Journal of Computer Applications
计算机应用

2022-01-14
http://www.joca.cn

文章编号:1001-9081(****)**-0000-00

您的稿件编号：2021111934
本文章有任何疑问时请联系
作者：李超
手机号： 13521687401
邮箱：li.chao@bjtu.edu.cn

doi:10.11772/j.issn.1001-9081.2021111934

基于区块链的联邦学习研究进展

孙睿[1,2]，李超[1,2*]，王伟[1,2]，童恩栋[1,2]，王健[1,2]，刘吉强[1,2]

(1.智能交通数据安全与隐私保护技术北京市重点实验室（北京交通大学） 北京 100044；
2.北京交通大学 计算机与信息技术学院 北京 100044)

(*通信作者电子邮箱 li.chao@bjtu.edu.cn)

摘 要: 联邦学习是一种能够实现用户数据不出本地的新型隐私保护学习范式。随着联邦学习相关研究工作的不断深入，其单点故障及可信性缺乏等不足之处逐渐受到重视。近年来，起源于比特币的区块链技术取得迅速发展。区块链开创性地构建了去中心化的信任，为联邦学习的发展提供了一种新的可能。本文对现有基于区块链的联邦学习研究工作进行了调查研究：首先，对现有基于区块链的联邦学习框架进行了对比分析；然后，深入讨论区块链与联邦学习相结合所解决的联邦学习痛点问题；最后，阐述基于区块链的联邦学习技术在物联网、工业物联网、车联网、医疗服务等多个领域的应用前景。

关键词: 联邦学习；区块链；结构框架；融合应用；隐私保护

中图分类号:TP309.2 **文献标志码：A**

Research progress of blockchain-based federated learning

SUN Rui[1,2], LI Chao[1,2*], WANG Wei[1,2], TONG Endong[1,2], WANG Jian[1,2], LIU Jiqiang[1,2]

(1. Beijing Key Laboratory of Security and Privacy in Intelligent Transportation (Beijing Jiaotong University), Beijing 100044, China;
2. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Federated Learning is a promising privacy-preserving learning paradigm that can keep users' data locally. With the deepening of the research on Federated Learning, the shortcomings of Federated Learning such as single point of failure and lack of credibility have been paid more and more attention. In recent years, blockchain technology, which originated from Bitcoin, has achieved rapid development. Blockchain has pioneered the construction of decentralized trust, providing a new possibility for the development of Federated Learning. Existing research work on blockchain-based Federated Learning was surveyed. First, frameworks for blockchain-based Federated Learning were compared and analyzed. Then, key points of Federated Learning solved by the combination of blockchain and Federated Learning were discussed. Finally, application prospects of implementing blockchain-based Federated Learning were presented in various fields, such as Internet of Things, Industrial Internet of Things, Internet of Vehicles, and medical services.

Keywords: federated learning; blockchain; structural framework; convergence application; privacy protection

0 引言

随着 GDPR（General Data Protection Regulation）等隐私保护法案的出现，用户私人数据在中央服务器的直接使用受到了极大限制[1]，能够保障用户隐私的机器学习方法开始获得工业界与学术界的广泛重视。2016 年，谷歌提出了联邦学习（Federated Learning, FL）这一新范式[2]，能够实现用户数据不出设备本地的模型训练。联邦学习的主要思想是使大量存储本地数据的用户设备（称为客户端）能够在本地协作地训练单个机器学习模型，而无需共享其原始数据。在联邦学习中，中央服务器负责聚合所有客户端提交的参数，对全局模型进行更新，并将更新后的模型反馈给客户端，使客户端能够在未来的迭代中继续进行本地训练，以不断从客户端的本地数据中获益[3]。由于其“数据不动模型动，数据可用不可见”的隐私特性，近年来，联邦学习受到研究者的高度关注。

然而，联邦学习也存在很多独特的挑战。首先，联邦学习通常需要大量用户的参与，这些用户身份背景各异，行为方式复杂，难以相互信任，且增加了诚实用户隐私泄露的风险。其次，在联邦学习过程中，全局模型的获得需要通过用户的多轮迭代进行模型更新，这产生了大量的通信成本，以及在网络传输过程中额外的存储开销。此外，不诚实的参与

收稿日期: 2021-11-14； **修回日期**:2021-12-08； **录用日期:** 2021-12-23。
基金项目: 国家重点研发计划课题(2020YFB2103802)，中央高校基本科研业务费专项资金(2019RC038)。
作者简介: 孙睿(1998—)，女，吉林扶余人，硕士研究生，主要研究方向：区块链技术；李超(1988—)，男，甘肃天水人，讲师，博士，CCF 会员（C0103M），主要研究方向：区块链技术；王伟（1976—），男，湖北英山县人，教授，博士，主要研究方向：网络与系统安全、工业互联网安全、区块链安全；童恩栋（1986—），男，山东聊城人，讲师，博士，主要研究方向：服务计算、人工智能安全；王健（1975—），男，山东烟台人，副教授，博士，主要研究方向：密码应用、区块链、网络安全；刘吉强（1973—），男，山东烟台人，教授，博士，主要研究方向：可信计算、隐私保护、物联网安全。

者和易受攻击的中央服务器可能会影响全局模型的安全性[4]。最后，本地设备可能是恶意的或易受攻击的，这可能导致传输的信息被泄露或篡改[5]。

近年来，起源于比特币的区块链技术取得迅速发展[6]。区块链建立在去中心化的对等网络之上，交易以数据形式由全网节点备份，交易顺序和内容的一致性和不可篡改性由共识机制保障。区块链开创性地构建了去中心化的信任，使人们可以选择信任区块链底层密码学技术的可靠性及对等网络中大部分节点的诚实性，而无需被迫信任单一实体[7]。区块链构建的去中心化信任实现方式为联邦学习的发展提供了一种新的可能。例如，联邦学习不仅可以利用区块链共识机制提供的一致性在不可信环境中建立可信交互，还可以利用区块链激励机制提供的经济属性有效促进联邦学习中信息的共享。

本文主要分为六个部分：第一部分简要介绍本文调研的概况，包括相关文献的获取方法和涉及内容；第二部分对现有基于区块链的联邦学习框架进行对比及总结；第三部分对当前基于区块链的联邦学习研究工作面向的痛点问题及相应的解决方法进行分析梳理；第四部分对基于区块链的联邦学习技术在多个不同领域的应用进行了讨论；第五部分对现存问题进行分析并对未来进行展望；最后是全文总结。

