计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(4):209−214 [doi: 10.15888/j.cnki.csa.009472]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

应用于供应链的区块链 PBFT 共识算法优化①

黄宇翔[1,2]

[1](西南林业大学 大数据与人工智能研究院, 昆明 650224)
[2](西南林业大学 森林生态大数据国家林业和草原局重点实验室, 昆明 650224)
通信作者: 黄宇翔, E-mail: yxhuang@swfu.edu.cn

摘　要: 目前, 区块链在供应链领域中的应用越来越受到业界的广泛关注. 但由于供应链中存在大量复杂性的事务, 这给可信的主节点选取工作带来了挑战. 因此, 在机器学习分类算法与 PBFT (practical Byzantine fault tolerance) 共识算法的基础上, 提出一种应用于供应链的区块链 PBFT 共识算法优化方法. 对构建供应链与区块链的集成框架进行分析, 根据供应链中参与共识的节点属性特征, 运用 K-近邻 (K-nearest neighbors) 来优化 PBFT 共识算法的主节点选取规则. 实验结果表明, 对共识节点进行信任评估分类可以较好地解决因视图切换所引发的效率问题, 从而提升区块链的吞吐量、时延、容错性等共识性能, 具有一定的实用性, 也给区块链在其他行业的应用提供了思路.

关键词: 区块链; 实用拜占庭容错; 供应链; K-近邻; 信任评估

引用格式: 黄宇翔.应用于供应链的区块链 PBFT 共识算法优化.计算机系统应用,2024,33(4):209–214. http://www.c-s-a.org.cn/1003-3254/9472.html

Optimization of Blockchain PBFT Consensus Algorithm for Supply Chain

HUANG Yu-Xiang[1,2]

[1](Institute of Big Data and Artificial Intelligence, Southwest Forestry University, Kunming 650224, China)
[2](Key Laboratory of State Forestry and Grassland Administration on Forestry and Ecological Big Data, Southwest Forestry University, Kunming 650224, China)

Abstract: Currently, the application of blockchain in the supply chain is receiving increasing attention from the industry. However, due to the presence of a large number of complex transactions in the supply chain, selecting trustworthy primary nodes poses a challenge. Therefore, based on the machine learning classification algorithms and PBFT (practical Byzantine fault tolerance), this study proposes a blockchain PBFT optimization method applied to the supply chain. The integrated framework for the supply chain and blockchain is analyzed, and K-nearest neighbors (K-NN) is applied to optimize the primary node selection rules of the PBFT consensus algorithm based on the features of participating nodes in the supply chain consensus. Experimental results show that trust evaluation classification of consensus nodes can effectively address efficiency issues caused by view switching, thereby improving the consensus performance of blockchain in terms of throughput, latency, fault tolerance, and other aspects. The proposed method is practical and provides ideas for the application of blockchain in other industries.

Key words: blockchain; practical Byzantine fault tolerance (PBFT); supply chain; K-nearest neighbor (KNN); trust evaluation

① 基金项目: 云南省教育厅科学研究基金 (2023J0697); 西南林业大学森林生态大数据国家林业和草原局重点实验室开放课题 (2022-BDG-03); 中央引导地方科技发展专项 (202307AB110009)
收稿时间: 2023-09-11; 修改时间: 2023-10-20, 2023-12-06; 采用时间: 2023-12-11; csa 在线出版时间: 2024-03-01
CNKI 网络首发时间: 2024-03-07

近年来, 电子商务作为一种新兴的贸易形式, 以其开放性和高效性的特点在现代经济生活中发挥着重要作用[1]. 供应链作为电子商务的重要组成部分, 在供应商、制造商、经销商、零售商和消费者等多个节点之间建立联系, 保证电子商务平台上各种交易的处理[2].

