计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202440456
61(10): 2570−2586, 2024

# 基于去中心化身份的跨域数据交易系统

郝嘉琨[1] 向 鹏[1] 何逸飞[1] 高健博[2] 关 志[3] 谢安明[1] 陈 钟[1,4]

[1](北京大学计算机学院 北京 100871)

[2](智能交通数据安全与隐私保护技术北京市重点实验室(北京交通大学) 北京 100044)

[3](北京大学软件工程国家工程研究中心 北京 100871)

[4](北京大学重庆大数据研究院 重庆 401329)

(jk.hao@pku.edu.cn)

# Cross-Domain Data Trading System Based on Decentralized Identity

Hao Jiakun[1], Xiang Peng[1], He Yifei[1], Gao Jianbo[2], Guan Zhi[3], Xie Anming[1], and Chen Zhong[1,4]

[1] (*School of Computer Science*, *Peking University*, *Beijing* 100871)

[2] (*Beijing Key Laboratory of Security and Privacy in Intelligent Transportation* (*Beijing Jiaotong University*), *Beijing* 100044)

[3] (*National Engineering Research Center for Software Engineering*, *Peking University*, *Beijing* 100871)

[4] (*Peking University Chongqing Big Data Research Institute*, *Chongqing* 401329)

**Abstract** In the digital era, data, as a critical factor of production, is essential for economic growth, technological innovation, and enhancing corporate competitiveness. Traditional centralized data trading systems pose a series of data and transaction security risks. Although decentralized data trading systems based on blockchain technology have emerged, which can guarantee security for data transactions through blockchain and smart contracts, systems based on blockchain still struggle with cross-regional, cross-domain, and international data transactions. They primarily face the following three challenges: 1) Consistency of cross-domain identity resolution results. 2) Availability of cross-domain identity resolution services. 3) Security and compliance issues for cross-domain data trading systems. To address these issues, we design a secure and efficient cross-domain identity resolution mechanism based on decentralized identity technology and construct a secure, regulated decentralized data trading system based on this mechanism. The identity resolution mechanism includes a set of cross-domain protocols that ensure the consistency of resolution results and an incentive mechanism that ensures the availability of services executing these protocols. The decentralized data trading system supports granular regulation based on identity attributes and ensures the atomicity of cross-domain data transactions. Experimental verification shows that this system meets the design targets. By batch constructing cross-domain resolution requests, the identity cross-domain resolution protocol and caching mechanism can ensure the correctness of the resolution results, with the resolution latency being only 45% of the cross-domain resolution latency. Additionally, the cache can increase the hit rate by 72% through active and passive update mechanisms. The system is also compatible with multiple blockchain systems, and cross-domain data transactions can be completed within an average of 2.3 blocks, making the system efficiency within an acceptable range.

**Key words** data element; decentralized identity (DID); blockchain; smart contract; incentive mechanism

**收稿日期**: 2024-06-04; **修回日期**: 2024-07-18

**基金项目**: 国家重点研发计划项目(2023YFB2704800); 国家自然科学基金项目(62202011, 62172010)

This work was supported by the National Key Research and Development Program of China (2023YFB2704800) and the National Natural Science Foundation of China (62202011, 62172010).

**通信作者**: 谢安明(xieanming@pku.edu.cn)

**摘　要**　在数字化时代，数据作为重要的生产要素，对经济增长、科技创新和企业竞争力提升至关重要. 传统的中心化的数据交易系统存在一系列数据安全、交易安全风险. 虽然当前已经出现了基于区块链实现的去中心化数据交易系统，可以为数据交易提供安全保障，但基于区块链的数据交易系统仍难以处理跨地区、跨地域、跨国家的数据交易，其主要面临 3 个挑战：1）跨域身份解析结果的一致性问题；2）跨域身份解析服务的可用性问题；3）面向跨域交易的数据交易系统的安全与合规问题. 为解决上述 3 个问题，基于去中心化身份技术设计了一套安全高效的身份跨域解析机制，并构建了一种安全、可监管的去中心化数据交易系统. 身份跨域解析机制包括确保解析结果一致性的身份跨域协议和保障该协议执行服务可用性的激励机制. 该去中心化数据交易系统支持基于身份属性的细粒度监管，并保证跨域数据交易的原子性. 实验验证表明，该系统满足设计目标，解析延迟仅为跨域解析延迟的 45%，缓存命中率通过主动和被动更新机制提高 72%. 系统兼容多种区块链，跨域数据交易平均在 2.3 个区块内完成，效率在可接受范围内.

**关键词**　数据要素；去中心化身份；区块链；智能合约；激励机制

**中图法分类号**　TP309

在当今数字化时代，数据已成为最宝贵的生产要素之一. 它在推动经济增长、促进科技创新和提升企业竞争力方面发挥着至关重要的作用. 作为一种新型的生产要素，数据的交易[1]、使用[2]和定价方式[3]与传统的生产要素有所不同. 因此，需要构建新型的数据要素市场，让数据能够更顺畅、及时地流向需求方，充分发挥其价值[4].

