基于信誉权重和可信分发的城市轨道交通入侵检测

尹雨晴[1]，张青苗[1]，赵军辉[1,2]，李汶佳[1]

（1. 华东交通大学信息与软件工程学院，江西 南昌，330013；2. 北京交通大学电子信息工程学院，北京，100044）

摘 要:在城市轨道交通系统中，基于通信的列车控制系统（Communication Based Train Control，CBTC）利用双向无线通信实现列车与轨旁设备间的实时数据交换来实现列车的正常运行。但 CBTC 开放的运行环境使其面临网络攻击的威胁，为此，本文提出一种基于信誉权重和可信分发的协同入侵检测方法。该方法首先采用差分隐私和秘密共享技术，支持本地训练检测模型，并通过主观逻辑动态评估节点可靠性；然后，结合信誉加权聚合算法有效抑制恶意攻击，提升系统稳定性；最后，引入区块链构建可信分发，确保模型更新安全。在 CBTCset 数据集上的仿真实验结果表明，所提方法的准确率最高可以达到 99.6%，在准确率、F1、精确度、召回率和时延等方面的性能均优于传统的加权聚合和隐私保护方法。

关键词：城市轨道交通，CBTC，入侵检测，区块链技术，可信分发，动态信誉机制

中图分类号：U285　　**文献标志码**：A

Intrusion Detection Based on Reputation Weight and Trusted Distribution in Urban Rail Transit

Yin Yuqing[1], Zhang Qingmiao[1], Zhao Junhui[1,2], Li Wenjia[1]

(1. School of Information and Software Engineering, East China Jiaotong University, Nanchang 330013, China; 2. School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China.)

Abstract: In urban rail transit systems, Communication-Based Train Control (CBTC) utilizes bidirectional wireless communication to enable real-time data exchange between trains and trackside equipment, thereby ensuring normal train operations. However, the open operating environment of CBTC exposes it to the threat of cyber attacks. To address this issue, this paper proposes a collaborative intrusion detection method based on reputation weight and trusted distribution. This approach first employs differential privacy and secret sharing techniques to support locally trained detection models, dynamically evaluating node reliability through subjective logic. Subsequently, it combines a reputation-weighted aggregation algorithm to effectively suppress malicious attacks and enhance system stability. Finally, blockchain technology is introduced to establish a trusted distribution mechanism, ensuring secure model updates. Simulation experiments on the CBTCset dataset demonstrate that the proposed method achieves an accuracy rate of up to 99.6%, outperforming traditional average-weighted aggregation and privacy-preserving methods in terms of accuracy, F1, precision, recall and time delay.

Keywords: Urban Rail Transit, CBTC, Intrusion Detection, Blockchain Technology, Trusted Distribution, Dynamic Reputation Mechanism

随着城市化进程的加快，城市轨道交通在缓解交通拥堵和提高运行效率方面发挥着关键作用[1]。基于通信的列车控制系统（Communication Based Train Control，CBTC）是城市轨道交通的中枢神经。它利用列车到地面（Train-to-Ground，T2G）的无线通信实现列车与轨旁设备之间的实时双向信息交互，从而保障列车运行的连续性与安全性[2]。然而，城市轨道交通开放的通信环境使得 T2G 无线通信面临拒绝服务攻击、会话劫持攻击和 MAC 地址欺骗等多种网络安全威胁[3-4]。而车地通信链路一旦遭到破坏，将直接影响列车运行的平稳性，严重的情况下，可能造成城市轨道交通运行中断，甚至引发公共安全事件[5]。因此，针对城市轨道交通无线通信系统开展安全入侵检测具有重要意义。

近年来，研究人员广泛采用机器学习的算法开展安全入侵检测的相关研究。例如，文献[6]提出了基于对抗性机器学习的网络入侵检测方法来提高网络入侵检测的检测性能以及泛化性。文献[7]提出了一种

收稿日期：2026-01-13

基金项目：国家自然科学基金项目（62261024）、高铁与城市轨道交通系统技术国家工程研究中心（2024YJ255）、国家重点研发项目（2020YFB1807204）

融合双向门控循环单元和注意力机制的网络入侵检测模型，对不同类型流量数据通过加权的方式进行重要程度的区分，从而整体提高该模型特征提取与分类的性能。文献[8]提出了一种混合时空预测框架，集成卷积神经网络（Convolutional Neural Network, CNN）、Informer 和 Softmax 分类器，实现正常和异常网络攻击的高精度检测。文献[9]提出了一种针对车地通信的双层入侵检测方法，以保证车地通信系统的安全性。在第一层引入机器学习算法对无线网络中的攻击进行检测；在第二层提出了基于状态观测器的入侵检测方法，在每个通信周期观测器计算真实的值与估计值之间的偏差。最后综合两层检测结果，采用不同的报警方式，提高了车地通信系统的检测效果。然而，基于机器学习的入侵检测模型通常依赖大规模数据的集中式训练，这与数据持有方因隐私顾虑而导致的数据孤岛现象存在根本性矛盾。

为确保数据隐私的前提下实现高效的模型训练，联邦学习提供了一种去中心化的思路。联邦学习将模型训练分散到各个本地端，同时再将分布式的小数据片段进行聚合，从而提高模型的安全性和隐私性[10]，受到了研究者的广泛关注。文献[11]提出了一种差分隐私联邦学习方法，通过向局部梯度注入一些噪声来保护局部梯度的隐私，从而防止对手推断任何局部信息。文献[12]提出了一种融合联邦学习与量子卷积神经网络的入侵检测模型来提高模型的检测性能和训练稳定性。文献[13]提出了一种基于联邦学习和注意力机制的物联网入侵检测模型，允许多个设备在保护其数据隐私的基础上协同训练全局模型，提取网络流量数据的关键特征，从而提高检测的准确率。为解决联邦学习面临的模型篡改的风险，一些研究人员提出区块链的方法。文献[14]利用区块链赋能的 CBTC 跨层防御方法，在网络层使用智能合约实现通信密钥的去中心化注册和更新，在物理层通过分析卡尔曼滤波器残差的分布，检测是否存在数据篡改攻击。文献[15]利用了区块链的透明性和智能合约的自动化解决了 V2V 通信中的身份信任和安全数据传输问题。文献[16]利用区块链赋能的分布式密钥管理与认证体系提供分布式的公钥管理和身份认证服务，同时使用深度强化学习，将区块生产者选择和安全节点切换决策建模为一个优化问题。

同时有研究人员提出了机器学习和区块链相结合的方法来提高检测的精确度和可信度。文献[17]在边缘计算三层架构中，通过联邦学习让车辆本地进行训练而不共享数据，利用区块链确保模型聚合过程的可靠性和可信度。文献[18]基于区块链和联邦学习的协作入侵检测机制，将机器学习模型的精度与区块链的共识机制和激励机制结合起来，创新性地鼓励参与者贡献高质量模型。文献[19]利用多种机器学习模型对网络流量进行二分类，区分正常流量与 DDoS 攻击流量，然后利用区块链构建一个可信的恶意节点身份管理系统，防止攻击者伪造或重复攻击。文献[20]利用联邦学习使每个边缘节点在本地训练模型。再利用区块链技术安全地聚合和验证模型更新，防止恶意节点上传不可靠的模型参数。

表 1 从隐私保护机制、鲁棒聚合策略、区块链赋能的可信审计、实时性约束四个维度对现有文献方法进行了系统评估。从表 1 可以看出，现有研究仅关注隐私保护、区块链可信性或入侵检测精度中的某一方面，缺乏面向 CBTC 系统实时性约束的综合设计。同时在鲁棒聚合机制方面采用简单平均或固定加权方式，缺少动态信誉驱动的自适应机制。此外，现有区块链融合方法大多用于模型存证或身份管理，未对模型更新分发的实时性与容错性进行严格建模。上述研究缺乏对节点持续行为的动态评估，同时基于简单的共识机制，可能导致系统收敛缓慢，并被少数恶意节点降低性能，无法实现高效、高质量的协同。

表 1 相关工作对比

Tab. 1 Comparison of existing works

文献	核心技术	隐私保护	鲁棒聚合	可信审计	实时性考虑
[9]	ML+观测器	×	×	×	√
[11]	差分隐私联邦学习	√	×	×	×
[13]	FL+注意力机制	×	×	×	×
[14]	区块链赋能跨层防御	×	×	√	√
[17]	区块链+FL	×	×	√	×

