计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202330899
61(9): 2246−2260, 2024

基于分片区块链的车联网数据共享方案

陈 骁[1] 黄牧鸿[1] 田一凡[1] 王 岩[1] 曹 晟[1,2] 张小松[1,2]
[1](电子科技大学计算机科学与工程学院 成都 611731)
[2](电子科技大学(深圳)高等研究院 广东深圳 518110)
(chenxiao_hey@std.uestc.edu.cn)

Internet of Vehicles Data Sharing Scheme via Blockchain Sharding

Chen Xiao[1], Huang Muhong[1], Tian Yifan[1], Wang Yan[1], Cao Sheng[1,2], and Zhang Xiaosong[1,2]
[1](*School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu* 611731)
[2](*Shenzhen Institute for Advanced Study , University of Electronic Science and Technology of China, Shenzhen, Guangdong* 518110)

Abstract Efficient and secure data sharing is crucial for the profound application of the intelligent Internet of vehicles, where achieving trusted data sharing between mutually untrusting vehicles has become a major focus of current research. With the characteristics of tamper resistance and traceability, blockchain has emerged as a primary approach to support data circulation in the intelligent Internet of vehicles. Existing blockchain-based data-sharing solutions for the Internet of vehicles suffer from low throughput and security vulnerabilities. In this paper, the blockchain sharding approach is introduced, where a machine learning-based sharding algorithm is utilized to partition road side unit (RSU) with geographical proximity into the same shard and iteratively optimize data-sharing loads within individual shards. This algorithm reduces intra-shard communication latency and subsequently improves throughput while balancing the data-sharing loads among different shards. To prevent bribery attacks within a single shard, a reputation-based intra-shard consensus protocol and the supervisor mechanism are proposed. The proposed protocol involves the election of RSUs with high reputation to participate in the intra-shard consensus process and dynamically calculates the latest reputation of RSUs. High-reputation RSUs are designated supervisors and regularly verify the legitimacy of blocks generated in different shards. Performance evaluation and security analysis demonstrate the scheme enhances the efficiency and security of data sharing in the intelligent Internet of vehicles.

Key words Internet of vehicles; blockchain sharding; consensus protocol; supervisor mechanism; data sharing

摘 要 高效安全的数据共享对于智能车联网的深度应用至关重要,在相互不信任的车辆之间实现可信的数据共享成为当前研究的热点.区块链技术以其防篡改、可追溯等特点,成为支撑智能车联网数据共享流通的主要途径之一.现有基于区块链的车联网数据共享方案,存在吞吐量小、安全性低等不足.引入区块链分片方法,提出基于机器学习的分片算法,将地理位置相近的路侧单元(road side unit, RSU)划分到同一分片,并迭代单个分片的数据共享最优负载,降低了片内通信延迟进而提高了吞吐量,平衡了不同分片之间的数据共享负载.为避免单个分片的贿赂攻击,提出了基于声誉的片内共识协议与监督人机制.选举具有高声誉的RSU参与片内共识过程,并动态计算RSU的最新声誉.设定声誉度高的RSU担任监督员,监督员可定期对不同分片产生的区块进行合法性验证.通过性能评估和安全性分析,证明方案有助于提

收稿日期:2023-11-01;**修回日期**:2024-04-25
基金项目:国家重点研发计划项目(2023YFB3105900);四川省重点研发计划项目(2023YFG0118)
This work was supported by the National Key Research and Development Program of China (2023YFB3105900) and the Key Research and Development Program of Sichuan Province (2023YFG0118).
通信作者:曹晟(caosheng@uestc.edu.cn)

升智能车联网数据共享的高效性和安全性.

关键词 车联网；分片区块链；共识协议；监督人机制；数据共享

中图法分类号 TP311.13；TP309

随着车联网中车辆数量的不断增加，车辆之间数据共享交换的速度和频率越来越高，安全高效的智能车联网数据共享方案已成为研究热点. 其中，基于区块链的方案持续涌现，成为车联网数据共享的重要途径. 目前，基于区块链的车联网数据共享方案主要有：基于区块链的联邦学习数据共享方案[1-2]，使用区块链代替联邦学习中的中心化服务器，提出了面向分布式终端的去中心化数据共享架构与系统；基于新型共识协议的数据共享方案[3-4]，针对不断变化的车联网环境，提出适用于不同场景的共识机制；基于区块链访问控制的数据共享方案[5-6]，提出新的加密方式实现对区块的细粒度访问控制. 以上方法所使用的区块链系统吞吐量较为有限，难以应对未来智能车联网中海量数据共享需求.

区块链分片技术[7]旨在提高吞吐量以实现区块链系统的可扩展性. 该技术将区块链网络节点分成多个分片(shard)，每个分片只需要处理一部分交易数据，多个分片并行执行. 主要的分片区块链技术，如Elastico[8]，RapidChain[9]，OmniLedger[10]，ChainSpace[11]，MVCom[12]，在提升系统吞吐量的同时，为了保证分片的安全性，采用了多种密码技术将区块链节点随机分配到各个分片中，以抵御敌手对片内节点的贿赂，防止片内恶意节点数量超过单个分片安全阈值.

将区块链验证节点随机地分配到不同分片，在理论上能最大程度地保证分片区块链的安全性，然而难以适配实际的车联网应用环境，原因在于：1)地理位置上接近的车联网区块链节点可能被分配到不同的分片当中，使得车辆之间的数据共享协同效率降低；2)随机分片的方案对车联网区块链分片的负载问题考虑较少，可能导致不同分片之间的负载存在较大差异，造成车联网基础设施资源(带宽、算力等)分配不均匀和利用不充分. 因此，亟需适用于实际车联网应用的区块链分片方案，该方案面临3个挑战：

1)现有的分片算法缺乏实现片内节点距离相近与片间负载均衡的能力，仅能够取得这2个要素中单一目标的较优解，无法同时优化2个要素；

2)各个分片内恶意路侧单元的数量未知，当分片内的恶意路侧单元(road side unit, RSU)数量过多时，由恶意RSU主导的共识产生区块上链；

3)为保证分片系统的安全性，现有方案需要定期对各个分片实行轮换和置乱等节点重新配置，带来了节点额外数据下载开销，且需要维持较大的单个片内节点数量，降低了片内共识过程效率.

针对上述3个挑战，本文设计了一种基于分片区块链的车联网数据共享方案. 主要贡献包括3个方面：

1)提出了一种基于机器学习的分片算法，该算法基于聚类实现片内节点在地理位置上的紧凑分布，借助平衡损失实现分片间在负载上的均匀分布，实现高吞吐量的车联网区块链架构；

2)提出了基于声誉的片内共识协议，通过片内各个RSU参与共识的行为评估其声誉，高于设定的声誉阈值的RSU可参与片内共识，确保高声誉节点主导片内共识，抵御片内恶意RSU数量过多导致的安全问题；

3)提出了监督人机制，选取声誉度较高的节点作为监督人，可以随机、定期验证不同分片节点交易验证的合法性，降低了对片内节点数量较大的要求，对提升分片系统的安全和效率都有重要作用.

1 相关工作

区块链是一种去中心化的分布式账本技术，通过将交易记录按照时间顺序连接成块，并使用密码学方法保证数据的安全性和完整性，实现了可信、不可篡改的数据存储. 将区块链技术应用于智能车联网，为车联网提供了去中心化的数据共享方式，减少了单点故障的风险. 在提升性能方面，文献[13]提出了一种车联网区块链资源管理方案，通过优化RSU和利用车辆的计算资源来提升吞吐量. 文献[14]设计了一种增强的委托权益证明(delegated proof of stake, DPoS)共识协议，加速了交易确认的同时缩短了数据共享时间，有效提升了车联网系统的吞吐量. 文献[15]提出了一种新型的区块验证方法，每个节点存储自身的交易验证记录，如果之前已验证过发送节点的区块，就简化验证过程，从而提升系统性能. 在增强安全性方面，文献[16]提出了一种基于区块链的车联网可信数据共享机制，利用基于Kademlia算法的流量数据转发方法来控制车联网中信道的拥

塞状态，保证了重要数据在传输过程中的安全性. 文献 [17] 设计了一种双链并行的车联网信息共享方案，通过智能合约的自动执行实现数据的安全流通，并与区域内其他用户共享车联网数据. 文献 [18] 提出一种基于区块链和加权密文的车联网数据安全共享方法，基于特征对链上数据进行访问控制，确保只有授权的访问者才能访问数据内容，实现了车联网中安全的数据共享. 文献 [19] 设计了基于零知识证明的区块链数据共享方案，以匿名方式将数据和车辆信息相结合，并允许其他车辆和 RSU 匿名验证数据记录，实现安全的数据共享.

文献 [13−19] 方案主要是将区块链技术引入车联网数据共享与安全领域，对区块链系统的吞吐量性能考虑较少. 分片机制是提升吞吐量的主要手段之一，它将区块链节点划分为多个分片，各个分片可以并行地、独立地处理各自的交易和数据，这使得区块链的吞吐量能够线性扩展，即吞吐量可以随着节点数的增加而线性增长. 因此，分片技术可以用来解决区块链的可扩展性问题，且同时保持较高的安全性和去中心化程度. 现有的区块链分片方案，一般从分片节点的随机分配以保障安全性[8-12]、降低跨分片开销以提升吞吐量[20-22]、分片之间负载均衡以避免热点分片[12,23-25]、分片安全与效率平衡[26-27]、智能合约分片[28-29] 等多个方面开展研究.

