去中心化的征信系统模型研究

陈春玲,沈 阳,余 瀚
(南京邮电大学 计算机学院,江苏 南京 210003)

摘 要: 为减少现代征信系统在数据存储上因单点故障和被恶意篡改而导致的安全性问题,针对数据的存储过于中心化、易伪造等隐患,提出了一种基于区块链存储结构和具有去中心化特征的征信系统模型。通过将存储数据的地址信息加密在区块链体内,采用广播方式在全网存储节点,并在校验数据时获取全网绝大部分的一致性区块链来保证信息的可靠性。实验结果表明,将用户的隐私数据保存在存储节点中,通过全网校验节点协作验证后将存储节点的地址信息写入加密的区块体内,保证了数据的隐秘性;在数据校验时,从全网获取最新副本更新至本地,保证了数据不被第三方恶意篡改;将用户的敏感数据使用区块链架构存储,能够大大提高数据的安全性,同时保证了存储效率。

关键词: 分布式;区块链;去中心化;征信模型
中图分类号: TP301.6 **文献标识码:** A **文章编号:** 1673-629X(2019)03-0122-05
doi: 10.3969/j.issn.1673-629X.2019.03.026

Research on Decentralized Model for Credit Information System

CHEN Chun-ling, SHEN Yang, YU Han
(School of Computer Science & Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: In order to reduce the security caused by single point failure and malicious tampering in data storage of modern credit system, we design a credit system model based on the blockchain storage structure and decentralized characteristics for the hidden dangers such as over-centralization and forgery of data storage. By encrypting the address information of the stored data in the blockchain body, using the broadcast to store the nodes in the entire network, and obtaining the most consistent blockchain of the entire network when verifying the data, the reliability of the information is ensured. The experiment shows that the user's private data is stored in the storage node, and the address information of the storage node is written into the encrypted block body after the cooperation verification of the whole network check node, ensuring the privacy of the data. During data validation, the latest copy from the entire network is updated locally, ensuring that the data is not maliciously manipulated by third parties. Storing sensitive data of the user by the blockchain architecture can greatly improve data security and ensure storage efficiency.

Key words: distributed; blockchain; decentralization; credit model

1 概 述

现代经济体系的发展,离不开信用的支撑。征信作为信用体系中的关键环节,奠定了金融信用风险管理的基础。征信有诸多益处,如:防范信用风险,促进信贷市场发展,服务其他授信市场,提高履约水平,加强金融监管和宏观调控,维护金融稳定,等等[1]。国内个人和企业征信信息尤为庞大,数据类型也从单一的文本数据到半结构化数据等多种多样。较为明显的是,绝大多数征信平台都使用了分布式数据存储来代替常规的单机或者其他网络存储系统[2],这仅仅改变了面临大数据量时的存储效率。但在新经济发展形势下,传统征信业中存在的信用信息不对称、数据采集渠道受限、数据隐私保护不力等问题依旧给现有的征信存储结构提出了重大的挑战。

传统分布式征信系统中,采用较多的就是一主多从(Master/Slaves)的模式(见图1)。在此模式下,数据分配和均衡由一个主服务器(Master)调度,多个从服务器(Slaves)负责元数据的存储。如现今应用广泛

收稿日期: 2018-04-13 **修回日期:** 2018-08-16 **网络出版时间:** 2018-12-20
基金项目: 国家自然科学基金(11501302)
作者简介: 陈春玲(1961-),男,教授,硕士,研究生导师,研究方向为软件工程、分布式组件技术、网络信息安全及其应用;沈 阳(1991-),男,硕士研究生,研究方向为网络与信息安全。
网络出版地址: http://kns.cnki.net/kcms/detail/61.1450.TP.20181219.1542.066.html

的是 Hadoop 上的 Zookeeper 分布式文件系统[3] 和 Google 的文件系统[4]。当主服务器发生故障或宕机等问题不能继续工作时,多个从服务器会依照自己的共有算法推选出新的 Master,所以分布式系统的可适应性较强。