[18]	区块链+FL	×	√	√	×
[20]	FL+区块链聚合验证	×	×	√	×

为解决上述问题，本文提出了一种基于信誉权重和可信分发的协同入侵检测方法。其主要贡献包括：

（1）面向 CBTC 系统的 T2G 通信，设计了隐私保护的分布式学习框架，通过差分隐私噪声注入和秘密共享技术，显著提升了通信数据保密性。同时列车、轨道设备等终端在本地进行模型训练时仅上传模型更新而非原始数据，从源头上保护数据隐私。

（2）针对列车、轨道设备等节点共同参与模型训练的特点，构建了一种基于主观逻辑的信誉机制，动态评估各参与节点的可靠性，并采用加权聚合算法有效抑制恶意更新，显著提升全局模型的鲁棒性。

（3）引入区块链技术，构建了可信分发，利用共识算法对模型更新进行安全聚合与存证，确保训练过程的不可篡改性。

1. 系统模型

1.1 CBTC 系统结构

如图 1 所示，CBTC 系统包括区域控制器（zone controller, ZC）、计算机联锁（computer interlocking, CI）、列车自动监控（automatic train supervision, ATS）、列车自动防护（automatic train protection, ATP）、列车自动驾驶（automatic train operation, ATO）等子系统。射频远程单元（Remote Radio Unit, RRU）是分布式基站架构的核心组件。同时部署边缘服务器（Edge Server, ES）来接收终端设备传输的数据分片，并对数据分片进行聚合和检测，来提高数据聚合和检测的速率。正常情况下，列车会向 ZC 发送自身状态信息，如位置、速度等；ZC 结合 ATS 的运行计划和 CI 的线路状态计算出移动权限（movement authority, MA），MA 规定了列车在当前时刻被允许占用的线路区段及其终止位置，是列车自动防护与运行控制的核心约束条件，用于保证列车运行过程中不发生追尾、冲突或越界等安全风险；ATP 根据 MA 生成最大速度曲线，并指导 ATO 控制列车运行。

在 CBTC 系统中，相邻列车需要按指定的路线和轨迹高效运行，当系统不受影响且按计划运行时，后车在每个通信周期中接收前车的新位置。然而，如果列车在通信过程中遭受安全攻击，例如，攻击者干扰 T2G 通信链路，造成列车从一个基站切换到另一个基站时切换失败或通信中断，导致后车不能及时收到前车的位置信息，启动不必要的列车制动，将严重降低 CBTC 的运行效率和乘客的乘车体验；或者攻击者建立与列车和地面设备的单独连接，在不被发现的情况下篡改 MA，使后车根据攻击者篡改的 MA 生成错误的操作命令，造成列车追尾事故，影响列车运行安全。图 1 红色曲线表示当列车被入侵者更改列车的状态和 MA 时的运行曲线，蓝色曲线表示列车的 ATP 紧急制动曲线和 ATO 服务制动曲线。

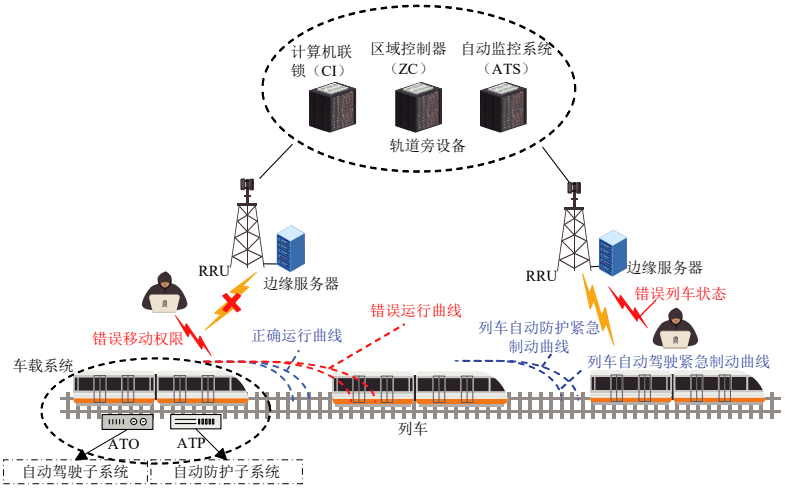

图 1 CBTC 系统

Fig.1 CBTC system

1.2 列车运行模型

根据物体运动定律，列车运行模型可表示为：

$$s\left(t+\Delta t\right)=s\left(t\right)+v\left(t\right)\cdot\Delta t+\frac{1}{2}a\left(t\right)\cdot\Delta t^{2},\tag{1}$$

$$v\left(t+\Delta t\right)=v\left(t\right)+a\left(t\right)\cdot\Delta t,\tag{2}$$

其中，$s\left(t\right)$表示列车在t时刻的位置，$v\left(t\right)$表示列车在t时刻的运行速度，$a\left(t\right)$则表示相应的加速度，Δt为时间步长。

在CBTC系统中，ATO根据接收到的MA信息计算目标速度曲线，并跟踪该曲线。MA的计算依赖于前车位置、线路条件及列车状态信息。MA定义了列车允许运行的安全距离$\hat{d}\left(t\right)$，确保列车在t时刻能够在障碍物前安全停车：

$$\hat{d}\left(t\right)=d\left(t\right)+\Delta d\left(t\right),\tag{3}$$

其中，$d(t)$为制动距离，$\Delta d\left(t\right)$为安全阈值。

因此，列车在t时刻的移动权限更新为：

$$d_{\mathrm{MA}}\left(t\right)=s\left(t+\Delta t\right)+\hat{d}\left(t\right).\tag{4}$$

1.3 列车通信模型

CBTC 系统的控制指令的实时性依赖连续、可靠的车地通信。本文采用当前CBTC系统常用的基于长期演进技术的物联网通信技术（Long Term Evolution for Machines, LTE-M）通信机制。LTE-M在物理层采用正交频分复用（Orthogonal Frequency Division Multiplexing, OFDM）技术，其在特定信道γ下的误码率为：

$$P_{\mathrm{e}}\left(\gamma\right)=Q\left(\sqrt{\frac{2E_{\mathrm{b}}}{N_0}}\right),\tag{5}$$

其中，E_b表示每比特的能量，N_0表示噪声功率谱密度，$Q\left(\cdot\right)$高斯Q函数，$Q\left(\cdot\right)=\frac{1}{2}\exp\left(\frac{\cdot}{\sqrt{2}}\right)$。对于一个长度为$L$比特的数据包，其在一个信道为$\gamma$的传输成功概率为：

$$P_{\mathrm{s}}\left(\gamma\right)=\left(1-P_e\left(\gamma\right)\right)^{L}.\tag{6}$$

则数据包丢失率为$P_{\mathrm{loss}}\left(\gamma\right)=1-P_{\mathrm{s}}\left(\gamma\right)$。考虑最大重传次数为$N_{\max}$，则一个数据包的成功传输延迟为：

$$T_{\mathrm{tx}}=T_{\mathrm{p}}+\sum_{N_s=1}^{N_{\max}}\left[P_{\mathrm{loss}}\cdot\left(N_s\cdot T_{\mathrm{out}}+T_{\mathrm{p}}\right)\right],\tag{7}$$

其中，T_{p}为传输时间，T_{out}为等待确认的超时时间，N_{s}为当前传输次数。

在CBTC系统中，列车在移动过程中需要在不同基站间执行切换。切换过程通常包含测量、决策、执行三个阶段。若切换过程因网络攻击而失败或延迟，将引入额外的通信中断时间T_{over}。此时间与无线信道条件、网络负载以及攻击强度相关。因此，总的车地通信中断时间T_{all}是传输延迟T_{tx}和切换延迟T_{over}的综合结果：