1 基于区块链的联邦学习研究现状

本文调查研究的对象包括从 2019 年至 2021 年在期刊、会议以及预印本上发表的论文。本文所收集的文献涵盖了区块链技术与联邦学习技术相互融合的近期研究工作。本研究主要遵循以下原则收集这些论文：（1）在现有的搜索引擎中搜索关键词，找到相关文献；（2）根据所选论文中的参考文献，再找到本研究所需要的论文。

在统计过程中可以发现，虽然目前基于区块链的联邦学习（Blockchain-based Federated Learning，BFL）技术还处于早期发展阶段，相关的文献研究较少，但是 BFL 架构正受到越来越多人的关注，处于持续发展的状态。这些 BFL 研究工作通常采用不同 BFL 框架，面向不同联邦学习中存在的痛点问题（如：单点故障、缺乏激励机制、投毒攻击），且面向不同应用领域（如：物联网，工业物联网，车联网，医疗服务领域）。

目前，已有少数几篇文献讨论了区块链与联邦学习的结合。文献[8]介绍了部分现有 BFL 研究工作所使用的区块链种类和平台，以及 BFL 框架的对比。文献[9]对部分现有 BFL 框架、BFL 基础设施以及 BFL 应用进行了对比总结。文献[10]进行了面向联邦学习的调查研究工作，其中所涉及的联邦学习架构、通信效率、奖励机制、隐私保护、安全聚合方案等方面的对比也包含了一些 BFL 架构。文献[11]对 BFL 进行了综述，介绍了联邦学习的不足，并从架构特点、资源分配等方面对现有联邦学习进行研究，最后总结了基于区块链的联邦学习架构在人工智能领域的未来发展。类似地，文献[12]也论述了现有联邦学习机制存在的问题和不足，并说明了将区块链技术与联邦学习结合之后对进一步完善联邦学习模型的展望。可以看到，这些综述工作主要集中于 BFL 框架及其在人工智能领域的应用前景，缺乏对 BFL 面向的痛点问题的深入分析和 BFL 在更广泛场景应用前景的展开讨论。因此，本文从 BFL 框架出发，深入讨论区块链与联邦学习相结合所解决的联邦学习痛点问题，进而阐述 BFL 在物联网、工业物联网、车联网、医疗服务等多个领域的应用前景，对现有 BFL 研究工作在基础架构、核心技术、应用前景三个层面进行了全面且完整的分析和对比工作，总结出区块链与联邦学习相结合的创新与应用方向。

2 现有 BFL 框架对比

本节对本文收集的文献中提出的 BFL 框架进行总结与对比，针对其不同的设计思路进行分析。图 1 展示了本文总结的 BFL 框架分类示意图。

首先，传统联邦学习框架通常由一个中央服务器（server）和多个用户（或设备、客户端）组成。早期的典型 BFL 框架普遍利用去中心化的区块链取代传统联邦学习框架中的中央服务器（server），其主要目的是克服中央服务器造成的单点信任及故障等问题[13-16]。该类框架的范例如图 2 所示，用户将本地模型提交给维护区块链的矿工，矿工执行交叉验证、模型聚合等步骤，并基于共识机制生成一致的全局模型，随后利用区块存储和传播该全局模型，使用户能够从区块中将一致的全局模型下载到本地，进行下一轮训练。除了利用区块链取代中央服务器，这一类典型框架还通常具备两个特性。在模型聚合前， 通过引入交叉验证等机制，能够保证参与全局模型更新的本地模型符合全局模型更新的方向，防止恶意用户利用恶意模型破坏全局模型的安全性。此外，通过引入奖励机制，能够激励用户贡献优质数据并积极参与训练，有效缓解联邦学习中的公平性问题，防止贡献不同的用户获得相似奖励，导致用户消极怠工。

在这一类典型框架的基础上，一些 BFL 框架在共识机制、奖励机制等方面进一步创新。文献[1]提出的 BFLC（Blockchain-based Federated Learning framework with Committee consensus）框架，没有采用常见的工作量证明（Proof of Work，PoW）共识机制，而是提出了委员会共识机制（Committee Consensus Mechanism，CCM）。这一机制的特色是利用由部分诚实节点组成的委员会，执行模型局部梯度验证和区块生成等步骤。由于只有部分节点参与本地模型验证及全局模型更新，联邦学习的整体效率获得了显著提升。该机制要求委员会以外的节点将本地模型发送给委员会节点进行验证打分，只允许合格的模型参与全局模型更新。在该

