2022 国际计算机、电力与通信会议(ICCPC) | 979-8-3503-9784-0/22/$31.00 ©2022 IEEE | DOI: DOI: IEEE10.1109 / ICCPC55978

使用区块链的安全增强联邦学习方法

S. reathy S. Sathya Priya

计算机科学系;计算机科学系工程学院;印度斯坦理工学院，印度斯坦理工学院，金奈。 钦奈。

srevathy@hindustanuniv.ac.in sathyap@hindustanuniv.ac.in|0>>>>

摘要在传统的机器学习方法中，从所有边缘设备收集的数据被发送到集中的服务器进行训练和预测输出。在集中式方式中，用户在将自己的数据共享给集中式服务器时，不得不在数据的隐私性和完整性上做出妥协。为了克服这一问题，引入了联邦机器学习方法，其中模型和数据分散，机器学习模型将在本地设备的数据上进行训练，参数将发送到云服务器进行共识更改，增强用户的数据隐私性。但是，在联邦机器学习中，节点到云服务器的身份验证和云服务器的身份验证仍然是一个需要解决的主要问题，因为恶意节点可以冒充身份验证的节点并与云服务器通信。在提出的模型中，节点身份验证使用基于以太坊的区块链和智能合约实现，从而增强了联邦机器学习方法的安全性。测量了节点认证的效率，并与机器学习算法进行了比较，准确率达到 99%。

关键词:区块链，集中式服务器，以太坊，联邦学习，安全，智能合约。

我的介绍。

技术的进步在提高生活水平方面起着至关重要的作用。技术进步要求开发人员和行业通过融合机器学习、边缘计算和区块链等最新技术来创新新的应用程序。各种设备的先进连接能力、低成本和小尺寸特性促使开发人员发明具有先进功能的应用程序。随着这些应用程序使用的增加，对这些设备生成的数据的隐私和机密性的担忧也在增加。

采用最新技术开发的智能应用[9]需要大量的用户数据进行培训和部署。这些应用程序大多将收集到的数据共享给集中服务器，集中服务器用于对多台设备收集到的数据进行聚合和分析，为最终用户提供个性化的服务。

979-8-3503-9784 /22/$31.00ⓧc 2022 ieee 50

因为它们不是存储在集中存储位置。联邦学习[2]还减轻了集中式服务器攻击和常见服务器故障问题。虽然这种方法有很多优点，但也存在主动攻击的可能性，比如 sybil 攻击，恶意节点携带已认证节点的身份，使得与它们通信的所有其他节点都容易受到攻击。

为了解决这个问题，为联邦机器学习引入了基于区块链的节点身份验证，其中服务器和节点通信使用智能合约进行身份验证。该系统使用以太坊区块链开发，确保了网络中的信任和可靠性，因为区块链中的交易是不可变和不可更改的[5]，从而增强了联邦机器学习方法的安全性。

云服务提供商将数据商业化给第三方应用程序，第三方应用程序需要通过出售其客户的数据向最终用户宣传其产品，或者通过应用程序编程接口在有限的访问权限下共享这些数据，这很容易受到数据推断攻击。

众所周知，推理攻击是一种数据挖掘技术，攻击者可以在不直接访问用户的情况下推断出有关用户的更敏感信息。它对数据进行分析，非法获取个人用户或数据库的信息，从而降低了数据的隐私性。这些设备收集的数据本质上是敏感的，如个人或行为信息[15]，包括地理信息、年龄、性别、健康信息和触摸屏输入，因此在保证用户隐私的前提下处理敏感数据至关重要。攻击者或入侵者的主要目标是通过发起主动和被动攻击来控制集中式云服务器并中断服务。