$$T_{\mathrm{all}}=T_{\mathrm{tx}}+T_{\mathrm{over}}.\tag{8}$$

由于列车的实时性要求，要保证列车的中断时间少于最大允许的传输时延$T_{\max}$，$T_{all}\le T_{\max}$。

2. 基于信誉权重和可信分发的协同入侵检测方法

2.1 入侵检测模型架构

本文提出了一种基于信誉权重和可信分发的协同入侵检测方法，图 2 展示了协同入侵检测方法的整体架构，该架构由本地训练模块、边缘聚合模块与区块链共识模块三部分组成。

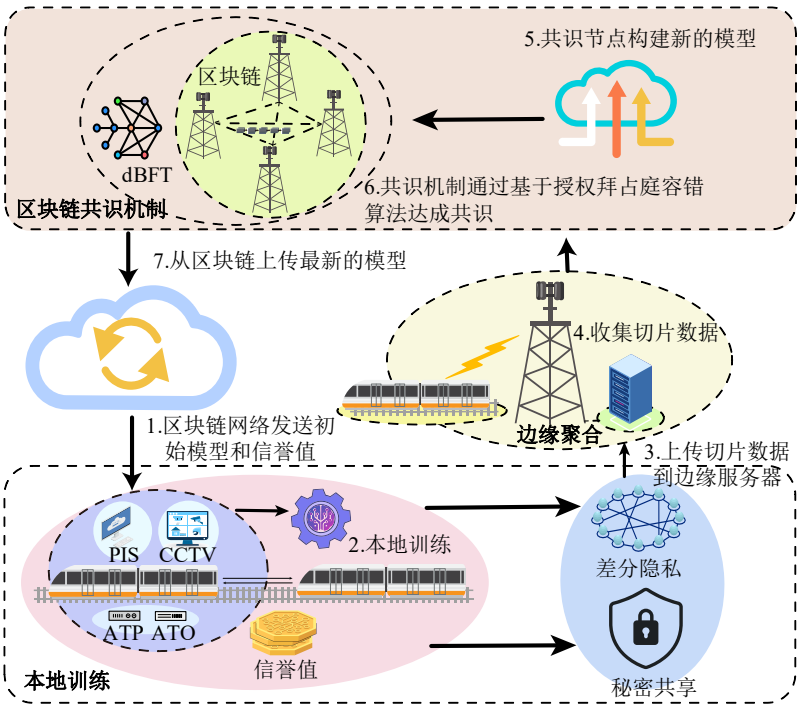

图 2 基于信誉权重和可信分发的协同入侵检测模型架构

Fig.2 A Collaborative Intrusion Detection Model Architecture Based on Reputation Weighting and Trusted Distribution

在本地训练模块中，区块链下发模型初始数据和信誉值，各终端节点在本地执行数据训练。为保障上传模型的可靠性，终端节点采用差分隐私向模型中注入校准噪声，并结合秘密共享机制对参数进行分片处理。在信誉评估阶段，系统基于主观逻辑对终端节点的历史行为、精确度以及交互成功率等多维指标进行动态量化后生成节点信誉值。该方法能够有效识别行为突变或存在潜在攻击风险的节点。

在边缘聚合模块中，各边缘服务器收集本地训练生成的模型参数分片进行聚合。聚合节点首先对接收到的参数进行一致性校验，并依据信誉评估结果对可靠节点赋予更高权重。随后，通过信誉加权聚合与两阶段滤波机制，对全局模型进行鲁棒聚合，确保聚合后模型的安全性与稳定性。

最后区块链共识模块负责实现模型更新的可信验证与安全分发。聚合后的参数首先被提交至区块链网络，由共识节点基于授权拜占庭容错算法（Delegated Byzantine Fault Tolerance, dBFT）进行一致性验证。经共识确认的模型参数被写入新区块并同步至各终端节点，从而形成不可篡改的模型更新记录，保证系统运行过程的可审计性与可追溯性。

与现有将隐私保护、区块链或信誉机制分别嵌入联邦学习框架的研究不同，本文提出一种动态信誉驱动的鲁棒联邦聚合与可信分发机制相协同。其核心不在于单一模块的引入，而在于构建一个统一的权重演化与可信分发相协同的系统。通过主观逻辑刻画节点的时序行为特征，将信誉的动态变化融入聚合权重的生成过程，通过模型更新的方向一致性校验，构建几何层面的鲁棒过滤机制。同时引入符合 CBTC 系统实时性要求的可信分发机制，最终实现系统安全、数据隐私与传输实时性的协同优化。该系统构建了一个安全、隐私保护且鲁棒的分布式防御框架。在列车运行过程中，系统能够动态检测 DoS 攻击、数据篡改等威胁，确保列车运行的安全可靠。接下来将分别介绍本地训练机制、基于主观逻辑的信誉评估模型、鲁棒加权聚合算法以及可信分发流程的设计与实现。

2.2 隐私保护的本地训练

本文选用联邦学习的架构，即在分布式节点上进行本地模型训练，并将模型更新数据聚合至中央服务器，实现各方的协作学习与性能提升。在联邦学习中，一般采用平均加权聚合算法，这种算法是根据各节点的数据量对其本地模型参数进行加权平均，以实现全局模型的协同更新。然而，在存在恶意节点或不可靠节点的场景下，单纯依赖数据量进行加权易受到攻击影响。由于 CBTC 系统的运行数据包含列车位置、速度和移动授权等敏感信息，攻击者可以利用这些数据制造恶意节点从而对列车进行攻击，加权聚合算法无法有效地保障数据的可靠性。因此，本文采用基于信誉的加权聚合算法来对模型进行聚合。该算法通过对各节点的历史行为、模型贡献度或可靠性进行量化评估，将信誉值作为权重参与模型聚合来提升模型的鲁棒性和安全性。在基于联邦学习的城市轨道交通入侵检测系统中，包括列车和轨旁设备利用本地数据训练检测模型。但是直接共享这些数据或模型更新会带来严重的隐私泄露风险，为了能够在实现协同检测的同时保护数据隐私，本文利用差分隐私和秘密共享设计了严格的隐私保护机制，形成端到端的隐私保护方案。

2.2.1 基于差分隐私的噪声注入

差分隐私是一种严格的、可量化的隐私保护框架，通过在输出中添加校准的噪声，使得任何单条数据记录的增减对最终发布结果的影响可忽略不计。在 CBTC 入侵检测场景中，这确保了攻击者无法通过分析模型更新来推断特定列车的运行状态或位置信息。

对于任意两个几乎完全相同、仅相差一条记录的列车数据集 $D, D' \in \mathcal{D}$，其中 $\mathcal{D}$ 所有可能数据集的集合。随机算法 $\mathcal{A}$ 满足 (ε, δ)，其中 $\varepsilon = 1.0$ 为隐私预算，用来控制隐私保护强度，且 $\varepsilon > 0$。当 ε 越小，隐私保护越强，但其模型效用可能降低；$\delta = 10^{-5}$ 为随机噪声。当且仅当对于算法所有可能的输出子集 O 满足 $O \subseteq Range(\mathcal{A})$，满足以下不等式：

$$\Pr\left[\mathcal{A}(D) \in O\right] \leq e^{\varepsilon} \cdot \Pr\left[\mathcal{A}(D') \in O\right] + \delta, \tag{9}$$

其中，$Range(\mathcal{A})$ 表示算法 $\mathcal{A}$ 所有可能输出结果的集合，$\Pr[\cdot]$ 表示概率算子，表示随机算法 $\mathcal{A}$ 在给定数据集输入条件下，其输出的结果落入指定输出集合的概率。$\mathcal{A}(D)$ 表示在数据集 D 上运行后得到的随机输出。对于噪声注入，高斯机制是实现差分隐私的一种常用方法，但其噪声强度并非任意设定，而是取决于待发布函数对单条数据变化的敏感程度。

设函数 $f: \mathcal{D} \to \boldsymbol{R}^d$ 为模型更新计算函数，$\boldsymbol{R}^d$ 表示 d 维实数向量空间。其 L_2 灵敏度 S_f 定义为：

$$S_f = \max_{D, D'} \| f(D) - f(D') \|_2, \tag{10}$$

其中，$\|\cdot\|_2$ 表示 L_2 范数。灵敏度规定了单条记录所能引入的最大影响。由于 CBTC 系统中列车运行数据的连续性和数值范围，梯度裁剪将所有本地模型更新 ΔW_i 的 L_2 范数限制在阈值 $C_{\max} = 1.0$ 内，即 $S_f = C_{\max}$。则给定函数 f 和灵敏度 S_f 情况下，对于高斯机制，随机算法 $\mathcal{A}$ 为：

$$\mathcal{A}(D) = f(D) + \mathcal{N}\left(0, \sigma^2 S_f^2\right). \tag{11}$$

该方法通过向真实输出 $f(D)$ 添加一个服从均值 $\mathcal{N}\left(0, \sigma^2 S_f^2\right)$ 为零，协方差矩阵为 $\sigma^2 S_f^2$ 的多维高斯分布来实现隐私保护。其中噪声尺度 σ 与隐私预算 ε 和随机噪声 δ 满足：

$$\sigma \geq \frac{\sqrt{2 \ln\left(\frac{1.25}{\delta}\right)}}{\varepsilon}. \tag{12}$$