尽管当前传统的中心化数据交易系统在交易效率上具有显著优势，例如便于用户快速寻找和获取所需数据、高效地进行数据交易、便于维护和管理等. 这是由于中心化的数据交易系统依赖于云存储和中心化的业务系统实现，当前的分布式系统技术可以实现高频、低延迟的交易，并提供可扩展性强的数据管理服务. 但在中心化模式下，如果数据交易系统的安全措施和监管措施不到位，极易导致数据泄露和数据滥用，侵犯用户数据权益[5]. 为了解决这些问题，新兴的去中心化数据交易系统使用区块链技术，提供防篡改和可追溯的安全保障，并通过智能合约实现交易逻辑，确保数据资产和交易的安全与隐私[6]. 去中心化的模式在交易效率和数据管理效率不如中心化的模式，但却可以在透明性上和安全性上提供更强的保证.

然而，即便是使用区块链技术实现的去中心化数据交易系统，在处理跨域数据交易时仍面临一系列问题. 当前的数据交易系统为了保证交易遵循系统制定的规则，或确保数据交易的安全可控，通常要求交易双方均在同一系统中注册账号，并且购买的数据资产也必须在当前系统上登记. 这导致在处理跨地域、跨行业和跨国数据交易时，交易流程复杂且受限. 这不仅不利于构建全国统一的数据要素市场，也限制了数据的全球化分享和利用. 在处理跨域数据交易场景中，当前数据交易系统架构面临 3 个挑战.

1）跨域身份解析结果的一致性问题. 数据交易双方的互信是交易顺利开展的基础. 在数据交易中，需要对交易双方身份建立信任，对交易的数据资产达成共识. 在跨域身份解析机制中，两域中的节点仅能通过受限的方式访问对方域中的区块链节点，获取区块链上信息[7]. 这使得攻击者更容易通过利用两域节点的信息差，进行钓鱼攻击、中间人攻击. 此外，缓存机制也会导致跨域身份解析结果不一致，将身份标识符解析到过时的身份信息或已经泄露私钥的身份上. 跨域身份解析结果的不一致会导致数据交易系统上交易的资金和数据资产被窃取，或导致买方购买到质量差或者和事先约定不匹配的数据资产[8].

2）跨域身份解析服务的可用性问题. 作为一个去中心化的数据交易系统，系统的参与者以一种去中心化的方式组织在一起. 跨域身份解析服务作为此系统的一部分，参与者自发地参与和提供服务[9]. 保证服务的可用性需要关注 2 个方面，如何保证参与者的解析服务的准确性和实时性，以及如何保证服务有足够的参与者支撑系统中弹性变化的跨域解析需求[10].

3）面向跨域交易的数据交易系统的安全与合规问题. 首先，由于难以直接进行跨域通信，导致两域难以对同一跨域数据交易进行协商. 如果使用中心化的跨域通信服务，当其产生单点故障时，会导致跨域数据交易无法继续执行，甚至会导致交易异常终止，产生资金和数据资产损失. 在更严重的情况下，

买方可以通过欺骗卖方已经支付，由于卖方无法直接查询买方是否已经支付，导致数据被骗取，卖方可以通过同样的方式骗取资金. 其次，由于缺乏身份跨域机制，智能合约无法通过对交易双方和数据资产的属性进行监管. 数据交易系统的身份机制仅基于身份标识符和公开的身份属性，无法满足不同地域、国家的法律法规的多样化的合规性要求[11]. 在一次跨域数据交易中，监管者应当能直接穿透域内外的壁垒，通过买卖双方的属性数据进行监管，判断双方是否存在违规的行为，判断此次交易是否合规.

本文的主要贡献包括4个方面：

1)为去中心化身份技术设计了一套安全、高效的去中心化身份(decentralized identity, DID)标识符跨域解析协议，使去中心化数据交易系统之间以去中心化的形式构建一个统一的身份层，保证用户身份和数据资产身份的跨域一致性.

2)为跨域身份解析协议设计了一套激励机制. 保证身份跨域解析协议的稳定运行，保证跨域数据交易的身份解析服务的可用性.

3)基于上述身份解析协议和激励机制，构建了一种安全、可监管的去中心化的数据交易系统. 此系统能保证跨域数据交易的原子性，并为跨域数据交易提供在身份属性细粒度的监管支持.

4)实现并验证了本系统的可行性和系统效率，为跨域数据流转提供了新的技术路径.