在分片区块链安全性方面，文献 [9] 利用有限布谷鸟原则以保证不同分片规模的平衡，并使用可验证秘密共享来生成无偏随机性，要求新加入节点解决工作量证明(proof of work, PoW)难题加入分片协议，从而抵御女巫攻击. 文献 [10] 中的每个共识周期，依赖可验证随机函数 (verifiable random function,VRF) 进行不可预测的领导者选举. 协议将领导者视为 RandHound 协议的客户来生成随机种子. 此外，还可以通过引入声誉机制来确保系统的安全与可信. 文献 [30] 提出一种基于区块链的隐私保护车辆数据共享框架，使用零知识证明技术以保护车辆的身份隐私，同时设计高效的分片协议降低系统的通信成本，增强系统的安全性. 针对分片内部共识安全问题，文献 [31] 基于车辆间历史交互信息和其他车辆的推荐信息评估该车辆的声誉，并赋予高声誉节点出块权力，从而确保数据共享的安全性和可追溯性. 然而，文献 [9−10, 30−31] 所述的方法要求对分片的高频率重构，且忽视了将参与共识的行为作为节点声誉的评估标准，难以适用于实际车联网动态变化环境，需要提出更简单高效的分片安全性保护方法和对节点的声誉评估方法.

2 预备知识

2.1 区块链分片

将区块链分片协议进行形式化. 给定一个区块链网络，其中包含n个节点，每个节点维护全网的一部分交易历史，其中有f个拜占庭节点. 目标是将该区块链网络分割成k个分片，其中$k \ll n$，以提高网络的可扩展性和性能. 每个分片$i(1 \leqslant i \leqslant m)$包含 1 组节点，记作$shard_i$，$shard = \bigcap_{i=1}^{m} shard_i$，且$shard_i \cap shard_j = \varnothing$，即所有节点被分片覆盖且分片间的节点互不相交. 区块链节点之间运行分片协议，输出一个集合X，其中包含k个不相交的分片或子集$X_i = \{x_i^j\}$ $(1 \leqslant j \leqslant |X_i|)$. 区块$i$中的交易$j$被定义为$x_i^j$. 节点可以通过使用函数$C$给定任何交易，输出 1 或 0 分别表示交易是否有效，以确定交易的合法性.

1)一致性. 对于$i \in \{1,2,\cdots,k\}$，所有诚实的节点对X_i达成一致.

2)有效性. 对于$i \in \{1,2,\cdots,k\}$和$j \in \{1,2,\cdots,|X_i|\}$，$C(x_{i,j}) = 1$.

3)可扩展性. k随着网络规模或n的增长而线性增长.

2.2 基于声誉的共识协议

共识协议的主要功能是存在拜占庭节点的情况下，确保诚实节点能够维护一个一致的账本. 声誉机制在现代社会中扮演着至关重要的角色. 在人们进行决策时，往往会依赖公开的且基于声誉的排名系统来做出选择. 将声誉机制与共识协议相结合可以实现 3 个重要目标：

1)对参与共识的节点的历史行为进行评估，从而实现对节点的良性行为进行褒奖和对不当行为进行追责[32-33]；

2)基于节点的声誉，公平地选举诚实的节点参与共识并且提升协议的可扩展性[34-35]；

3)将节点投票权重转换为声誉，当系统中的恶意节点超过容错上限时，能使共识协议在一定时间内保证安全性与活性[32,34].

3 基于分片区块链的车联网数据共享方案

3.1 系统模型

本文设计了一个支持分片的车联网区块链架构，由车载单元(on board unit, OBU)、RSU 和可信机构

(trusted authority, TA)组成. 系统模型及组件之间的关系如图 1 所示.

1)OBU. OBU 通常依附于智能汽车，具备计算、存储和无线通信能力，可以与其他 OBU、RSU、TA 进行通信，在通信之前需要向 TA 发起并完成注册.

2)RSU. RSU 是分片内参与共识的节点. 其主要职责是接收来自 OBU 的数据，并与其他 RSU 达成共识后，最终将数据上传至本地分布式账本. 在此过程中，假设每个分片中各个 RSU 的公钥是相互知晓的.

3)TA. TA 为合法 OBU 和 RSU 颁发证书，并通过安全通道分发车辆和 RSU 的公钥和私钥.

OBU 之间的通信以及 OBU 与 RSU 之间的通信采用 IEEE 802.11p[36] 或 DSRC[37] 协议，以上协议在数据传输过程中具有低延迟、高稳定的性质.

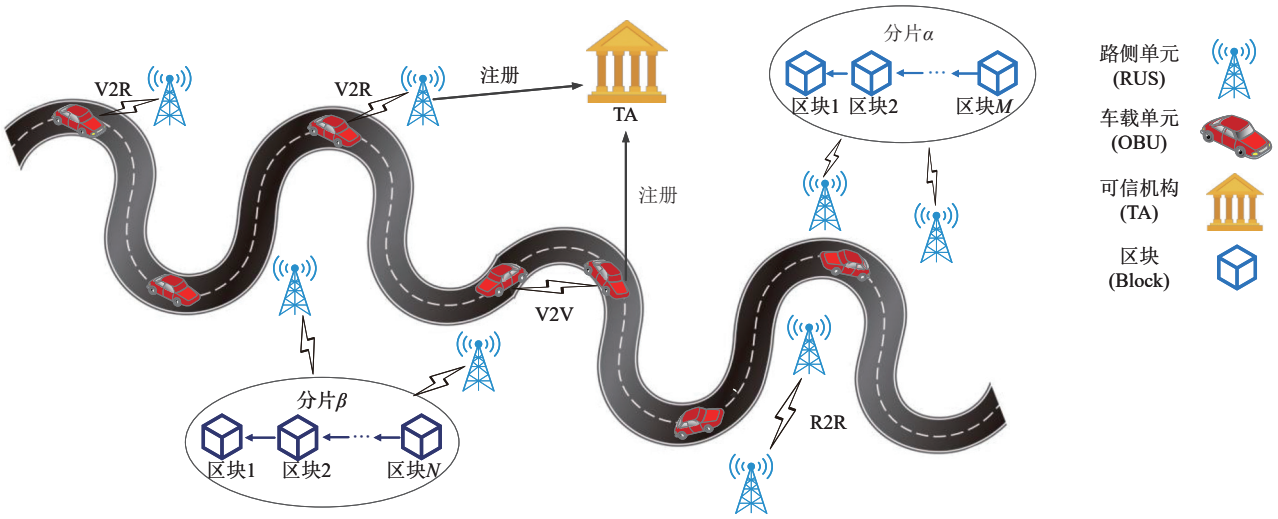


Fig. 1 System model

图 1 系统模型

3.2 基于机器学习的分片算法

区块链中，节点数据可以被转化基于属性的表示形式，以便机器学习算法进行分析和应用于分片任务. 首先需要对基于机器学习的车联网分片问题进行定义.

定义 1. 基于机器学习的车联网分片. 给定一个由 N 个 RSU 组成的集合：

$$V=\{c_i=(x_i,y_i),w_i|x_i,y_i\in\mathbb{R},w_i\in(0,+\infty),i=1,2,\cdots,N\}. \tag{1}$$

其中 c_i 代表 RSU 的 2 维位置坐标，表示 RSU 所在的地理位置；w_i 代表 RSU 的负载，表示 RSU 一定时间内的平均负载. 基于机器学习的车联网分片需要通过机器学习模型，最小化以下目标函数：

$$\arg\min\sum_{i=1}^{k}\sum_{v\in X_i}w+\sum_{i=1}^{k}\sum_{v\in X_i}\|c-\mu_i\|, \tag{2}$$

其中 v 代表 RSU 节点，k 代表分片数，X_i 代表第 i 个分片，μ_i 代表归属于第 i 个分片的所有 RSU 的位置均值，$\|c_i-\mu_j\|$代表两点之间的欧氏距离.

现有的聚类算法一般是通过将邻近节点聚合成簇实现分片，满足片内节点相近的需求. 在车联网区块链分片中，若分片间负载不均衡，会导致基础设施资源分配失衡，进而造成资源浪费. 例如，在下班高峰时段，某个分片的车流量激增，导致该分片的 RSU 负载显著增加，无法高效地处理车辆数据，而其他分片的计算资源却处于闲置状态. 因此，实现负载均衡的分片能避免资源浪费，并提高数据共享效率. 本文提出了一种面向车联网区块链分片的均衡聚类算法，该算法在 K-Means++[38] 算法上使用正则项约束簇间负载以尽可能保持均衡，其损失函数为

$$L(c_i,w_i,\mu_j)=\|c_i-\mu_j\|+w_i(1+\lambda)^{T-M+w_i}, \tag{3}$$

$$\lambda=\exp\left(\frac{\ln\bar{w}-\ln M}{\bar{w}-M}-1\right), \tag{4}$$

$$T=\sum_{v\in X_j}w, \tag{5}$$

$$M=\frac{\sum w}{k}, \tag{6}$$

其中 T 代表第 j 个簇内所有节点负载之和，M 代表每个簇的最优负载值，λ代表损失函数的惩罚系数.

由式(3)可知，该损失函数由 2 部分组成：第 1 部分是距离损失$\|c_i-\mu_j\|$，其代表某个点与最近聚类中心的欧拉距离，与 K-Means 算法中的距离损失相同；

第 2 部分是正则项 $w_i(1+\lambda)^{T-M+w_i}$, 其代表分片负载均衡的惩罚项. 采用贪心算法为节点分配聚类簇, 若分配的簇加入该节点后权重超过簇的最优负载值, 惩罚项则会呈现指数增大, 进而实现负载均衡的目标; 而 λ 作为自适应权重, 能够根据节点的权值与质心数自动调整, 使得惩罚处于合理的范围内.

基于该损失函数的优化过程如算法 1 所示.

算法 1. 均衡聚类算法.