然而, 由于每个企业都各自独立维系着一条记录自身业务数据的数据库, 致使信息相互割裂, 无法满足多方互信问题[3]. 出于风控、资质和信用的考虑, 金融机构仅愿意为核心企业提供融资服务, 而不愿意再投入额外的人力、物力去校验中小企业的数据信息的真实性. 在考虑供应链安全性前提下, 文献[4]通过使用基于物联网架构的区块链, 来构建食品供应链. 文献[5]利用区块链解决了药品安全问题, 实现了药品的可追溯. 对于供应链的按需服务, 文献[6]提出了基于区块链的汽车供应链, 重点强调区块链技术有能力提高供应链同行之间的可信度, 以减少交易欺诈.

在共识处理性能方面, 区块链去中心化特征带来的代价之一就是区块链的性能. 文献[7]提出了一种基于分层模型的多中心的共识机制, 与其他共识机制不同的是设计了主从链, 用以提高区块链系统的吞吐量. 文献[8]提出了一种针对投票过程的奖惩方案及其相应的节点信用评估方案. 文献[9]提出了基于跳跃 Hash 的共识机制, 引入分片概念, 为解决分片带来的隐患, 介绍了异步共识的模式. 文献[10]根据节点的服务质量来划分不同的区域, 进而基于此提出了一种更加优秀的共识协议, 在这个协议中, 整个网络被划分为小区域, 每个区域根据其 QoS 指定一个节点. 文献[11]提出了时间阈值的概念, 即各节点每隔一定时间会对区块的请求日志进行清除, 从而减少 PBFT 中垃圾回收的网络开销.

上述文献提出的解决方案, 仅是借助区块链技术特性应用在供应链领域, 并没有关注区块链+供应链中主节点的选取方法存在弊端和网络中广播消息需要频繁地验证所引发的共识性能问题. 基于此, 本文提出并设计了一种应用于供应链的区块链 PBFT 共识算法优化方法, 以实现供应链应用场景下, 区块链共识的高效与安全性.

1 预备知识

1.1 区块链

区块链是集分布式存储、智能合约、共识算法、密码学等关键技术为一体的全新分布式记账技术, 具有多中心、难以篡改、可追溯等天然特性[12]. 具体地说, 一个区块是由块首和块体两部分所构成, 块首存放前一区块的散列值; 块体则存放数据[13]. 区块与区块之间依据出块时间依次勾稽相连, 便形成一条完整的区块链, 如图 1 所示.

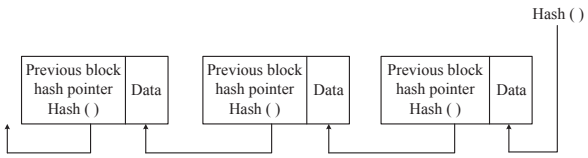

图 1 区块链示意图

依据节点共识的准入机制可将区块链分为公有链、私有链和联盟链. 由于供应链企业的交易业务特性, 本文采用 Hyperledger 项目下的 Fabric 联盟链. Fabric 不仅支持实用拜占庭容错算法 (practical Byzantine fault tolerance, PBFT), 而且是面向企业级的分布式账本平台[14].

1.2 PBFT 共识算法

2002 年, Castro 等人提出了 PBFT 共识算法, 并将其应用于解决拜占庭将军问题的分布式数据管理系统中[15]. PBFT 可以在区块链网络中 1/3 以内共识节点恶意出错的情况下, 通过三阶段协议来保证系统的安全性与活性[16]. 在区块链网络中, 存在$|R|$个共识节点, 当且仅当$|R|=3f+1$ 时, PBFT 方可正常运行, 并通过式 (1), 来进行主节点 p 的选取工作.

$$p = v \bmod |R| \tag{1}$$

其中, f 为非诚实节点的个数, v 表示当前主节点所运行的视图号. PBFT 运行流程如图 2 所示.