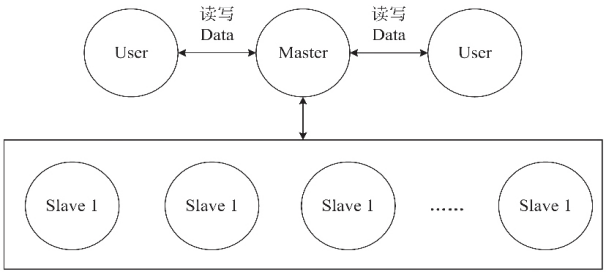

图 1　Master/Slaves 分布式存储

在分布式数据存储过程中,元数据通常包括了存储数据的物理节点信息等,被保存在从节点中。在存储和读取数据时,都需要访问主节点来获取元数据信息。这种存储元数据的方式存在以下问题:

(1) 容易造成单点故障,即如果主节点发生故障或宕机,那么元数据可能在一定时间内无法被读取。虽然通过节点选举可以一定程度上避免该问题,但节点之间切换效率较低;

(2) 存储元数据的节点可信性很难保证,即中心节点的较高权限可以直接篡改元数据;

(3) 数据在主节点校验和中转的过程中,信息的完整性和安全性得不到保障[5]。

针对传统分布式征信系统中存在的问题,结合区块链的数据可追溯性、不可修改性、共识机制等技术特点,提出了一种去中心化的征信系统模型(decentralized credit information system,DCIS)。

2　预备知识

2.1　区块链

区块链(Blockchain)是比特币的底层技术,它是一串使用密码学方法相关联产生的数据块,每一个数据块中包含了一次比特币网络交易的信息,用于验证其信息的有效性和生成下一个区块,其本质上是一个去中心化的数据库。区块链技术与传统数据库最主要的区别在于,传统数据库是集中且相互封闭隔离的,而区块链数据库由于去中心化,分布于整个网络所有计算机,使用分布式账本技术进行核算,没有行使管理权限的第三方。区块链技术的主要特点包括:

(1) "去中心化"。区块链技术依托于分布式记账、分布式传播及分布式存储三大技术,所有信息的记录、传播都在网络所有节点上进行,没有中心服务器进行操控。

(2) 开放及匿名性。由于公钥、私钥等现代密码技术的应用,数据库中除了私人信息被加密,其他信息均是公开透明的,网络任何节点均可进行查询。

(3) 不可篡改性。区块链中,任何一个节点想要更改数据,必须得到整个网络中 51% 节点认可方可实现,单个节点对数据库的修改无效[6]。

(4) 可追溯性。区块链中每个区块都会被盖上一个时间戳,以说明信息何时被写入,并据此建立一个可根据时序追根溯源的大账本,从而可以验证可疑数据的真实可靠性[7]。另外,区块链这种可溯性特点还可以行使类似现实生活中审计的角色,不仅保证了监管部门对数据进行有效检索和监察,还可以有效杜绝虚假数据的干扰。

(5) 支持可编程智能合约[8]。区块链智能合约相当于现实生活中的交易合同,可编程则是指可以根据交易内容的不同对交易合同进行修改。正是由于可编程智能合约的存在,使区块链技术可应用场景被无限放大,通过对智能合约的设定和编程,可以满足人们对于不同交易类型的需求。

2.2　区块链结构

在区块链中,没有中心节点的互通,每一个节点地位平等,整个区块链是一个对等的点对点网络,所有参与计算的节点都保存了区块链的副本,每个节点在特定时段会更新本地区块链[9-10],上一个区块的哈希信息存储在下一个区块的区块体之中,如图 2 所示。

