数据管理的一个重要需求就是对数据进行共享利用。在信息时代，各机构拥有的数据量呈井喷式增长，数据共享的需求也越来越强烈，如何在数据共享过程中保持数据完整性、防止数据泄露成为研究热点问题。一些工作利用区块链的安全特性，建立了基于区块链的数据共享系统，实现了区块链上的数据安全查询、多方数据可信交换以及可靠数据传输。目前，区块链上的数据安全共享主要包含以下三个方面：

1. 单方数据可信查询

|  |  |  |  |
| --- | --- | --- | --- |
| 区块链 | 关系型语义 | 数据完整性保护 | 数据隐私保护 |
| ChainSQL | 强 | 弱 | 弱 |
| BigchainDB | 弱 | 弱 | 弱 |
| EtherQL | 弱 | 弱 | 弱 |
| SEBDB | 强 | 强 | 弱 |
| BlockchainDB | 弱 | 强 | 弱 |
| vChain | 弱 | 强 | 强 |
| Searchain | 弱 | 弱 | 强 |
| 文献[Cai2017] | 弱 | 弱 | 强 |
| 文献[ChenL2019] | 弱 | 强 | 强 |

在数据共享过程中，存在某一节点查询区块链上其他节点数据并进行分析使用的情况。一些工作[ChainSQL2017, McConaghy2016, LiY2017, ZhuYC2019, Cai2020, Jiao2019, EI-HindiM\_S2019, EI-HindiM\_V2019]研究将数据库技术与区块链相结合, 实现了区块链数据查询。ChainSQL[ChainSQL2017]将数据存储在底层的一个关系型数据库中, 并将数据库操作日志存储在区块链上, 可以根据日志记录来查询任意时刻数据库表记录的数据. BigchainDB[McConaghy2016]将区块链去中心化和防篡改等特性应用到分布式数据库上, 提出了一种基于管程的一致性策略来保证数据安全, 并提供NoSQL语句实现数据查询.EtherQL[LiY2017]可以实现范围查询和Top-k查询, 其方法是实时监听链上数据并同步到MongoDB中, 利用MongoDB进行查询操作.

当用户进行数据查询时，考虑到不能保证区块链上每个节点都是可信的，需要采取方法确保查询到的数据没有被篡改，并能验证用户获得的数据是否正确。文献[Jiao2019]提出一种可查询且不可篡改的区块链数据库, 设计了一种基于哈希指针的树型索引结构, 利用该索引可以快速查询区块内数据. SEBDB[ZhuYC2019]将关系数据语义添加到区块数据中, 设计了类似于SQL的查询语句, 将每种交易类型表示为一张关系表, 重新定义并实现了区块链上关系表的建立、选择、连接等操作, 以此来支持验证查询得到数据是否被篡改. 文献[Cai2020]在SEBDB基础上, 增加了在区块链上建立视图操作, 提高了查询效率. BlockchainDB[EI-HindiM\_S2019, EI-HindiM\_V2019]将区块链作为存储层, 在存储层之上构建数据库层, 利用分片等数据管理技术扩展区块链, 并提供数据查询接口, 可以保证数据查询的一致性和可信性. 该系统节点间存在多个分片, 每一分片维护一个区块链用来存储部分数据, 并在数据库层记录数据存储在哪些分片. 当用户进行数据查询时, 需提交数据所在分片的ID, 之后BlockchainDB会在对应分片检索数据.BlockchainDB提出Online和Offline两种验证数据正确性的方式，对于Online验证方法，用户将其查询得到的数据进行广播, 之后所有存储该数据的节点都会进行验证. 当大多数节点验证数据与本地存储的相符时, 则认为查询结果是可信的；否则, 则认为数据有被恶意篡改的可能.对于Offline验证方法，系统预先设定一个延迟验证参数, 当用户验证请求达到该参数时, 系统开始验证.

文献[XuC2019]提出了一个区块链框架vChain, 可以高效验证查询结果的正确性. vChain提出一个轻量级查询格式, q=<[ts, te],[a, b],y>.其中[ts, te]表示一个时间段. [a, b]表示数字类型属性v的范围. y表示一个布尔类型表达式. 保存在区块链中的数据对象表示形式为<ti, vi, wi>，v中存储的是多个数字类型的属性，w中存储的是多个非数字类型的属性。vChain在区块头中增加一个额外的AttDigest字段来实现属性查询. AttDigest记录区块上所有数据对象的属性通过多集合加密累加器作用生成的结果，该累加器可以判断两个集合交集是否为空[XuC2018].利用该性质, 若查询条件可以表示为集合的形式, 那么就可以通过比较返回结果记录的AttDigest与查询条件集合的AttDigest来验证查询结果是否符合查询条件.vChain还利用前缀树将该方法推广到范围查询。