输入: 节点集 V, 分片数 k;

输出: 聚类结果.

① 用式(4)~(6)初始化 λ, T, M;
② for *iter* in *max_iters*
③ clear *cost*[k];
④ *previous_assignment* = *assignment*;
⑤ for each v_i sampled from V
⑥ assign v_i to k_j by minimizing 式(3);
⑦ *cost*[k_j] += w_i;
⑧ end for
⑨ for each k_j in k
⑩ *centers*[k_j] = *mean*(c_i where *assignment* == k_j);
⑪ end for
⑫ if *previous_assignment* == *assignment*
⑬ break;
⑭ end if
⑮ end for

算法 1 采用贪心策略搜索最优的分片策略. 每次遍历将为当前节点分配损失最小的簇, 并更新对应簇的负载值, 直至算法收敛. 在搜索前期, 惩罚项指数为负, 该阶段的分配策略是距离优先; 随着节点的不断分配, 部分簇的负载已接近最优负载值, 惩罚项会逐步增大, 该阶段的分配策略是混合分配; 在搜索末期, 绝大多数簇已达到最优负载, 分配策略是负载优先. 这种策略能够确保算法在前期形成簇状聚类结构, 并在搜索过程中不断优化, 进而在维持聚类结构的同时实现负载均衡. 值得注意的是, 由于 RSU 的数据负载随着时间不断变化, 因此每隔一定时间需重新运行均衡聚类算法以生成新的区块链分片.

3.3 基于声誉机制的片内共识

通过基于机器学习的分片算法将 RSU 分入若干个分片后, 为了保证片内共识协议安全地运行, 需要对片内 RSU 进行筛选, 从而防止恶意的 RSU 主导共识过程. 本节对片内 RSU 选举的方式和参与共识的过程进行描述.

3.3.1 设计概览

单个分片内安全地运行共识协议存在 2 个挑战: 1)基于机器学习的分片无法保证片内节点多数诚实, 单个分片可能会被贿赂; 2)当每个分片内 RSU 数量过多时, 需要共识协议具备一定的可扩展性, 从而保持系统性能的稳定.

为保证共识协议的最大容错能力, 假设系统处于同步网络模型中, 消息在 2 个 RSU 间的传播的延迟上限为 Δ, 能容忍小于 1/2 的拜占庭节点. 为了保证分布式账本确定的一致性, 需要在分片中设计一个同步拜占庭容错(Byzantine fault-tolerant, BFT)协议. 同时, 由于各个分片中诚实的 RSU 数量未知, 无法确保片内共识协议能安全运行. 对此, 需要对每个 RSU 进行声誉的评估, 通过声誉显式表达 RSU 的可信度. 此外, 需要基于各个 RSU 的声誉选举高声誉的 RSU 参与片内共识. 基于声誉的状态机复制(reputation-based state machine replication, Reputation-based SMR)协议[32]是一个同步 BFT 协议, 能对参与共识的节点行为进行评估, 从而为每个节点评估所有节点公认的声誉. 当 RSU 作为参与共识的节点时, 该协议能针对 RSU 的扣留区块(即 RSU 作为领导者拒绝提议区块)、提议冲突区块(即 RSU 作为领导者提议 2 个及 2 个以上的区块)、恶意投票(即 RSU 投票给非领导者区块)、恶意区块提议(即非领导者 RSU 提议区块) 4 个恶意行为对 RSU 的声誉进行削减. 同时, 该协议能针对 RSU 的正常区块提议与投票增加 RSU 的声誉, 因此该协议拥有普适性, 能够很好地适用于车联网环境.

在 Reputation-based SMR 协议上设计片内共识节点选举函数, 选择声誉高于阈值的 RSU 参与共识, 避免声誉低的恶意节点主导共识. 同时, 选举高声誉的节点参与共识而非分片内所有节点降低了协议的通信复杂度并提高了可扩展性, 协议的性能不会随着片内节点的增多而明显下降.

3.3.2 协议描述

1)协议组件

基于声誉的片内共识协议将 Reputation-based SMR 协议与片内共识节点选举函数作为协议组成部分. Reputation-based SMR 协议中的声誉机制能基于节点参与共识的行为, 对产生不当行为的拜占庭节点声誉削减, 并给予产生良性行为的节点声誉褒奖. 此外, Reputation-based SMR 协议的证据广播阶段保证了声誉一致性.

定义 2. Reputation-based SMR 协议. 由领导者选

举函数、拜占庭广播原语和声誉机制组成. Reputation-based SMR 协议满足 3 个属性：

1)安全性. 任意 2 个诚实节点在同一高度上不会提交不同的区块.

2)声誉一致性. 在同一区块高度，任意 2 个诚实节点不会对某一节点产生不同的声誉视图.

3)活性. 如果 1 笔交易被 1 个诚实节点接受，那么这笔交易最终会被包含在所有诚实节点的账本上.

基于 Reputation-based SMR 协议构建片内共识节点选举函数，确保高声誉的节点参与共识. 具体过程如算法 2 所示.

算法 2. 片内共识节点选举函数.

输入：分片内 RSU 的总数 m，分片内 RSU 的公钥集合 PK，分片内 RSU 的声誉集合 U，当前轮次数 q；

输出：高声誉 RSU 的公钥集合 PK'.

① 初始化 $PK' \leftarrow \varnothing$；

② $x \leftarrow H(pk_i, q)$；

③ if $\tau r_i - \dfrac{x}{2^{\eta}} > 0$ then

④ $\quad PK' \leftarrow PK' \cup p_i$；

⑤ else

⑥ $\quad PK' \leftarrow PK'$；

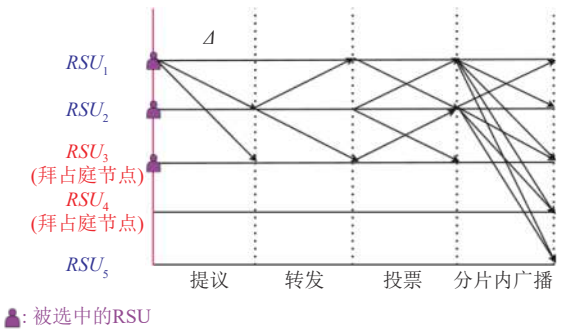
⑦ end if

算法 2 中，r_i代表任意RSU_i的声誉；τ代表难度系数，控制各个节点被选举的难度；η代表所选加密哈希函数的位长. 可以看出，RSU 声誉越高，越容易超过给定阈值，从而担任共识节点. 反之亦然.

2)基于声誉的片内共识协议

通过片内共识节点选举函数与 Reputation-based SMR 协议构建基于声誉的片内共识协议. 协议流程如图 2 所示.

基于声誉的片内共识协议在由片内共识节点选举函数选举出的高声誉节点中执行提议、转发和投票步骤. 此外，基于声誉的片内共识协议通过分片内广播的共识结果与分片内所有 RSU 共享. 下面描述该协议的详尽过程.



注：各个节点更新自身当前轮次声誉，并执行片内共识节点选举函数.

Fig. 2 Flow chart of reputation-based intra-shard consensus protocol

图 2 基于声誉的片内共识协议流程图

1)节点更新声誉. 此步骤与 Reputation-based SMR 协议下的节点更新声誉步骤相同. RSU 执行 Reputation-based SMR 协议中附带的声誉函数更新当前轮次的声誉并开启一个 4Δ的计时器.

2)区块生成. 分片内的 RSU 执行片内共识节点选举函数，选举高声誉 RSU 参与共识，并且被选举的高声誉 RSU 被公知. 在此基础上，通过 Round-Robin 选举领导者来进行区块提议. 如果某个 RSU 被片内共识节点选举函数选中，它将参与共识协议；否则，它将等待分片内广播传播的区块与各类证据.

3)片内广播. 这个阶段与 Reputation-based SMR 协议的证据广播阶段相似. 当计时器仅剩Δ时，参与共识流程的 RSU 将在提议、转发和投票过程中产生的区块和各类证据(如扣留区块、良性区块提议等)广播给分片内所有节点. 分片内所有诚实 RSU 将会对这些内容达成共识.

4)区块提交. 此阶段与 Reputation-based SMR 协议的区块提交阶段相同. RSU 根据在此轮共识中是否出现扣留区块证据或冲突区块证据来判定是否提交空块.

然而，当分片内恶意 RSU 数量过多时(如超过 1/2 时)，无法以极大概率选举多个诚实的 RSU 参与共识. 此时，恶意 RSU 将有可能终止甚至主导片内的共识过程，并产生非法区块，从而影响分布式账本上区块的合法性. 在 3.4 节中将用监督人机制来解决这个问题.

3.4 监督人机制

分片机制将区块链节点划分为多个分片，每个分片内的分片规模远小于整个区块链网络的规模. 通常情况下，在相同的节点规模的基础上，分片数量的增加，即每个分片内节点数量的减少，有助于提高系统的吞吐量.

然而，片内节点数的减少会降低分片系统的安全性，例如：单个分片容易遭受贿赂攻击[39]. 节点规模为 4 000 个，未分片的区块链系统，敌手需要攻击区块链系统全网的 1/3 节点(1 667 个节点)；分片后，在 16 个分片情况下，敌手仅需攻击分片系统中单个分片内的 1/3 节点(84 个节点).

同时，对片内共识协议能够支持的节点数量提出了新的要求. 例如：使用 PBFT 作为片内共识协议

时, 由于该算法的通信复杂度为$O(n^2)$, 当分片内节点数量增多时, 其共识效率会大幅度降低, 导致分片系统的总吞吐量快速下降. 因此, 现有分片系统往往需要选择能够支持更多节点的片内共识协议[10].