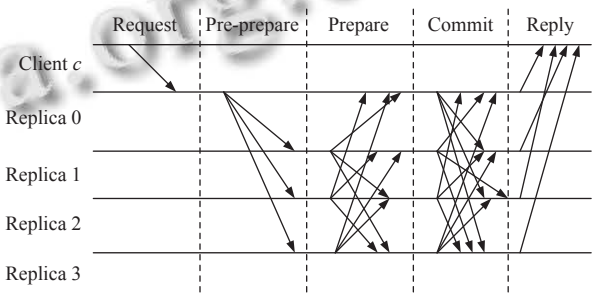

图 2 PBFT 运行流程

1.3 K-近邻

K-近邻算法是一种基本分类和回归方法[17]. 即给定一个训练数据集, 对新的输入实例, 在训练数据集中找到与该实例最邻近的 k 个实例, 这 k 个实例的多数属于某个类, 就把该输入实例分类到这个类中[18]. 常见度量最近邻有以下几种方式.

设特征空间χ是n维实数向量空间R^n, x_i, $x_j \in \chi$, $x_i = (x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(n)})^{\mathrm{T}}$, $x_j = (x_j^{(1)}, x_j^{(2)}, \cdots, x_j^{(n)})^{\mathrm{T}}$, x_i, x_j的

L_P距离定义为:

$$L_P(x_i,x_j)=\left(\sum_{l=1}^{n}\left|x_i^{(l)}-x_j^{(l)}\right|^p\right)^{\frac{1}{p}} \tag{2}$$

此处$P\geqslant 1$, 当$P=2$时, 为欧氏距离, 即:

$$L_2(x_i,x_j)=\left(\sum_{l=1}^{n}\left|x_i^{(l)}-x_j^{(l)}\right|^2\right)^{\frac{1}{2}} \tag{3}$$

当$P=1$时, 为曼哈顿距离, 即:

$$L_1(x_i,x_j)=\sum_{l=1}^{n}\left|x_i^{(l)}-x_j^{(l)}\right| \tag{4}$$

当$P=\infty$时, 它是各个坐标距离的最大值, 即:

$$L_\infty(x_i,x_j)=\max_l\left|x_i^{(l)}-x_j^{(l)}\right| \tag{5}$$

在第 4 节中, 将给出本文选择 K-近邻分类算法的依据.

2 区块链与供应链集成的框架设计

由于区块链在隐私保护和防篡改上具有独特优势, 其可以满足供应链中供应商和买方之间的信任需求. 因此, 本文构造了区块链+供应链的集成框架. 在这个框架下, 区块链为供应链提供了一个良好的信息共享渠道, 节点与节点之间可以充分信任, 如图 3 所示.

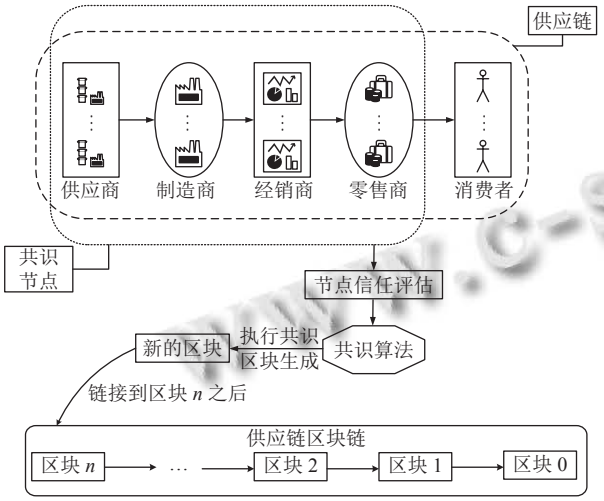

图 3 区块链与供应链集成的框架

在图 3 中, 区块链网络中的共识节点是由供应链各参与主体所构成. 同时, 根据参与主体的信任评估来确定主节点和备份节点的选择. 在经过多方严格共识的基础上, 一个新的区块得以生成, 并链接到区块 n 之后, 从而保证供应链中存储的信息的安全性和隐私性.