文献[Ramachandran2017]利用智能合约实现对数据溯源查询. 当创建一个数据文件时, 创建者可在区块链中部署一个智能合约. 利用该合约可以更改数据拥有者, 添加可访问用户. 由于所有的修改均需调用部署的智能合约, 因此通过查询调用过的智能合约, 便可实现对数据的溯源查询. LineageChain[Ruan2019]研究在区块链环境下查询数据的起源和更新变化. 该系统中每个数据具有初始版本号, 数据发生变化时其版本号也进行更新, 数据版本号与区块同步递增. LineageChain将这种递增关系构建出Merkle DAG来代替Merkle Tree, 并利用Merkle DAG的拓扑序实现溯源查询. 同时, 该系统利用智能合约完整、安全的记录数据发生的变化, 并提供接口来查询这些数据变化.

部分工作注重区块链数据查询的隐私性。Searchain[Jiang2020]是一种基于区块链的关键字搜索系统, 实现了对授权关键字进行私密搜索. 该系统通过OKSA构建检索协议, 可以授权和验证关键字, 保证了数据检索的有序性、可扩展性和隐私安全.文献[Cai2017]提出了一种基于区块链的针对可信私有关键字进行搜索的隐私保护机制, 该机制采用一种可搜索的加密方案, 实现了高效且安全的对加密数据进行搜索. 文献[ChenL2019]提出一种基于区块链的电子病例搜索加密系统. 该系统利用复杂的逻辑表达式来构建电子病历的索引并存储在区块链中, 用户可以利用该表达式来搜索索引. 在该系统中, 用户有权决定谁可以访问他们的电子病历数据, 保证了电子病历指标的可信性和可追溯性.另有一些工作利用安全多方计算实现数据私密查询共享，将在下一小节详细介绍。

1. 多方数据安全共享

区块链上的数据共享中，数据通常由多方提供，需要额外注意多方数据的隐私保护问题。

文献[ZhengB2018]提出了一种基于区块链的可扩展且保护隐私的数据共享方案，将同态加密系统Paillier应用在区块链上，可以有效保护敏感信息。在该方案中，可以对共享数据进行交易，并利用(p,k)阈值Paillier加密机制保护交易信息，实现共享数据的保密性。为了解决大数据过度中心化存储带来的信任问题，文献[YueL2017]提出一种基于区块链的去中心化大数据流通安全系统，为数据提供者和数据查询方提供了数据安全共享平台。该平台实现了区块链上数据溯源查询，并通过智能合约的自动执行保证数据共享的安全，保证了信息采集过程的透明度，有效保护用户隐私和知情权。文献[ShenM2019]提出一种基于区块链的保护物联网数据隐私的支持向量机训练模型。该模型利用Paillier加密系统设计了一些安全的构建块，如安全多项式乘法和安全比较等，构造了一个安全的支持向量机训练算法，该算法不需要可信的第三方。该模型将物联网数据加密存储在区块链上，为多个数据提供者提供了安全可靠的数据共享平台。文献[ZhangA2018]提出一种基于区块链的个人医疗信息共享方案. 该方案构建了私有链和联盟链两种类型的区块链. 私有链存储医疗数据, 联盟链存储数据的安全索引，通过公钥可搜索加密实现了安全共享。

区块链数据共享过程中同样需要抵抗恶意攻击，保证数据不被篡改。文献[Zikratov2017]首次讨论了使用区块链存储、共享文件时防止数据被篡改的方法，提出了针对交易信息、区块信息的完整性保护方案。文献[LiuCH2018]以智能移动终端为基础，结合以太坊区块链和深度强化学习提出了一个高效数据收集和安全数据共享的框架。该框架利用深度强化学习可以大规模的收集数据，并利用区块链技术抵抗Dos攻击、DDos攻击等，实现安全可靠的数据共享。文献[LiangX2017]提出一个基于区块链的以用户为中心的个人健康数据共享方案，为了保持数据完整性，能够从云数据库检索每条记录完整性和验证的证明，并在区块链上存储证明的索引。该方案还采用Fabric支持的通道概念来实现特定场景的隔离通信。

区块链上的数据共享存在这样一种情况，提供数据的各方希望利用共享的数据完成一些计算、分析，但是不希望其他用户知道自己提供的数据。针对这种“数据孤岛”现象，安全多方计算提供了一种解决方案，可以实现数据的隐秘共享。文献[YueX2016]分析了在区块链上利用安全多方计算实现隐私数据共享的优势，但是没有提出具体的方案。文献[Raman\_arXiv]提出一个基于区块链的可信多方计算与验证平台。该平台公开透明的记录已验证的计算，可以检测计算发生错误的源头，能够跟踪计算的所有权，在不信任的节点之间实现高效安全的数据、模型、结果共享。

iCube是全球首个安全多方计算区块链金融项目，内置自主开发的安全多方计算算法沙盒。iCube在区块链底层添加了支持联合计算并保护参与者隐私的协议，从而实现了参与节点在信息隐私保护的前提下实现数据联合共享计算的功能。点融区块链云服务平台也推出了商业化的安全多方计算服务，该服务以SaaS服务的形式提供，在联盟链上采用秘密共享和同态加密算法等技术，实现各个节点在保障各自数据隐私安全的前提下，利用各方数据进行加密条件下的安全共享计算。为解决东南亚金融机构间数据安全共享问题，Defi提出基于区块链的安全多方计算系统，采用差分隐私与数据混淆技术，支持数据不用不可见，保护用户隐私安全。