采用超几何分布来计算每个分片在不同分片规模下, 随机挑选出一个分片, 其恶意节点数量大于1/2的概率(分片内恶意节点数量超过分片安全阈值).

$$Pr\left[X \geqslant \left\lfloor \frac{m}{2} \right\rfloor\right] = \sum_{x=\left\lfloor \frac{m}{2} \right\rfloor}^{m} \frac{\binom{t}{x}\binom{n-t}{m-x}}{\binom{n}{m}}, \tag{7}$$

其中m代表分片规模, n代表区块链网络中的总节点数, t代表区块链网络中恶意节点的数量, x则是被随机选中组成一个分片的恶意节点数量.

当$t=n/3$时, 即系统容忍不超过1/3的恶意节点. 图3展示了使用超几何分布模拟在4 000个节点的场景下, 随机选择一个分片, 并计算该分片中恶意节点比例超过1/2的概率. 由图3可以看出, 分片规模越大, 选取分片失败的概率越低, 区块链系统的安全性越高, 当分片规模超过250个节点时, 其失败概率是可以接受的. 例如: RapidChain[9]在网络规模为4 000个节点时, 将每个分片的节点数量限制在250个节点以保障极高的安全性, 平均每4 500年才会出现1次故障.

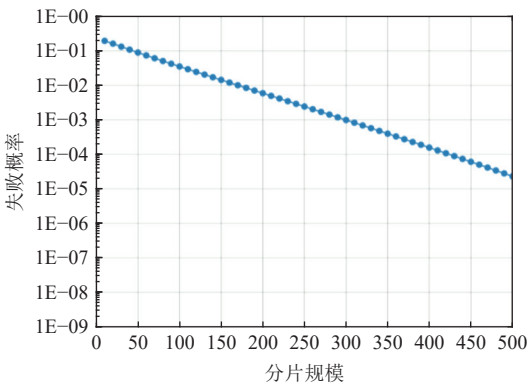


Fig. 3 Random selection shard failure probability

图3 随机挑选分片失败概率

现有的保证系统安全性的分片方案通常是定期地将各个分片中的节点进行全部打乱或者部分打乱. 然而, 这种做法会破坏分片内节点之间的地理位置关系, 不利于3.2节分片算法的实施. 此外, 还引入了节点更换分片时数据移动的额外开销. 当1个节点从分片α转移到分片β时, 它需要重新下载分片β节点所存储的所有历史数据, 以便能够验证分片β后续的数据.

重新组建整个分片需要耗费大量时间和额外成本, 但仅进行部分重新配置无法确保足够的安全性. 例如, 在RapidChain[9]中, 系统周期性地对分片进行重新划分, 以确保大多数成员仍属于活跃分片. 当有新节点加入时, 参考分片会根据一定规则随机地将新节点分片到某个活跃分片中, 并将一定数量的节点从该活跃分片随机分配到其他的消极分片中, 实现了新节点加入系统, 控制了活跃分片和消极分片数量, 达到了一种活跃节点间的平衡. 这种方式旨在平衡活跃分片数量以及不活跃分片数量, 从而提高了整个区块链分片系统的活跃性、增加了敌手共谋攻击的成本, 间接提高了系统的安全性. 但由于仅是对各个分片的部分节点进行重新配置, 因此系统安全性提高的程度非常有限.

为了解决以上这些问题, 本文提出了监督人机制, 以代替现有的分片重构方案.

根据3.3节得到每个节点的声誉度, 选取每个分片中声誉度较高的节点作为系统监督人, 它们可以监督和验证分片系统中的数据. 这意味着每个数据同时被1个分片的所有节点以及监督人所验证, 管理者可以通过调控奖惩机制来控制监督人的数量, 达到所期望的性能与安全性的平衡.

车辆在行驶过程中, 会不断地与周围的RSU进行通信. 每个RSU需要接收和传播数据, 如图1所示. RSU接收到了新的车辆数据时, 它会将带有车辆签名的数据传输给地理位置接近的若干个RSU. 在车辆数据产生时, RSU首先将数据发送至TA, TA向所有RSU广播数据, 通过车辆单元的移动和RSU之间的交互, 促使各分片之间可以共享数据, 监督人也进而具备监督并验证所有分片数据的能力. 如果发现其他分片对某次数据进行作恶, 任何一个验证了该分片数据的监督人都可以将其验证结果通知给分片系统, 从而协同决定原始分片是否存在恶意行为, 并进一步做出处理.

当节点加入系统时, 每个节点需要缴纳一部分的押金, 该押金由TA进行管理, 用于防止节点在后续处理车联网数据时出现恶意行为. 并且, 有利于避免恶意监督人以增加系统运行负担为目的而提交无效监督, 提高恶意行为门槛. 若监督人发现非法数据, 表明该数据所属分片已经被损坏, 即分片内的恶意节点数量已经超过安全阈值. 在这种情况下, 监督人提交一笔特殊交易, 特殊交易包含监督信息以及用于防止恶意举报行为的押金. 区块链节点将根据这一特殊交易验证举报信息是否有效, 验证结果分2

种情况：1)如果举报有效，则返回监督人的举报押金，扣除恶意数据中包含的所有参与签名的节点的押金，并报告至 TA, TA 根据一定规则对监督人以及作恶节点进行奖惩，并降低作恶节点声誉；2)若举报无效，说明监督人提交了恶意的举报，则扣除监督人的举报押金以及基础押金，降低监督人节点的声誉.

监督人可以在恶意分片中、正常分片中、数据发起用户或者区块链网络之外的第三方. 它们可以从共享数据中存储数据，从而拥有对分片系统中所有分片数据验证的能力. 显然，节点保存选择验证分片系统中的数据越多，监督成功并获得的奖励就越多. 这一机制会鼓励更多的节点提高声誉度以及担任监督人的角色，形成积极循环，从而进一步增强车联网系统的安全性.

4 安全性分析

4.1 共识协议安全属性分析

本节将证明由片内共识节点选举函数选举的 RSU 满足诚实多数. 此外，证明基于声誉的片内共识协议满足安全性、活性.

引理 1. 诚实多数. 如果单个分片内诚实 RSU 数量大于 1/2，则片内共识节点选举函数选出的 RSU 中至少一半是诚实的.

证明. 在最坏的情况下，所有节点的声誉值均为 1. 片内共识节点选举可以看作一个独立的随机抽样问题，节点被选中的概率为

$$Pr[R_j \text{ is selected}] = \frac{sr_j}{\sum_{i=1}^{m} r_i} = \frac{s}{m}, \tag{8}$$

其中 s 是被选中 RSU 的数量，m 是分片内 RSU 的总数，r_j 是某个 RSU 的声誉值.

假设 $\rho = \frac{m-f^*}{m}$ 表示选举的节点为诚实节点的概率，$\sigma = \frac{f^*}{m}$ 表示选举的节点为拜占庭节点的概率，其中 f^* 代表分片内恶意节点的数量. 随机变量 A 表示被选中 RSU 中诚实节点的数量，随机变量 B 表示选中 RSU 中拜占庭节点的数量. 随机变量 A 和 B 满足以下二项式分布：

$$Pr[A=k] = \binom{m}{k}\left(\rho\frac{s}{m}\right)^k\left(1-\rho\frac{s}{m}\right)^{m-k}, \tag{9}$$

$$Pr[B=k] = \binom{m}{k}\left(\sigma\frac{s}{m}\right)^k\left(1-\sigma\frac{s}{m}\right)^{m-k}, \tag{10}$$

设 X 和 Y 分别表示为

$$X = \left\{A : 0 \leqslant A \leqslant \left\lfloor \frac{s}{2} \right\rfloor\right\}, \tag{11}$$

$$Y = \left\{B : \left\lceil \frac{s}{2} \right\rceil \leqslant B \leqslant m\right\}, \tag{12}$$

当 X 和 Y 都未成立，被选中的 RSU 诚实大多数的性质成立. $Pr[X \cup Y]$ 代表 X 和 Y 至少有 1 个成立，其概率为

$$\begin{aligned} Pr[X\cup Y] =& Pr[X] + Pr[Y] - Pr[X\cap Y] = Pr\left[A \leqslant \left\lfloor\frac{s}{2}\right\rfloor\right] + \\ & Pr\left[B \geqslant \left\lceil\frac{s}{2}\right\rceil\right] = F\left(\left\lfloor\frac{s}{2}\right\rfloor, m, \frac{\rho s}{m}\right) + \\ & F\left(m - \left\lceil\frac{s}{2}\right\rceil, m, 1-\frac{\sigma s}{m}\right), \end{aligned} \tag{13}$$

其中 $F(k,m,p) = Pr[X \leqslant k]$ 是随机变量 $X \sim B(m,p)$ 的累积分布函数，根据二项式尾界定理可知

$$Pr[X\cup Y] \leqslant 2\exp\left(-2m\left(\frac{\lfloor s/2\rfloor - \sigma s}{m}\right)^2\right) = \epsilon(s), \tag{14}$$

被选举的 RSU 满足诚实多数. 证毕.

引理 2. 安全性. 如果片内共识节点选举函数选举出的 RSU 满足诚实多数且 Reputation-based SMR 协议满足安全性，那么基于声誉的片内共识协议也满足安全性.

证明. 假设基于声誉的片内共识协议不满足安全性，即这里存在某个出块周期，在这个周期中 2 个诚实 RSU 提出了 2 个不同的区块 b 与 b'. 通过引理 1，被选举出的 RSU 符合诚实多数的性质，且 Reputation-based SMR 协议满足安全性性质，对于任意 2 个被选中的诚实 RSU，它们在同一个周期会对相同区块达成共识并提交相同的区块到本地的区块链上. 这与假设相反，因此基于声誉的片内共识协议满足安全性. 证毕.