3 区块链与供应链集成的框架设计

3.1 节点信任评估

本文研究所需供应链数据集是由课题合作单位所提供的真实数据, 单位管理人员已对数据进行初步处理. 根据数据驱动的方法, 利用 K-近邻分别从{资产负债、应收账款周转率、总资产周转率、存货周转率、销售利润率、净值产收益率、净利润增长率}等 7 个属性特征来评估供应链中共识节点的信任情况, 选择信任最高的节点担任主节点.

为了保证主节点的可信和共识效率, 每个交易周期都会重新评估基于最新交易数据的所有节点的信任情况. 根据新数据, 重新挑选出新的主节点.

3.2 优化后的 PBFT 执行流程

本文经过优化后的 PBFT 共识算法在保持原有三阶段广播协议的基础上, 增加了一个初始化处理过程, 如图 4 所示.

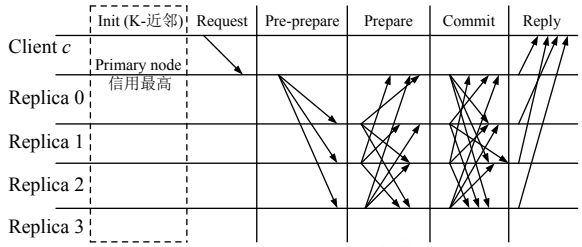

图 4 优化后的 PBFT 运行流程

(1) 初始化阶段 (init): 对区块链网络中的共识节点通过 K-近邻分类模型进行信任评估. 依据信任评估结果挑选主节点, 从而替换依据式 (1) 的选主方式. 此时, 区块链网络中共识节点的初始化工作完成. 如果供应链中没有新的节点加入或旧的节点退出, 则就不需要重新启动共识节点信任评估操作.

(2) 预准备阶段 (pre-prepare): 主节点 (Replica 0) 给接收到客户端 (Client c) 的请求分配提案编号 n, 并将预准备消息<<PRE_PREPARE, v, n, digest>, message> 列表多播至各个备份节点 (Replica 1, Replica 2, Replica 3). 其中, message 为客户端的请求消息, digest 为 message 摘要.

(3) 准备阶段 (prepare): 备份节点接收到主节点发送的预准备消息列表后, 开始检验 message 摘要的合法性, 并确保提案编号 n 满足水线的取值范围. 如果验证通过, 则向全网多播准备消息<PREPARE, v, n, digest, k >, 并将预准备消息和准备消息写入消息日志中; 否

则, 备份节点不做任何操作. 共识节点收集来自其他节点的准备消息, 如果共识节点集齐超过全网一半的准备消息时 (也包含自己的), 则该节点进入准备就绪状态; 否则, 将一致性共识验证失败的消息<FAILURE, v, t, c, k >发送至客户端 c.

(4) 确认阶段 (commit): 进入准备就绪状态的共识节点, 则向全网多播一条确认消息<COMMIT, v, n, $D(m)$, k >, 并收集来自其他节点的确认消息. 如果共识节点集齐超过全网一半的确认消息时 (也包含自己的), 则该节点就进入提交状态, 执行客户端 (Client c) 的请求, 最后将处理成功的结果<SUCCESS, v, t, c, r, k >返回至客户端 c, 其中 r 为最终的执行结果; 否则, 将一致性共识验证失败的消息<FAILURE, v, t, c, k >发送至客户端 c.

3.3 算法描述

本文优化后的 PBFT 共识算法伪代码如算法 1 和算法 2 所示.

算法 1. 信任评估选取主节点 p

输入: 供应链节点特征属性数据集
输出: 主节点 p

```
1) for int k = 0 to |R|–1 do
2)     K-近邻 (对节点 k 进行信用分类);
3)     将 k 节点信用等级存入集合 R 中;
4) end for
5) R = {R_0, ⋯, R_|R|–1};
6) if R_k 是信用最高的节点 then
7)     供应链中第 k 个节点为主节点;
8) end if
9) return p
```