因此，对于给定的隐私参数(ε,δ)，终端节点i的带噪声模型更新为：

$$\Delta\tilde{W}_i=\Delta W_i+n_i,\quad n_i\sim\mathcal{N}\left(0,\sigma^2{C_{\max}}^2\right), \tag{13}$$

其中，i为第i列列车，$\triangle W_i$为原始模型，n_i为从指定高斯分布中采样的噪声向量，确保了每个终端节点的本地更新满足(ε,δ)差分隐私。攻击者观察到的最终输出，即加噪后的模型更新$\triangle\tilde{W}_i$，在数据集D和D'下出现的概率是非常接近的。因此，攻击者无法通过观察输出来推断任何单条记录是否存在于数据集中。

2.2.2 基于秘密共享的分布式存储

上一节提到的差分隐私技术确保了数据内容的隐私性。然而，经过差分隐私处理后的模型在存储和传输过程中仍面临单点泄露风险。为分散此风险，本文引入秘密共享方案，将完整的模型更新分散为多个无意义的分片，分别存储于不同的边缘服务器上，该方法与差分隐私相结合，共同构成了端到端的防御体系。在确保各端数据隐私的前提下，提高了 CBTC 入侵检测系统的可靠性。

对于终端节点上传的、已加噪的模型更新$\triangle\tilde{W}_i$，我们将其编码为有限域$\boldsymbol{F}$的元素序列，采用 Shamir (t,n)阈值秘密共享方案。其核心思想是将秘密值即模型更新的每个参数作为多项式的常数项，随机选择$t-1$个系数构成一个$t-1$次多项式，然后将多项式在n个不同点上的取值作为分片分发至n个边缘服务器。任意少于t个分片无法恢复多项式，从而无法获得秘密。其分布式存储与恢复流程如下：

a.分片生成

将更新后的模型$\triangle\tilde{W}_i$的每个元素独立作为秘密，生成多个分片。设阈值参数为$k_{\min}$，即至少需要$k_{\min}$个分片才能恢复数据。总边缘服务器数量为n，对于第k个元素s_k，其中$k=1,2,...,k_{\max}$。构造一个$k_{\min}-1$随机多项式：

$$f_k(x)=s_k+a_1x_1+a_2x_2^2+\cdots+a_{k_{\min}-1}x_n^{k_{\min}-1}, \tag{14}$$

其中，$a_1,a_2,...,a_{k_{\min}-1}$为从有限域中随机选择的系数，使多项式的取值呈现无规律分布，攻击者无法通过少量分片反推多项式形式及原始秘密，用来保证秘密的随机性与不可预测性。$x_1...x_n$表示边缘服务器对应的节点编号，是基于有限域的非零互异元素。对于任何少于$k_{\min}$个不同的点都无法确定唯一的多项式$f_k(x_j)$，因此无法得到常数项s_k。

b.分片分发

计算n个分片:

$$\{(x_j,f_k(x_j))\}_{j=1}^{n}=\{(x_j,f_1(x_j)),(x_j,f_2(x_j)),...,(x_j,f_k(x_j))\}. \tag{15}$$

该式表示为将各分片$(x_j,f(x_j))$发送至对应的边缘服务器j。

c.分片重构

当中央聚合器需要恢复原始模型更新以进行聚合时，需向边缘服务器请求分片。设收集到的有效分片索引集合为J，且$|J|\geq k_{\min}$。当收集到至少$k_{\min}$个分片时，通过拉格朗日插值可恢复原始秘密：

$$f_k(x)=\sum_{j\in J}f_k(x_j)\cdot\ell_j(x), \tag{16}$$

其中，$\ell_j(x)$为拉格朗日多项式，表示为：

$$\ell_j(x)=\prod_{\substack{j'\in J\\ j'\neq j}}\frac{x-x_{j'}}{x_i-x_{j'}}. \tag{17}$$

则 s_k 即为多项式在 $x=0$ 处的值：

$$s_k=f_k(0)=\sum_{j\in J}f_k(x_j)\cdot\ell_j(0). \tag{18}$$

在此重构过程中，只要收集到至少 $k_{\min}$ 个正确的分片，拉格朗日插值就能唯一确定原始多项式，进而恢复出 s_k。所以当部分边缘服务器不可用或被攻击者控制，只要仍有足够数量的服务器正常运作，系统就能正确恢复模型更新并进行后续的聚合操作。

2.3 基于主观逻辑的信誉机制

为应对本地训练中可能存在的恶意节点或低质量数据问题，本节引入基于主观逻辑的信誉机制，建立一套动态的信誉评估体系。通过量化评估节点的行为与贡献，动态调整其可信度，从而提升系统的整体安全性与鲁棒性。

主观逻辑表达对节点可靠性的信任度原理如下：设意见三元组 $\omega_i=\left(b_i,h_i,u_i\right)$，$b_i$、$h_i$ 和 u_i 分别表示节点可靠、节点不可靠和不确定性，满足 $b_i+h_i+u_i=1,\quad b_i,h_i,u_i\in[0,1]$。节点的信誉值 R_i 由意见的期望值计算：

$$R_i=E\left(\omega_i\right)=b_i+\partial_i\cdot u_i, \tag{19}$$

其中，∂_i 是先验基率，表示在完全不确定时的默认可靠概率，∂_i 的取值反映了系统对未知节点的先验信任程度，通常设为 0.5 以体现无偏性，即在不掌握任何历史信息时，认为节点可靠与不可靠的可能性相等。在系统初始阶段，由于缺乏历史交互信息，所有节点被赋予相同的初始意见，$\omega_i^{(0)}=\left(b_i^{(0)},h_i^{(0)},u_i^{(0)}\right)=(0,0,1)$，这一初始化表示完全不确定状态，其中节点可靠、节点不可靠值均为 0，不确定性为 1。因此初始信誉值为 $E\left(\omega_i^{(0)}\right)=0.5$。

在评估过程中，系统通过多维度评估来动态更新节点的信誉。这种多维评估能够全面反映节点的行为特征，提高恶意节点检测的准确性。信誉更新基于多次交互观察，交互历史 $H_i=\left(r_i^1,r_i^2,r_i^3\right)$ 是记录节点成功和失败的次数，其中 r_i^1 为成功贡献有效更新的次数，r_i^2 为贡献无效或恶意更新的次数，r_i^3 为记录因节点行为导致的安全距离预警或通信超时事件。该交互历史反映了节点过去行为对系统稳定性与安全性的影响，并作为信誉值的长期累积记忆。在 CBTC 入侵检测场景中，由于误报可能导致列车不必要的紧急制动，严重影响运营效率，因此设置数据质量 $A_i=\alpha_1\cdot P_i+\alpha_2\cdot T_{delay}$，其中 $\alpha_1+\alpha_2=1$，P_i 为测试节点的数据精确度。$T_{delay}=\exp\left(-\frac{T_{all}}{T_{\max}}\right)$为时间延迟的评分。

接着要衡量节点行为与群体共识的偏差，设置行为一致性 $C_i=\frac{1}{N-1}\sum_{i'\neq i}\mathrm{sim}\left(\Delta W_i,\Delta W_{i'}\right)$，该式子用来计算该节点更新与其他大多数节点更新的平均相似度，来检测潜在的异常行为。其中 $\mathrm{sim}\left(\mathbf{a},\mathbf{b}\right)=\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$，其中 $i\neq i'$，$\triangle W_{i'}$ 为其他列车的本地更新模型。

在定义了信誉评估的维度后，在每一轮训练交互即模型下发、节点本地训练与参数上传的完整过程结束后，边缘服务器根据节点贡献与行为表现动态更新其信誉值。

节点 i 在第 k' 轮交互中的表现，通过其在数据质量 Q_i 和行为一致性 C_i 上的得分，并结合交互是否成功完成，得到一个综合评分：

$$F_i^{k'} = \lambda_1 \cdot A_i + \lambda_2 \cdot C_i + \lambda_3 \cdot I_{\text{success}}, \tag{20}$$

其中，$\lambda_1 + \lambda_2 + \lambda_3 = 1$，$I_{\text{success}}$ 为本次交互成功指示函数，成功为 1，失败为 0。