## 1 相关工作

在数字经济高速发展的当下，数据安全成为了一个不可忽视的重要议题. 随着《中华人民共和国数据安全法》的推行，规范数据处理活动、提高数据安全能力成为了行业的必然要求[12]. 针对数据安全治理，技术应用在数据的全生命周期中发挥着关键作用，尤其是在数据的价值实现环节. 例如，可搜索加密技术不仅保障了密文检索的安全性，还通过避免高成本运算，大大降低了实际应用中的技术门槛[13-14]. 同时，隐私保护数据审计方案的提出，基于零知识证明构造，不仅简化了证书管理，还增强了第三方审计过程中的数据隐私保护[15]. 为了解决隐私和审计带来的效率下降问题，使用区块链存储优化技术和隐私计算等技术共同保证了安全的数据共享方案的效率[16]. 这些技术的应用和发展，为数据安全治理提供了强有力的支持，是推动数据产业健康持续发展的关键因素.

但当前的研究大多集中在数据使用环节的隐私和监管，而构建一个数据交易系统涉及对多领域技术的综合运用，需要确保系统在身份认证、数据存储和数据共享等关键方面的安全性和高效性. 在用户和数据资产的身份认证部分，可以采用基于区块链的方法来确保身份的真实性和交易的不可篡改性[17]. 在数据交易通信的物理层消息认证技术也被广泛应用于保证信息传输的安全性[18]，保证数据交易请求和数据资产内容传输的安全[19]. 在数据资产的元信息访问和数据资产内容的流转方面，基于区块链的去中心化框架[20]，用于安全的数据交换和任务分配，文献[21]则详细探讨了代理重加密技术，这是一种在不解密的情况下实现数据所有者之间安全数据共享的有效方法，可以用于进行安全的数据资产验证. 对于托管在数据交易系统中的数据资产，一种基于属性的加密技术，允许数据的加密和解密过程基于用户的属性进行，且包含外包解密的可验证性，可以防止托管的数据资产内容被泄露[22]. 为了保护基于云存储构建的数据交易系统，提出的一系列针对云存储环境设计的加密解决方案，以保障托管在云平台上的数据安全[23].

为解决去中心化数据交易系统主要面临的3个挑战. 当前的数据交易系统通过多种机制在某些层次上解决了上述问题，但大多局限于域内的数据交易，难以扩展到跨域数据交易的场景[24-26].

身份系统是一个用于管理和验证个体或实体身份的框架，为用户提供身份注册、认证、管理、授权等服务. 交易双方相互之间通过身份系统完成互认，买方对数据的元信息建立起信任，这是数据交易的基础. 使用非对称密钥对作为用户的身份标识是一种在数据交易系统中常用的身份系统，一些基于区块链的方案，直接使用区块链上的账户地址作为用户的标识符，以及使用数据的哈希值作为数据的唯一标识符[27-28]. 这些方案难以清晰地描述身份信息和难以支撑具体的数据交易系统. 当前大多使用中心化的身份系统，如CA[29]，或使用DID[30-31]，为数据交易系统上的用户和数据赋予唯一标识符，并添加身份属性描述. 虽然DID在安全性上具有优势，但在没有统一的身份解析机制的情况下，这些系统难以保证身份跨域的一致性. DID的身份信息存储在区块链上，由于域外节点难以直接访问域内的区块链节点，信息的更新会造成身份解析到过期身份上. 此外，用户也难以验证和发现恶意节点返回的错误身份解析结果，这进一步加剧了缺乏一致性保证造成的危害.

当前使用DID的数据交易系统方案，在跨域场

景中，大多依赖于中心化的身份跨域服务来实现身份跨域. 如文献 [31]，虽然使用了 DID 的方案，在域内能够通过区块链保证身份解析服务的可用性. 但在跨域场景中，仍需要依赖于第三方服务，这会造成可用性问题，例如单点故障. 文献 [29] 在多个数据交易系统之上构建了一条联盟链，并在联盟链上构建身份层实现身份的可信跨域. 但其在数据系统内部使用了中心化的身份机制，身份机制是身份系统中用于管理、认证和验证身份的一系列方法、技术和流程，因此中心化的身份机制可能导致身份系统和数据交易系统出现单点故障等问题. 此外联盟链的参与者较为固定，如果参与者出现故障，也会造成跨域身份解析服务不可用的问题. 文献 [30] 中通过信任系统，为参与者赋予信誉值，通过信誉值可以提高市场的透明度和参与者的信任. 这种方案也能促使系统中参与跨域身份解析节点的积极参与解析，提高跨域身份解析服务的可用性.

交易的原子性是指买方支付资金后，应当能获取到应得的数据所有权，并获取数据内容. 原子性的数据购买部分可以通过智能合约保证，支付资金后通过修改区块链上状态获得相应的数据权益. 在跨域数据交易中，交易的原子性出现了新的挑战. 在跨域场景中，交易应当在 2 个域中同时批准交易、完成资金支付才算完成数据购买，数据完成跨域传输后才算完成数据交付. 数据购买部分的原子性可以通过当前一系列区块链跨链方案解决，包含公证人、哈希时间锁、中继链[32] 等一系列方案，可以实现跨域支付和所有权转移的原子性. 如果系统缺乏数据可用性保障机制[27,29,33]，那么在跨系统的数据交付中难以保证原子性. 例如在文献 [28, 30–31] 中，使用了去中心化存储、数据缓存等技术，保证了跨域数据传输时，不会因为域内的恶意节点而导致数据无法被正确、实时地交付.

