基于区块链技术的数据共享与隐私保护研究

王 磊[1]
WANG Lei

摘 要 区块链技术在数据共享和隐私保护领域得到广泛应用。基于区块链的数据共享模型的设计需求包括支持多角色参与、确保数据的安全性和隐私性、实现数据的高效共享和可追溯性等。为实现这些需求，构建了一个包含数据层、区块链层、应用层和隐私保护层的数据共享模型，并利用智能合约和隐私保护算法来保护数据隐私。此外，研究了一种基于区块链的隐私保护算法，综合了区块链技术、同态加密和智能合约，为数据的隐私保护提供了强大的技术支撑，不仅展示了区块链技术在数据共享和隐私保护方面的巨大潜力，也为相关领域的研究和应用提供了有价值的参考。

关键词 区块链技术；数据共享；隐私保护；智能合约

doi：10.3969/j.issn.1672-9528.2024.08.038

0 引言

随着信息技术的日新月异，数据共享已然成为推动社会进步和经济发展的核心驱动力。数据的流通与共享不仅促进了科研的突破、企业的创新，还极大地提升了政府治理的效能和公共服务的水平。然而，数据共享在带来诸多益处的同时，也面临着严重的隐私保护挑战。传统的数据共享模式往往依赖于中心化的数据存储和管理系统，存在数据泄露、篡改和滥用的高风险。由于数据集中存储在少数几个机构或服务器中，一旦这些机构或服务器受到攻击或内部出现漏洞，就可能导致大量敏感数据被非法获取和滥用。此外，中心化的管理系统还可能因为利益驱动或权力滥用而对数据进行不当处理，进一步加剧了隐私泄露的风险。

区块链技术以其去中心化、分布式设计，解决了数据共享中的隐私保护问题。数据不再由单一机构管理，而是由网络多个节点共同维护，增强了安全性和可靠性。区块链结合密码学算法和共识机制，确保数据完整性和不可篡改性。其透明可追溯特性使所有交易和数据交换都被公开记录，防止数据滥用或篡改。此外，区块链支持智能合约，自动执行数据共享协议，提高了效率和安全性。

综上所述，区块链技术以其去中心化、不可篡改和透明可追溯的鲜明特性，为解决数据共享过程中长期存在的隐私保护难题提供了坚实的技术支撑。鉴于其在确保数据安全与隐私保护方面的显著优势，论文基于这一背景，深入探讨了区块链技术在数据共享隐私保护领域的应用，旨在进一步挖掘其潜力，为数据的安全共享与利用提供更为有效的解决方案。

1. 闽南理工学院信息管理学院 福建石狮 362700

[基金项目] 2021 年福建省教育厅中青年教师教育科研项目（JAT210505）

1 区块链技术

区块链技术是一种前沿的分布式数据库技术，它通过独特的密码学算法确保了数据在存储和传输过程中的高度安全性和不可篡改性。这种技术不仅允许网络中的参与者直接进行安全、可靠的数据交换，而且无需依赖任何传统的第三方信任机构，极大地提升了数据交换的效率和安全性。区块链由一系列按时间顺序排列的数据块组成，每个数据块包含一定数量的交易记录和指向前一个数据块的指针。通过共识机制，网络中的参与者可以共同维护一个全局一致的区块链账本。区块链的核心技术包括共识机制、加密算法和智能合约等[1-2]。

1.1 共享机制

区块链技术的关键在于其独特的共识机制，它让网络中的参与者通过特定算法和规则共同维护区块链账本。目前，主流共识机制包括工作量证明[3]（proof of work，PoW）和权益证明[4]（proof of stake，PoS）。PoW 通过解决计算难题来确保网络安全性，但存在能源浪费问题；而 PoS 基于参与者持有的权益进行挖矿，更节能，但可能导致富者更富。委托权益证明（delegated proof of stake，DPoS）作为 PoS 的改进，允许持有者委托代表参与共识，提高了效率，但可能引发中心化风险。这些共识机制各有优劣，适用于不同场景。随着区块链技术的发展，未来还将有更多创新的共识机制涌现。

1.2 加密算法

除了共识机制，区块链技术还依赖多种加密算法来确保数据安全性。哈希算法是其中之一，它可以将任意长度的数据转化为独特的哈希值，从而验证数据的完整性和真实性。每当新数据块添加到区块链，都会通过哈希算法与前一个数据块相连，确保整个区块链的不可篡改性。此外，椭圆曲线加密算法（elliptic curve cryptography，ECC）[5] 在区块链中扮演着加密和解密数据的角色。这种算法利用椭圆曲线上的点进行加密操作，不仅提高了安全性，还降低了计算资源的消耗。在区块链中，ECC 保护用户的私钥和交易数据，确保数据在传输过程中的机密性。这两种算法的结合应用，为区块链技术提供了强大的数据安全保障。哈希算法保证了数据的完整性和真实性，而 ECC 确保了数据的机密性，使区块链成为一种可靠的数据存储和传输方式。