一个节点 i 的信任状态由一个信息对 $\left(z_i, \bar{z}_i\right)$ 表示，其中 z_i 表示正向信息，包括行为一致、节点可信或交互成功，$z_i^{k'} = G\left(r_i^1\right)$。其中 $G(\cdot)$ 为历史聚合函数，用于对历史统计信息进行归一化映射，以生成可用于信誉更新与决策分析的证据进行量化处理。$\bar{z}_i^{k'}$ 是负向信息，包含行为不一致、表明节点不可信或交互失败，$\bar{z}_i^{k'} = G\left(r_i^2, r_i^3\right)$。同时为增强稳定性，引入遗忘因子 $\hbar \in [0,1]$ 对历史证据进行指数衰减，则当前信息对为：

$$z_i^{k'} = G\left(r_i^1\right) = (1-\hbar)\sum_1^{k'} \hbar^{k'} \frac{r_i^1}{r_i^1 + 1}, \tag{21}$$

$$\bar{z}_i^{k'} = G\left(r_i^2, r_i^3\right) = (1-\hbar)\sum_1^{k'} \hbar^{k'} \left(\alpha_1 \frac{r_i^2}{r_i^2 + 1} + \alpha_2 \frac{r_i^3}{r_i^3 + 1}\right), \tag{22}$$

则更新后的信息对 $\left(z_i, \bar{z}_i\right)$ 为：

$$z_i^{k'+1} = z_i^{k'} + F_i^{k'}, \quad \bar{z}^{k'+1} = \bar{z}_i^{k'} + \left(1 - F_i^{k'}\right), \tag{23}$$

基于更新后的证据，计算节点新的主观意见三元组 $\omega_i^{k'+1} = \left(b_i^{k'+1}, h_i^{k'+1}, u_i^{k'+1}\right)$：

$$b_i^{k'+1} = \frac{z_i^{k'+1} + \alpha \cdot A_i}{z_i^{k'+1} + \bar{z}_i^{k'+1} + \alpha + \beta}, \tag{24}$$

$$h_i^{(\mathbb{k}+1)} = \frac{\bar{z}_i^{(\mathbb{k}+1)} + \beta \cdot (1 - C_i)}{z_i^{(\mathbb{k}+1)} + \bar{z}_i^{(\mathbb{k}+1)} + \alpha + \beta}, \tag{25}$$

$$u_i^{(\mathbb{k}+1)} = \frac{\alpha + \beta}{z_i^{(\mathbb{k}+1)} + \bar{z}_i^{(\mathbb{k}+1)} + \alpha + \beta}, \tag{26}$$

其中，α, β 为平滑超参数，控制信誉更新的敏感度。较大的 α, β 值使得信誉变化更为平滑，增强系统稳定性；较小的值使得系统对节点行为变化更为敏感。

最终，节点的信誉值 $R_i^{k'+1}$ 由其主观意见的期望值决定，反映了对该节点可靠性的综合量化评估：

$$R_i^{k'+1} = E\left(\omega_i^{k'+1}\right) = b_i^{k'+1} + \partial_i \cdot u_i^{k'+1}. \tag{27}$$

为平滑信誉变化，防止突变，采用指数移动平均进行平滑处理。定义本轮的信誉增量为 $\Delta R_i = R_i^{k'+1} - R_i^{k'}$，引入衰减因子 $\eta \in [0,1]$，η 控制了历史信誉的权重，η 越大，信誉变化越平稳。则信誉更新后的信誉值为：

$$R_i^{k'+1} \leftarrow \eta \cdot R_i^{k'} + (1-\eta) \cdot \left(R_i^{k'} + \Delta R_i\right) = \eta \cdot R_i^{k'} + (1-\eta) \cdot \Delta R_i. \tag{28}$$

2.4 鲁棒加权聚合算法

在节点完成本地模型的隐私保护训练与信誉评估后，为安全、高效地聚合这些分布式模型更新，保障全局模型性能与鲁棒性。本文设计了一种两阶段过滤的鲁棒加权聚合算法。该算法以节点的动态信誉值为

核心依据，并引入模型更新方向一致性检测，从而在恶意攻击环境下仍能保证全局模型的正确收敛。

首先根据节点信誉值进行初步筛选，快速排除显著不可信的节点，缩小可靠节点的范围。设第k'轮联邦学习中有N个节点参与，基于信誉阈值$\theta_{\max}$对参与节点进行初步筛选，节点集合$S_{\max}$定义为：

$$S_{\max}=\{i|\ R_i^{k'}\geq\theta_{\max},i=1,2,...,N\}. \tag{29}$$

通过初步筛选的节点，其模型更新仍可能存在与全局优化方向不符的细微偏差。因此需要计算节点i的模型更新$\Delta\tilde{W}_i$与候选集$S_{\max}$中其他节点更新的平均余弦相似度，以评估其方向一致性：

$$\mathrm{sim}_i=\frac{1}{|S_{\max}|-1}\sum_{i'\in S_{\max},i'\neq i}\frac{\Delta\tilde{W}_i\cdot\Delta\tilde{W}_{i'}}{|\Delta\tilde{W}_i|\cdot|\Delta\tilde{W}_{i'}|}. \tag{30}$$

设定方向一致性阈值θ_{low}。若$\mathrm{sim}_i<\theta_{\mathrm{low}}$，则判定节点$i$为异常节点，并将其从最终聚合集合中删除，最后得到可信任节点集合：

$$S_{\mathrm{turst}}=\{i|\ sim_i\geq\theta_{\mathrm{low}},i=1,2,...,N\}. \tag{31}$$

综合上述两次筛选，最终的全局模型更新$\Delta W_{\mathrm{global}}$由可信任节点集$S_{trust}$中节点的模型更新。同时根据其信誉值加权计算得出：

$$\Delta W_{\mathrm{global}}=\frac{\sum_{i\in S_{\mathrm{trust}}}R_i^{k'}\cdot\Delta\tilde{W}_i}{\sum_{i\in S_{\mathrm{trust}}}R_i^{k'}}. \tag{32}$$

2.5 区块链共识与可信分发流程

为了确保整个系统的可信性，本文设计了基于授权拜占庭容错算法（Delegated Byzantine Fault Tolerance, dBFT）的区块链层，为系统提供不可篡改、可追溯、可验证的可信分发服务。区块链共识层作为系统的信任基石，不仅负责安全地聚合和分发模型参数，还通过分布式账本技术记录完整的训练历史，为系统审计和故障追溯提供可靠依据。为了适应 CBTC 系统需求的区块链网络，本文设计了分层式的节点架构，根据其功能和权限将节点分为三个层次：

a.验证节点$\mathcal{C}=\{c_1,c_2,...,c_O\}$：负责执行 dBFT 共识算法，验证交易有效性并参与新区块的生成。节点验证从边缘节点中选举产生，其数量满足$P=3p_{\max}+1$，其中$p_{\max}$为系统最大容错节点数。

b.协作节点$\mathcal{W}=\{w_1,w_2,...,w_Q\}$：负责收集终端设备的模型更新分片，执行初步验证，并作为验证节点与终端设备之间的通信桥梁。

c.参与节点$\mathcal{G}=\{g_1,g_2,...,g_R\}$：即参与联邦学习的 CBTC 终端设备，负责执行本地训练并提交模型更新。

dBFT 共识算法为确保系统在存在恶意节点时的安全运行，要求参与共识的各类节点数量满足$|\mathcal{C}|\leq|\mathcal{W}|\leq|\mathcal{G}|$。上述数量约束不仅提高了共识成功率，同时也确保了系统的可扩展性。基于此约束，dBFT 共识流程在每个阶段都设计了相应的验证机制和超时处理，以适应 CBTC 系统的实时性要求。

在满足节点数量约束的前提下，共识过程由多个节点协同完成。首先，参与节点p_i将其在当前轮次中经过差分隐私和秘密共享处理的模型更新分片的交易记录发送至连接的协作节点。该交易记录定义为：

$$\mathcal{TX}_i=\left\{\mathcal{H}\left(\Delta\tilde{W}_i\right),t_{\mathrm{send}},\sigma_i,s_k\right\}, \tag{33}$$

其中，$\mathcal{H}(\cdot)$为哈希函数，用于生成模型更新的唯一摘要。t_{send}为时间戳，σ_i为数字签名。协作节点对交易的合法性进行初步验证后，会将有效交易广播至验证节点，并进入 dBFT 共识流程。在共识阶段，验证

