基于区块链和工作量证明的数据抗篡改技术研究

徐 勇[1] 杜 明[1]
XU Yong DU Ming

摘 要 随着信息化技术的不断深化应用，数据安全和抗篡改愈发重要。在各类信息系统建设中，不论是科学研究的相关数据，还是设备管理的设备履历数据，都需要数据真实可信，而现实中经常出现需要数据时临时补造数据的情况。为防止这种问题，需要设计一个低成本运行，对数据无入侵，由中心化组织维护，但抗中心化篡改的技术方案。基于区块链和工作量证明，设计一个最小可支持单机系统运行的抗篡改技术 TDBBP（technology of tamper detection based on blockchain and pow），通过链式哈希块关联各数据日志块，通过工作量证明配合审计，使得稍长时间的数据日志块没有足够时间伪造重建，从而提供抗篡改能力。

关键词 抗篡改；区块链；工作量证明；哈希算法；日志链

doi：10.3969/j.issn.1672-9528.2024.10.039

0 引言

随着信息化的不断深入，信息化的重心从数据的采集、传递和保存逐渐转移到对数据的分析利用，例如数据驱动的工厂、媒体、科研和 AI 模型的训练等，都需要可信的数据。而临时补充数据的造数行为十分常见，这样的补造数据往往导致错误的分析结果和决策。对此，部分系统采用数字水印、数字签名、时间戳签名、区块链等技术完成抗篡改，但这些技术有其适用的专用场景，或存在中心化篡改风险，或成本高昂。

传统数字水印为了嵌入水印信息，需要采用各种方法或多或少地对原图像数据进行修改，从而限制了数字水印的鲁棒性和安全性。近年来，各国学者密切关注一种不对原图像数据进行修改的“零水印”的新数字水印技术[1]，并引发致力于零水印研究的热潮，提出了一些新算法。如使用 Hu[2] 根据代数不变理论提出的七个几何不变矩作为零水印，但该类算法存在只能抵抗简单的几何攻击（包括旋转、平移、缩放），不能抵抗常规信号处理攻击（如有损压缩、滤波、噪声干扰等），也不能应对复杂的几何攻击，如扭曲攻击、镜像翻转等问题。且几何不变量的构造大都基于全局图像信息，对图像的形状和纹理信息有一定要求，算法效率较低。针对空域零水印算法不足，很多学者[3-6]提出了变换域的零水印算法，获得了较好的性能改善。这种方法需要侵入数据中，不适用于常见的业务系统数据。

数字签名基本原理是发送者以秘密解密密钥运算报文传输结果于接收者。接收者既知发送者公开加密密钥获取原报文[7]。数字签名即公钥数字签名，以公钥密码系统为载体，可划分为三类：其一，以大整数分解问题为载体的公钥密码系统，例如 RSA 加密体制；其二，以有限域离散对数问题为载体的公钥密码系统，例如 DSA 加密体制；其三，以椭圆曲线离散对数问题为前提的公钥密码加密系统，常见的是椭圆曲线密码加密体制[8]。数字签名的安全性具体要求为可信、不可伪造、不可复制、消息不可变、无法消除行为。可信即基于验证签名，接收者可判断签名有效性；不可伪造即除了合法签名者没有人可以拥有能够成功通过接收者验证的消息签名对；不可复制即消息签名复制到其他消息是失效状态；消息不可变即签名消息被篡改之后，接收者一旦收到消息便可迅速发现；无法消除行为即签名者若是签名某消息，便不能否认或者抵赖签名行为。这种方法依赖于数字密钥的管理安全，面对众多的数据和系统，密钥和密码管理已经成为人们的重要负担，其风险越来越高，且存在中心化篡改的风险。

区块链的底层技术并不是一种全新的技术，而是根据比特币系统需要的特性将已有的技术结合起来，如密码学、分布式系统、P2P 网络、博弈论等。比特币的发明过程借鉴了 Adam 在 Hashcash[9] 中设计的工作量证明机制、Haber 和 Stornetta[10] 提出的用于保证数字文件安全的时间截方法、Dai 在 B-money[11] 中设计的奖惩机制。迄今为止，比特币系统已经上线成功运行 10 年时间，显示了高度的稳定性和可

1. 吉林建筑大学电气与计算机学院 吉林长春 130119

靠性。随着比特币大获成功，其底层区块链技术也越来越受到关注。

基于区块链的可追溯、抗篡改能力，李峰[12]针对假冒伪劣产品日益增多，给消费者和厂家带来了极大的信任危机的问题，提出基于区块链溯源的包装防伪信息追溯方法。该方法利用区块链技术的不可篡改性、分布式共识和透明性，实现了商品生命周期的全程可追溯和防伪保障；马海峰等人[13]提出区块链的数字文凭的流通解决方案，以联盟链为基础，使用分布式文件存储系统方案，构造数字文凭共享方案；卫泽晨等人[14]针对软件版本一致性问题，提出了一种基于改进默克尔树与区块链的电网调度自动控制软件版本一致性管控方法。