1.3 智能合约

智能合约是一种基于区块链技术、能够自动执行的程序代码。这些合约在编写时设定了特定的条件和规则，一旦满足这些条件，合约便能自动执行数据共享协议中的相关条款和条件。智能合约的引入极大地提升了数据共享的效率和准确性。传统的合同和协议需要人工介入来执行，这不仅耗时费力，而且容易受到人为操作错误和欺诈行为的影响。而智能合约的自动执行特性则完全避免了这些问题，降低了人为因素的干扰，提高了数据共享的效率和准确性。同时，智能合约的执行过程也是公开透明的，所有参与者都可以查看合约的执行情况和结果，进一步增强了数据共享的透明度和可信度 [6-7]。

2 基于区块链的数据共享模型设计

2.1 数据共享模型需求分析

在数据共享过程中，参与者通常包括数据提供者、数据请求者和数据管理者等角色。他们需要在一个安全、可靠和可信的环境中进行数据交换和共享。因此，数据共享模型需要满足以下需求：支持多角色参与、保证数据的安全性和隐私性、实现数据的高效共享和追溯等。图 1 为数据共享模型。

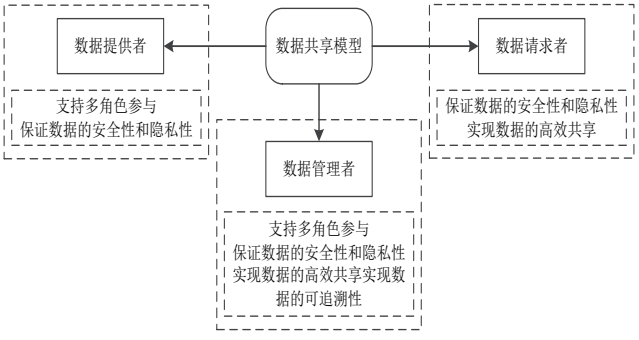

图 1 数据共享模型

（1）支持多角色参与

数据共享模型需要能够支持多个角色的参与，包括数据提供者、数据请求者和数据管理者。数据提供者通常拥有原始数据，并以某种形式进行共享；数据请求者希望获取并使用这些数据；数据管理者则负责管理和维护整个数据共享平台，确保数据的安全和有效共享。

（2）数据的安全性和隐私性

数据的安全性和隐私性是数据共享过程中的核心问题。数据提供者需要确保数据在共享过程中不会被未经授权的第三方访问或篡改，同时数据请求者也需要保证获取的数据是真实和可信的。数据管理者需要采取适当的安全措施，如加密、访问控制和审计等，来保护数据的安全和隐私。

（3）数据的高效共享

数据共享模型需要能够高效地处理数据请求和响应，确保数据在需要时能够迅速、准确地传递给数据请求者。这要求数据共享平台具备高性能、可扩展性和灵活性等特点，能够处理大量数据请求，完成复杂的数据处理任务。

（4）数据的可追溯性

在数据共享过程中，数据的可追溯性至关重要。一旦数据出现问题或争议，就需要能够追溯数据的来源和流向，以便查明原因并采取相应的措施。因此，数据共享模型需要设计合理的数据追溯机制，确保数据的来源和流向可追踪、可验证。

总之，一个成功的数据共享模型需要综合考虑多角色参与、数据的安全性和隐私性、数据的高效共享和数据的可追溯性等需求。通过精心设计和实施这样的模型，可以促进数据的有效共享和利用，推动各领域的创新和发展。

2.2 数据共享模型构建

本文设计了一个基于区块链的数据共享模型。该模型包括数据层、区块链层、应用层和隐私保护层四个部分。数据层负责存储和管理数据资源；区块链层提供去中心化、不可篡改和透明可追溯的账本服务；应用层实现数据共享的各种功能和操作；隐私保护层则采用加密算法和隐私保护算法来保护数据隐私。

在数据共享模型的实现过程中，采用智能合约技术来定义和执行数据共享协议中的条款和条件。通过智能合约的自动执行功能，可以实现数据共享的高效、准确和安全。同时，还采用零知识证明和同态加密等隐私保护算法来保护数据隐私，确保数据在共享过程中不被泄露或滥用。