引理 3. 活性. 如果片内共识节点选举函数选举出的 RSU 满足诚实多数且 Reputation-based SMR 满足活性，则基于声誉的片内共识协议满足活性.

证明. 假设基于声誉的片内共识协议不满足安全性，即这里存在一个交易被诚实的 RSU 接受，但最终没有被所有诚实的 RSU 提交到本地的账本上. 通过引理 1，被选举出的 RSU 符合诚实多数的性质，在此基础上通过轮询调度算法选举的领导者要么是诚实 RSU，要么是恶意 RSU(拜占庭节点). 如果领导者是诚实 RSU，那么领导者会提议一个区块到所有节点，并且在片内广播后，所有诚实节点将在本地提交这一区块. 如果领导者是恶意 RSU，那么该 RSU 要么遵循协议，要么背离协议. 当 RSU 遵循协议时，结果与诚实 RSU 担任领导者时相同. 反之，当 RSU 背离协议时，因为 Reputation-based SMR 协议满足活性与片内广播，最终所有的交易都将提交到每个分片中

节点的账本上. 这与假设相反, 因此基于声誉的片内共识协议满足活性. 证毕.

引理 4. 声誉一致性. 如果基于声誉的片内共识协议满足安全性, 则基于声誉的片内共识协议满足声誉一致性.

证明. 假设某一共识轮次结束时, 分片内存在 2 个诚实的 RSU1 和 RSU2 对 RSU3 的声誉视图不同. 通过引理 2, 基于声誉的片内共识协议满足安全性, 这表明在任意一个周期, 任意 2 个诚实的 RSU 提交由诚实的领导者 RSU 提议的区块或空块. 如果提交空块, 这表明领导者 RSU 是恶意的并且扣留区块或提议冲突区块, 在当前共识轮次的计时器还剩 1 个 Δ时, 这 2 种行为都会被委员会内至少 1 个诚实的 RSU 所检测. 委员会内所有诚实的 RSU 会将检测到的行为与区块广播给片内所有 RSU, 片内所有诚实的 RSU 会根据收到的行为更新在其视图上各个节点的声誉, 从而使所有诚实的 RSU 拥有相同的声誉视图, 这与假设相反. 如果提交的区块是正常区块, 那么存在 2 种情况: 1)协议的执行中不存在不当行为; 2)协议执行的过程中存在不影响共识安全性的不当行为, 即恶意区块提议与恶意投票行为. 对于情况 1, 所有诚实的 RSU 通过声誉函数一致地更新其本地声誉视图, 这与假设相反. 对于情况 2, 它与诚实节点提交空块的情况相同, 在片内广播后, 所有诚实的 RSU 将获得相同的节点声誉视图, 这与假设相反. 证毕.

4.2 监督人机制安全性分析

假定分片系统中存在n个监督人, 分片系统在运行时, 每当产生t个由不同分片生成的区块时, 监督人需要对这些区块进行验证以确认其数据合法性. 每个监督人至少需要验证 1 个区块, 同时, 为了增加系统的安全性, 验证区块的选择必须具备一定的不可预测性, 以确保潜在的敌手无法轻易获悉下一个需要验证的区块.

监督人从t个块中选取的规则为: 首先, 每个监督人以不可预测的方式选取 1 个区块. 接着, 对于剩下的$t-1$个区块, 监督人对每个区块的选择验证的概率均为x.

本文引入了以下区块选择规则选择必选区块, 以增强监督人区块选择的不可预测性.

$$\text{区块序号} = H(\text{最新区块哈希值}\|\text{唯一标识符}) \bmod t, \tag{15}$$

其中区块序号是监督人从t个区块中选取的, $H()$是一个安全的、单向的哈希函数, 其将最新的区块哈希值和监督人标识符转换为一个确定的值. 其次, $\|$表示字符串连接操作符, mod t表示对结果取模运算. 最后, 这个规则的应用范围是分片系统中需要验证的区块.

通过引入这个区块选择规则, 可以增加系统的安全性, 因为敌手无法轻易预测每一个监督人需要验证的区块, 这使得潜在的敌手难以进行恶意行为, 从而保护了整个分片系统的安全性和稳定性.

根据以上规则, 在必选验证环节, 某个区块不被所有监督人随机以不可预测选中验证的概率为

$$Pr(C) = \left(\frac{t-1}{t}\right)^n, \tag{16}$$

对于剩下的$t-1$个区块, 某个区块不被某个监督人随机选择的概率可以用$1-x$表示. 此外, 各个区块不被所有监督人选择是相互独立的, 符合二项分布, 可表示为

$$Pr(D) = (1-x)^n, \tag{17}$$

最终, 某个区块不被监督人所验证的概率为

$$Pr(\text{block is not selected}) = Pr(C) \times Pr(D) = \left(\frac{t-1}{t}\right)^n (1-x)^n, \tag{18}$$

式(18)反映了监督人机制的安全性. 假设$t=10$, 系统共有$n=80$个监督人, $x=0.3$, 一个区块不被监督人验证的概率为3.86×10^{-12}. 可以得知, 分片系统中每个区块都以高概率被监督人所验证. 随着监督人数量的增加, 某个特定区块不被选中的概率逐渐减小, 从而增大了分片系统的安全性. 一方面涉及监督人必选验证环节的不可预测性, 另一方面涉及监督人随机验证区块的不可预测性. 敌手难以预测哪个监督人将验证哪个区块, 保证了分片系统的安全性.

5 实验评估

5.1 分片方案性能评估

本节对所提出基于机器学习的分片算法进行实验评估. 本实验基于 Python 语言实现均衡聚类、*K*-Means++[38]、层次聚类[40] 和随机分配等算法, 并评估它们的簇间负载方差与聚类距离成本. 结果显示均衡聚类算法得到的分片结果在距离成本几乎没有增加的情况下, 簇间负载方差显著低于其他算法. 此外, 在 RSU 节点负载不均衡、呈现指数分布的情况下, 均衡聚类算法仍然能够维持较低的簇间负载方差与聚类距离成本.

通过实验, 评估以下 2 个问题:

1)RSU 节点负载呈现均匀分布下, 随着 RSU 节点负载的增加, 均衡聚类算法和其他 3 个算法的聚

类距离成本与簇间负载方差的变化趋势；

2）RSU 节点负载呈现指数分布下，随着 RSU 节点负载的增加，均衡聚类算法和其他 3 个算法的聚类距离成本与簇间负载方差的变化趋势.

5.1.1 实验设置与实现

实验基于 Python3.10 实现，采用 numpy1.62.2 实现均衡聚类算法和随机分配算法，采用 scikit-learn1.3.2 实现 *K*-Means++与层次聚类算法，采用 matplotlib 3.8.2 与 seaborn0.13.0 实现数据可视化. 随机分配算法是目前绝大多数区块链分片方案所采用的算法；层次聚类算法相较 *K*-Means++算法具有更强的可解释性，被广泛应用于无监督聚类. 对于地理位置呈现随机分布的 RSU 节点，*K*-Means++算法与层次聚类算法能够将相近的 RSU 节点划分至同一分片，相较随机分配算法更加适用于车联网区块链.

实验模拟了 200 个 RSU 并将其划分为 15 个分片，其中每个 RSU 拥有位置信息与负载状态 2 种属性. 实验从标准均匀分布中采样 200 个 2 维数据作为 RSU 的位置信息，分别从均匀分布与指数分布中采样 200 个 1 维数据作为 RSU 的负载状态，且分布的期望值线性增大，进而模拟 RSU 的负载状态在不同分布、不同大小下分片算法的性能.

实验中涉及的性能指标为簇间负载方差和聚类距离成本，以每个 RSU 与其分片中心的平均欧拉距离为单位. 簇间负载方差能够直观反映各个分片间负载的差异性，方差越大则代表算法负载均衡的能力越差. 而每个 RSU 与其分片中心的平均欧拉距离（以下简称簇均值）能够反映簇内紧凑性. 采用以上 2 种指标对片内节点距离相近和片间负载均衡 2 个要素进行评估.

5.1.2 均衡 *K*-Means 性能评估

图 4 展示了 4 种算法簇间负载方差随 RSU 节点平均负载的变化. 相对于其他 3 种算法，均衡 *K*-Means 簇间方差更低. 随着 RSU 负载逐步提升，其余 3 种算法的簇间负载方差均指数级增长；而均衡 *K*-Means 线性增长，表明该算法具有更强的负载均衡能力.

图 5 与图 6 分别展示了 4 种和 3 种算法聚类距离成本随 RSU 节点平均负载的变化. 随机分配算法的簇均值显著高于其他算法，表明同一分片的节点在地理位置上分布稀疏，车辆之间的数据共享协同效率更低. 除随机分配算法以外的 3 种聚类算法均将地理位置相近的节点划分至同一分片，均衡 *K*-Means 与另外 2 种算法在簇均值上相近，均保持在较低水平，表明该算法满足车联网区块链分片的 2 个要素.