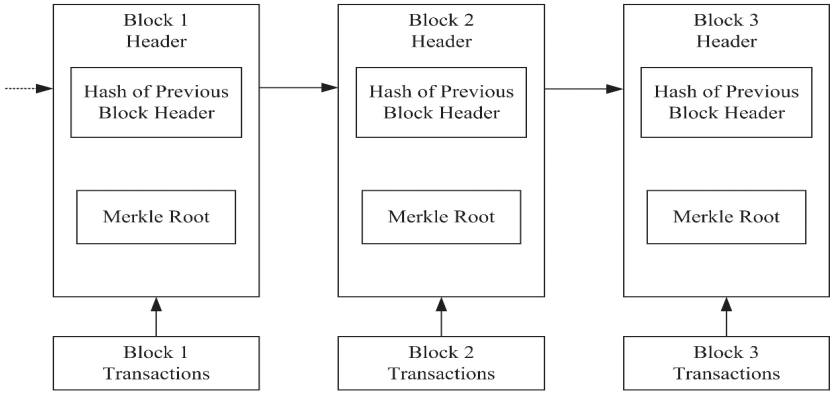

图 2　区块体组成

每一个区块由区块头(header) 和区块体(body) 组成,数据的操作记录作为原始数据记录在区块体中,并且加密保存。一个区块体中包含多条数据,组成 hash 树根(Merkle Root) 。在区块头中保存着时间戳、前一个区块的 hash 值、Merkle Root[11] 。其中 hash 是一种算法: 一段明文经过哈希算法处理后,将转化成为一段长度较短、位数固定的散列数据。这个加密过程是不可逆的,无法通过输出的散列数据倒推原本的明文。同时,输出的散列数据和输入的明文是一一对应的,任何一个输入信息的变化都将使得输出的散列数据发生变化。时间戳给区块链里的每一个信息都标记了其发生时间,证明交易记录的真实性。任何人都无法篡改时间戳,区块链通过时间戳保证了每个区块依次顺序相连。

如果想恶意篡改区块中的内容,那就必须重新计算当前以及所有前驱区块的哈希值,具有很高的计算复杂性。通过这一特点可以保证区块链信息内容的完整性,同时也能解决传统分布式存储中主节点失效切换的问题。当任一节点的本地区块链和全网大多数节点的区块链不一致时,直接同步最新的区块链至本地,防止信息篡改[12] 。

3 去中心化的征信系统模型

3.1 模型单点构造

DCIS 单节点的模型结构如图 3 所示。

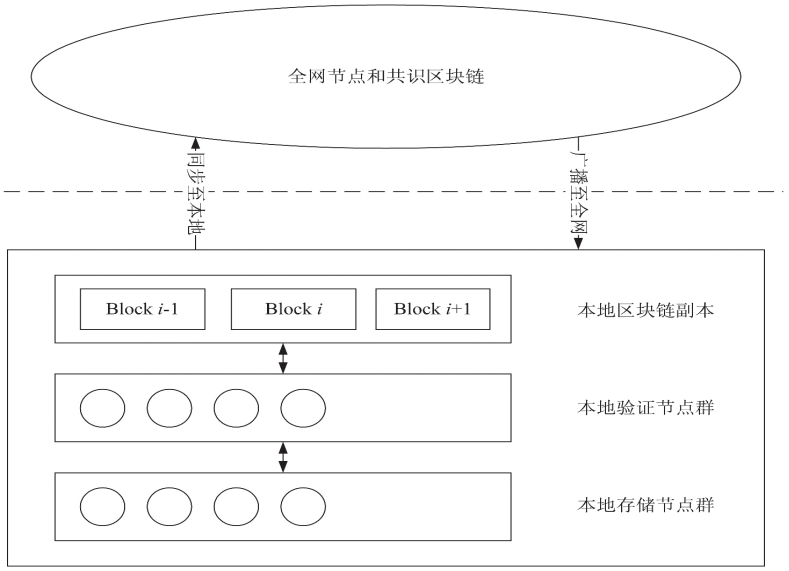

图 3 DCIS 单节点的模型结构

图中上半部分为与外网全节点交互,下半部分为单个节点构造。其中单节点由三部分构成: 本地区块链副本、本地验证节点群、本地存储节点群。在真实的部署环境中可以将本地验证节点和本地存储节点部署在外网,加强单机存储性能。