明确数据共享过程中的各种需求和规则，如数据所有者、

数据访问者、访问权限、访问费用等。使用 Solidity 等智能合约编程语言编写合约代码。定义数据共享的结构体、状态变量、函数和事件。

（1）结构体设计

```
struct DataItem {
    address owner;
    string dataHash; // 数据的哈希值
    mapping(address => bool) accessPermissions; // 访问权限映射
    uint256 accessFee; // 访问费用
}
```

其中，DataItem 结构体在 Solidity 中定义了一个数据项的基本信息，具体包含以下信息。

owner：数据项的拥有者地址。

dataHash：数据的哈希值，用于验证数据的完整性。

accessPermissions：一个映射，记录哪些地址（账户）有权访问这个数据项。

accessFee：访问这个数据项需要支付的费用（如果有的话）。

（2）函数设计

```
function requestAccess(uint256 _dataId) external {
    DataItem storage data = dataItems[_dataId];
    require(data.accessPermissions[msg.sender]==false, "Already has access");
    require(payable(msg.sender).send(data.accessFee), "Payment failed");
    data.accessPermissions[msg.sender] = true;
    emit AccessGranted(msg.sender, _dataId);
}
```

函数 requestAccess 接收一个参数 _dataId，用于指定要访问的数据项。从 dataItems 映射中根据 _dataId 获取数据项，并存储在局部变量 data 中。检查请求者（msg.sender）是否已有访问权限，若有，则报错。若请求者无权限，则尝试向其支付访问费用 data.accessFee，支付失败则报错。支付成功后，赋予请求者访问权限。触发 AccessGranted 事件，通知外部监听者请求者已获得访问权限。

此函数允许用户请求访问数据项，前提是用户无权限且支付成功。成功后更新权限并通知外部。

（3）智能合约部署与测试

使用 Truffle、Remix 等工具部署智能合约到区块链网络。进行单元测试、集成测试和模拟测试，确保合约逻辑正确、功能完整。

（4）智能合约交互

数据所有者通过区块链钱包与智能合约交互，设置数据访问权限和费用。数据访问者发起访问请求，支付费用，智能合约自动验证并执行。

3 基于区块链的隐私保护算法研究

本算法通过结合区块链技术、同态加密和智能合约来实现数据的隐私保护。敏感数据使用同态加密算法进行加密，确保数据在区块链上的存储和传输过程中保持隐私性。智能合约则定义了数据访问权限、计算逻辑和隐私保护策略。算法流程如下。

（1）数据预处理

在数据处理中，需细致预处理原始数据，包括清洗、脱敏和标准化。清洗去除错误、重复及异常，确保数据准确。脱敏加密或隐藏敏感信息，保护隐私与机密。标准化统一数据格式，便于后续处理。识别敏感数据是关键，定义标准后自动或辅助识别并标记，支持安全存储与合规使用。

（2）数据加密

选择 Paillier 加密作为同态加密算法。对敏感数据 m 进行 Paillier 加密，生成密文 c。其数学公式为：

$$c = g^m \cdot r^n \bmod n^2 \quad (1)$$

式中：g 和 n 是公钥的一部分，r 是随机选择，m 是明文数据，c 是加密后的密文。将密文 c 与非敏感数据一同存储，或将密文的哈希值存储在区块链上。

（3）智能合约部署

编写智能合约，定义数据访问权限、计算逻辑和隐私保护策略。在智能合约中嵌入解密和计算逻辑（对于计算，可以使用密文下的同态计算）。

使用 Solidity 编译器（如 solc）将合约代码编译为字节码。使用区块链开发工具（如 Truffle、Hardhat、Remix 等）或区块链平台的命令行工具（如 Ethereum 的 ethereum-cli）将编译后的字节码部署到区块链上。在部署过程中，需要提供初始参数（如公钥）和任何必要的 gas 费用。在部署之前和之后，使用单元测试、集成测试和模拟环境来验证合约的功能和安全性。确保合约按预期工作，并且没有安全漏洞或错误。随着时间的推移，可能需要更新或修复智能合约中的错误或添加新功能。考虑使用版本控制来跟踪合约的更改，并确保在更新之前充分测试新版本的合约。

（4）数据上链

将加密后的敏感数据（或其经过高级加密技术处理后

的哈希值）以及非敏感数据精心策划后存储在区块链上，这一举措不仅保障了数据的机密性，还促进了数据的可追溯性和透明度。区块链的去中心化特性确保了数据不被单一实体控制，从而增强了数据的安全性。同时，通过将智能合约的地址与数据存储的具体位置（如区块链上的区块编号、交易哈希等）紧密关联，我们构建了一个自动化且高效的数据管理系统。智能合约能够根据预设的条件和规则，自动执行数据访问、验证或解密等操作，极大地简化了数据处理的流程，提高了系统的响应速度和安全性。这种结合区块链与智能合约的存储方案，为数据的隐私保护、安全存储和高效利用提供了强有力的技术支撑。