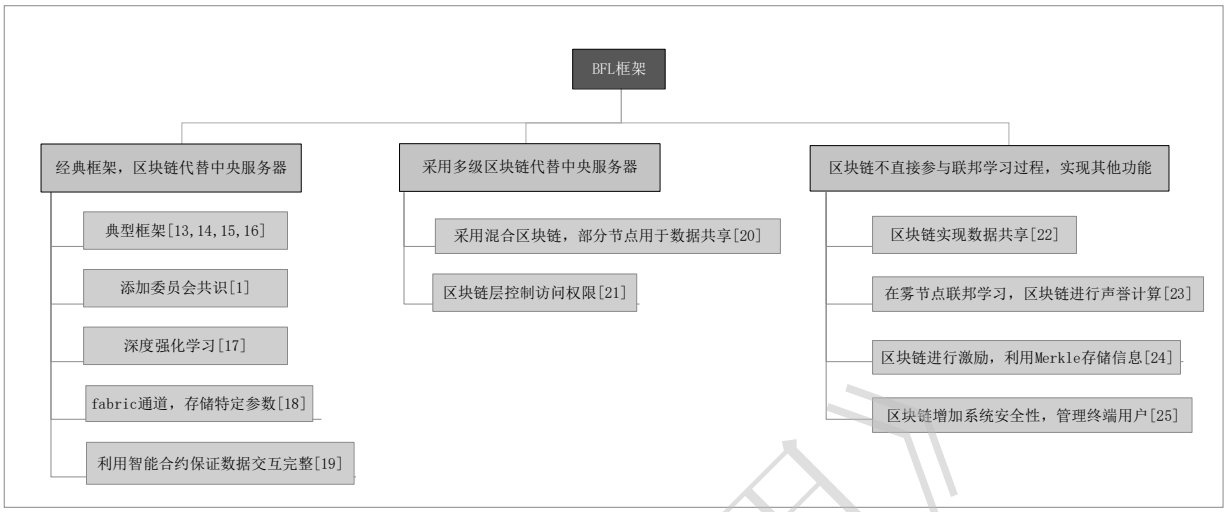

图1 BFL 框架分类

Fig. 1 Classification of BFL framework

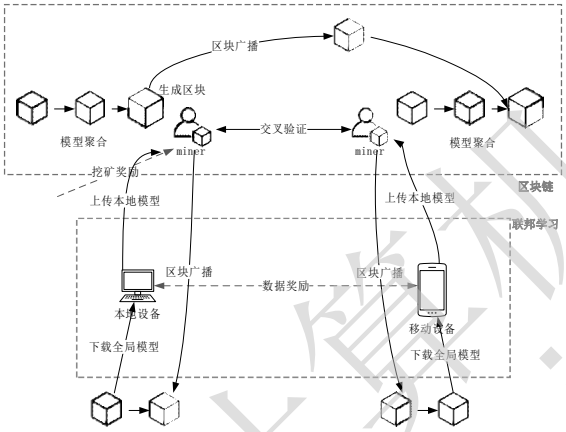

图2 典型 BFL 框架

Fig. 2 Typical BFL framework

机制中，为提升安全性，基于节点历史表现分数及智能合约，委员会成员会进行周期性更替。文献[17]引入深度强化学习（Deep Reinforcement Learning），寻求能够最小化系统时延、能耗及奖励总额的最优系统参数，包括用户训练本地模型时的建议数据量和能耗，以及区块生成速率。文献[18]提出了 FLChain（Federated Learning via mec-enabled blockChain network）框架，该框架同时包含移动设备和边缘设备，其中移动设备主要负责使用设备上的数据样本进行本地模型更新，边缘设备一方面为资源受限的移动设备提供较充足的网络资源，另一方面充当 FLChain 区块链网络中的节点，对区块链进行维护。FLChain 使用了联盟区块链超级账本 fabric 中的通道技术，利用通道的隔离特性，增强全局模型训练的安全性，并提供一定程度的数据隐私保护。文献[19]提出了一种名为 CrowdSFL（Crowd computing Secure framework based on blockchain and Federated Learning）的众包 BFL 框架，其只要目的是降低众包过程中用户开销并保障众包安全性。在 CrowdSFL 中，整个众包系统基于区块链构建，每一个参与者都拥有独立的区块链账户。CrowdSFL 提出了由智能合约控制的数据交互模式，能够保证数据以正确的格式上传并保存到区块中。

上述 BFL 框架均采用单一类型的区块链取代传统联邦学习中的中央服务器，近年来，少数工作提出了由更复杂的多级区块链取代中央服务器的 BFL 框架。文献[20]提出了一种基于混合区块链的 BFL 框架 PermiDAG（Permissioned blockchain and the local Directed Acyclic Graph），该框架中的混合区块链以运行在路边单元（Road Side Unit，RSU）的许可区块链作为主链，同时允许车辆节点组成多个本地有向无环图（Directed Acyclic Graph，DAG）。作为主链的许可区块链负责记录数据共享的相关信息以及全局模型聚合的相关参数。车辆节点组成的多个本地 DAG 则用来提升数据共享的效率，并将数据共享事件以及训练过的模型参数作为交易存储在块中。同时，基于本地 DAG，邻近的车辆节点相互通讯，获得附近车辆的本地模型，并利用这些模型提升自己的本地模型，实现异步学习的过程。此外，文献[21]提出的 BFEL（Blockchain-empowered Federated Edge Learning）框架也采用了多级区块链的结构。该框架包含应用层和区块链层两部分，其中应用层主要负责执行联邦学习过程。该框架中的区块链层同时包含一条基于公有区块链的主链，和多条基于联盟区块链的子链。通过利用多个子链设置访问权限，该框架能够提升数据隐私保障能力，实现性能隔离。

上述 BFL 框架均采用单一或多级区块链取代传统联邦学习中的中央服务器，然而，在一些 BFL 框架中，区块链既没有直接取代中央服务器，也没有直接参与到传统联邦学习过程中。例如，在文献[22]中提出的 BFL 框架中，区块链仅用于实现数据共享的功能。该框架中的区块链内存在三种不同类型的交易：（1）检索交易，允许节点将请求的模型信息

通知其他节点；（2）模型交易，允许节点将模型训练的数据传给其他节点；（3）数据共享交易，允许节点将共享的数据返回给请求者。具体来说，数据请求者（requester）将数据共享请求发送给区块链，区块链进行检索交易，查看缓存中是否已包含相应数据。若存在，区块链将查询结果以及所请求的数据模型直接返回给请求者，并生成数据共享交易；若不存在，区块链将进行多方信息检索，组件模型训练委员会，利用模型交易进行模型训练，生成请求者所需要的模型，在将模型返回给请求者的同时进行缓存，以备未来之需。在文献[23]中提出的 fine-grained FL 框架中，联邦学习的执行主要发生在云节点和雾节点，该框架中的区块链不直接参与联邦学习，主要负责计算和存储参与联邦学习的各节点声誉。此外，在文献[24]提出的用于工业物联网内设备故障检测的 BFL 框架中，区块链主要用于对客户端数据进行可信存储和验证。在该框架中，客户端定期创建 Merkle 树组织传感器收集的数据，并将 Merkle root 存储在区块链中。在未来发生争议时，存储在区块链中的 Merkle root 可被用作证据帮助解决争议。文献[25]提出的一种基于联盟区块链的 BFL 框架，目的是在数字孪生无线网络模型中提升边缘计算能力。该框架由多类终端用户组成，如：物联网设备、移动设备、基站以及宏基站。其中，基站负责执行联邦学习的本地训练，宏基站则充当联邦学习的中央服务器。由于联邦学习无法解决孪生终端用户之间缺乏信任的问题，该框架引入联盟区块链增强系统安全性，利用区块链记录数字孪生过程中的数据，并通过控制访问权限来管理用户。

表 1 对不同 BFL 框架进行总结对比。

表1 不同 BFL 框架对比

Tab. 1 Comparison of different BFL frameworks

结构	种类	共识机制	激励机制
BlockFL[13]	公链	PoW	数据和挖矿奖励
FL-Block[14]	公链	PoW	数据和挖矿奖励
BFL[15]	公链	PoW	数据和挖矿奖励
BFL[20]	私链	代理权益证明	行动的影响
FLChain[18]	公链	拜占庭/PoW	-
BFEL[21]	公链/联盟链	验证证明	贡献
BFLC[1]	联盟链	CCM	根据贡献利益共享
FL[26]	联盟链	权益证明/拜占庭	多重 Krum 和声誉