(1) 本地区块链副本。该副本始终与全网绝大多数节点同步数据,当本节点数据丢失或者被篡改时,可以通过校验与全网的 hash 比较,即可恢复数据。并且链中的每个区块数据写入后不可更改(只读 Read-Only) ,如果用户修改数据且合法,则修改内容将记录在新生成的区块中,再全网广播同步。

(2) 本地验证节点群。由若干个校验节点代替传统分布式数据库中的单一主节点,在用户存储数据时,先同步全网区块链至本地,多个验证节点同时校验用户戳和数据,优先完成数据校验的节点负责调度本地存储节点进行数据存储,其他节点回到初始状态,新的信息更改记录由验证节点写入本地区块链副本,全网广播; 在数据校验时,先同步全网区块链至本地,验证节点接受外部请求并校验权限,检索存储节点中数据并返回。

(3) 本地存储节点群。由若干存储节点构成,负责对接验证元数据写入和数据查询功能,也可将其部署在远程。

3.2 数据存储

当一条新的征信信息将要写入(新建或修改) 时,本地和全网的验证节点需要先校验其合法性。当签名的 hash 值验证通过后,再构造区块体,并入链尾。分为以下五步:

(1) 用户对待存储的数据进行数字签名操作组合成<签名,数据>对。

(2) 将一定时间内不同用户生成的数据整合成数据对集 $D=\{<sign_1,data_1>,<sign_2,data_2>,\cdots,<sign_m,data_m>\}$ 。

(3) 将 D 发送给全网验证节点集 Club_Check,节点并行校验数据集。最先完成校验的节点 i 具有优先记录权,其他节点回到初始状态。

(4) 验证节点 i 将当前周期内需要记录的所有信息<sig,data>,计算组成 Merkle 树,并构造一个新的区块写入本地。

(5) 全网广播,将本地区块链更新至全网。

值得说明的一点: 不同于比特币等常规虚拟货币的 POW(工作量证明) 等机制的计算(验证) 节点之间的竞争关系,优先完成签名校验的节点具有记录权,其他节点回到初态,此机制可提高节点验证效率[13]。

3.3　共享和校验

当数据一旦写入区块并全网同步后,数据就永久保存在链中无法修改历史记录,只能对数据进行进一步操作。当数据需要共享和校验时分为以下四步:

(1) 当全网请求最终到达数据拥有者时,由数据的拥有者发起请求。将带有签名信息 sign 的请求发送给全网验证节点集 Club_Check。

(2) 验证节点检测本地与全网区块链的一致性,如果不一致,则更新到最新状态。

(3) 当全网验证节点通过签名校验后检索 sign 对应的 data,此时所有验证节点将进行检索。

(4) 全网绝大多数节点获得相同的 data 值之后,第一个得到数据<sig,data>的节点返回给发起请求的用户的同时,广播其他节点暂停检索回到初始状态。

3.4　模型分析

区块链表示存储数据的区块,且随着时间有序排列。其中存储的是每一个用户及其征信信息的修改(或新建)。所以当提供用户的有效签名 sign 和对应的时间过滤后可得到唯一的数据信息。所以保证了数据的可追溯性。

当需要读取或者校验数据时,为避免本地篡改或者读脏数据[14]等行为,验证节点会定期更新同步本地区块链。同步过程中,验证节点会对比全网区块链和本地区块链的待同步区块头信息,因为任何的信息单方面修改都会导致块头信息不一致,此时需要替换本地区块。所以保证了数据的完整性。