1. 数据隐蔽传输

区块链上各节点共享的数据在传输过程中要做到隐蔽传输，即将待共享数据隐藏在实际通信数据中[Cachin2004]。由于采用P2P网络结构，基于区块链的隐蔽通道有一个缺点：接收方通过固定标签识别包含隐蔽消息的事务，会降低系统的可用性和隐私性。文献[TianJ2019]提出一种区块链隐蔽通道构建方案DLchain。该方案用动态标签替代固定标签，设计了一种基于真实交易数据统计分布的动态标签生成算法，具有隐蔽性、抗追溯性和强鲁棒性等特点。

文献[SiCX2020]同样采用动态标签，提出一种区块链数据隐蔽传输机制。该机制利用哈希函数生成含有标签信息的动态标签地址，发送方分享带有隐私信息的特殊交易，接收方筛选出区块链上的特殊交易并提取出隐蔽信息。

文献[GaoF2020]提出了一种秘密数据传输机制，可以在开放式区块链网络中建立秘密通道。通过使用kleptography算法修改区块链中的签名算法和交易过滤机制，可以在签名数据中嵌入特殊信息，以便接收者可以通过检测签名数据来过滤特殊交易。这种机制可以有效地解决恶意用户，恶意接收者和网络窃听者的威胁，并可以有效地提高隐蔽性。

文献[Partala2018]提出了一个简化的理想区块链模型，并设计了一种协议使双方能按照该模型通过区块链进行隐蔽通信。该模型提出一种将私密消息安全的嵌入到区块链中的方法，该方法满足基础加密哈希函数的随机oracle模型。该模型还根据计算不可分性为此类协议的安全性和隐蔽性制定了严格的定义。

文献[Basuki2019]设计一种基于区块链的事务隐写和图像隐写的联合隐写技术，提出一种新的具有高信息容量和安全、保密介质的隐蔽通信方案。事务隐写提供安全、保密的介质，用于存储隐写图像的索引。图像隐写提供大容量的存储，并通过多播传输进行分区，以提高效率。该方案还保留了自然事物模式，从而不会产生异常交易，保证了系统安全。

[ChainSQL2017] Beijing PeerSafe Technology Co., Ltd. White paper for blockchain database application platform. 2017. http://blockchain.peersafe.com/PDF/ChainSQL-whitepaper.pdf

[McConaghy2016] McConaghy T, Marques R, Müller A, Jonghe DD, McConaghy TT, McMullen G, Henderson R, Bellemare S, Granzotto A. BigchainDB: A scable blockchain database. 2016. https://www.bigchaindb.com/whitepaper

[LiY2017] Li Y, Zheng K, Yan Y, Liu Q, Zhou XF. EtherQL: A query layer for blockchain system. In: Proc. of the Int’l Conf. on Database Systems for Advanced Applications. Suzhou: Springer-Verlag, 2017. 556−567. [doi: 10.1007/978-3-319-55699-4\_34]

[ZhuYC2019] Zhu YC, Zhang Z, Jin CQ, Zhou AY, Yan Y. SEBDB: Semantics empowered blockchain database. In: Proc. of the 35th ICDE. 2019. 1820−1831. [doi: 10.1109/ICDE.2019.00198]

[Cai2020] 蔡磊,朱燕超,郭庆兴,张召,金澈清.面向区块链的高效物化视图维护和可信查询.软件学报,2020,31(3):680−694. http://www.jos.org.cn/1000-9825/5914.htm

[Jiao2019] 焦通,申德荣,聂铁铮,寇月,李晓华,于戈.区块链数据库:一种可查询且防篡改的数据库.软件学报,2019,30(9):2671−2685. http://www.jos.org.cn/1000-9825/5776.htm

[EI-HindiM\_S2019] El-Hindi M, Heyden M, Binnig C, Ramamurthy R, Arasu A, Kossmann D. Blockchaindb—Towards a shared database on blockchains. In: Proc. of the 2019 Int’l Conf. on Management of Data (SIGMOD 2019). New York: ACM, 2019. 1905−1908.