主动攻击是指直接引起的攻击
对网络资源的破坏。被动攻击是指间接窃听网络并收集流量信息，从而导致主动攻击的攻击。 一些活跃的攻击是

1. *DDOS 攻击*——恶意节点向通过认证的节点不断发送请求，使其资源无法被其他节点使用。
2. *Sybil 攻击*——恶意节点携带被认证节点的身份[10]，破坏资源。3.*蠕虫洞攻击*-在这种攻击中，攻击者将恶意节点放置在网络中，使其与经过身份验证的节点一起发送和接收信息。

被动攻击不破坏任何网络资源，但通过窃听其他被认证节点的通信并分析流量信息来实现被动主动攻击。

为了克服集中式存储中的数据泄露问题[1]，引入了联邦机器学习方法，其中集中式服务器必须将全局机器学习模型的副本发送到所有设备。本地设备必须下载模型的副本，并使用它们的设备数据单独训练它们而不将它们共享到集中存储。经过训练后，参数将被共享到服务器上，以进行共识更改。这种方法增强了用户数据的隐私性

本文的其余部分组织如下，第二节描述了区块链，第三节介绍了区块链联邦学习方法的相关工作。第四节包含联邦学习以及传统机器学习方法与联邦学习方法的比较，第五节说明了基于区块链的联邦学习节点认证，第六节讨论了实验和结果，最后第七节涵盖了结论。

1.

第二区块链

区块链被定义为在网络节点之间共享的分散和分布式数据库或分类账技术。数据以块的形式存储在区块链中。区块链中的块具有头和体的结构。头包含

前一个块的哈希值、时间戳和事务数据。一旦交易使用共识成功验证，区块就会链接在一起形成链。它用于增强网络之间的安全性、信任度和透明度。

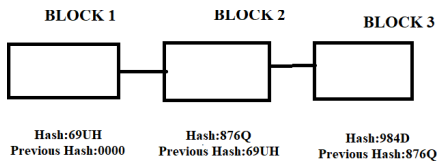

图 1 所示。区块链结构

A. 特征

区块链是一个可信的、安全的网络，具有以下特点:

不变的:

区块链中的数据是不可改变的和永久的，因此它们被称为不可变的。

透明的:

网络中的交易对所有人都是可访问的，因此它们被称为透明的。

分布:

区块链中的信息是分布式的，每个节点都维护着交易的副本[11]。区块链的这一特性使它们能够抵抗集中式节点故障。因此它们被称为分布式的

分散:

区块链没有集中的服务器来存储信息，也没有集中的管理机构。网络中的事务被分发到每个节点，因此网络中的任何异常或更新都会广播到所有参与节点。*可核查:*

网络中的每个人都可以验证信息的有效性，因此它们是可验证的。

安全:

存储在区块链中的信息使用加密算法进行加密，并以散列值存储。因此他们是安全的。

B. 区块链的类型:

区块链根据其部署性质分为三种类型[12]。

- 公共区块链
- 私人区块链
- 财团区块链

公共区块链:

没有访问限制的区块链被称为公共区块链。任何人都可以发送交易并作为验证者参与。它对所有人开放。

私人区块链:

在组织内部或组织之间使用的区块链对指定成员的访问有限，称为私有区块链。

财团区块链

区块链是有许可和无许可的结合，被称为联盟区块链

区块链使用共识，这是一种算法，它定义了必须由系统中所有节点同意的协议。目前流行的共识算法有 PoW、PoS 和 PoA。它保证了网络中对等体之间的真实性和信任度[13]。智能合约用于基于以太坊的区块链，用于在满足条件时自动验证交易。区块链是一项很有前途的技术，它被部署在供应链管理、医疗保健和智能电网等各种应用中[4]，以保护网络免受安全攻击。

联合学习和区块链两种技术的融合增强了联合学习方法的安全性框架。联邦学习方法被认为是基于技术的行业的福音，因为它解决了用户的数据隐私问题，因为它采用分散的方法将模型发送给用户

局部设备和全局机器学习模型训练模型。但是，不确定服务器将通信发送给授权节点还是恶意节点。提议的基于区块链的节点身份验证确保集中式服务器使用智能合约与经过身份验证的节点进行通信。