任何非用户本人的数据直接请求,最终都会转发至用户,由用户决定是否授权进一步获取原始数据。恶意获取隐私行为都会在最终授权这关被及时发现。所以保证了数据的安全性。

4　实验与结果分析

4.1　实验设计

实验硬件环境: 处理器 Intel(R) Core(TM) i5-6500 CPU @ 3.20 GHz,内存 8 G,系统为 Win10×64。该系统之上建立三台虚拟机生成 Zookeeper 分布式环境,一台 master 节点服务器,两台 slaves 存储服务器。虚拟机与本机采用桥接网卡形式模拟真实环境。开发语言 Java(jdk1.8),验证节点采用多线程模拟。其中征信数据为常规文档数据。

4.2　实验结果分析

实验中采用控制变量法,即某几个变量变化时,其他变量保持不变,从而观察性能和稳定性。

实验表明存储不同大小数据时在 DCIS 和 Zookeeper 分布式系统中,存储时间随着数据量的增加而增加,且 DCIS 额外会多损耗一点时间,主要原因是 DCIS 模型在存储征信数据时需要校验用户签名和生成区块操作,会耗费少量额外时间。说明该模型可以最大程度保证数据的存储效率。且当数据量为 0 时,系统依旧会损耗部分时间,因为验证节点的校验签名操作,在无数据存储但有数据存储请求的情况下,依旧会执行。

图 4 是“用户数和验证节点数量对存储时间的影响”的实验结果。表明随着用户数和验证节点数量的增加,数据存储时间也会增加且后期趋于线性。主要因为数据量和验证节点数一定时,用户数变多会导致签名数量的增加,延长了校验的时间长度。另一方面,验证节点同步各自本地区块链时由于环境原因是同步操作,节点越多,则平均节点校验时间相应增加。“用户数和数据量变化对存储时间的影响”的实验结果与图 4 相似。

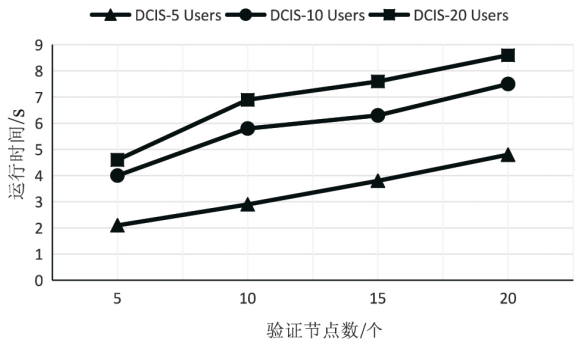

图 4　验证节点数和用户数变化时 DCIS 效率

图 5 是“数据量和验证节点数量变化对存储时间的影响”的实验结果。表明随着数据量和验证节点数量的增加,数据存储时间也会增加。

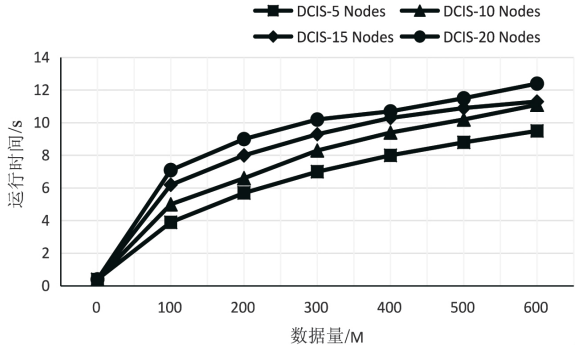

图 5　验证节点数和数据变化时不同存储方式的效率

图 6 是“主节点失效时系统恢复能力”的实验结

果。表明 DCIS 模型和 Zookeeper 模型在部分节点无法工作的情况下恢复能力差别较小。原因在于 DCIS 在节点失效时新节点的替换需要同步全网区块链副本至本地,这取决于通信状况。