3 BFL 面向的痛点问题

本节总结了 BFL 面向的联邦学习中的痛点问题，以及相应的解决办法或是对联邦学习相应不足之处的改善方法。

3.1 联邦学习存在的痛点问题

联邦学习主要存在以下五类痛点问题：

（1）单点故障：联邦学习的中央服务器容易受到恶意更新的影响，致使全局模型更新产生缺陷，从而影响所有本地模型更新，破坏局部模型更新的准确性。另外，联邦学习需要本地设备上传本地模型更新到中央服务器，当同时传输的设备模型过多时，会导致网络过载。

（2）缺乏激励机制：联邦学习通常假定每个本地设备都自愿将数据资源贡献给全局模型，但是这并不符合实际。参与者缺乏激励机制，就会影响参与者的动力，甚至一些参与者不贡献数据也获得相应奖励，导致了不公平的经济补偿。

（3）投毒攻击：恶意用户可能通过故意上传精心计算的恶意本地训练模型以影响全局模型训练，有目的性地破坏机器学习的预测结果。这主要是因为联邦学习缺乏审计恶意用户或是恶意模型更新的能力。

（4）缺少隐私政策：尽管训练数据资源存储在本地设备中，联邦学习框架也可能会导致训练数据的隐私泄露。

（5）效率低下：由于联邦学习需要客户端与服务器进行通信，传输本地学习模型，进行多轮模型训练迭代后进行局部或全局模型更新等等，客户端与服务器的通信效率以及模型训练的效率也会影响联邦学习的性能。

3.2 痛点问题的解决方案

为了解决单点故障问题，BFL 通常采用区块链分布式节点代替联邦学习中的中央服务器。在很多文章中，所设计的结构框架都提到解决单点故障的问题[13][26-28]。例如，文献[14]通过将区块链与联邦学习结合，实现了去中心化隐私保护，从而防止边缘计算场景中的单点故障问题。在文献[29]提出的方案中，区块链可以选择特定的工业 4.0 设备来领导一轮学习，从而代替原本的中央服务器，以有效防止单点故障。该方案在原本的联邦学习架构中，引入矿工节点，由矿工节点进行本地设备的模型更新。每个本地设备计算并将本地模型更新上传到相关矿工节点上，但并不把原始数据上传给矿工节点。接着，通过共识机制记录并验证本地模型更新，聚合本地设备上传的模型更新，并将更新区块附加到区块链，本地设备可以从区块链中的区块下载模型更新。此外，现有的 BFL 工作采用了多种不同类型的区块链共识机制保障一致性，包括：工作证明（PoW），拜占庭容错（Byzantine Fault Tolerance，BFT）[18]，权益证明（Proof of Stake，PoS），代理权益证明（Delegated Proof of Stake，DPoS）[20],验证证明（Proof of Verifying，PoV）[21]，训练质量证明（Proof of Training Quality，PoQ）[22]，权利证明（Proof of Authority，PoA）[26]，联邦证明（Proof of Federation，PoF）[30]，委员会共识（CCM）[1]。为了解决缺乏激励机制问题，BFL 通常利用区块链技术构建面向期望行为的激励机制，或面向异常行为的惩罚机制，从而激发本地用户产生对全局模型更新作出

贡献的积极性。例如，文献[13-15]中所提出的激励机制包括数据奖励和挖矿奖励两个方面，终端设备的数据奖励从其相应的矿工那里接收，奖励的数量与其数据样本的大小成线性正比。当矿工完成模型聚合并生成区块后，能够从区块链网络中获得挖矿奖励，挖矿奖励的数量和其所绑定的终端设备的聚合数据样本数量成线性正比。文献[1]提出了一种贡献利益共享的激励方式，在每一轮模型更新聚合后，管理者们会根据提交的模型更新的分数向相应的节点分配奖励。文献[26]提出基于声誉的激励机制，在上传本地模型后，验证者使用预先约定的方法计算声誉，并消除不满意的更新。如果提交的更新通过了验证，该提交者的声誉就会增加，反之声誉会减少，最后根据提交者的声誉，进行奖励。文献[24]根据在本地模型训练中使用的客户数据的大小和质心距离，设计了一种基于智能合约的激励机制。在文献[31]提出的方案中，诚实的训练者可以根据训练好的模型的贡献获得可观的分区利润，而恶意模型会被及时发现并受到严厉的惩罚，并通过评估训练者的可靠性和贡献进行奖励。文献[8]引入一种具有竞争力的模型更新方法，从而使 worker 的利润最大化，实现激励的目的。该方法也要求在某一轮更新中，所选择的每个 worker 选择前一轮 worker 提交的最优模型更新，并用它们更新自己的模型，每一个 worker 的奖励由下一轮的 worker 投票结果决定。

为了解决投毒攻击问题，BFL 通常利用部署在区块链中的共识机制，进行模型更新验证，从而有效防止投毒攻击。文献[14]提出用区块链系统取代中央服务器，以利用区块链不可篡改的特性消除投毒攻击。文献[21]提出了验证证明（PoV）共识机制，用来协作验证预定义矿工之间的本地模型的更新质量。在该方案中，只有经过验证的模型更新，才可以存储在块中，从而防止投毒攻击。为了减少恶意的投毒模型更新，文献[26]提出了一种基于声誉的众包激励机制。在这种机制下，如果用户被检测到是恶意更新，其更新模型会被拒绝，不仅在本轮更新中不能收获奖励，还会降低其声誉，从而影响未来收益，得到惩罚。文献[32]提出了一种计分机制来判断设备是否为会进行投毒攻击的恶意设备，从而选择参与模型更新的训练者，以抵御投毒攻击。文献[30]提出了一种 multi-Krum（multiple-Krum）的共识机制，会拒绝与大多数模型更新方向相反的模型更新。在每一轮更新中，由多数投票选出验证同行委员会，委员会使用 multi-Krum 拒绝恶意的模型更新，从而防止投毒攻击。文献[33]提出了一种分散验证机制来验证局部模型更新，该机制对每个模型的有效性进行投票，通过投票结果来排除潜在的恶意设备。