三、相关工作

网络中的集中式服务器充当中央集线器，存储从边缘设备接收到的信息。这些设备共享的信息本质上是敏感的，比如用户的出生日期、身份信息、家庭住址和健康状况。在少数情况下，它接收组织的机密信息。单点故障导致整个网络中断。目前大多数网络攻击都是针对集中式服务器的。

联邦机器学习方法被认为是一种有前途的隐私保护技术。研究人员[8]分析了联邦机器学习方法的安全威胁和漏洞，讨论了这种方法也容易受到中毒攻击，其中更新可能被入侵者中毒，并提出了改进本地更新中毒的技术。作者在论文[6]中讨论了另一种攻击，其中讨论了模型在服务器中的放置以及在后门中攻击模型的方法。作者[7]认为数据中毒攻击不仅会毒害设备数据的数据，还会毒害机器学习模型。

本文用联邦学习方法对敌方攻击者进行了分类[14]。

诚实但好奇的联邦学习(FL)服务器:

FL 服务器很想接收来自客户机的更新，并推断关于私有客户机数据集的信息。

恶意 FL 服务器:

这类服务器有意图对客户端的私有数据集发起攻击。

诚实但好奇的客户:

这种类型的服务器是诚实的，但好奇地知道从服务器的连续全局参数，并试图推断它。

恶意的客户:

恶意客户端试图推断其他客户端的更新并构造自己的全局参数，并获取有关其他客户端的私有数据集的附加信息。

为了减轻这种安全风险，提出了联合学习和区块链技术的集成
[14]这种方法用于保护移动设备中的数据不被泄露。区块链和联邦

提出了保护雾计算的学习集成方法[16]，保证了区块链可以与联邦学习方法一起部署，以增强网络的安全性。这就需要安全基础设施来保护联邦学习方法免受各种导致数据推理攻击的主动攻击。为了防止联邦学习方法中恶意服务器和恶意客户端的数据泄露，提出了基于区块链的节点认证的联邦学习方法，该方法在保护隐私的同时提高了服务器和客户端的安全性。

四、联邦机器学习:

联邦机器学习是机器学习的一种形式，用于在去中心化数据上训练中心模型[16]。中心模型的副本被分发给节点，这些节点在中心模型上训练数据，并将结果更新聚合回集中式服务器。它使用 Fedavg 算法来优化结果更新。

A. 联邦学习的优势:

联邦机器学习方法的优点是

- 高度个性化的
- 隐私保护
- 低延时

经典机器学习 Vs 联邦机器学习

在经典的机器学习方法中，从所有边缘设备收集数据并共享到集中服务器进行训练。机器学习模型将在收集的数据上进行训练，然后预测测试数据或新数据的输出这种传统的机器学习方法对用户数据的隐私构成了威胁用户数据被收集并存储在集中的服务器上进行训练。为了克服这种问题，引入了联邦机器学习方法，其中集中式机器学习模型训练将在分散的数据上进行。用户不需要将自己的数据共享给集中式服务器，而是由服务器将机器学习模型的副本共享给用户的设备，在每个设备本地独立进行训练，结果共享给集中式服务器，减轻了集中式服务器的负担，同时保护了用户数据的隐私性。集中式服务器聚合来自所有设备的结果并更新模型以提高性能。

传统的机器学习方法要求所有的数据都存储在集中的存储系统中。通过数据推理攻击，影响用户的数据隐私。在联邦机器学习方法中，模型在分散的数据上进行训练，消除了数据推理攻击。单点故障是集中式机器学习方法中的主要问题，如果服务器受到攻击，它会影响连接到它们的其他节点的安全性和功能。在