中心化的系统具有速度快、效率高的优点，同时也有不透明和舞弊的风险。而去中心化的系统有不易舞弊的优点，但同时其成本高、效率低。

总体来说，抗篡改技术的研究在国内外都取得了显著的进展，但更广泛的数据存在于区块链系统之外，面临着中心化组织的安全风险。

鉴于以上情况，需要一种低成本、非侵入、由中心组织维护但抗篡改的技术。

首先，区块链具有数据块关联修改的特点，为数据修改提高了难度，具有抗篡改的特点，这是本研究内容需要的。其次，对于大规模部署的点对点系统，区块链系统分布广泛，采用激励机制保障运行和安全，运行成本高昂，这点对本研究内容并不适用。最后，对于单节点的区块链，需要权益等支撑的算法难以具备条件，而 PoW 类算法具有难度高、运行条件少的优点，对于本研究内容更加适用，因此本设计基于 PoW 类算法提高数据篡改难度。

综上所述，本技术 TDBBP 进行以下设计。

（1）基于区块链设计链式存储结构，保障数据完整性的同时，提高数据篡改难度。

（2）基于 PoW 类算法进行区块挖掘，进一步提高数据篡改的难度。

（3）设计数据审计算法，对数据进行验证，发现与区块链不一致的数据。

本设计中，没有分布式区块链的大量验证和保存节点，设计为通过审计算法完成区块链的篡改审计，从而对抗中心化的篡改；对于数据的可靠存储，采用常规中心化系统的技术即可，如备份和灾备机制，在系统建设时根据可靠程度需求进行相应的设计即可，本设计不予涉及。

1 TDBBP 之链式存储结构的日志链设计

对管理数据的中心组织来说，可以对数据进行任意改动，只要去除修改痕迹，则无法发现改动。为实现对数据的修改记录，本设计对数据建立变更日志，按变更日志进行重放即可得到数据内容。如果数据日志实现了不可篡改，而数据本身完成了篡改，则对数据日志重放后，即可对比发现差异，从而控制篡改的发生。

本设计的日志链设计包含如下 3 个设计要点。

设计要点 1：支持自动重放的变更日志

变更日志的重放过程，对于不同的日志内容，可采用人工或自动重放，如仓库管理的出入库记录日志，结构化数据库管理的 sql 语句变更日志，均可以进行按日志顺序自动重放，得到数据结果后与当前数据进行自动校核，发现差异；而某些数据变更日志只能人工理解，无法进行自动重放和对比，建议在具体的系统设计中，尽量建立变更日志的规律，以方便自动重放和对比。这里举两个例子说明支持自动重放的变更日志。

支持自动重放的变更日志示例一：采用账本的记账方式，建立数据和变更日志。

支持自动重放的变更日志示例二：采用数据库表结构保存数据，采用数据库管理系统的 sql 增删改语句建立变更日志。

设计要点 2：采用 hash 值关联日志块，提高篡改难度

如图 1 所示，本日志链通过日志块的序号和日志时间识别顺序，通过 hash 进行前后块内容校核，如其中一块的内容发生变化，则其本块 hash 发生变化，其后的各块均需要更新前块 hash 和本块 hash，由此，早先的日志数据变更需要进行所有的后续数据块变更，故篡改影响和难度增大。

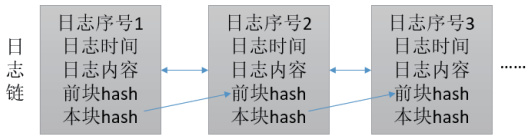

图 1 日志链

设计要点 3：日志数据块按业务速度进行生成，不使用 PoW，增强日志的吞吐能力

本日志链区别于下一章设计的证明链，不进行 PoW 限制，从而保证系统日志的写入速度与业务速度相同。

综上所述，变更日志由 hash 进行链接，日志内容尽量采用支持自动重放的设计，且本日志链并无 PoW，生成变更日志的速度可以保证。

2 TDBBP 基于 PoW 工作量证明类算法的证明链设计

本章的证明链设计如图 2，证明链与日志链一样由序号标识顺序，由摘要（hash 值）进行前后区块关联。

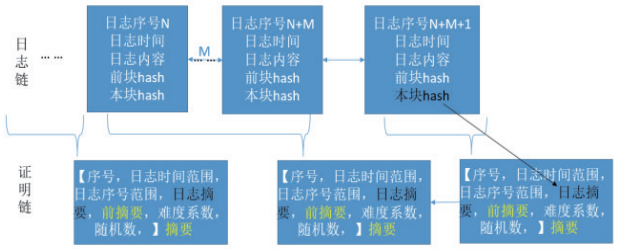

图 2 证明链

证明链生成规则如下。

（1）区块以单线程顺序生成。

（2）无待关联的日志块时，不予生成，保持监控日志链中出现待关联的日志块。

（3）出现待关联的日志块即开始生成证明链的下一区块。

（4）生成证明链的区块时，采用 PoW 进行难度限制，通过寻找随机值，使块摘要的 2 进制数前 N 位为 0，增加大量生成区块的时间成本。具体的 N 值，在具体项目中建议设定为 0.5 ～ 1 h 为佳。