为了防止隐私泄露，一些 BFL 方案会设计额外的隐私政策。例如，文献[34]中通过使用同态加密方法来保证训练模型的隐私。文献[22]将差分隐私集成到联邦学习中来保护数据隐私。同样，在文献[30]中也使用了差分隐私来保护数据隐私安全，该文献提出的系统还进一步使用了可验证的秘密共享方案来进行模型安全聚合，从而保护个人隐私安全。文献[35]通过结合区块链和联邦学习技术，设计面向轻量级网络设备的入侵检测算法，从而在共享数据的同时保护网络用户的数据隐私。文献[36]利用区块链去中心化、防篡改等特点，将数据记录及其他重要信息存储在区块链上，而完整数据则加密存储在分布式数据库，从而实现用户数据的安全存储，防止用户隐私数据泄露。

最后，为了解决效率低下问题，BFL 方案通常采用各类方法降低需要传输的数据量。文献[14]提出的方案将具体的相关数据保存在链外的分布式哈希表中，仅在区块链上存储指针，从而降低传输数据量。文献[20]提出了边缘数据学习模型的异步联邦学习方案，通过选择参与节点，进一步提高联邦学习的学效率。文献[1]提出了委员会共识机制（CCM），在将局部梯度附加到链上之前验证局部梯度。在这种机制下，只有几个节点来验证模型更新，而无需向每个节点广播并达成协议，从而提高模型验证的效率。文献[21]中提出了一种梯度压缩方案，该方案能够在不影响学习精度的情况下提高由区块链授权的联邦边缘学习的通信效率。另外，文献[37]提出了一种在联邦学习中添加超参数优化和弹性权重合并的方法，从而提高模型训练的精度和效率。

表 2 对 BFL 解决的痛点问题进行了总结。

表2 BFL 解决的问题

Tab. 2 solved issues by BFL

解决的问题	解决方法
激励机制	数据奖励和挖矿奖励[13-15]
	贡献利益共享[1]
	多重 Krum 计算声誉[26]
	智能合约[24]
投毒攻击	验证证明共识机制[21]
	多重 Krum 共识机制[30]
	认知计算[28]
	计分机制[32]
提高学习效率	从边缘数据学习模型的异步学习[20]
提高模型验证效率	委员会共识机制[1]
提高通信效率	梯度压缩[21]
提高模型训练效率	超参数优化和弹性权重合并[37]

4 BFL 在不同领域中的应用

目前，基于区块链的联邦学习（BFL）技术已经被应用到许多行业领域，本文对当前 BFL 技术在物联网、工业物联网、车联网、医疗服务等多个领域的应用前景进行了总结。

在物联网领域，很多敏感信息会存储在物联网设备当中。文献[22]构建了一个分布式多方数据共享的模型，并通过差分隐私进一步保护数据的真实性，使得设备可以安全准

确地检索数据。不同于常见的 PoW 共识算法，文献[22]中使用了 PoQ 共识算法用于验证训练模型，以提升计算资源的利用效率。面向工业物联网(Industrial Internet of Things，IIoT)的故障检测场景,文献[24][38]中提出了一种名为质心距离加权联邦平均的联邦平均算法。此算法考虑到每个客户端数据集的正类和负类之间的距离，能够减少 IIoT 设备故障检测中数据异构问题的影响。为了帮助家电制造商提高服务质量并优化家电功能，文献[26]提出了分层众包的联邦学习系统，利用区块链技术防止恶意模型更新。为了使 6G 网络更加安全高效地应用到物联网中，文献[39]提出了区块链和联邦学习相结合的架构，结合移动边缘计算和设备到设备（Device to Device, D2D）通信，应对 6G 网络面临的挑战。

BFL 也能够给医疗服务带来巨大进步。通常来说，对于患者的远程检测或是某些人工智能辅助诊断都需要很多患者的疾病信息。然而，很多医疗信息中包含着患者的敏感信息，这些数据对某些攻击者具有很高的内在价值。因此，BFL 正被逐步应用到医学领域。文献[40]中为医疗保健联盟提出了一种 BFL 方案，该方案建立了一套兼容以太坊生态系统的企业级区块链组件,且集成了一系列隐私保护技术。文献[40]还提出了一种新的安全聚合协议，将其作为运行在 AMD 公司的可信硬件环境 SEV（Secure Encrypted Virtualization）中，以保证隐私数据的安全。文献[41]在医疗领域提出了基于区块链的联邦学习框架，将智能合约应用于联邦学习算法的数据聚合过程中，以确保数据共享时的透明度和使用许可，并通过基于大量患者信息的训练来预测糖尿病的风险。此外，文献[42]开发了基于 BFL 的面向医疗物联网（Internet of Medical Things，IoMT）设备的轻量级安全和隐私算法。文献[43]不仅提出了可以应用于医疗物联网的可信任的 BFL 框架，还设计了面向数据分类的新冠病毒应用程序，可以学习全球的新冠病毒诊断相关模型。该方案不仅包含可信的、防篡改的梯度挖掘方法和基于去中心化共识的聚合器，还为负责聚合的区块链节点增加额外的安全性。文献[44]也提出了应用于医疗保健的 BFL 方案，其目标是实现患者医疗信息的保护和共享，并构建一个全球实时应用模型。另外，文献[45]提出了一个 BFL 框架,该框架可以使用最新数据，通过基于胶囊网络的分割和分类扫描肺部 CT 图像，在医院之间共享数据来提高新冠病毒的识别率，同时实现患者的隐私保护。为了实现医疗健康数据的安全可靠共享和智能处理，文献[46]提出了一种基于 BFL 的健康医疗共享体系。文献[47]提出的方案首先利用区块链构建虚拟数据中心，可以使各个数据拥有方上传相关的医学数据，该数据并不是原始数据，而是包含原始数据一些特征属性的元数据。接着，在进行模型训练和学习后，研究机构就可以获取相关数据的研究结果，从而获得高质量数据，并将数据应用到慢性疾病的预防管理中。