算法 2. 本文优化后的 PBFT 算法

输入: 客户端 c 请求消息
输出: 处理结果 r

```
1) if (供应链中有节点数据更新) then
2)     for int k = 0 to |R|–1 do
3)         由算法 1, 选出主节点 p;
4)     end for
5)     p 接收的请求消息, 并分配编号 n;
6)     执行 PBFT 三阶段协议;
7) end if
8) if (共识节点通过验证) then
9)     return <SUCCESS, v, t, c, r, k >;
10)    生成一个新的区块;
11) else
12)    return <FAILURE, v, t, c, k >;
13) end if
```

3.4 算法的时间复杂度分析

由于 K-近邻的时间复杂度为 O(m×n). 其中, m 为共识节点数据集的特征个数, n 为共识节点的个数; PBFT 共识算法的时间复杂度为 O(n^2). 经过上面小节处理流程分析, 可以计算出本文经过优化之后的 PBFT 共识算法最坏情况下的时间复杂度为 O(n^2+m×n). 虽然本文算法最坏情况下的时间复杂度有所增加, 但仍处于平方阶级别.

4 实验验证与分析

4.1 分类算法的选择

本文使用 WEKA 开源软件对 C4.5[19]、K-近邻[20]、朴素贝叶斯[21]、支持向量机[22]、贝叶斯网络[23]等机器学习分类算法建立分类模型, 并采用十折交叉验证法 (10-fold cross validation) 测试数据集. 最后, 通过式 (6) 计算出分类准确率 (*Precision*):

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

其中, TP 和 FP 分别为分类正确和分类错误的实例, 测试数据集分类准确率如表 1 所示.

表 1 5 种分类算法的模型分类准确率比较

分类算法	准确率 (%)
C4.5	70.86
K-近邻	73.14
朴素贝叶斯	68.37
支持向量机	67.42
贝叶斯网络	65.78

由表 1 可知, K-近邻分类算法的准确率略高于其他 4 种. 因此, 本文选择其作为共识节点信任评估分类的方法.

4.2 实验环境

本文使用 Caliper 区块链性能测试工具, 来展开性能测试与分析. 首先, 在 PC 机上配置好 Ubuntu 操作系统和虚拟机, 并为虚拟机分配内存和磁盘大小. 然后, 针对 Hyperledger Fabric 的实现所需要的基础编程环境 Go 语言及其相关工具进行安装配置. 最后, 编译安装 fabric-peer、fabric-orderer、fabric-ca 组件. 实验使用的硬件配置如表 2 所示.

4.3 实验分析

为了和优化前的 PBFT 共识算法作对比, 本文选取区块大小在[50, 350]范围内的交易数量进行模拟测

试, 经过 10 次实验计算出平均值, 得到了不同区块大小下的交易时延, 如图 5 所示. 并选取 40 次实验的平均值作为优化后的 PBFT 共识算法的 TPS 值, 并将其与目前其他成熟的区块链平台的吞吐量作比较, 对比结果如图 6 所示.

表 2 实验环境配置表

参数	配置
虚拟机	VMware Workstation 12 Pro
内存大小	16 GB
CPU	Intel(R) Core(TM) i7-8550U@3.70 GHz
Linux内核	4.16
虚拟机操作系统	Ubuntu 16.04.2 LTS (64 bit)
硬盘	521 GB