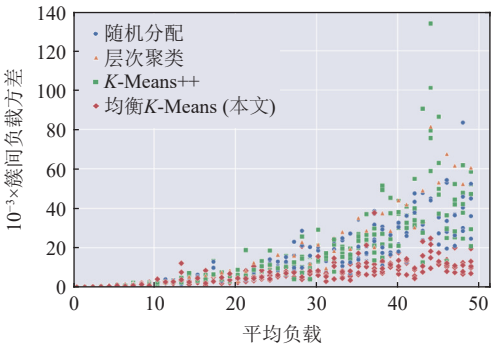


Fig. 4 Comparison of load variance between clusters of four algorithms

图 4 4 种算法的簇间负载方差对比

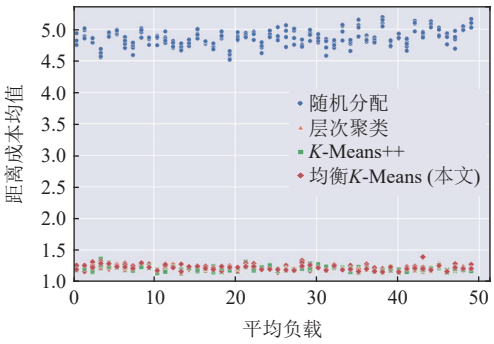


Fig. 5 Comparison of clustering distance costs of four algorithms

图 5 4 种算法的聚类距离成本对比

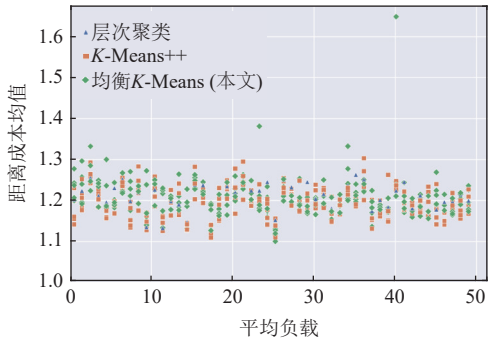


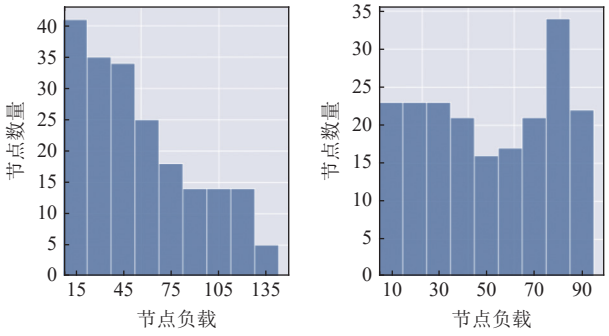
Fig. 6 Comparison of clustering distance costs of three algorithms

图 6 3 种算法的聚类距离成本对比

图 7 展示了负载呈指数分布与均匀分布的差异. 相对于均匀分布，指数分布下的 RSU 负载不均衡程度显著增加. 图 8 展示了除随机分配以外的 3 种算法在该实验设置下的聚类簇间方差. 在 RSU 负载不平衡的情况下，最高方差超过 600 000，表明 *K*-Means++与层次聚类算法的片间负载存在极大的差异. 均衡 *K*-Means 方差最高未超过 100 000，显著优于其他算法. 由于 RSU 负载不平衡，维持分片间节点均衡将会付出更多的距离成本. 图 9 展示了除随机分配以外的 3

种算法在该实验设置下的聚类距离成本. 相对其他算法, 均衡 *K*-Means 的簇均值略有上升, 仅提升 0.2 左右, 表明其仍然保持较为紧凑的聚类结构.

图 10 对比了 4 种算法的聚类可视化结果. 相较随机分配算法, 其他 3 种算法生成的聚类结构边界清晰, 表明它们能够将相邻的节点划分至同一分片, 满足片内节点相近的需求. 而 *K*-Means++与层次聚类算法仅考虑片内节点相近, 难以满足车联网区块链分片中负载均衡的要求. 相较 *K*-Means++与层次聚类算法, 均衡聚类算法簇间负载方差更低, 更加适用于车联网区块链分片. 聚类结构同样反映了均衡聚类算法特点, 即部分分片间节点出现“交叉”. 由于节点分配时需要同时考虑负载与距离 2 个因素, 部分节点为了实现负载均衡, 未加入距离最近的簇, 导致“交叉”的出现, 表明均衡聚类算法更加符合车联网区块链分片的需求.



(a) 指数分布 (b) 均匀分布

Fig. 7 Difference between balanced load distribution and exponential distribution

图 7 负载呈均匀分布与指数分布的差异

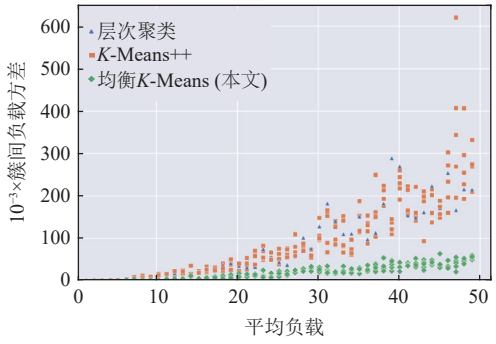


Fig. 8 Inter-cluster variance of clusters under exponential load distribution

图 8 负载呈指数分布下的聚类簇间方差

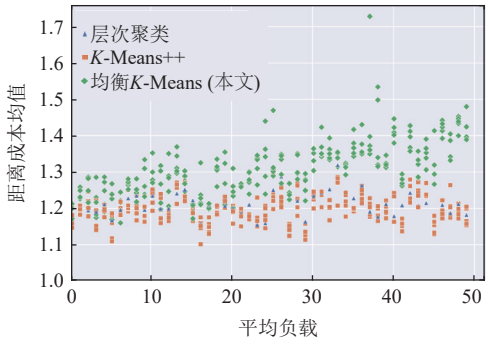
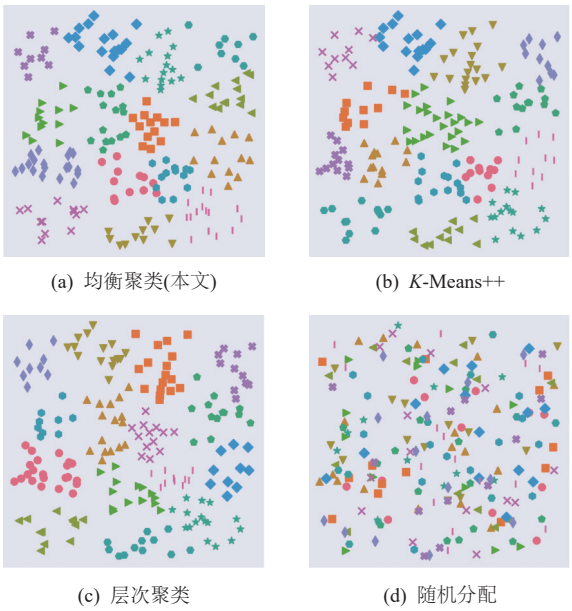


Fig. 9 Clustering distance cost under exponential load distribution

图 9 负载呈指数分布下的聚类距离成本



(a) 均衡聚类(本文) (b) *K*-Means++

(c) 层次聚类 (d) 随机分配

Fig. 10 Clustering visualization results of four algorithms

图 10 4 种算法的聚类可视化结果

5.2 片内共识性能评估

本节对所提出片内共识协议进行实验评估. 该实验通过 Go 语言实现基于声誉的片内共识协议, 并在云实例上评估它们的吞吐量与共识时延. 结果显示在多个分片中基于声誉的片内共识协议的共识时延与 Reputation-based SMR 基本相同. 且随着分片中 RSU 数量的增加, 基于声誉的片内共识协议的吞吐量下降幅度低于 Reputation-based SMR 协议与 Sync-HotStuff 协议[41]. 此外, 随着分片数量的增长, 基于声誉的片内共识协议相较于 Reputation-based SMR 吞吐量更优, 且相较于 Sync-HotStuff 协议没有显著的吞吐量差距.

通过实验, 评估以下 3 个问题:

1)在单个分片内, 随着区块中包含的交易的增加, 基于声誉的片内共识协议的性能变化趋势;

2)在单个分片内, 随着 RSU 数量的增加, 基于声誉的片内共识协议与 Reputation-based SMR 共识时延和吞吐量的变化趋势;

3)随着分片数量的增长, 基于声誉的片内共识

协议共识时延和吞吐量的变化趋势.

5.2.1 实验设置与实现

本节的实验利用 libp2p 完成点对点网络的搭建，采用 protobuf 进行信息的序列化与反序列化以便在网络中传递数据，采用 secp256k1 椭圆曲线算法来执行数字签名，采用 SHA256 作为哈希函数. 使用 Go 语言成功实现了 Sync-HotStuff、Reputation-based SMR 以及基于声誉的片内共识协议. 在协议的实现中，每个区块包含由 OBUs 发送的多个交易. 将每个区块包含的最多的交易数量称为该区块的大小，设置同步网络上限时延Δ=0.25 s，即基于声誉的片内共识协议和 Reputation-based SMR 的时延计时器的设定均为 1 s. 而对于 Sync-HotStuff，相较于以上 2 个协议缺少证据广播阶段和片内广播阶段，因此其时延计时器设定为 0.75 s.

基于声誉的片内共识协议部署在阿里云实例上，参数如表 1 所示.

Table 1 Elastic Compute Service Configuration Parameters

表 1 云服务器配置参数

参数	配置环境
实例规格	ecs.c7a.8xlarge
处理器核心数	32
内存/GB	256
本地存储/GB	100
处理器主频/GHz	3.5
内网带宽/GBps	25
内网收发包/PPS	1 200 万
处理器型号	AMD EPYC™ Milan 7T83
镜像操作系统	Ubuntu18.04 64 位

实验中涉及的性能指标主要为共识时延与吞吐量.