[EI-HindiM\_V2019] El-Hindi M, Carsten B, Arasu A, Kossmann D, Ramamurthy R. Blockchaindb—A shared database on blockchains. VLDB Endowment, 2019,12(11):1597−1609.

[XuC2019] Xu C, Zhang C, Xu J. vChain: Enabling verifiable Boolean range queries over blockchain databases. In: Proc. of the 2019 Int’l Conf. on Management of Data (SIGMOD 2019). New York: ACM, 2019. 141−158.

[XuC2018] Xu C, Chen Q, Hu H, Xu J, and Hei X. 2018. Authenticating aggre-gate queries over set-valued data with confidentiality. IEEE Trans. Knowl. Data Eng., 30, 4, 630–644.

[Ramachandran2017] Ramachandran A, Kantarcioglu M. Using blockchain and smart contracts for secure data provenance management. CoRR, vol. abs/1709.10000, 2017.

[Ruan2019] Ruan P, Chen G, Dinh TTA, Lin Q, Ooi BC, Zhang M. Fine-Grained, secure and efficient data provenance on blockchain systems. Proc. of the VLDB Endowment, 2019,12(9):975−988.

[Jiang2020] Jiang P, Guo F, Liang K, et al. Searchain: Blockchain-based private keyword search in decentralized storage[J]. Future Generation Computer Systems, 2020, 107: 781-792.

[Cai2017] Cai C, Yuan X, Wang C. Towards trustworthy and private keyword search in encrypted decentralized storage[C]//2017 IEEE International Conference on Communications (ICC). IEEE, 2017: 1-7.

[ChenL2019] Chen L, Lee W K, Chang C C, et al. Blockchain based searchable encryption for electronic health record sharing[J]. Future Generation Computer Systems, 2019, 95: 420-429.

[ZhengB2018] Zheng B K, Zhu L H, Shen M, et al. Scalable and privacy-preserving data sharing based on blockchain[J]. Journal of Computer Science and Technology, 2018, 33(3): 557-567.

[ZhangA2018] Zhang A, Lin X. Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain[J]. Journal of medical systems, 2018, 42(8): 140.

[LiuCH2018] LIU C H, LIN Q X, WEN S L. Blockchain-enabled data collection and sharing for industrial IoT with deep reinforcement learning[J]. IEEE Transactions on Industrial Informatics, 2018, 15(6): 3516-3526.

[YueL2017] Yue L, Junqin H, Shengzhi Q, et al. Big data model of security sharing based on blockchain[C]. 2017 3rd International Conference on Big Data Computing and Communications (BIGCOM). IEEE, 2017: 117-121.

[ShenM2019] SHEN M, TANG X Y, ZHU L H, et al. Privacy-preserving support vector machine training over blockchain-based encrypted IoT data in smart cities[J]. IEEE Internet of Things Journal, 2019, 6(5):7702-7712.

[Raman\_arXiv]Raman R K, Vaculin R, Hind M, et al. Trusted multi-party computation and verifiable simulations: A scalable blockchain approach[J]. arXiv preprint arXiv:1809.08438, 2018.

[YueX2016] YUE X , WANG HJ , JIN DW. et al .Healthcare data gateways : found healthcare intelligence on blockchain with novel privacy risk control[J]. Journal of Medical Systems, 2016, 40(10): 218.

[Zikratov2017] Zikratov I, Kuzmin A, Akimenko V, et al. Ensuring data integrity using blockchain technology[C]//2017 20th Conference of Open Innovations Association (FRUCT). IEEE, 2017: 534-539.

[LiangX2017] Liang X, Zhao J, Shetty S, et al. Integrating blockchain for data sharing and collaboration in mobile healthcare applications[C]//2017 IEEE 28th annual international symposium on personal, indoor, and mobile radio communications (PIMRC). IEEE, 2017: 1-5.

[Cachin2004] Cachin C. An information-theoretic model for steganography[J]. Information and Computation, 2004, 192(1): 41-56.

[GaoF2020] Gao F, Zhu L, Gai K, et al. Achieving a Covert Channel over an Open Blockchain Network[J]. IEEE Network, 2020, 34(2): 6-13.

[TianJ2019] Tian J, Gou G, Liu C, et al. DLchain: A Covert Channel over Blockchain Based on Dynamic Labels[C]//International Conference on Information and Communications Security. Springer, Cham, 2019: 814-830.

[Partala2018] Partala J. Provably secure covert communication on blockchain[J]. Cryptography, 2018, 2(3): 18.

[SiCX2020] 司成祥,高峰,祝烈煌,巩国鹏,张璨,陈卓,李锐光.一种支持动态标签的区块链数据隐蔽传输机制[J].西安电子科技大学学报,2020,47(05):94-102.

[Basuki2019] Basuki A I, Rosiyadi D. Joint Transaction-Image Steganography for High Capacity Covert Communication[C]//2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA). IEEE, 2019: 41-46.