（5）数据访问与计算

授权用户通过智能合约请求访问敏感数据或执行计算任务。智能合约验证用户的身份和请求的有效性。如果请求被批准，智能合约就会在加密状态下执行计算。例如，对于加法运算，可以直接在密文上执行公式（2）～（4）。

$$c_1 = g^{m_1} \cdot r_1^{n} \bmod n^2 \tag{2}$$

$$c_2 = g^{m_2} \cdot r_2^{n} \bmod n^2 \tag{3}$$

$$c_{\text{sum}} = c_1 \cdot c_2 \bmod n^2 \tag{4}$$

式中：c_{sum} 是加密后的和。

（6）结果验证与输出

用户收到结果后，如果需要解密，则使用 Paillier 解密算法进行解密。解密公式为：

$$m = \lambda \cdot L(c^{\lambda} \bmod n^2) \bmod n \tag{5}$$

式中：m 是明文消息。c 是 m 经过 Paillier 加密后的密文。n 和 λ 是 Paillier 加密算法的公钥参数。在 Paillier 加密中，公钥通常由 (n, g) 组成，其中 n 是两个大素数的乘积，而 λ 是 n 的某个与加密和解密相关的参数（具体是 n 的 Carmichael 函数值）。$L(x)$ 是一个函数，用于 Paillier 解密算法。对于整数 x，其定义为：

$$L(x) = \frac{x-1}{n} \tag{6}$$

但 $L(x)$ 在解密时，其输入 x 是经过模 n^2 运算后的结果，即 $x = c^{\lambda} \bmod n^2$。

本算法综合了区块链技术、同态加密和智能合约，以实现数据的隐私保护。通过 Paillier 加密算法对敏感数据进行加密，保证数据在区块链上的存储和传输安全。智能合约定义了数据访问和计算规则，实现了在加密状态下进行安全计算的功能。算法包括数据预处理、加密、智能合约部署、数据上链、访问与计算以及结果验证与输出等步骤，确保数据在整个过程中的隐私性和安全性。利用区块链的不可篡改性和智能合约的自动化执行，本算法为隐私数据的保护提供了强有力的技术支撑。

4 总结

随着数字化时代的到来，数据共享和隐私保护成为社会发展的重要议题。区块链技术以其独特的分布式数据库特性、去中心化、不可篡改性和智能合约等功能，为数据共享和隐私保护提供了新的解决方案。论文在深入分析区块链技术的基础上，设计了一个基于区块链的数据共享模型，并通过智能合约和隐私保护算法来保护数据隐私。此外，还研究了一种基于区块链的隐私保护算法，该算法综合了区块链技术、同态加密和智能合约，为数据的隐私保护提供了强有力的技术支撑。本文的研究不仅有助于推动区块链技术在数据共享和隐私保护领域的应用和发展，也为相关领域的研究和应用提供了有价值的参考。未来，随着区块链技术的不断发展和完善，其在数据共享和隐私保护方面的应用将会更加广泛和深入。

参考文献：

[1] 高航，俞学劢 . 区块链技术在跨域数据共享与隐私保护中的应用研究 [J]. 中国管理信息化 ,2023,26(22):185-187.

[2] 刘朗，王卓 . 基于区块链技术的双链数据共享系统研究 [J]. 信息技术与信息化 ,2023(6):125-128.

[3] 黄润 . 基于工作量证明的区块链的赞助区块截留攻击研究 [D]. 广州：华南理工大学 ,2024.

[4] 王捷 . 区块链的权益证明共识算法改进研究 [D]. 南宁：广西民族大学 ,2022.

[5] 冯云霞，王西贤 . 基于椭圆曲线加密算法的工业物联网数据隐私保护方案 [J]. 智能计算机与应用 ,2022,12(12):110-113+121.

[6] 江鄞，李嘉兴，武继刚 . 基于区块链智能合约的异构服务器安全去重 [J/OL]. 郑州大学学报 (工学版),1-9[2024-07-10].https://doi.org/10.13705/j.issn.1671-6833.2024.02.010.

[7] 樊骁翔 . 区块链智能合约与财务共享服务的整合 [J]. 中国集体经济 ,2024(7):145-148.

【作者简介】

王磊（1987—），男，福建泉州人，本科，工程师，研究方向：区块链、数据挖掘和大数据分析、计算机。

（收稿日期：2024-05-15）