BFL 方案也已广泛应用于车联网（Internet of Vehicles，IoV），以实现数据共享和自动驾驶。Pokhrel 等[15]提出了一种完全去中心化的的 BFL 框架，该框架实现了 IoV 中端到端的可信通信，并且通信延迟也在可接受范围内，从而促进自动化车辆的有效通信。文献[15]采用 BFL 验证车载机器学习（on-Verhicle Machine Learning，oVML）的模型更新，提高了自动化车辆的性能和隐私安全性。Lu 等[20]提出了一种 BFL 框架，由路边单元维护的主许可区块链和车辆运行的本地有向无环图（DAG）组成，用于在 IoV 中进行高效数据共享。另外，文献[20]还提出了基于边缘数据的异步联邦学习方案，通过委托权益证明（DPoS）选择优化的参与节点，从而提高联邦学习的效率。文献[48]提出了一种基于区块链的分层联邦学习算法，减少了存储消耗，提高了训练精度。所提出的基于 BFL 的知识共享方式，能够提高车内网络的可靠性和安全性，并利用学习证明（Proof of Learning，PoL）共识机制，实现了一种轻量级区块链，避免了计算能力浪费。

此外，BFL 正被逐步推广到其他多个领域。在内容缓存领域，文献[49]提出了一种新的算法，即称为 CREAT（blockchain-assisted Compressed algoRithm of fEderated leArning applied for conTent caching）的采用区块链技术辅助的联邦学习算法，以预测缓存文件，并提高缓存命中率。在位置预测领域，文献[50]提出的方案采用 BFL 在用户的移动设备上进行本地培训，从而在利用这些数据进行更好的位置预测的同时保护用户的隐私。在移动众感知领域，文献[51]提出了 SFAC（Secure Federated learning for uAv-assisted Crowdsensing）架构，一种用于无人机辅助移动众感知（Mobile Crowd Sensing，MCS）的安全联邦学习框架，并利用局部差分隐私保护数据提供方的隐私。另外，BFL 也被应用到灾难响应领域，文献[52]提出一种由区块链授权的 BFL 架构，将在未来 6G 网络中利用无人机上的无线移动模块实现灾难响应系统。此外，BFL 也被应用到新闻推荐领域，文献[53]提出了一种基于联邦学习的云边端协同过滤算法推荐系统。该系统利用差分隐私技术向训练模型中添加噪声，进一步防止了数据隐私的暴露。

表 3 对不同 BFL 在不同领域的应用进行了对比。

表3 BFL 的应用总结

Tab. 3 Summary of applications of BFL

应用领域	功能
物联网	保证数据安全[22]
	工业物联网的故障检测[24]
	优化家电功能[26]
医疗服务	保护病人隐私[40]
	预测糖尿病风险[41]
	医疗物联网设备的隐私算法[42]
	新冠病毒诊断共享[43]
	医疗信息保护和共享[44]
	提高新冠病毒识别率[45]
车联网	预测自动车辆间的有效通信[15]

	车辆运行数据共享[20]
	提高车内网络可靠性和安全性[48]
内容缓存	提高缓存命中率[49]
位置预测	更准确的位置预测和隐私保护[50]
移动众感知	保护无人机学习后的隐私[51]
灾难响应	无人机和 6G 网络灾难响应系统[52]
新闻推荐	过滤信息，精准推荐，隐私保护[53]

5 现存问题与未来展望

区块链的引入有助于解决传统联邦学习存在的部分痛点问题。然而，区块链与联邦学习的结合也面临区块链本身带来的问题，有待研究者不断探索。

首先，多数区块链系统缺乏足够强力的隐私保护手段，因而 BFL 框架需要引入差分隐私、同态加密等隐私保护技术，对放置在区块链上的数据进行额外保护。例如，文献[22][30]中均在模型提取的过程重使用差分隐私来加入噪声，从而保护个体数据的隐私。文献[30]还提出了一种可验证的秘密共享方案来进行模型安全聚合。除此之外，文献[34]中通过使用同态加密算法对训练数据进行加密来保证训练模型的隐私。对于现存 BFL 框架来说，如何更好的平衡隐私保护的开销与训练的准确性仍是亟待解决的关键问题。

其次，尽管 BFL 框架可以通过合理设计共识机制对投毒攻击进行一定程度的抵御，多数区块链共识算法却面临本身的安全风险。例如，在最常见的 PoW 共识算法中，由于各个矿工接收区块存在时延，很有可能发生分叉（forking）问题。文献[13]通过引入 ACK(ACKnowledge character)，在等待时间内确定是否发生分叉，如发生分叉就重新进行挖矿，以此缓解该问题。近期的一些研究提出了新的共识算法，但是这些共识算法的安全性普遍缺乏理论证明和实际验证。因此，如何构建可证明安全的 BFL 共识算法仍是亟待解决的关键问题。

6 结语

本文介绍了区块链技术与联邦学习技术相结合的基于区块链的联邦学习（BFL）研究领域发展现状，通过广泛调查研究现有 BFL 领域的相关科研文献，对现有 BFL 工作在基础架构、核心技术、应用前景三个层面进行了全面、完整的分析和对比工作，总结出区块链与联邦学习相结合的创新与应用方向。目前来看，BFL 领域仍处于初期发展阶段，多数研究工作仅通过引入区块链技术解决联邦学习的单点信任问题，缺乏在隐私、效率、公平性等方面的进一步探索。此外，多数研究工作仍停留在理论阶段，且部分工作提出的 BFL 框架并不完整，导致这些工作提出的 BFL 技术在实际场景中的可用性可能存在不足。随着区块链和联邦学习这两个热门领域的快速发展，作为其交叉领域的 BFL 能够汲取这两个领域的技术精华，进而交融出独特的创新性技术反哺这两个领域，建立一种可信的隐私保护学习范式，为多个应用领域的相关行业带来巨大变革。

参考文献

[1] LI Y, CHEN C, LIU N, et al. A blockchain-based decentralized federated learning framework with committee consensus[J]. IEEE Network, 2020, 35(1): 234-241.

[2] MCMAHAN B,MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR 2017:1273-1282.

[3] KONKEY J, MCMAHAN H B, YU F X, et al. Federated learning: Strategies for improving communication efficiency[EB/OL]. (2017-10-30)[2021-07-13]. https://arxiv.org/pdf/1610.05492.pdf

[4] LI L ,FAN Y,TSE M , et al. A review of applications in federated learning[J]. Computers & Industrial Engineering, 2020, 150(5):106854.

[5] LI T,SAHU A K,TALWALKAR A , et al. Federated Learning: Challenges, Methods, and Future Directions[J]. IEEE Signal Processing Magazine, 2020, 40(3):50-60.