节点根据预先设定的轮换规则 $c_m = c_{(r \bmod n)+1}$ 选择新区块提案节点，其中 r 为共识轮次，n 为验证节点的总数。当本轮收集到的所有交易均被验证后，生成新区块提案：

$$\mathcal{B}' = \left\{ \mathcal{H}(\mathcal{B}), \{\mathcal{TX}_i\}_{i=1}^{N}, r, c_m \right\}, \tag{34}$$

其中，$\mathcal{B}_n$ 表示已确认的区块，$\{\mathcal{TX}_i\}_{i=1}^{N}$ 表示当前轮次中收集到的所有有效交易集合。该区块提案将广播至所有验证节点来完成区块确认流程并最终达成一致性。

节点之间的广播过程如图 3 所示。完整交互过程可分为请求、预准备、准备、提交和回复五个阶段。主要流程如下：

请求阶段：客户端向主节点提交区块生成请求，主节点在接收并验证请求后启动共识流程。

预准备阶段：主节点对所生成的提案区块进行签名，通过广播预准备消息将区块提案发送至所有验证节点。

准备阶段：各验证节点独立校验提案区块的有效性与合法性。验证通过后，节点向其他验证节点广播准备消息，以表达对该提案的一致性认可。在此阶段，各节点仅传播区块摘要及签名而非完整区块内容，从而有效降低网络通信开销。当某验证节点累计收到超过 $2f+1$ 条有效准备消息后，表明该提案已获得多数验证节点支持。

提交阶段：节点广播提交消息，并在收到超过 $3p_{\max}+1$ 有效提交消息后，确认该提案区块已达成全局共识，可安全地写入本地账本。

回复阶段：验证节点向客户端返回执行结果。客户端在接收到超过 $3p_{\max}+1$ 条一致回复后，即可确认其提交的交易或请求已被系统成功处理，从而完成整个 dBFT 共识流程。

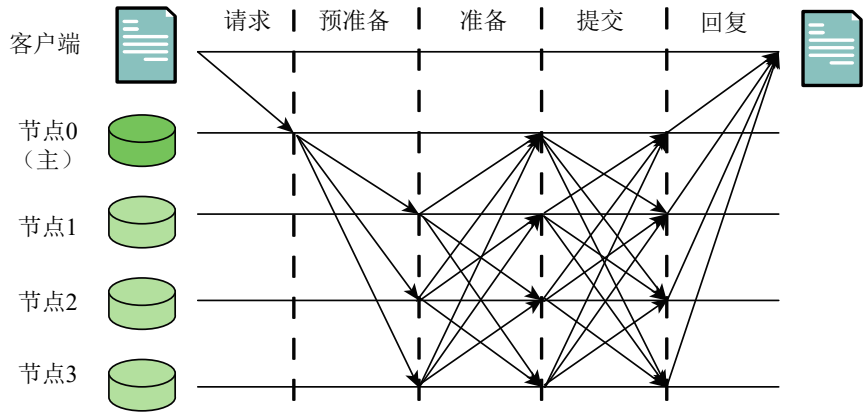

图 3 dBFT 共识算法流程图

Fig.3 **dBFT Consensus Algorithm Flowchart**

完成共识流程后，更新全局状态：

$$\mathcal{S}_{\text{global}}^{(r+1)} = \left\{ W_{\text{global}}^{(r+1)}, \{\rho_i^{(r+1)}\}_{i=1}^{N}, \mathcal{H}(\mathcal{B}), \mathrm{t}_{\text{com}} \right\}, \tag{35}$$

其中，t_{com} 为提交确认时间戳，节点最终确认区块提交的时间点。$\{\rho_i^{(r+1)}\}_{i=1}^{N}$ 所有节点的本地参数集合。更新后的状态通过协作节点分发给所有参与节点，开启新一轮联邦学习训练。

为了满足 CBTC 系统的实时性约束，dBFT 共识的完成时间上界表示为：

$$T_{\text{cc}} \le T_{\text{com}} - T_{\text{send}}. \tag{36}$$

同时根据 CBTC 的实时性要求，所有过程的时延要满足最大允许的传输时延，即 $T_{\text{round}} = T_{\text{train}} + T_{\text{agg}} + T_{\text{cc}} \le T_{\max}$，其中 T_{agg} 为聚合过程的时间。

3. 实验与评估

为验证本文提出的基于信誉权重和可信分发的城市轨道交通入侵检测方法的有效性与优越性，本节基于公开 CBTCset 数据集搭建仿真实验平台，明确实验设置，从检测性能和实时性两个维度与主流基线方法对比，并通过量化分析验证方法优势。

3.1 实验设置

本文选用公开数据集 CBTCset 仿真实验。CBTCset 是一个专为 CBTC 系统中网络异常行为攻击检测设计的公开参考数据集，由 TSMset 和 WOMset 两个子集构成，其中 TSMset 包含 1,158,951 条列车状态消息，记录列车发送给地面单元的列车状态消息；WOMset 包含 1,156,272 条轨旁操作消息，包含地面单元发送给列车的轨旁操作信息。该数据集通过 TrainSec 仿真框架生成，模拟了真实 CBTC 通信场景，包含五种攻击类型，如表 1 所示，其中 TSM Delay 为列车通过添加随机延迟来延迟发送列车状态消息；WOM Delay 为地面单元通过添加随机延迟来延迟发送轨旁操作消息；False Position 为列车通过在其实际位置上添加随机偏移量来发送错误位置信息；False Speed 为列车通过在其实际速度上添加随机偏移量来发送假速度信息；False LMA 为地面单元向轨旁操作信息中的移动权限限制值添加随机偏移量[21]。

表 2 攻击类型

Tab.2 Attacks Type

攻击	对象
为列车添加随机延迟(TSM Delay)	列车
为地面单元添加随机延迟(WOM Delay)	地面单元
在列车实际位置上添加随机偏移量(False Position)	列车
在列车实际速度上添加随机偏移量(False Speed)	列车
为地面单元向移动权限限制值添加随机偏移量(False LMA)	地面单元

本文基于以上数据进行仿真，并与平均加权聚合方法和隐私增强加权聚合算法进行了对比。其中，平均加权聚合方法是经典联邦学习聚合方法，采用迭代式的本地训练与服务器端聚合机制，其主要流程包括服务器端全局模型初始化、客户端本地更新、参数上传以及加权聚合等四个关键阶段。隐私增强加权聚合算法在模型更新过程中引入高斯噪声并结合梯度裁剪技术，以控制敏感度并实现差分隐私保护。其隐私保护框架涵盖噪声注入、隐私预算管理与隐私会计等核心模块，通过本地差分隐私处理确保单个数据样本不会显著影响输出模型。但仍然采用固定加权聚合，无动态信誉评估与鲁棒过滤机制。

实验核心参数配置如表 3 所示。

表 3 实验参数配置

Tab.3 experimental parameter configuration

参数名称	符号	取值
边缘服务器数量	n	7
重构阈值	$k_{\min}$	5
信誉筛选阈值	$\theta_{\max}$	0.8
验证节点数	P	4
可容忍节点数	$p_{\max}$	1

结合 CBTC 系统入侵检测高检测精度、强实时性的核心需求，选取准确率、精确度、F1 和时延作为评估指标，覆盖检测性能和系统实时性两大维度。

3.2 实验结果与分析

图 4 展示了在不同数据量下所提出的协同入侵检测方法的准确率变化趋势。当数据规模一样时，随着轮次的增加，准确率也在提升。这是由于在检测过程中根据信誉权重筛选出了恶意节点，因此在后续的训

练过程中这些恶意节点对模型无法造成有效的威胁。同时当轮次一样时，随着数据规模的增加，精确度不断上升并趋于平稳。数据量越小的曲线，其收敛速度越慢。这是因为当训练数据规模较小时，模型对特征空间的覆盖不足，无法判定攻击行为。随着数据量持续增加，模型收集更多攻击样本与正常样本，学习更稳定、泛化更强的判别特征，从而可以降低误判并显著提升精确度。总体而言，该变化趋势是由于信誉加权聚合策略使得高信誉节点的模型更新在全局模型中占据更大权重，从而在大规模数据条件下显著提升全局模型性能。同时，区块链可信分发的安全同步机制保证了模型更新的完整性与一致性，使得系统在不同参与规模下仍具有收敛能力。其中，在数据量达到 110 万且轮次最高时的精确度最高可达到 99.6%。体现了本文所提出的基于差分隐私的本地训练机制、基于信誉加权聚合策略和共识算法的有效性。