根据数据交易系统所处的行业、地域、国家不同，所应遵守的监管规则不同，因此数据交易系统具备适应多种场景下的合规性要求的能力. 跨域数据交易中，监管要求比域内数据交易更为严格，但跨域的数据交易应当在 2 个域中同时满足合规性. 当前的方案大多可以通过在数据交易的智能合约中内嵌监管规则，基于买卖双方的身份信息、数据的元数据进行细粒度的监管. 当前数据交易系统因缺乏统一的身份层，身份信息无法可信地跨域传递，导致交易系统无法对跨域数据交易进行细粒度的监管. 文献 [32] 未使用区块链和智能合约，而文献 [27, 31, 34] 等，仅使用公钥作为用户身份标识，这些方案在跨域场景中难以实现对用户身份属性的细粒度监管；如文献 [30–31] 使用 DID 技术实现了身份标识符、身份属性的跨域互认，保证了跨域数据交易在身份属性粒度的可监管性.

## 2 系统概述

### 2.1 系统模型

本文设计参考了现有的数据交易模式的构造. 在本文设计的去中心化数据交易系统中，存在 4 种角色，分别是用户、区块链节点、数据托管节点、边缘节点. 如图 1 所示，在跨域数据交易场景中，存在 2 个去中心化数据交易系统，分别处于不同的域中，以 $D_p$ 和 $D_q$ 表示.

1）用户. 用户是数据资产的买卖双方. 卖方将数据资产上传到数据交易系统中，并在系统中定价；买方在系统中挑选数据，选中后支付资金，在交易在系统通过后下载数据. 本文使用 $U_s$ 表示卖方，用 $U_b$ 表示买方.

2）区块链节点. 在去中心化数据交易系统中，数据资产的交易通过区块链节点验证执行，交易逻辑包含身份验证、合规性验证、资金验证等，通过区块链上的智能合约实现. 本文使用 *BN* 表示区块链节点.

3）数据托管节点. 由于区块链本身的透明性、冗余存储等特性，数据内容不适合直接存储到区块链上. 在去中心化数据交易系统中，数据资产的内容被托管到区块链下的节点中，然后使用数据的哈希值锚定在区块链上保证数据存储和传输的完整性. 卖方会将数据资产托管到数据托管节点中，买方在购买到数据后，会从数据托管节点处下载数据. 本文使用 *DN* 表示数据托管节点.

4）边缘节点. 为保证去中心化数据交易系统的安全和数据安全，系统内部的用户、区块链节点、数据托管节点应当限制对系统外访问的权限. 为支持跨域的数据交易，本系统设置边缘节点用于转发对域外的请求和数据传输，同时接受外部请求和数据传输. 边缘节点可以同时作为区块链节点或仅作为一个可访问区块链节点、转发请求、接收响应的服务器. 本文使用 *EN* 表示边缘节点.

如图 2 所示，上述系统角色可以将去中心化数据交易系统划分为 3 层，分别是身份层、交易层、数据层.