[6] NAKAMOTO S. Bitcoin: A peer-to-peer electronic cash system[EB/OL]. (2010-12-12)[2021-07-15]. https://bitcoin.org/bitcoin.pdf

[7] ZHENG Z, XIE S, DAI H N, et al. Blockchain challenges and opportunities: A survey[J]. International Journal of Web and Grid Services, 2018, 14(4): 352-375.

[8] TOYODA K, ZHANG A N. Mechanism design for an incentive-aware blockchain-enabled federated learning platform[C]//Proceedings of the 2019 IEEE International Conference on Big Data (Big Data). Piscataway:IEEE, 2019: 395-403.

[9] HOU D, ZHANG J, MAN K L, et al. A Systematic Literature Review of Blockchain-based Federated Learning: Architectures, Applications and Issues[C]//Proceedings of the 2nd Information Communication Technologies Conference (ICTC). Piscataway:IEEE, 2021: 302-307.

[10] WAHAB O A, MOURAD A, OTROK H, et al. Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems[J]. IEEE Communications Surveys & Tutorials, 2021, 23(2): 1342-1397.

[11] 李凌霄,袁莎,金银玉.基于区块链的联邦学习技术综述[J/OL](2021-06-25)[2021-08-28].https://doi.org/10.19735/j.issn.1001-3795.2021.04.0094.(LI L X,YUAN S,JIN Y Y.Review of blockchain-based federated learning[J/OL].(2021-06-25)[2021-08-15] .https://doi.org/10.19735/j.issn.1001-3795.2021.04.0094.)

[12] 邵俊,蔺静茹.基于区块链的联邦学习应用研究[J].中国新通信,2021,23(5):124-125.(SHAO J,LIN J R. Application Research of Federated Learning based on blockchain[J]. China New Telecommunications,2021,23(5):124-125)

[13] KIM H, PARK J, BENNIS M, et al. Blockchained On-Device Federated Learning[J]. IEEE Communications Letters, 2020, 24(6):1279-1283.

[14] QU Y, GAO L, LUAN T H, et al. Decentralized Privacy Using Blockchain-Enabled Federated Learning in Fog Computing[J]. IEEE Internet of Things Journal, 2020, 7(6):5071-5083.

[15] POKHREL S R, CHOI J. Federated learning with blockchain for autonomous vehicles: Analysis and design challenges[J]. IEEE Transactions on Communications, 2020, 68(8): 4734-4746.

[16] AWAN S, LI F, LUO B, et al. Poster: A reliable and accountable privacy-preserving federated learning framework using the

blockchain[C]//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security.New York:ACM,2019: 2561-2563.

[17] HIEU N Q ,TT A,LUONG N C , et al. Resource Management for Blockchain-enabled Federated Learning: A Deep Reinforcement Learning Approach[EB/OL]. (2020-05-01)[2021-08-20].https://arxiv.org/pdf/2004.04104v2.pdf.

[18] MAJEED U, HONG C S. FLchain: Federated learning via MEC-enabled blockchain network[C]//Proceedings of the 20th Asia-Pacific Network Operations and Management Symposium (APNOMS). Piscataway: IEEE, 2019: 1-4.

[19] LI Z, LIU J, HAO J, et al. CrowdSFL: A secure crowd computing framework based on blockchain and federated learning[J]. Electronics, 2020, 9(5): 773.

[20] LU Y, HUANG X, ZHANG K, et al. Blockchain empowered asynchronous federated learning for secure data sharing in internet of vehicles[J]. IEEE Transactions on Vehicular Technology, 2020, 69(4): 4298-4311.

[21] KANG J, XIONG Z, JIANG C, et al. Scalable and communication-efficient decentralized federated edge learning with multi-blockchain framework[C]//Proceedings of the 2nd International Conference on Blockchain and Trustworthy Systems. Cham:Springer, 2020: 152-165.

[22] LU Y, HUANG X, DAI Y, et al. Blockchain and federated learning for privacy-preserved data sharing in industrial IoT[J]. IEEE Transactions on Industrial Informatics, 2019, 16(6): 4177-4186.

[23] REHAMN M ,SALAH K,DAMIANI E , et al. Towards Blockchain-Based Reputation-Aware Federated Learning[C]// Proceedings of the 2020 IEEE INFOCOM Conference on Computer Communications Workshops. Piscataway: IEEE, 2020:183-188.

[24] ZHANG W, LU Q, YU Q, et al. Blockchain-based federated learning for failure detection in industrial IoT[J]. IEEE INTERNET OF THINGS JOURNAL,2021:8(7):5926-5937.

[25] LU Y, HUANG X, ZHANG K, et al. Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks[J]. IEEE Transactions on Industrial Informatics, 2020, 17(7): 5098-5107.

[26] ZHAO Y, ZHAO J, JIANG L, et al. Privacy-preserving blockchain-based federated learning for IoT devices[J]. IEEE Internet of Things Journal, 2020, 8(3): 1817-1829.

[27] MA C, LI J, DING M, et al. When Federated Learning Meets Blockchain: A New Distributed Learning Paradigm[EB/OL]. (2021-06-04)[2021-08-21]. https://arxiv.org/pdf/2009.09338.pdf

[28] NAGAR A. Privacy-preserving blockchain based federated learning with differential data sharing[EB/OL]. (2019-12-10)[2021-08-21]. https://arxiv.org/ftp/arxiv/papers/1912/1912.04859.pdf

[29] QU Y, POKHREL S R, GARG S, et al. A Blockchained Federated Learning Framework for Cognitive Computing in Industry 4.0 Networks[J]. IEEE Transactions on Industrial Informatics, 2020, PP(99): 2964-2973.

[30] SHAYAN M, FUNG C, YOON C J M, et al. Biscotti: A blockchain system for private and secure federated learning[J]. IEEE Transactions on Parallel and Distributed Systems, 2020, 32(7): 1513-1525.

[31] BAO X, SU C, XIONG Y, et al. Flchain: A blockchain for auditable federated learning with trust and incentive[C]//Proceedings of the 5th International Conference on Big Data Computing and Communications (BIGCOM). Washington, DC: IEEE Computer Society, 2019: 151-159.

[32] ZHANG K, HUANG H, GUO S, et al. Blockchain-Based Participant Selection for Federated Learning[C]//Proceedings of the 2020 International Conference on Blockchain and Trustworthy Systems. Cham:Springer, 2020: 112-125.