证明链与日志链不同的地方包括以下方面。

（1）证明链中的区块，关联到日志链的多个连续区块，并保存最后一块的摘要，从而对关联区块的修改时必须对证明区块及后续区块进行修改。

（2）证明链中的区块，采用 PoW 进行难度限制。

综上所述，证明链通过 PoW 生成链式区块，并关联日志链，从而保证日志链的大量修改需要大量的时间，在短时间篡改的条件下，难以具备条件，最终提供了抗篡改的能力。假如相关人员进行大量篡改，同时由于无法生成大量证明区块，选择生成少量证明区块关联篡改的日志，则长时间的日志链关联了少量的证明区块，在审计中会被审计算法发现。

3 TDBBP 数据审计算法的设计

在比特币的区块链中，通过众多的验证节点，采用共识机制保证区块的合理性。在本研究的应用场景中，由数据的使用者运行区块审计程序，进行日志链和证明链区块数据的审计。

审计的对象包括实际数据、日志链、证明链，从中发现数据篡改问题或疑问。

（1）重放日志链，生成数据，比较生成数据和实际数据是否存在偏差，如存在偏差，则实际数据有篡改。

（2）对日志链从序号 1 开始计算本块摘要，比较摘要是否一致，比较上块摘要与本块中的上块摘要是否一致，不一致说明日志链有篡改。

（3）对证明链从序号 1 开始计算本块摘要，比较摘要是否一致，比较难度是否匹配，比较上块摘要与本块中的上块摘要是否一致，不一致说明证明链有篡改。比较本块中的日志链时间范围、序号范围、日志摘要是否正确，不正确说明日志链或证明链有篡改。

（4）统计证明链中的每个区块包含的日志链区块数，从而得到日志链区块生成速度偏快的证明区块，此类区块存在篡改生成的风险，需要从业务角度进行解释短时间内产生大量日志区块的原因。

（5）统计证明链中的每个区块包含的业务日志时间范围，正常的情况下，各区块的数据时间范围平均为块的生成周期 T，如果大于该值 100%，即 $2T$ 时，则可认定为由篡改所致。

4 TDBBP 抗篡改分析及对比分析

工作量证明在 1993 年就已被提出，后来它被应用于抗 DDOS 攻击和反垃圾邮件。像互联网巨头公司微软就将其应用在 Hotmail、Exchange、Outlook 等电邮服务上，要求所有收到的邮件都使用强 PoW 附件，以此预防大量垃圾邮件发出。其对低频业务影响甚小，而对高频业务则产生巨大的资源消耗，从而对抗高频次行为。

4.1 证明链单块生成分析

证明链区块中的难度系数 N，为区块 hash 值的前 N 位。区块的摘要满足难度条件的概率为 2 的 N 次幂分之一，即对于一个区块 S，其满足难度 N 的概率 P 为：

$$P(\text{区块}(N, S)) = 1/2N \quad (1)$$

即平均需要计算 $2N$ 次 SHA256 的 hash 值方可完成区块生成。

4.2 证明链篡改时高频次生成块行为的分析

在业务数据篡改的场景中，日常生成数据时，通过调整难度系数 N，使运行的区块生成时间设为 t=30 min，则每天区块数据约为 48 个区块，考虑每日的空闲时段为 50%，则每日为 24 个区块，则一个月内的区块数量为 30×24=720 块。日常运行生成区块时，只需生成一块即可，为低频业务，且由于证明链区块包括多块日志链区块对，可以保证面对高频业务数据时，证明链的区块仍旧为低频业务；而在篡改数据的情况下，由于区块的链接关系，早期区块的修改要完成后续所有区块的重新生成，如果需篡改影响 1 个月的区块，即 720 块，需要耗时 720×30=21 600 min，即 360 h。由于篡改一般是突发需求，实际需要在半天或 1 天内生成，这种半天或 1 天时间内串行生成 720 块即为高频业务，受工作量证明算法的限制，根本不具备所需的时间条件，因此无法完成篡改。难度 N 时的篡改耗时分析见表 1。