身份层管理用户和数据资产的身份标识及属性. 可信的身份认证机制是数据交易和数据传输的安全

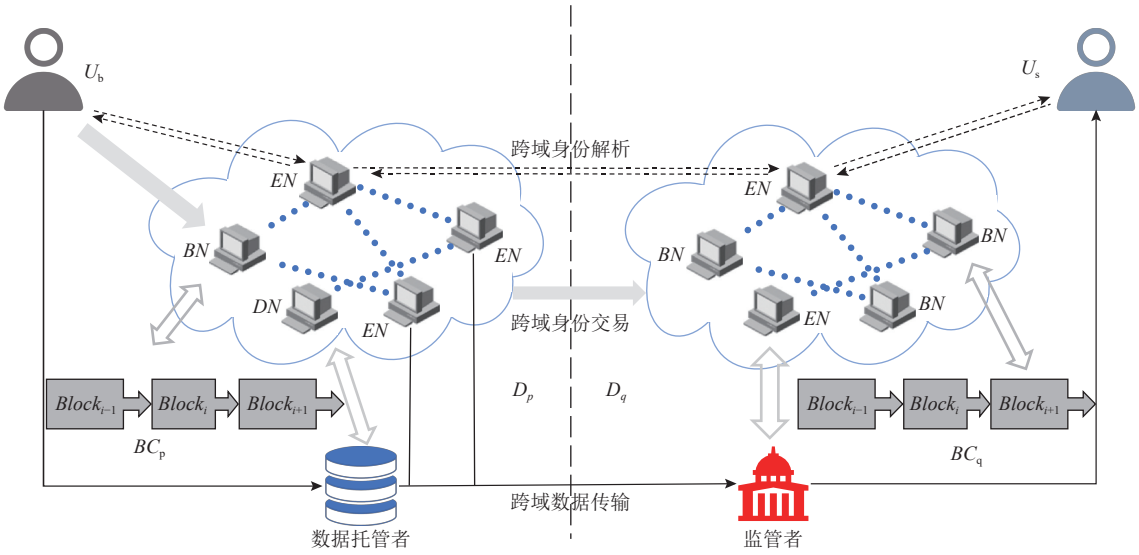



Fig. 1 System model

图 1 系统模型

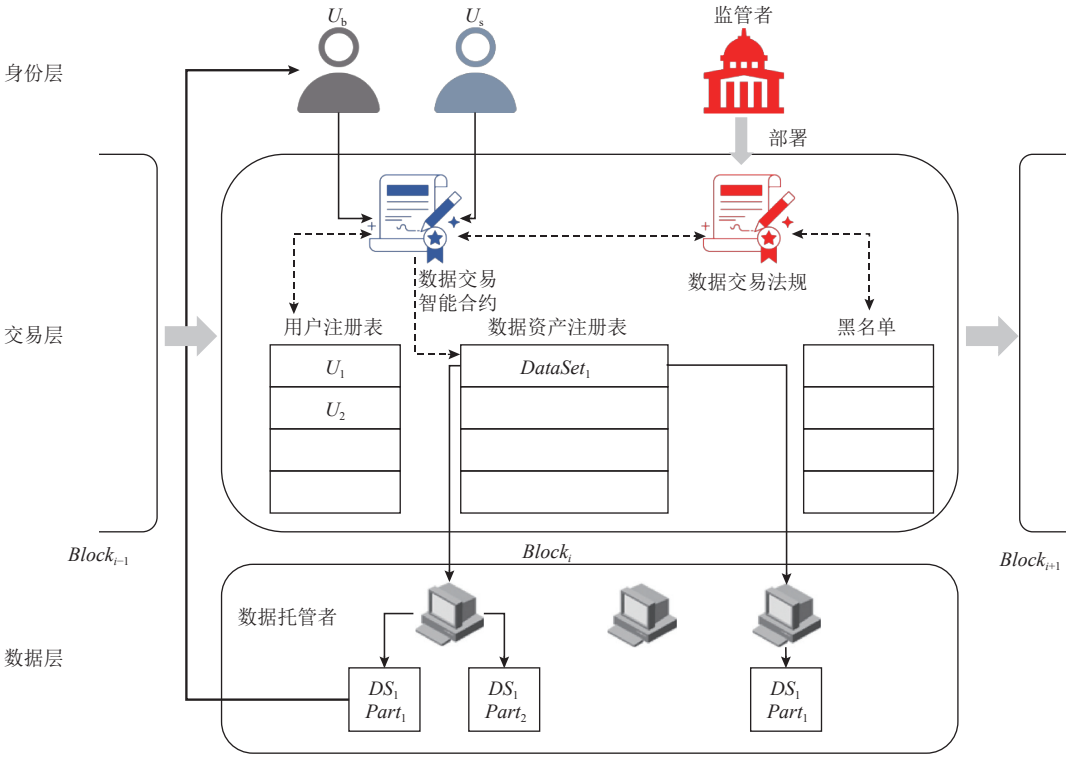



Fig. 2 System architecture

图 2 系统架构

保证，也是数据交易监管的基础. 交易层负责执行数据交易，包括身份验证、合规性验证、资金验证等，根据执行结果决定是否批准交易. 数据层负责存储数据和传输数据. 在跨域数据交易的场景下，这 3 层都将通过跨域机制和域外系统进行交互，实现跨域的身份认证、交易执行、数据传输.

## 2.2 设计目标

1)跨域身份解析结果一致性. 在跨域解析场景中一致性是指，身份的解析结果在所有的域中保证一致. 在时刻 $t$，域 $D_p$ 中的用户 $U_b$ 和域 $D_q$ 中的用户 $U_s$，同时发起对域 $D_p$ 中的某 DID 标识符的解析，则二者的解析结果应当相同.

2)跨域身份解析服务可用性. 在跨域解析场景中，可用性是指域内外的用户发起的身份解析操作，应当在一定时间范围内获得正确的解析结果.

3)跨域数据交易原子性. 跨域数据交易原子性

有2个方面的要求，在数据购买阶段，数据交易中，当且仅当资金支付到位后数据才应当被转移所有权和进行数据传输. 在数据交付阶段，当资金支付完毕后，必须获得约定好的数据资产的所有权，并获得完整的数据资产内容.

4)细粒度的跨域数据交易监管. 在跨域数据交易中，两域应当支持对交易双方身份属性、数据资产元信息的监管.