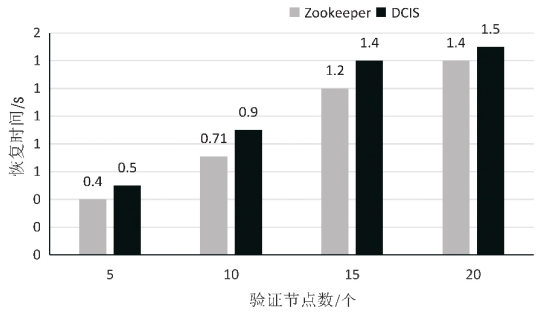

图 6 验证节点个数变化时的系统恢复效率

综上所述,使用去中心化征信系统模型在存储数据时,在牺牲少量时间的基础上,能保证较高的系统数据安全性,弥补现代征信系统的部分短板。

5 结束语

结合区块链的去中心化、共识机制、不可修改记录等特点,提出了一种基于去中心化的系统模型 DCIS。实验结果表明,DCIS 具有能够追溯数据的完整流水记录,能防止恶意篡改,极符合征信行业的业务需求[15],且任何查看用户征信信息的行为都需要得到数据拥有者的校验授权,极大地保护了用户的隐私数据。实验模型中使用协同机制,可以在多线程环境中多节点同时验证。且以 Zookeeper 分布式模型进行实验参考,表明 DCIS 也具有较好的存储性能,能够在牺牲少量性能的基础上极大地加强了数据的安全性。但是,区块链也存在私钥丢失或泄漏、对用户数据“被遗忘”、现有信息系统管理建设条例和征信监管体系等不适应的问题,这也是今后要做的主要工作。

参考文献:

[1] ØLNES S,UBACHT J,JANSSEN M. Blockchain in government: benefits and implications of distributed ledger technology for information sharing [J]. Government Information Quarterly,2017,34(3):355-364.

[2] 叶湘榕. 互联网金融背景下的征信体系完善研究 [J]. 华北金融,2015(6):48-53.

[3] 谭玉靖. 基于 ZooKeeper 的分布式处理框架的研究与实现 [D]. 北京:北京邮电大学,2014.

[4] GHEMAWAT S,GOBIOFF H,LEUNG S T. The Google file system [C]//Proceedings of 9th ACM symposium on operating systems principles. Bolton Landing,NY,USA: ACM,2003:29-43.

[5] 陈天伟,彭凌西. 基于 ZooKeeper 的一种分布式系统架构设计与实现 [J]. 通信技术,2018,51(1):87-91.

[6] LARIOS-HERNANDEZ G J. Blockchain entrepreneurship opportunity in the practices of the unbanked [J]. Business Horizons,2017,60(6):865-874.

[7] HUCKLE S,BHATTACHARYA R,WHITE M,et al. Internet of things,blockchain and shared economy applications [J]. Procedia Computer Science,2016,98:461-466.

[8] YEOH P. Regulatory issues in blockchain technology [J]. Journal of Financial Regulation and Compliance,2017,25(2):196-208.

[9] KSHETRI N. Blockchain's roles in strengthening cybersecurity and protecting privacy [J]. Telecommunications Policy,2017,41(10):1027-1038.

[10] LEMIEUX V L. Trusting records: is blockchain technology the answer? [J]. Records Management Journal,2016,26(2):110-139.

[11] 陈兰香,邱林冰. 基于 Merkle 哈希树的可验证密文检索方案 [J]. 信息网络安全,2017(4):1-8.

[12] VRANKEN H. Sustainability of bitcoin and blockchains [J]. Current Opinion in Environmental Sustainability,2017,28(10):1-9.

[13] 庹小忠. 基于区块链的加密信息备份系统研究与设计 [D]. 南充:西南石油大学,2017.

[14] DINH T T A,WANG Ji,CHEN Gang,et al. Blockbench: a framework for analyzing private blockchains [C]//Proceedings of the 2017 ACM international conference on management of data. [s. l.]: ACM,2017:1085-1100.

[15] 蔡维德,郁 莲,王 荣,等. 基于区块链的应用系统开发方法研究 [J]. 软件学报,2017,28(6):1474-1487.