5.2.2 基于区块大小的协议性能评估

将车辆产生的每条数据近似看成区块中的 1 笔交易，在区块大小分别为 400, 800, 1 600 笔交易的条件下，固定每笔交易大小为 64 B，测试了基于声誉的片内共识协议的性能. 通过控制 OBU 节点发送命令的时间间隔来模拟不同的负载，OBU 节点发送每个命令的速度越快，客户端在给定时间拥有的未完成命令池就越大. 如图 11 所示，最开始增加 OBU 发送请求的速率不会对延迟产生影响，因为节点的带宽还没有达到饱和状态. 随着节点负载的增加，吞吐量会增加，但延迟仍然保持稳定，直到系统的带宽达到饱和. 一旦带宽饱和，随着负载的增加，吞吐量将保持不变或稍微下降，而延迟则会上升. 由图 11 可知，基于声誉的片内共识协议的吞吐量在 720 tps 左右达到带宽饱和，将以 800 区块大小与 64 B 交易大小作为后续实验的固定参数.

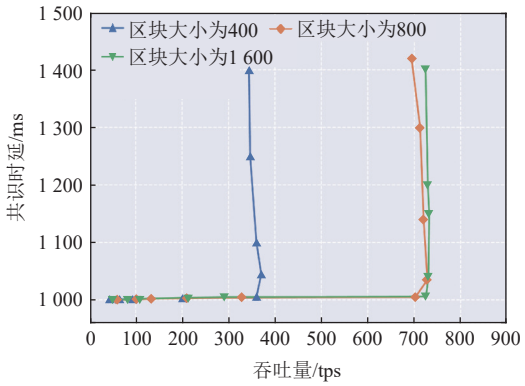


Fig. 11 Impact of different block sizes on protocol performance

图 11 不同区块大小对协议性能的影响

5.2.3 基于 RSUs 节点数量的协议性能评估

图 12 与图 13 展示了在单个分片中，随着 RSU 数量的增加，Reputation-based SMR 协议、Sync-HotStuff 协议与基于声誉的片内共识协议在共识时延与吞吐量 2 个性能指标上的对比. 设置 RSU 个数分别为 5, 9, 17, 33，从图 12 与图 13 中可以观察到，随着 RSU 数量的增加，基于声誉的片内共识协议的共识时延增长速率与吞吐量的下降速率明显低于 Reputation-based SMR 协议. 且当网络中 RSU 数为 33 时，基于声誉的片内共识协议的时延和吞吐量分别约为 1 600 ms, 490 tps，相较于 Reputation-based SMR 协议的时延和吞吐量分别为 2 100 ms 和 280 tps 左右，基于声誉的片内共识协议有着更优的性能. 而对于 Sync-HotStuff,

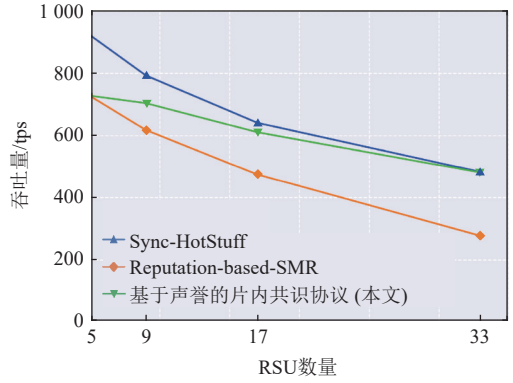


Fig. 12 Impact of the number of RSU on the throughput of protocol

图 12 RSU 数量对协议吞吐量的影响

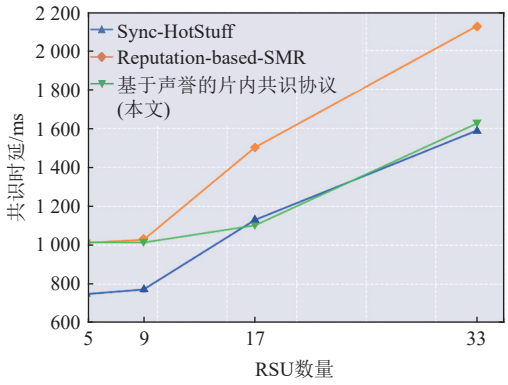


Fig. 13 Impact of the number of RSU on consensus latency

图 13 RSU 数量对共识时延的影响

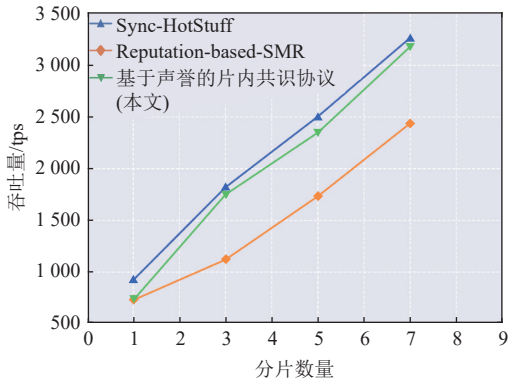


Fig. 14 Impact of the number of shards on the throughput of protocol

图 14 分片数量对协议吞吐量的影响

该协议不具有对 RSU 不当行为追责的能力，但有更低的共识时延与更高的吞吐量. 随着 RSU 数量的增加，Sync-HotStuff 协议的性能明显下降，当 RSU 数量为 33 时，Sync-HotStuff 协议的时延和吞吐量分别为 1 590 ms, 492 tps 左右，这表明随着 RSU 数量的增加，基于声誉的片内共识协议与 Sync-HotStuff 协议在性能上的差距会逐步减小. 其原因是基于声誉的片内共识协议通过小规模 RSU 替代分片内所有 RSU 参与共识，从而带来了更低的通信损耗，这意味着基于声誉的片内共识协议的性能不会随着片内 RSU 的增多而明显下降.

5.2.4 基于分片数量的协议性能评估

为了尽可能体现基于声誉的片内共识协议能对高声誉节点进行选举，从而保证片内共识协议的正常运行，本节设计了如下实验. 固定每个分片中 RSU 节点数为 11，测试分片数量分别为 1, 3, 5, 7 的情景下 2 个协议的吞吐量. 当分片数量为 3 时，设定其中一个分片的恶意 RSU 为 6 个(为了简化实验的进行，通过代码沉睡这 6 个节点，使其不参与片内共识协议). 同时，因为恶意节点不参与共识过程，它们的声誉远小于其他节点，将这样的分片称为恶意分片. 同理，当分片数量为 5 时，恶意分片数量为 2；当分片数量为 7 时，恶意分片数量为 3.

图 14 展示了 Sync-HotStuff, Reputation-based SMR 和基于声誉的片内共识协议在不同分片数量下的吞吐量变化趋势. 随着分片数量的增加，系统吞吐量明显上升. Reputation-based SMR 协议的吞吐量低于基于声誉的片内共识协议. 而对于 Sync-HotStuff, 因为其在单个分片内拥有更高的吞吐量，当分片数量为 1 且不存在恶意分片时，该协议有着最优的性能. 随着恶意分片数量的增多，基于声誉的片内共识协议与 Sync-HotStuff 协议在吞吐量上并未体现出明显的差距. 其原因是，在恶意分片中 Sync-HotStuff 与 Reputation-based SMR 无法通过足够的签名来完成共识. 而对于基于声誉的片内共识协议，它通过片内共识节点选举，使片内诚实 RSU 提供签名并完成共识.

6 结 论

本文提出了一种基于机器学习分片的车联网区块链数据共享方案，每个分片负责存储和管理特定类型的车联网数据，使得设备间的数据共享可以并发执行，提高了数据共享的效率. 为了保证分片方案的安全性，本文提出了基于声誉的共识协议和监督人机制，通过片内共识节点选举函数尽可能地保证高声誉的诚实 RSU 参与并完成共识. 监督人机制使节点从共享数据中存储数据并获得验证其他分片数据的能力，进一步有助于提高系统的安全性. 性能评估和安全分析证明，本文方案可以实现安全高效的车联网数据共享.

作者贡献声明：陈骁、黄牧鸿负责方案整体设计并撰写论文；田一凡、王岩负责部分算法思路和实验方法；曹晟、张小松提出指导意见并修改论文.

参 考 文 献

[1] Jiang Wenxian, Chen Mengjuan, Tao Jun. Federated learning with blockchain for privacy-preserving data sharing in Internet of vehicles[J]. China Communications, 2023, 20(3): 69−85

[2] Hu Xiaoya, Li Ruiqin, Wang Licheng, et al. A data sharing scheme based on federated learning in IoV[J]. IEEE Transactions on Vehicular Technology, 2023, 72(9): 11644−11656

[3] Diallo E, Dib O, Al Agha K. An improved PBFT-based consensus for

securing traffic messages in VANETs[C]//Proc of the 12th Int Conf on Information and Communication Systems (ICICS). Piscataway, NJ: IEEE, 2021: 126−133

[4] Lee S, Seo S H. Design of a two layered blockchain-based reputation system in vehicular networks[J]. IEEE Transactions on Vehicular Technology, 2021, 71(2): 1209−1223

[5] Jiang Bingcheng, He Qian, Liu Peng, et al. Blockchain empowered secure video sharing with access control for vehicular edge computing[J]. IEEE Transactions on Intelligent Transportaison Systems, 2023, 24(9): 9041−9054

[6] Sun Jianfei, Xiong Hu, Zhang Shufan, et al. A secure flexible and tampering-resistant data sharing system for vehicular social networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(11): 12938−12950

[7] Huang Huawei, Kong Wei, Peng Xiaowen, et al. Survey on blockchain sharding technology[J]. Computer Engineering, 2022, 48(6): 1−10(in Chinese)
(黄华威, 孔伟, 彭肖文, 等. 区块链分片技术综述[J]. 计算机工程, 2022, 48(6): 1−10)

[8] Luu L, Narayanan V, Zheng Chaodong, et al. A secure sharding protocol for open blockchains[C]//Proc of the 23rd ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2016: 17−30

[9] Zamani M, Movahedi M, Raykova M. RapidChain: Scaling blockchain via full sharding[C]//Proc of the 25th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2018: 931−948