## 3 DID跨域解析协议

为了保证跨域身份解析结果的一致性，本节设计了一个跨域身份解析协议. 身份解析协议是用于解析和验证身份信息的通信协议. 它定义了一组标准化的规则和方法，使得不同系统之间能够互操作，并对身份信息进行验证和解析. DID框架可以保证域内解析的一致性，本协议将基于DID框架，设计跨域身份解析协议，实现跨域身份解析的一致性. 为了解决缓存机制带来的跨域身份解析不一致的问题，本文设计可以保证一致性的缓存机制，在保证一致性的前提下，提升跨域解析的效率.

### 3.1 DID框架

本系统的身份层基于DID框架实现. DID是W3C提出的一套去中心化的身份标准，DID可以基于区块链构建DID解析网络，基于身份解析结果构建身份凭证交换通道，实现身份的互认[35]. DID可以保证域内身份解析服务的可用性和一致性，作为去中心化数据交易系统的身份层，为用户和数据资产提供身份服务.

DID框架依托于一系列技术和标准，包括DID标识符、DID文档、可验证凭证(verifiable credential, VC)[36]以及可验证数据注册表(verifiable data registry)等关键组成部分. 如图3所示，通过这些组件，DID框架为用户提供了一个高效而灵活的身份标识符解析和身份属性验证机制.

DID的身份标识符解析机制涉及DID标识符、DID文档这2种数据结构和可验证注册表模块. 用户通过和基于区块链实现的可验证注册表进行交互，根据区块链上信息，从DID标识符查询对应的DID文档. 由于在不同的场景下，所需要的身份结构不同，因此存在多种不同类型的DID方法. W3C DID工作组维护了一个DID方法的注册表[37]，此注册表用于确保不同DID方法的互操作性和一致性. 该注册表包含所有已知的DID方法的全球参数、属性和值，定义了DID解析器如何应用CRUD操作(创建、读取、更新和撤销)来管理DID文档. 这有助于开发者在全球范围内实现统一的DID使用标准.

DID的身份属性验证机制涉及DID文档和VC数据结构，主要角色包括颁发者(Issuer)、持有者(Holder)、验证者(Verifier). Issuer是具权威身份的实体，负责基于对Holder的了解，签发并用其公钥签名验证凭证. Holder作为凭证的所有者，负责安全存储凭证并在需要时出示. Verifier接收Holder提供的凭证，并通过DID文档中的信息(如Issuer的公钥)验证凭证的签名和内容，以确定凭证的真实性和Holder的身份属性.

验证过程首先通过DID标识符解析出DID文档，获取公钥和加密方案等信息确保通信安全. Issuer生成并签署可验证凭证，Holder接收并存储凭证. 当Holder需要证明身份时，向Verifier展示凭证. Verifier

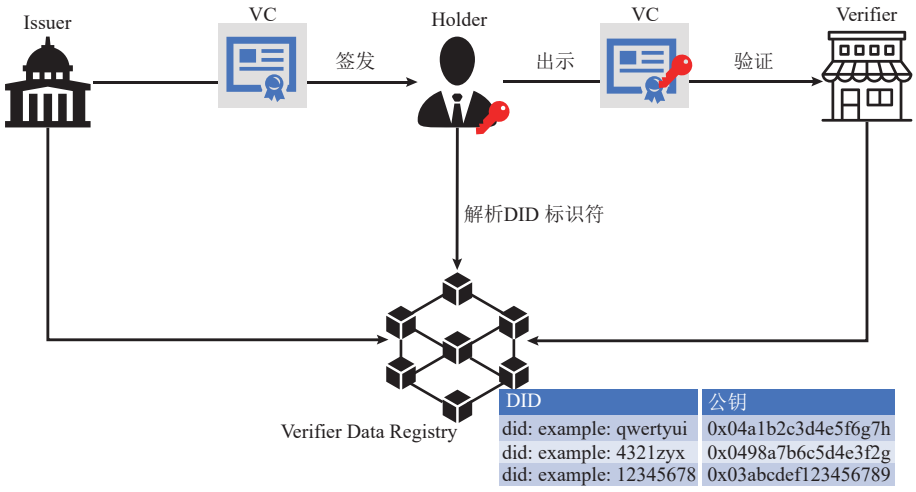



Fig. 3 Decentralized identity framework

图3 去中心化身份框架

通过 DID 文档获取 Issuer 的公钥，对凭证签名和内容进行验证，检查凭证的完整性、有效性及 Issuer 的权威性，从而确认 Holder 的身份属性.

DID 的身份解析、身份认证、身份注册等身份操作都依赖于 DID 的用户对域内的区块链的直接访问. 在 DID 规范中，缺乏对域外身份解析机制的设计. 当前存在部分依赖于中心化的 DID 解析服务的身份跨域方案，但都无法保证身份解析结果的一致性和身份解析服务的可用性. Universal DID 解析器[38]和注册器[39]是用于管理和解析 DID 的工具. Universal DID 解析器负责将 DID 解析为对应的 DID 文档，提供统一的接口以便与不同 DID 方法进行互操作. Universal DID 注册器则负责 DID 的创建、更新和停用，确保不同 DID 方法的一致性和安全性. 上述方案依赖于第三方的解析服务实现跨域的身份解析，即便可以通过将解析服务分散到多个节点上构成一个解析联盟的方式来避免恶意的跨域身份解析者造成的后果，仍无法解决服务不可用、错误的解析结果的问题，也无法对恶意的解析者进行追责.