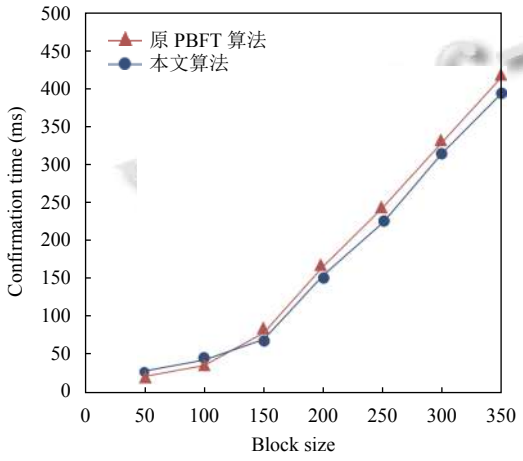

图 5 不同区块大小的交易时延

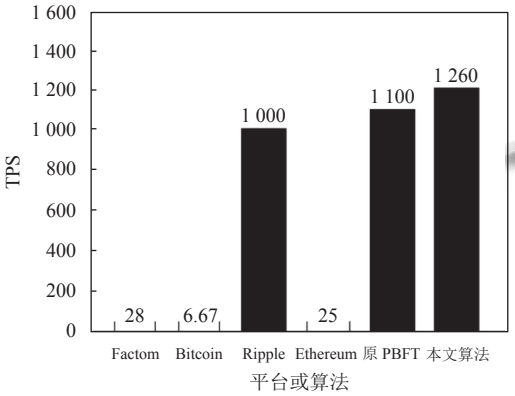

图 6 本文方法与其他主流区块链平台共识算法 TPS 比较

本文分别依次模拟取 f 的值为 0, 1, 2, 3, 4, 5 来进行实验比较, 每个取值实验 10 次, 取 10 次实验的平均值作为该次实验的最终处理结果, 并与原 PBFT 共识算法的实验结果作对比, 如图 7 所示.

由图 7 可知, 本文算法若出现 3 个以上拜占庭节点时, 时延会趋向于正无穷, 即无法在有限的时间内正常完成一致性共识验证工作. 相较于原 PBFT 共识算法, 可以容忍更多区块链网络中的共识节点出错.

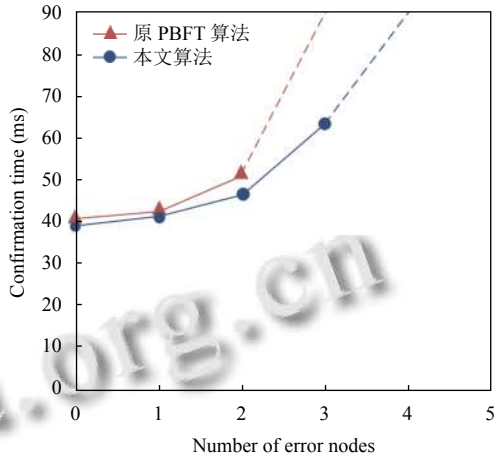

图 7 时延与出错节点数目关系

5 结论与展望

本文提出一种应用于供应链的区块链 PBFT 共识算法优化方法. 根据数据驱动的方法, 运用 K-近邻来优化 PBFT 共识算法的主节点选取规则, 从而解决区块链共识过程中, 因视图切换所引发的效率问题. 实验结果表明, 与现有方法相比, 本文方法在吞吐量、时延、容错性等共识性能上均有一定的提升, 验证了所提方法的有效性. 下一步, 将区块链与数字农业相结合进行应用研究.

参考文献

1 Cui Y, Mou J, Cohen J, *et al*. Understanding information system success model and valence framework in sellers' acceptance of cross-border e-commerce: A sequential multi-method approach. Electronic Commerce Research, 2019, 19(4): 885–914. [doi: 10.1007/s10660-019-09331-0]

2 Guo WJ, Qi H, Guo QJ. An E-commerce based supply chain model of chain catering enterprises. Proceedings of the 2009 International Conference on Management of E-commerce and E-government. Nanchang: IEEE, 2009. 322–325.

3 Wang YY, Hua ZS, Wang JC, *et al*. Equilibrium analysis of markup pricing strategies under power imbalance and supply chain competition. IEEE Transactions on Engineering Management, 2017, 64(4): 464–475. [doi: 10.1109/TEM.2017.2693991]

4 Mondal S, Wijewardena KP, Karuppuswami S, *et al*. Blockchain inspired RFID-based information architecture for

food supply chain. IEEE Internet of Things Journal, 2019, 6(3): 5803–5813. [doi: 10.1109/JIOT.2019.2907658]

5 Kumar R, Tripathi R. Traceability of counterfeit medicine supply chain through Blockchain. Proceedings of the 11th International Conference on Communication Systems & Networks. Bengaluru: IEEE, 2019. 568–570.