[33] CHEN H, ASIF S A, PARK J, et al. Robust Blockchained Federated Learning with Model Validation and Proof-of-Stake Inspired Consensus[EB/OL]. (2021-01-09)[2021-08-23]. https://arxiv.org/pdf/2101.03300v1.pdf

[34] MARTINEZ I, FRANCIS S, HAFID A S. Record and reward federated learning contributions with blockchain[C]//Proceedings of the 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). Piscataway: IEEE, 2019: 50-57.

[35] 任涛,金若辰,罗咏梅.融合区块链与联邦学习的网络入侵检测算法[J].信息网络安全,2021,21(7):27-35.(REN T,JIN R C,LUO Y M. Network Intrusion Detection Algorithm Integrating Blockchain and Federated Learning[J]. Netinfo Security,2021,21(7):50-57.)

[36] 冯涛,焦滢,方君丽,等.基于联盟区块链的医疗健康数据安全模型[J].计算机科学, 2020, 47(4): 305-311.(FENG T,JIAO Y,FANG J L, et al. Medical Health Data Security Model Based on Alliance Blockchain[J].Computer Science, 2020, 47(4): 305-311.)

[37] KUMAR S, DUTTA S, CHATTURVEDI S, et al. Strategies for enhancing training and privacy in blockchain enabled federated learning[C]//Proceedings of the 6th IEEE International Conference on Multimedia Big Data (Big MM). Washington, DC: IEEE Computer Society, 2020: 333-340.

[38] 中国石油大学(华东). 一种基于区块链和联邦学习的物联网设备故障检测方法:CN202011087722.7[P]. 2021-01-05.(China University Of Petroleum. A failure detection method for Internet of things device based on blockchain and federated learning:CN202011087722.7[P]. 2021-01-05.)

[39] 代玥玥,张科,张彦.区块链赋能 6G[J].物联网学报, 2020, 4(1):111-120.(DAI Y Y,ZHANG K,ZHANG Y. Blockchain empowered 6G[J].Chinese Journal on Internet of Things,4(1):111-120.)

[40] PASSERAT PALMBACH J, FARNAN T, MILLER R, et al. A blockchain-orchestrated federated learning architecture for healthcare consortia[EB/OL].(2019-10-12)[2021-08-24].https://arxiv.org/pdf/1910.12603.pdf

[41] EL RIFAI O, BIOTTEAU M, DE BOISSEZON X, et al. Blockchain-based federated learning in medicine[C]//Proceedings of the 2020International Conference on Artificial Intelligence in Medicine.,Cham:Springer, 2020: 214-224.

[42] POLAP D, SRIVASTAVA G, JOLFAEI A, et al. Blockchain technology and neural networks for the internet of medical things[C]//Proceedings of the 2020-IEEE conference on computer communications workshops (INFOCOM WKSHPS). Piscataway: IEEE, 2020: 508-513.

[43] RAHMAN M A, HOSSAIN M S, ISLAM M S, et al. Secure and provenance enhanced Internet of health things framework: A blockchain managed federated learning approach[J]. IEEE ACCESS, 2020, 8: 205071-205087.

[44] AICH S, SINAI N K, KUMAR S, et al. Protecting Personal Healthcare Record Using Blockchain & Federated Learning Technologies[C]//Proceedings of the 23rd International Conference on Advanced Communication Technology (ICACT). Piscataway: IEEE, 2021: 109-112.

[45] KUMAR R, KHAN A A, KUMAR J, et al. Blockchain-federated-learning and deep learning models for covid-19 detection using ct imaging[J]. IEEE Sensors Journal, 2021,21(14):16301-16314.

[46] 邢丹,徐琦,姚俊明.边缘计算环境下基于区块链和联邦学习的医疗健康数据共享模型[J].医学信息学杂志,2021,42(2):33-37.(XING D,XU Q,YAO J M.Medical and Health Data Sharing Model Based on

Blockchain and Federated Learning in the Edge Computing Environment[J]. Journal of Medical Informatics,2021,42(2):33-37.)

[47] 李铮.一种支持隐私与权益保护的数据联合利用系统方案[J]. 信息与电脑, 2020(14):4-7.(LI Z. A Scheme of Data Joint Utilization System Supporting Privacy and Rights Protection[J]. China Computer & Communication,2020(14):4-7.)

[48] CHAI H,LENG S,CHEN Y , et al. A Hierarchical Blockchain-Enabled Federated Learning Algorithm for Knowledge Sharing in Internet of Vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, PP(99):1-12.

[49] CUI L,SU X X,MING Z X , et al. CREAT: Blockchain-assisted Compression Algorithm of Federated Learning for Content Caching in Edge Computing[J]. IEEE Internet of Things Journal, 2020, PP(99):1-1.

[50] HALIM S M D, KHAN L, THURAISINGHAM B. Next-location prediction using federated learning on a blockchain[C]//Proceedings of the 2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI). Piscataway: IEEE, 2020: 244-250.

[51] WANG Y,SU Z,ZHANG N , et al. Learning in the Air: Secure Federated Learning for UAV-Assisted Crowdsensing[J]. IEEE Transactions on Network Science and Engineering, 2020, PP(99): 1055 - 1069.

[52] POKHREL S R. Federated learning meets blockchain at 6g edge: A drone-assisted networking for disaster response[C]//Proceedings of the 2nd ACM MobiCom Workshop on Drone Assisted Wireless Communications for 5G and Beyond. New York: ACM,2020: 49-54.

[53] WANG Y,TIAN Y,YIN X , et al. A trusted recommendation scheme for privacy protection based on federated learning[J]. CCF Transactions on Networking, 2020, 3(3-4):218-228.

This work is partially supported by the National Key Research and Development Program of China (2020YFB2103802), the Fundamental Research Funds for the Central Universities (2019RC038).

SUN Rui, born in 1998, M. S. candidate. Her research interests include blockchain.

LI Chao, born in 1988, Ph. D., assistant professor. His research interests include blockchain, data privacy.

WANG Wei, born in 1976, Ph. D., professor. His research interests include network and system security, industrial Internet Security and blockchain security.

TONG Endong, born in 1986, Ph. D., assistant professor. His research interests include service computing, artificial intelligence and security.

WANG Jian, born in 1975, Ph. D., associate professor. His research interests include password application, blockchain and network security.

LIU Jiqiang, born in 1975, Ph. D., professor. His research interests include trusted computing, privacy protection and cloud computing security.