联邦机器学习方法的数据不共享到集中式服务器，连接节点只在用全局模型训练本地数据后共享参数。

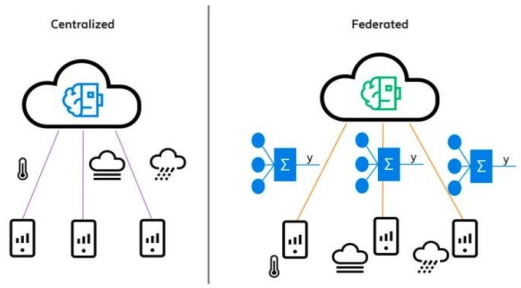

图 2 所示。传统机器学习 Vs 联邦机器学习训练

V 基于区块链的联合学习节点认证

基于区块链的节点身份验证是使用以太坊区块链构建的。联邦学习方法具有集中式服务器，具有全局机器学习模型。基于开发人员对问题的定义，全局机器学习模型可以是有监督的或无监督的。

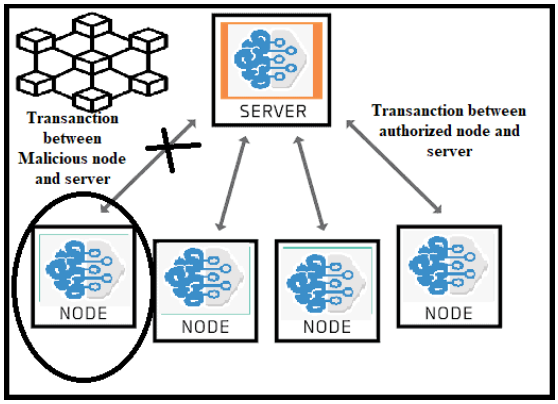

图 3 所示。联邦学习的节点身份验证

图 1.3 展示了联邦学习方法的节点认证，其中服务器将全局机器学习模型分发给节点。节点下载模型的副本并使用它们的设备数据进行训练。在将机器学习模型的副本发送到节点之前，将部署节点身份验证智能合约，以确保节点与经过身份验证的节点进行交互。

一旦服务器开始启动与节点的通信块将被创建。智能合约验证服务器和认证节点之间是否启用交互。如果它是有效的，那么区块将被添加到链中，否则交互将被中止。

节点授权智能合约：用于节点授权的伪代码检查在经过身份验证的服务器和经过身份验证的节点之间是否启动了交互。

```
//定义一个合约 ServerRole 来管理这个角色-添加，删除，检查合约 ServerRole {

 using Roles for Roles;

 //定义 2 个事件，一个用于添加，另一个用于删除

 事件 NodeAdded(地址索引帐户);事件

 noderdeleted(地址索引帐户);

  //在构造函数中添加节点 constructor()

 public {

   _addNode (msg.sender);

 }

 //定义一个修饰符

 modifier onlyNode(address onlyNode ID)

   {require(isValidNode(onlyNode ID));;}

 //定义一个函数'isServer'来检查这个角色

 函数 isServer(地址帐户)公共视图返回(bool) {

   返回 nodes.has(账户);

 }

 //定义一个函数'addServer'来添加角色函数

 addServer(地址帐户)public {

   _addServer(账户);

 }

 //定义一个函数

 函数 renounceServer() public {

   _removeServer (msg.sender);
```

}

//定义一个内部函数'_addServer'来添加这个角色，由'addServer'调用

函数_addServer(地址帐户)内部{server.add(帐户);

发出 serverAdded(账户);

} //定义一个内部函数'_removeServer'来移除这个角色，由'removeServer'调用

函数_removeServer(地址帐户)内部{server.remove(帐户);

发出 serverRemoved(账户);

}}

节点身份验证完成后，节点仅向服务器发送训练模型的更新
，而不是共享全部数据。服务器将使用 fedag 算法改进从节点和模型收集的更新的加权平均值。在下一个 epoch，将改进模型的副本发送到所有节点。

Vi 实验设置及结果

基于区块链的节点身份验证使用以太坊框架与松露构建，并使用 Ganache 提供用于测试的醚。使用财团区块链，它是无许可和许可的结合。集中式服务器和节点的角色由管理员预定义，因此具有一定的权限。另一方面，任何节点都可以与服务器交互，因此它们是无权限的。交易一旦通过用坚实语言编写的智能合约进行验证。

