数据环境下协同安全检测的可信运行基础。

3 实验设计与效果评估

3.1 实验设置与数据说明

为验证所提出的基于联邦学习的大数据隐私保护与协同网络安全检测模型的可行性与性能优势，搭建了仿真联邦实验平台，采用 Python 与 TensorFlow Federated 框架构建系统环境，运行平台为 Intel Xeon Gold 5218 CPU、256GB 内存，模拟部署 10 个边缘节点与 1 个中央聚合服务器。各节点模拟不同组织单位的数据环境，采用非独立同分布（Non-IID）数据划分策略，进一步贴近现实场景。实验所用数据集包括 UNSW-NB15 与 CIC-IDS2017，涵盖多种常见网络攻击类型与正常流量行为。其中，UNSW-NB15 提供了包括探测攻击、

DoS、后门等在内的九类威胁样本，CIC-IDS2017 则包括 BotNet、Web 攻击等多维度标签，数据样本结构复杂、特征维度较高。所有原始数据在上传训练前均进行标准化处理，包括缺失值填补、异常值清除、类别标签 One-hot 编码与数值特征归一化等，确保模型训练的稳定性与收敛性。

3.2 实验方案与评估指标

实验设计包括三种对比模型：传统集中式深度检测模型（Centralized-DNN）、不带隐私保护机制的联邦检测模型（Plain-FL）与本研究所提出的差分隐私嵌入联邦检测模型（DP-FedSec）。各模型在同样数据条件下进行训练，并分别测试其在准确率、召回率、F1 值、误报率等检测性能指标上的表现。同时，采集通信延迟与参数传输开销，对系统部署效率进行横向对比。为进一步评估隐私机制带来的安全增益，设计了模拟对抗攻击实验，包括推理攻击与模型中毒两类典型威胁，比较各模型在隐私泄露风险与鲁棒性上的差异。

表 2 三种模型在不同评估指标下的对比结果

模型类型	准确率（%）	召回率（%）	F1 值（%）	误报率（%）	通信开销（MB）
Centralized-DNN	94.8	91.2	92.9	4.7	0
Plain-FL	92.6	88.5	90.4	6.1	320
DP-FedSec	93.4	90.1	91.7	5.2	348

从表 2 中可见，DP-FedSec 模型在保证相对高精度的同时，有效控制了误报率，且虽略高于 Plain-FL 的通信负担，但整体性能均衡性更优，体现出联邦架构在兼顾精度与资源利用方面的优势。

在对抗攻击模拟实验中，进一步测试三种模型面对梯度推理攻击与模型投毒操作时的表现，相关结果整理如下：

表 3 模型在攻击场景下的鲁棒性与隐私保护能力评估

模型类型	推理攻击成功率（%）	模型中毒后准确率下降（%）	异常节点识别成功率（%）
Centralized-DNN	78.4	36.5	无防护机制
Plain-FL	42.7	24.1	48.6
DP-FedSec	11.5	9.8	87.3

表 3 数据显示，DP-FedSec 在隐私攻击防护与模型鲁棒性方面表现最为优越，推理攻击成功率显著降低，模型在受到投毒时仍能维持较高稳定性，且具备较强的异常节点识别能力，验证了所引入的差分隐私与防御机制的实际效能。

为进一步观察模型训练过程的稳定性与收敛效率，记录不同模型在迭代过程中的准确率变化曲线，如表 4 所示：

表 4 不同模型在迭代过程中的收敛速度（轮数与准确率变化）

迭代轮数	Centralized-DNN	Plain-FL	DP-FedSec
5	81.3	76.5	78.1
10	87.9	83.2	85.4
15	91.2	86.8	89.7
20	94.1	90.2	92.5
25	94.8	92.6	93.4

如表 4 所示，虽然 DP-FedSec 在早期迭代阶段收敛速度略慢于集中式模型，但最终达到的准确率接近，并优于普通联邦模型，证明该方法在保证训练效率的基础上，兼顾了隐私保护与检测效果，具备良好的工程可行性。

3.3 结果分析与模型优势验证

综合前述实验数据可知，所提出的 DP-FedSec 模型在多个关键指标上均表现优越，不仅准确率高、误报率低，而且在面临复杂攻击与非独立分布数据时仍保持较强稳定性，说明其在实际部署中具有较

高的实用价值。通过对不同模型在训练过程中准确率随轮次变化趋势的记录，可进一步观察其收敛效率与性能增长幅度。如表 5 所示：

表 5 模型训练轮数与准确率增长趋势对比

训练轮数	Centralized-DNN	Plain-FL	DP-FedSec
5	81.3%	76.5%	78.1%
10	87.9%	83.2%	85.4%
15	91.2%	86.8%	89.7%
20	94.1%	90.2%	92.5%
25	94.8%	92.6%	93.4%

从表 5 可以看出，DP-FedSec 模型在早期轮次中的准确率上升幅度稳定，虽然略慢于集中式模型，但收敛后达到的最终准确率几乎持平，并显著优于未加隐私保护的 Plain-FL 模型，验证了其在融合隐私保护机制的同时仍具备强竞争力。此外，在面对推理攻击和参数投毒测试时，DP-FedSec 模型保持了较高的鲁棒性与识别准确性，进一步说明其不仅在性能维度具备优势，也在安全维度展现出可持续防御能力，具备在多源异构环境中推广应用的可行性与适应性。

4 结论

本研究围绕大数据环境下的隐私保护与网络安全检测双重需求，构建了一种融合联邦学习机制的协同检测模型。该模型通过分布式本地训练与加密参数聚合的方式，实现了多数据主体之间的协同防御目标，同时规避了原始数据集中带来的隐私泄露风险。在系统设计中嵌入差分隐私机制、随机扰动策略及对抗攻击识别模块，有效提升了模型在复杂应用场景下的安全性与鲁棒性。实验部分基于 UNSW-NB15 与 CIC-IDS2017 两类典型数据集开展对比验证，结果显示该模型在检测准确率、召回率等核心指标上表现良好，且在通信资源控制、对抗攻击防护、异常节点识别等方面具备明显优势，展现出广泛的实用推广价值。

参考文献

[1]齐歌.大数据技术在计算机网络信息安全管理中的应用[J].中国宽带,2025,21（05）:40-42.

[2]李建华,银鹰,李思源,等.大数据安全与隐私计算技术综述[J].网络空间安全科学学报,2024,2（06）:1-15.

[3]赵一畅.大数据时代的网络空间安全风险与防御[J].数字通信世界,2024,（10）:75-77+80.

[4]郭书群.大数据环境下高职计算机网络安全防护面临的挑战与对策研究[J].信息与电脑（理论版）,2024,36（15）:130-132.

[5]段昕汝,陈桂茸,陈爱网,等.联邦学习中的信息安全问题研究综述[J].计算机工程与应用,2024,60（03）:61-77.

作者简介：李琛，女，（1984-12-2），汉，河北省，出版中级，硕士，研究方向：计算机科学与技术