表 1 篡改耗时分析

难度 N	生成 1 块 hash 次数均值（正常运行时）	生成 720 块 hash 次数均值（异常篡改时）
20	220≈1.05×106	220×720≈7.55×108
30	230≈1.07×109	230×720≈7.73×1011
40	240≈1.1×1012	240×720≈7.92×1014

如果强行采用少量的区块进行保存，例如仅生成 24 块，则审计时，可以发现区块包含的业务数据时长平均约为 360/24=15 h，即区块对应了超长时间范围的日志数据，而在设计中，证明链中的区块对应的时间范围符合出块所需时长，正常围绕 t=30 min 波动，这种异常十分容易检出，从而发现篡改。

基于以上分析可知，本方法对于临时起意的，半个月以上的数据篡改的场景具有良好的检出能力，从而完成抗篡改目的。

4.3 各种抗篡改技术对比分析

本章对数字水印、数字签名、数字摘要、时间戳签名、区块链和本技术进行对比如表 2。从表 2 中可以看出，本技术拥有适用广泛、成本低、中心化风险中低的综合优势。

表 2 抗篡改技术对比分析

抗篡改技术	适用数据类型	系统规模	成本	中心化风险	其他风险
数字水印	多媒体数据	所有规模	低成本	高	水印破解
数字签名	所有数据	所有规模	中低成本	高	密钥安全
数字摘要	所有数据	所有规模	低成本	高	
时间戳签名	所有数据	所有规模	中等成本	中	
区块链	所有数据	超大规模系统	高成本	低	
本技术 TDBBP	所有数据	所有规模	低成本	中低	“矿场”类技术加速篡改

5 总结与展望

从各类抗篡改技术对比分析来看，本技术 TDBBP 拥有广泛的适用场景、较低的成本和中心化篡改风险，是一种有广泛应用潜力的抗篡改技术。同时，本技术存在以下改进的方向：本技术在 PoW 算法上，可以继续深入研究，以寻找证明工作量的同时更加有益的 PoW 算法，进一步减少计算资源的浪费；目前的设计尚无法严格规定证明链区块生成的耗用时间，而这个时间对于抗篡改能力评估和节约计算资源十分重要，可以进一步进行探索；在日志链的设计选择上，可以考虑采用其他方式进一步增强日志链的并行生成能力，从而支持海量业务的应用场景。

参考文献：

[1] 温泉 , 孙锬锋 , 王树勋 . 零水印的概念与应用 [J]. 电子学报 , 2003, 31(2): 214-216.

[2]HU M. Visual pattern recognition by moment invariants[J].IRE transactions on information theory, 1962, 8(2): 179-187.

[3] 罗丹妮 . 抗几何攻击的数字图像零水印算法研究 [D]. 西安 : 西安建筑科技大学 , 2013.

[4] 叶天语 , 马兆丰 , 钮心忻 , 等 . 强鲁棒零水印技术 [J]. 北京邮电大学学报 , 2010, 33(3): 126-129.

[5] 马建湖 , 何甲兴 . 基于小波变换的零水印算法 [J]. 中国图象图形学报 , 2008, 12(4):581-585.

[6] 杨树国 , 李春霞 , 孙枫 , 等 . 小波域内图象零水印技术的研究 [J]. 中国图象图形学报 : A 辑 , 2004, 8(6): 664-669.

[7] 陈亚茹 , 丛培强 , 陈庄 . 一种椭圆曲线数字签名的改进方案 [J]. 信息安全研究 ,2019, 5(3): 217-222.

[8] 汪潇潇 , 程鸿芳 . 浅析椭圆曲线数字签名的研究与发展 [J]. 科技风 , 2020(34): 90-91.

[9] 何蒲 , 于戈 , 张岩峰 , 等 . 区块链技术与应用前瞻综述 [J]. 计算机科学 ,2017, 44(4): 1-7.

[10]HABER S, STORNETTA W S. How to time-stamp a digital document[J].Journal of cryptology, 1991,3:99-111.

[11]SHAO Q, JIN C, ZHANG Z, et al. Blockchain:architecture and research progress[J].Chinese journal of computers, 2018, 41(5): 969-988.

[12] 李峰 . 基于区块链溯源的包装防伪信息追溯方法 [J]. 计算机测量与控制 , 2024, 32(6):220-226.

[13] 马海峰 , 高永福 , 薛庆水 , 等 . 基于区块链的数字文凭认证及共享方案 [J]. 计算机工程与设计 , 2024, 45(2): 376-382.

[14] 卫泽晨 , 李立新 , 刘金波 , 等 . 基于改进默克尔树与区块链的电网调度自动控制软件版本一致性管控方法 [J]. 电网技术 , 2024, 48(3):1273-1280.

【作者简介】

徐勇（1973—），男，吉林长春人，博士，副教授，研究方向：大数据、多元统计分析、人工智能。

（收稿日期：2024-07-12）