6 Sharma PK, Kumar N, Park JH. Blockchain-based distributed framework for automotive industry in a smart city. IEEE Transactions on Industrial Informatics, 2019, 15(7): 4197–4205. [doi: 10.1109/TII.2018.2887101]

7 She W, Gu ZH, Liu W, *et al*. A channel matching scheme for cross-chain. International Journal of Embedded Systems, 2020, 12(4): 500–509. [doi: 10.1504/IJES.2020.107646]

8 Wang YH, Cai SB, Lin CL, *et al*. Study of blockchains's consensus mechanism based on credit. IEEE Access, 2019, 7: 10224–10231. [doi: 10.1109/ACCESS.2019.2891065]

9 Sonkamble RG, Phansalkar SP, Potdar VM, *et al*. Survey of interoperability in electronic health records management and proposed blockchain based framework: MyBlockEHR. IEEE Access, 2021, 9: 158367–158401. [doi: 10.1109/ACCESS.2021.3129284]

10 Li HD, Gao P, Zhan Y, *et al*. Blockchain technology empowers telecom network operation. China Communications, 2022, 19(1): 274–283. [doi: 10.23919/JCC.2022.01.020]

11 Liang ZH, Huang YX, Cao ZC, *et al*. Creativity in trusted data: Research on application of blockchain in supply chain. International Journal of Performability Engineering, 2019, 15(2): 526–535.

12 黄宇翔, 梁志宏, 张梦迪, 等. 面向学分银行的区块链学习成果管控模型. 计算机工程, 2019, 45(5): 18–24.

13 徐恪, 凌思通, 李琦, 等. 基于区块链的网络安全体系结构与关键技术研究进展. 计算机学报, 2021, 44(1): 55–83.

14 Androulaki E, Barger A, Bortnikov V, *et al*. Hyperledger fabric: A distributed operating system for permissioned blockchains. Proceedings of the 13th EuroSys Conference. Porto: ACM, 2018. 30.

15 Castro M, Liskov B. Practical byzantine fault tolerance and proactive recovery. ACM Transactions on Computer Systems, 2002, 20(4): 398–461. [doi: 10.1145/571637.571640]

16 黄宇翔, 梁志宏, 黄芯, 等. 基于区块链的供应链可信数据管理. 计算机系统应用, 2018, 27(12): 9–17. [doi: 10.15888/j.cnki.csa.006674]

17 Cover T, Hart PE. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 1967, 13(1): 21–27. [doi: 10.1109/TIT.1967.1053964]

18 丁世飞, 徐晓, 王艳茹. 基于不相似性度量优化的密度峰值聚类算法. 软件学报, 2020, 31(11): 3321–3333. [doi: 10.13328/j.cnki.jos.005813]

19 Mantas CJ, Abellán J. Analysis and extension of decision trees based on imprecise probabilities: Application on noisy data. Expert Systems with Applications, 2014, 41(5): 2514–2525. [doi: 10.1016/j.eswa.2013.09.050]

20 Al-Yaseen WL, Othman ZA, Nazri MZA. Intrusion detection system based on modified K-means and multi-level support vector machines. Proceedings of the 1st International Conference on Soft Computing in Data Science. Putrajaya: Springer, 2015. 265–274.

21 Aljawarneh S, Aldwairi M, Yassein MB. Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. Journal of Computational Science, 2018, 25: 152–160. [doi: 10.1016/j.jocs.2017.03.006]

22 Hashem SH. Efficiency of SVM and PCA to enhance intrusion detection system. Journal of Asian Scientific Research, 2013, 3(4): 381–395.

23 刘鹤, 季宇, 韩建辉, 等. 面向阻变存储器的长短期记忆网络加速器的训练和软件仿真. 计算机研究与发展, 2019, 56(6): 1182–1191. [doi: 10.7544/issn1000-1239.2019.20190113]

(校对责编: 牛欣悦)