以太坊数据集[3]包含正常和异常事务的地址，用于评估节点身份验证，节点身份验证的角色在智能合约中定义。气体消耗以以太(ETH)表示。

该数据集包含 Hash、Nonce、Transaction index、Block number、Time stamp、from address 和 toaddress 的特征。前 500 个哈希包含 80%的认证地址和 20%的异常地址。

正常地址和异常地址在数据集中的分布如图 4 所示。

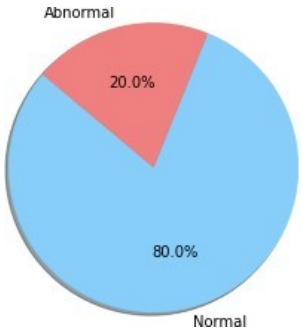

图 4 所示。正常与异常数据分布

使用 k -最近邻(KNN)算法和随机森林算法对节点认证性能进行评估，这两种算法被认为是分类问题的最佳分类算法。KNN 算法根据其邻居的分类方式对数据进行分类。随机森林算法采用集成方法进行更好的分类。

KNN 和 Random forest 的分类准确率分别为 95.7 和 99.6。在此随机森林优于 KNN 算法。

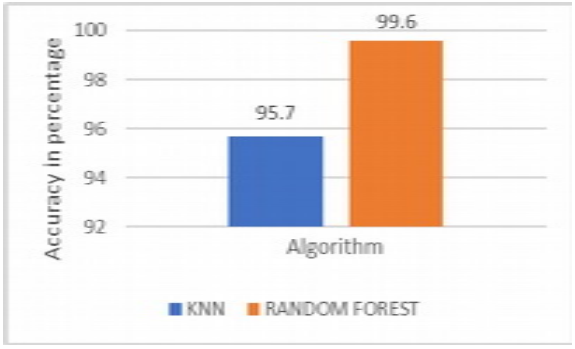

图 5 所示。性能比较 KNN 和随机森林

图 6 描述了每个事务的节点身份验证的 gas 消耗。

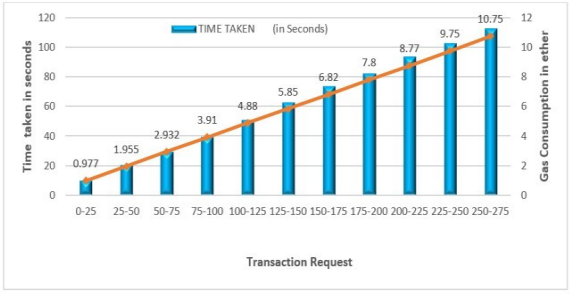

图 6 所示。节点认证的 Gas 消耗

实验使用了 275 个样本数据，包括合法地址和异常地址。利用 KNN 算法和随机森林算法对认证节点和异常节点的授权进行评估，分别获得了 95.7 和 99.6 的授权值。

第七章结论

技术的进步扩大了日益增长的网络中的安全威胁。物联网和银行业的架构主要依赖于其集中式服务器进行数据存储和决策。然而，单点故障破坏了整个经济和数据丢失。一个有希望的解决方案是采用分散的方法，如联邦学习，它可以保护数据的隐私，防止数据丢失。在我们提出的模型中，通过使用智能合约实现基于区块链的节点身份验证来解决服务器到节点身份验证的主要问题。随机森林算法和 KNN 算法的性能评价结果分别为 95.7%和 99.6%。虽然正确率更高，但节点认证的验证时间和 gas 消耗更高，这将集中在未来的工作中来减少。因此，基于区块链节点认证的安全增强联邦学习方法可以用于网络增强隐私保护和服务器屏蔽。