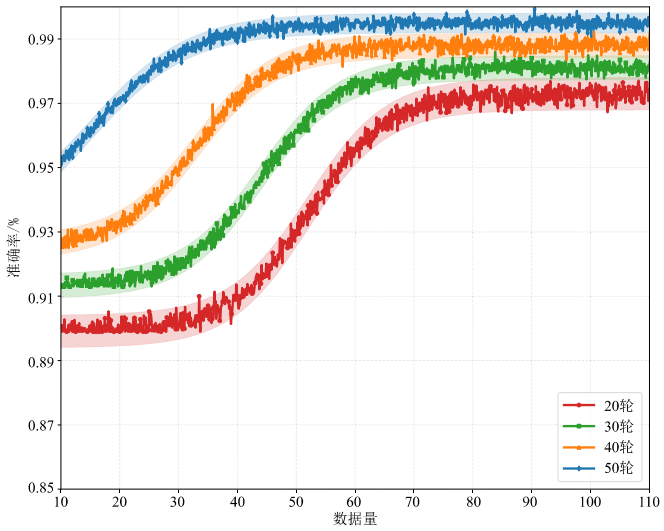

图 4 准确率随数据量和轮次变化趋势

Fig.4 Accuracy trends with data volume and epochs

图 5 为多方案的性能对比结果，分别对比了仅差分隐私和平均加权聚合、差分隐私和区块链、差分隐私和信誉加权三种方案。从图中可以看出本文的方法在四项核心检测指标上均优于其他方案。其中无信誉机制、无区块链的方案，即差分隐私和平均加权聚合方法，因缺乏恶意节点防御机制易受模型污染和因差分隐私噪声注入引入额外方差，整体检测性能处于较低水平；单一引入区块链可信分发或动态信誉机制的方案，检测性能均有明显提升，其中动态信誉机制通过恶意节点过滤与鲁棒加权聚合，成为检测性能提升的核心驱动力，其性能增益显著高于单一区块链模块，而区块链则通过解决模型更新传输篡改、单点故障等问题，为检测性能提升提供了安全支撑。本文方法性能优势并非源于单一技术模块的简单叠加，而是实现了差分隐私和秘密共享的双层隐私保护、动态信誉机制的鲁棒聚合、区块链可信分发的深度协同，构建起 “隐私保护-鲁棒聚合-可信分发” 的有机闭环，让高信誉节点的有效模型更新得以安全、高效地同步与聚合，既提高了准确率和精确度，又实现了低误报、高召回的检测效果。

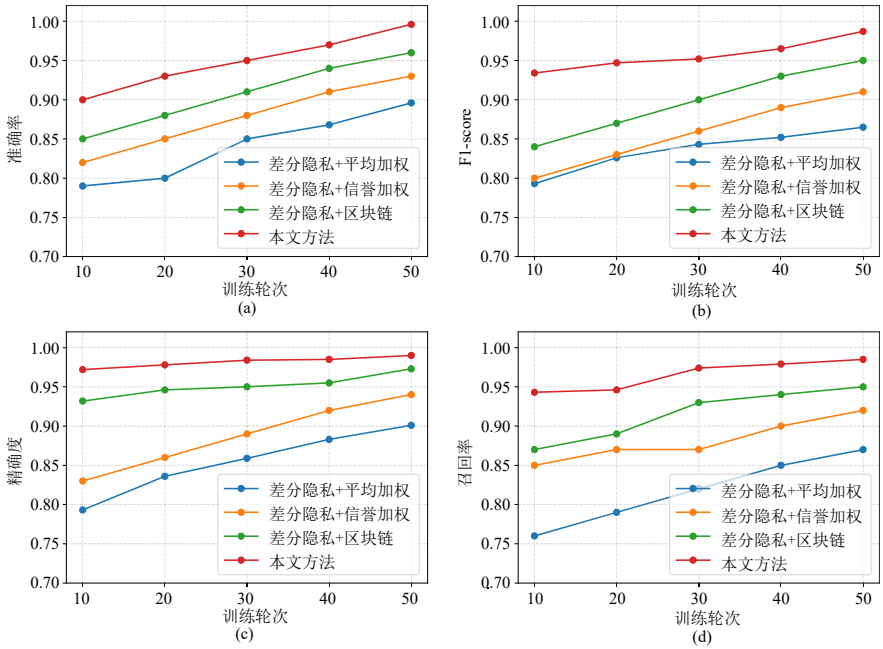

图 5 不同方案下的性能对比：（a）准确率，（b）F1-score，（c）精确度，（d）召回率

Fig. 5 Performance comparison under different schemes: (a) Accuracy, (b) F1-score, (c) Precision, (d) Recall

图 6 展示了平均加权聚合、隐私增强加权聚合以及本文提出的基于信誉权重与可信分发的聚合方法在准确率、F1 和精确度方面的性能比较结果。仿真结果表明，本文提出的基于信誉权重与可信分发的聚合在三个方面的性能均优于其他两种方法。其中准确率 92%、F197.4%、精确度 99.1%；而平均加权聚合仅为 93.6%、95.7%、98.7%，隐私增强加权聚合为 91.0%、96.8%、97.3%，本文方法在三项关键指标上均高于传统方法。这是因为平均加权聚合方法缺乏安全防护机制，容易受到恶意节点攻击，从而导致通信中断或模型污染。隐私增强加权聚合算法由于噪声注入导致额外方差，从而限制了模型的性能上限，且对动态攻击的防御能力不足。而本文所用方法引入动态信誉评估机制，有效衡量节点贡献度并优先聚合高信誉节点的更新结果。同时结合区块链技术构建可信分发，来保障模型参数在传输过程中的安全性与可追溯性。再根据信誉权重依据主观逻辑模型计算，通过节点的历史交互数据评估其可靠性，从而过滤低质量或潜在恶意的更新信息。最后运用可信分发进一步确保了聚合过程的完整性与一致性，实现了隐私保护、安全防御与模型性能提升的动态结合。

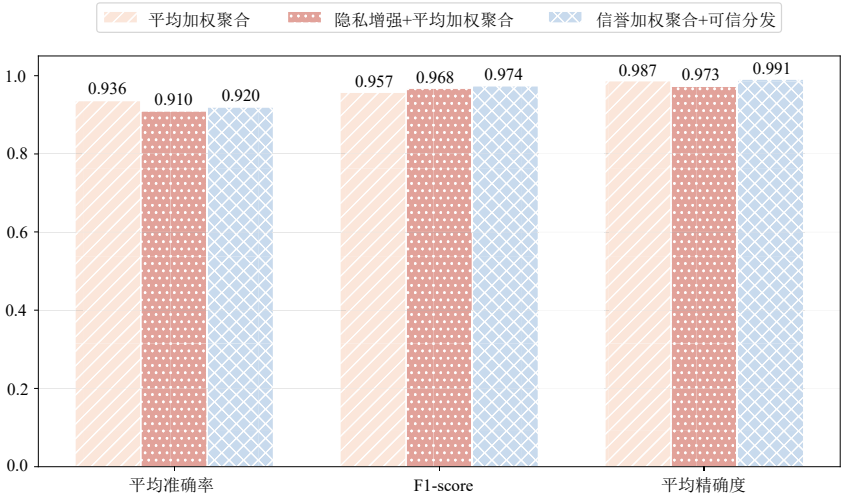

图 6 准确率、F1-score 和精确度对比

Fig.6 Performance Comparison of Accuracy, F1-Score, and Precision

图 7 展示了平均加权聚合、隐私增强加权聚合以及本文提出的基于信誉权重与可信分发的聚合方法在不同训练轮次下的时延变化趋势。仿真结果表明，本文提出的基于信誉权重与可信分发的聚合方法在训练

过程中表现出时延持续下降的趋势，最高轮次时延仅 0.05s，较初始时延降低 84%。其主要原因是信誉权重机制对节点可靠性的动态评估，使系统能够优先选择高信誉节点的更新结果，从而有效减少无效通信。可信分发在聚合过程中通过共识优化与安全传输机制的改进，降低了模型更新的重传风险与共识开销。对于平均加权聚合方法，其初始时延较低，仅需 0.11s，但随着训练轮次的增加呈现波动上升的趋势，最高轮次时延为 0.43s。这主要是由于该方法采用简单的平均聚合机制，缺乏安全优化与鲁棒性设计，在面对攻击者不断变化的攻击模式时难以准确识别其行为意图，导致模型收敛过程不稳定。为保证模型稳定性，系统在后续轮次中需进行重新训练与策略调整，从而显著增加通信时延与计算开销。相比之下，隐私增强加权聚合的时延表现较为平稳，但整体呈现逐步上升的趋势，初始时延 0.14s，最终轮次时延增至 0.38s，增幅达 171%。其原因在于差分隐私机制的引入增加了计算复杂度。在训练轮次不断增加的情况下，系统需采用更严格的梯度裁剪与更精细的噪声校准策略，以维持隐私预算的合理分配。同时，节点本地资源消耗逐渐增加，且缺乏对节点可信性的快速判断机制，进一步导致通信时间延长。