### 3.2 身份跨域解析协议

由于域外节点无法访问域内的区块链节点，这导致 DID 的解析结果验证方案无法通过默克尔证明验证身份的解析结果的正确性和实时性. 因此，有必要设计一种可以确保身份跨域解析的一致性身份跨域解析协议.

当前存在多种 DID 的跨域解析方案，但都依赖于中心化的跨域解析器，破坏了 DID 的去中心化特性. 为设计身份跨域机制，本文将在 DID 框架和本文提出的数据交易系统模型的基础上定义 2 种新的操作，实现一致性的身份的跨域解析.

为保证域内用户可以验证域外 DID 标识符的解析结果，需要将域外的区块链最新根哈希写入到区块链上. 此操作应由可以访问域外信息的边缘节点执行，定义为 $BC_p.Set(BC_q,rootHash)$. 边缘节点需要发起或响应域外请求，例如更新根哈希或解析 DID 标识符等，本文将边缘节点发起的域外请求定义为 $EN_p.request(BC_q,req) \rightarrow resp$.

使用上述定义的方法，可以构建一套保证解析结果一致性的身份跨域解析协议. 如图 4 所示，此机制可以分为 3 个阶段，解析开始前的数据准备阶段、解析过程中跨域请求及响应阶段、解析完成后的解析结果验证阶段.

各阶段具体过程为：

1)数据准备阶段

① 域 $D_p$ 内的区块链应当及时处理用户的 DID 标识符和 DID 文档的注册和修改请求，保证域内身份信息的正确性和实时性. 域 $D_p$ 内的用户 $U_b$ 可以根据自己身份信息的更新情况，调用 $BC_p.Set(DID, DIDDoc)$ 更新区块链上的 DID 文档信息.

② 域内的边缘节点集合 $\{EN_1, EN_2, \cdots, EN_n\}$ 应定期对本节点所负责的外域区块链的根哈希进行同步，将对应的根哈希同步到本域的区块链上. 假设 $EN_i$ 有权限访问域 $D_q$，则 $EN_i$ 应当定期发起请求 $EN_i.request(BC_q,QueryLatestRootHash) \rightarrow \{BlockNumber,rootHash\}$，并调用 $BC_p.Set(BC_q,rootHash)$ 更新区块链上域 $D_q$ 的根哈希.

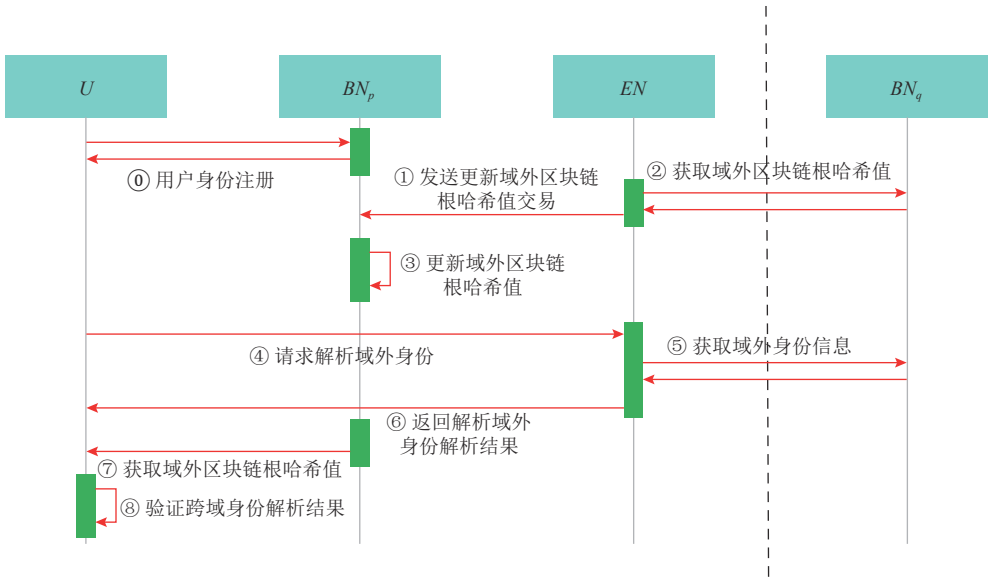



Fig. 4 Cross-domain identity resolution protocol

图 4 跨域身份解析协议

2）跨域请求及响应阶段

① 用户 $U_b$ 找到负责对应外域的边缘节点 $EN_i$，发起对域外的身份解析请求和跨域解析请求 $U_b.request(EN_i,\{BC_q, DID_s\})$，并等待响应.