[10] Kokoris-Kogias E, Jovanovic P, Gasser L, et al. OmniLedger: A secure, scale-out, decentralized ledger via sharding[C]//Proc of the 39th IEEE Symp on Security and Privacy (SP). Piscataway, NJ: IEEE, 2018: 583−598

[11] Al-Bassam M, Sonnino A, Bano S, et al. ChainSpace: A sharded smart contract platform[C]//Proc of the 25th Network and Distributed System Security(NDSS). Reston, VA: the Internet Society, 2018: 536−552

[12] Huang Huawei, Huang Zhenyi, Peng Xiaowen, et al. MVCom: Scheduling most valuable committees for the large-scale sharded blockchain[C]//Proc of the 41st Int Conf on Distributed Computing Systems (ICDCS). Piscataway, NJ: IEEE, 2021: 629−639

[13] Ni Weiquan, Asheralieva A, Maple C, et al. Throughput-efficient blockchain for Internet-of-vehicles[C/OL]//Proc of the 64th IEEE Globecom Workshops (GC Wkshps). Piscataway, NJ: IEEE, 2021 [2024-05-07].https://ieeexplore.ieee.org/abstract/document/9681973

[14] Cui Jie, Ouyang Fenqiang, Ying Zuobin, et al. Secure and efficient data sharing among vehicles based on consortium blockchain[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7): 8857−8867

[15] Kumar A, Yadav A S, Kushwaha D S. VChain: Efficient blockchain based vehicular communication protocol[C]//Proc of the 10th Int Conf on Cloud Computing, Data Science & Engineering (Confluence). Piscataway, NJ: IEEE, 2020: 762−768

[16] Yuan Mingyang, Xu yang, Zhang Cheng, et al. TRUCON: Blockchain-based trusted data sharing with congestion control in Internet of vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(3): 3489−3500

[17] Zhao Yang, Du Kai. An application of the IoVs information sharing scheme based on blockchain technology[C]//Proc of the 2nd Int Conf on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI). Piscataway, NJ: IEEE, 2022: 617−620

[18] Anish T P, Syed Musthafa A R. Elavarasu, et al. Block chain based secure data transmission among Internet of vehicles[C]//Proc of the 2nd Int Conf on Innovative Practices in Technology and Management (ICIPTM). Piscataway, NJ: IEEE, 2022: 765−769

[19] Kouicem D E, Bouabdallah A, Lakhlef H. An efficient and anonymous blockchain-based data sharing scheme for vehicular networks[C/OL]//Proc of the 25th IEEE on Computers and Communications (ISCC). Piscataway, NJ: IEEE, 2020[2024-04-24]. https://ieeexplore.ieee.org/document/9219641

[20] Hellings J, Sadoghi M. Byshard: Sharding in a Byzantine environment[J]. Proceedings of the VLDB Endowment, 2021, 14(11): 2230−2243

[21] Que Qifeng, Chen Zhihao, Zhang Zhao, et al. A coordinator-free cross-shard transaction execution for sharded permissioned blockchains[J]. Journal of Computer Research and Development, 2023, 60(11): 2469−2488(in Chinese)
(阙琦峰, 陈之豪, 张召, 等. 面向分片许可链的无协调者跨片交易处理[J]. 计算机研究与发展, 2023, 60(11): 2469−2488)

[22] Zhang Jianting, Chen Wuhui, Luo Sifu, et al. Front-running attack in sharded blockchains and fair cross-shard consensus[C]//Proc of the 31st Network and Distributed System Security(NDSS). Reston, VA: the Internet Society, 2024: 196−213

[23] Zhang Yuanzhe, Pan Shirui, Yu Jiangshan. Txallo: Dynamic transaction allocation in sharded blockchain systems[C]//Proc of the 39th Int Conf on Data Engineering (ICDE). Piscataway, NJ: IEEE, 2023: 721−733

[24] Li Mingzhe, Wang Wei, Zhang Jin. LB-Chain: Load-balanced and low-latency blockchain sharding via account migration[J]. IEEE Transactions on Parallel and Distributed Systems, 2023, 34(10): 2797−2810

[25] Wang Keyan, Jiang Xin, Jia Linpeng, et al. Throughput model of starlike sharding structure for blockchains and its applications[J]. Journal of Software, 2023, 34(9): 4294−4309(in Chinese)
(王柯元, 姜鑫, 贾林鹏, 等. 区块链星型分片架构通量模型及应用[J]. 软件学报, 2023, 34(9): 4294−4309)

[26] David B, Magri B, Matt C, et al. Gearbox: Optimal-size shard committees by leveraging the safety-liveness dichotomy[C]//Proc of the 29th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2022: 683−696

[27] Xu Yibin, Zheng Jingyi, Düdder B, et al. A two-layer blockchain sharding protocol leveraging safety and liveness for enhanced performance[J]. arXiv preprint, arXiv: 2310.11373, 2023

[28] Tao Yuechen, Li Bo, Jiang Jingjie, et al. On sharding open blockchains with smart contracts[C]//Proc of the 36th In Conf on Data Engineering (ICDE). Piscataway, NJ: IEEE, 2020: 1357−1368

[29] Wu Kaidong, Ma Yun, Cai Huaqian, et al. BETASCO: Consortium blockchain system based on smart contract-oriented sharding[J]. Journal of Software, 2023, 34(11): 5042−5057(in Chinese)

(吴恺东, 马郭, 蔡华谦, 等. BETASCO: 面向智能合约分片的联盟区块链系统[J]. 软件学报, 2023, 34(11): 5042−5057)

[30] Huang Junqin, Kong Linghe, Wang Jingwei, et al. Secure data sharing over vehicular networks vased on multi-sharding blockchain[J/OL]. ACM Transactions on Sensor Networks, 2024, 20(2): 1550−4859. [2024-05-07]. https://dl.acm.org/doi/abs/10.1145/3579035

[31] Kang Jiawen, Xiong Zehui, Niyato D, et al. Towards secure blockchain-enabled Internet of vehicles: Optimizing consensus management using reputation and contract theory[J]. IEEE Transactions on Vehicular Technology, 2019, 68(3): 2906−2920

[32] Huang Muhong, Han Runchao, Du Zhiqiang, et al. Reputation-based state machine replication[C]//Proc of the 21st IEEE Int Symp on Network Computing and Applications(NCA). Piscataway, NJ: IEEE, 2022: 225−234

[33] Cao Sheng, Dang Sixuan, Du Xiaojiang, et al. An electric vehicle charging reservation approach based on blockchain[C/OL]//Proc of the 63rd IEEE Global Communications Conf. Piscataway, NJ: IEEE, 2020[2024-05-07]. https://ieeexplore.ieee.org/abstract/document/9322093

[34] Yu Jiangshan, Kozhaya D, Decouchant J, et al. Repucoin: Your reputation is your power[J]. IEEE Transactions on Computers, 2019, 68(8): 1225−1237

[35] Kleinrock L, Ostrovsky R, Zikas V. Proof-of-reputation blockchain with nakamoto fallback[C]//Proc of the 21st Int Conf on Cryptology. Berlin: Springer, 2020: 16−38

[36] Zhao Li, Fang Jiayi, Hu Jinling, et al. The performance comparison of LTE-V2X and IEEE 802.11p[C]//Proc of the 87th Vehicular Technology Conf (VTC Spring). Piscataway, NJ: IEEE, 2018: 2577−2465

[37] Xia Zhenchang, Wu Jia, Wu Libing, et al. A comprehensive survey of the key technologies and challenges surrounding vehicular ad hoc networks[J]. ACM Transactions on Intelligent Systems and Technology, 2021, 12(4): 1−30

[38] Arthur D, Vassilvitskii S. K-Means++: The advantages of careful seeding[C]//Proc of the 18th ACM-SIAM Symp on Discrete Algorithms. New York: ACM, 2007: 1027−1035

[39] Bonneau J, Felten E W, Goldfeder S, et al. Why buy when you can rent? Bribery attacks on bitcoin consensus[C]//Proc of the 20th Int Conf on Financial Cryptography and Data Security. Berlin: Springer, 2016: 19−26

[40] Johnson S C. Hierarchical clustering schemes[J]. Psychometrika, 1967, 32(3): 241−254

[41] Abraham I, Malkhi D, Nayak K, et al. Sync-HotStuff: Simple and practical synchronous state machine replication[C]//Proc of the 39th IEEE Symp on Security and Privacy (SP). Piscataway, NJ: IEEE, 2020: 106−118



Chen Xiao, born in 1998. PhD candidate. His main research interest includes blockchain and network security.

陈　骁, 1998 年生. 博士研究生. 主要研究方向为区块链与网络安全.



Huang Muhong, born in 1997. PhD candidate. His main research interest includes blockchain and network security.

黄牧鸿, 1997 年生. 博士研究生. 主要研究方向为区块链与网络安全.



Tian Yifan, born in 1999. Master candidate. His main research interest includes blockchain and network security.

田一凡, 1999 年生. 硕士研究生. 主要研究方向为区块链与网络安全.



Wang Yan, born in 2000. Master candidate. His main research interests include machine learning and deep learning.

王　岩, 2000 年生. 硕士研究生. 主要研究方向为机器学习、深度学习.



Cao Sheng, born in 1981. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include blockchain, network security, and intelligent education.

曹　晟, 1981 年生. 博士, 研究员, 博士生导师. CCF 高级会员. 主要研究方向为区块链、网络安全、智能教育.



Zhang Xiaosong, born in 1968. PhD, professor, PhD supervisor. Member of CCF. His main research interests includes blockchain and network security.

张小松, 1968 年生. 博士, 教授, 博士生导师. CCF 会员. 主要研究方向为区块链与网络安全.