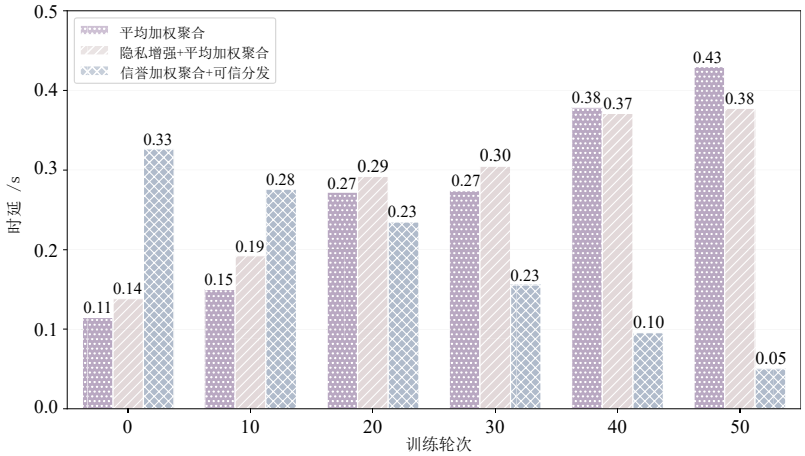

图 7 不同训练轮次下的时延变化

Fig.7 Time Delay Variation Across Different Training Rounds

4. 结论

本研究提出了一种基于信誉权重和可信分发的协同入侵检测方法，针对当前列车控制系统通信过程中面临的网络安全威胁进行检测。得出以下结论：

1) 该方法融合了联邦学习的高效分布式特性、信誉机制的动态可信评估能力以及区块链技术的去中心化与不可篡改特征，提升了系统在隐私保护、安全防御及鲁棒性方面的综合性能。
2) 结合 CBTCset 数据集开展了仿真实验。实验结果表明，该方法能够有效识别并抑制潜在的恶意节点攻击行为，从而提升列车通信系统的安全性与稳定性，同时满足列车的实时性要求。

未来可以探索轻量化区块链结构来增强系统在边缘计算环境中的部署效率。

参考文献

[1] ZHANG Q, ZHANG C, ZHAO J, WANG D, and XU W. Dynamic resource allocation for multi-access edge computing in urban rail transit[J]. IEEE Transactions on Vehicular Technology, 2025, 74(2): 3296–3310.

[2] 赵军辉, 张丹阳, 贺林. 智慧城轨交通通信技术的分析与展望[J]. 电信科学, 2024, 37(4): 1-13.
ZHAO J H, ZhANG D Y, HE L. Analysis and prospect of communication technology in smart urban rail[J]. Telecommunications Science, 2024, 37(4): 1-13.

[3] LAKSHMINARAYANA S, KARACHIWALA J S, CHANG S Y, et al. Signal jamming attacks against communication-based train control: Attack impact and countermeasure[C]//Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks. 2018: 160-171.

[4] 周天清, 王博博. UDN 中面向安全卸载的多目标联合优化算法研究[J]. 华东交通大学学报, 2024, 41(01): 70-77.
ZHOU T Q, WANG B B. Research on multi-objective joint optimization algorithm for secure offloading in UDN[J]. Journal of East China Jiaotong University, 2024, 41(01): 70-77.

[5] WANG X, LIU L, ZHU L, et al. Joint security and QoS provisioning in train-centric CBTC systems under sybil attacks[J]. IEEE Access, 2019, 7: 91169-91182.
[6] 沈华, 田晨, 郭森森, 等. 基于对抗性机器学习的网络入侵检测方法研究[J]. 信息网络安全, 2023, 23(08): 66-75.
SHEN H, TIAN C, GUO S S, et al. Research on adversarial machine learning-based network intrusion detection method[J]. Netinfo Security, 2023, 23(08): 66-75.
[7] 杨晓文, 张健, 况立群, 等. 融合 CNN-BiGRU 和注意力机制的网络入侵检测模型[J]. 信息安全研究, 2024, 10(03): 202-208.
YANG X W, ZHANG J, KUANG L Q, et al. A network intrusion detection model integrating CNN-BiGRU and attention mechanism[J]. Journal of Information Security Research, 2024, 10(03): 202-208.
[8] YUAN H, WANG S, BI J, et al. Hybrid and spatiotemporal detection of cyberattack network traffic in cloud data centers[J]. IEEE Internet of Things Journal, 2024, 11(10): 18035-18046.
[9] GAO B, BU B, ZHANG W, et al. An intrusion detection method based on machine learning and state observer for train-ground communication systems[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7): 6608-6620.
[10] 肖雄, 唐卓, 肖斌, 等. 联邦学习的隐私保护与安全防御研究综述[J]. 计算机学报, 2023, 46(05): 1019-1044.
XIAO X, TANG Z, XIAO B, et al. A survey on privacy and security issues in federated learning[J]. Chinese Journal of Computers, 2023, 46(05): 1019-1044.
[11] CUI L, QU Y, XIE G, et al. Security and privacy-enhanced federated learning for anomaly detection in IoT infrastructures[J]. IEEE Transactions on Industrial Informatics, 2021, 18(5): 3492-3500.
[12] 李冬芬, 向秋雨, 胡志康, 等. 融合联邦学习与量子卷积神经网络的入侵检测模型[J]. 计算机研究与发展, 2025, 62(10): 2512-2522.
LI D F, XIANG Q Y, HU Z K, et al. An intrusion detection model integrating federated learning and quantum convolutional Neural Networks[J]. Journal of Computer Research and Development, 2025, 62(10): 2512-2522.
[13] 尹春勇, 王珊. 基于联邦学习和注意力机制的物联网入侵检测模型[J]. 信息安全研究, 2025, 11(09): 788-796.
YIN C Y, WANG S. Internet of things intrusion detection model dased on federated learning and attention mechanisms[J]. Netinfo Security, 2025, 11(09):788-796.
[14] LIANG H, ZHU L, YU F R, et al. A cross-layer defense method for blockchain empowered CBTC systems against data tampering attacks[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 24(1): 501-515.
[15] DAS D, BANERJEE S, CHATTERJEE P, et al. A secure blockchain enabled v2v communication system using smart contracts[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 24(4): 4651-4660.
[16] ZHU L, LIANG H, WANG H, et al. Joint security and train control design in blockchain-empowered CBTC system[J]. IEEE Internet of Things Journal, 2021, 9(11): 8119-8129.
[17] ABOU El H Z, MOUDOUD H, BRIK B, et al. Blockchain-enabled federated learning for enhanced collaborative intrusion detection in vehicular edge computing[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(7): 7661-7672.
[18] LIU H, ZHANG S, ZHANG P, et al. Blockchain and federated learning for collaborative intrusion detection in vehicular edge computing[J]. IEEE Transactions on Vehicular Technology, 2021, 70(6): 6073-6084.
[19] El S A I, ABDELAZIZ M, HUSSEIN M, et al. DDoS mitigation in IoT using machine learning and blockchain integration[J]. IEEE Networking Letters, 2024, 6(2): 152-155.
[20] ABDEL-B M, MOUSTAFA N, HAWASH H, et al. Federated intrusion detection in blockchain-based smart transportation systems[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(3): 2523-2537.
[21] FAKHERELDINE A, ZULKERNINE M, MURDOCK D. Cbtcset: a reference dataset for detecting misbehavior attacks in cbtc networks[C]//2023 IEEE 34th international symposium on software reliability engineering workshops (ISSREW). IEEE, 2023: 57-62.

第一作者：尹雨晴（1999—），女，硕士研究生，研究方向为城市轨道交通信息安全。E-mail: tiko0104@163.com

通信作者：赵军辉（1973—），男，北京交通大学电子与信息工程学院教授，研究方向包括无线和移动的通信及相关应用。E-mail: junhuizhao@hotmail.com.