参考文献

[1] Abhishek Agarwal, Ayush Prasad, Rishabh Rustogi, Sweta Mishra，“使用深度学习方法检测和缓解云中的欺诈性资源消耗攻击”，信息安全与应用学报，第 56 卷，2021 年，102672,ISSN 2214-2126。

[2] Alberto Blanco-Justicia, Josep Domingo-Ferrer, Sergio marty ínez, David Sánchez, Adrian Flanagan, Kuan Eeik Tan，“在联邦学习系统中实现安全和隐私: 综述、研究挑战与未来方向”，《人工智能工程应用》，第 106 卷，2021,104468,ISSN 0952-1976。

[3] Al-E 'mari, S, Anbar, M.， Sanjalawe, Y.， Manickam, S. (2021)。以太坊网络上基于标记交易的数据集。: Anbar, M, Abdullah, N.， Manickam, S.(编)《网络安全进展》。ace 2020。通信在计算机和信息科学，卷 1347。施普林格、新加坡。https://doi.org/10.1007/978 - 981 - 33 - 6835 - 4 _5

[4] 区块链理事会，核桃，https://www.blockchain-council.org/blockchain/top-10-promising-blockchain-use-案例/从 2019 年开始提供。

[5] Dharmin Dave, Shalin Parikh, Reema Patel, Nishant Doshi，区块链技术及其解决方案综述，《计算机科学进展》，第 160 卷，2019 年，第 740-745 页，ISSN 1877-0509。

[6] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin 和 V. Shmatikov，“如何后门联邦学习”，《AISTATS proc》，2020，第 1-10 页。

[7] 孙国光，丛勇，董军，王强，吕磊，刘杰，“基于联邦机器学习的数据中毒攻击”；*国际计算机、电力和通信会议(ICCPC*

IEEE *物联网学报，第 9 卷，第 2 期。13, pp. 11365-* 11375, 2022 年 7 月 1 日，doi: 10.1109 / JIOT.2021.3128646。

[8] N. Bhagoji, S. Chakraborty, E. Mittal 和 S. Calo，“通过对抗镜头分析联邦学习”，[在线]。网站:https://arxiv.org/abs/1811.12470.， 2018。

[9] N. C. Will，“使用可信执行环境的雾/云增强物联网应用的隐私保护数据聚合方案”;2022 IEEE 国际系统会议(SysCon)， 2022, pp. 1-5, *doi:* 10.1109 / SysCon53536.2022.9773838。

[10] Pim Otte, Martijn de Vos, Johan Pouwelse,TrustChain: 抗 Sybil 的可扩展区块链，未来一代计算机系统，2017,ISSN 0167-739X。

[11]S. Revathy 和 S. S. Priya，《基于区块链的农民生产者-消费者模型》;2020 第四届计算机、通信*与信号处理国际会议(ICCCSP)， 2020, pp. 1-5, doi:* 10.1109 / ICCCSP49186.2020.9315214。

[12] 《区块链的真相》，哈佛商业审查,https://hbr.org/2017/01/the-truth-about-blockchain。

[13]T. M. Fernández- caramés 和 P. Fraga-Lamas，《区块链在物联网中的应用综述》;电子工程学报，vol. 6, pp.32979-33001,2018。

[14] 王伟龙，王英杰，黄艳，穆春晓，孙子策，童向荣，蔡志鹏，基于区块链和边缘计算的移动众包隐私保护联合学习系统，计算机网络，vol . 215,2022,109206,ISSN 1389-1286。

[15] 龚晓霞，陈勇，王强，王明李生，“针对云的私有数据推理攻击”模型、技术与研究方向 IEEE communications magazine, doi: 10.1109 / MCOM.004.2100867, 2022 年

[16] Y. *Qu* 等人，“在雾计算中使用区块链支持的联邦学习去中心化隐私”;IEEE 物联网学报，第 7 卷，第 7 期。6, *pp. 5171-5183, June 2020, doi:* 10.1109 / JIOT.2020.2977383