② 边缘节点转发用户的身份解析请求 $EN_i.request(BC_q,\{DID_s\}) \rightarrow \{DID_s, DIDDoc_s, MerkleProof_s\}$，并进行跨域身份解析.

③ 边缘节点响应用户的身份解析请求，解析结果为 $\{DID_s, DIDDoc_s, MerkleProof_s, Sig_{EN_i}\}$. 响应的 DID 文档和默克尔证明字段应当使用边缘节点 $EN_i$ 的 DID 标识符对应公钥进行签名保证解析结果的不可抵赖性.

3）解析结果验证阶段

① 用户发起请求 $BC_p.Get(BC_q, rootHash, blockNumber) \rightarrow rootHash$，从域内区块链上获取解析结果和 $BC_q$ 的根哈希值.

② 用户首先验证解析结果中的 DID 文档是否已经过期、是否和 DID 标识符对应，以及是否有边缘节点的签名. 验证通过后，使用获取到的根哈希值验证解析结果的默克尔证明，以保证解析结果的完整性.

### 3.3 身份跨域解析的缓存机制

为加速身份跨域解析，在边缘节点上设置缓存是一个必然的举措[40]. 但在跨域身份解析的场景中，节点可能返回过时的解析结果，或者节点缓存被污染后返回错误的解析结果，破坏解析协议的一致性[41].

为确保解析结果的一致性，节点缓存记录的结构应包括 6 个关键字段：DID 标识符、DID 文档、Merkle Proof、区块号、生存时间（time to live, TTL）、和最近使用时间（least recently used, LRU）. 除了 DID 标识符和其对应的 DID 文档，还需要存储 Merkle Proof 和对应的区块号，使用户能够验证解析结果的完整性.

解析结果的 TTL 是由 DID 文档设定的生存时间、2 个域之间规定的缓存记录 TTL，以及全局设定的 TTL 中的最小值决定，这是为了尽可能保证数据的实时更新. 每条记录还会记录其最近一次使用的时间，这有助于评估数据的活跃度. 结合 TTL 和 LRU 两个字段，系统可以决定哪些数据在过时后应被淘汰、哪些需要更新，在缓存容量有限时，确定数据的置换优先级高低.

基于上述记录结构，本文设计了一种身份解析缓存更新算法. 该算法结合区块链上存储的最新根哈希和本地缓存数据，主动更新本地缓存.

**算法 1.** 身份解析缓存更新操作.

输入：缓存表 *IDCache*，最新被解析的身份集合 *RDIDs*，最新跨域解析的身份记录集合 *Rs*，区块链上存储的块号 *Bn*，丢弃延时 *t*，*IDCache* 容量阈值 *V*；

输出：待更新的身份集 *UDIDs*，新丢弃延时 *t*.

① set *UDIDs*=∅; /* 如果缓存达到容量阈值，缩小记录缓存时间，减少缓存条数 */
② if *IDCache.size*>*V*
③ *t*=*t*−1;
④ end if
/* 更新记录最近使用时间 */
⑤ for each record *r* in *Rs*
⑥ if *r.DID* in *IDCache* and *Bn*≥*r.Bn* > *IDCache[r.DID].Bn*
⑦ replace *IDCache[r.DID].Record* by *r*;
⑧ *IDCache[r.DID].LRU*=*r.Bn*;
⑨ end if
⑩ end for
/* 移除过期记录 */
⑪ for each record *r* in *IDCache*
⑫ if *r.Bn*+*r.ttl*<*Bn*
⑬ delete *IDCache[r.DID]*;
⑭ end if
⑮ if *r.LRU*+*t*>*Bn*
⑯ put *r.DID* into *UDIDs*;
⑰ end if
⑱ end for
⑲ return *UDIDs*, *t*.

算法 1 会在每一个区块上链后被调用，并根据返回的 *UDIDs* 主动更新缓存的解析结果. 在主动更新之外，缓存也会被动地被更新. 一次跨域解析请求，会同时涉及域内和域外的 2 个边缘节点. 如在一次跨域解析请求中，域 $D_p$ 的 $EN_p$ 向 $D_q$ 的 $EN_q$ 请求身份解析 $DID_t$ 后，当 $DID_t$ 对应的 DID 文档信息发生变化后，$EN_p$ 会主动地向 $EN_q$ 发起更新请求，更新 $DID_t$ 的缓存信息. 通过主动和被动这 2 种更新策略减少了解析中需要跨域请求的概率，提升了解析效率，同时确保了缓存数据的时效性和准确性.

## 4 身份跨域解析可用性保障机制

身份跨域解析和缓存机制可以保证跨域身份解析的一致性，可以保证用户获得的解析结果的完整性和实时性. 此机制建立在正确且实时更新的域外

区块链根哈希的基础上,边缘节点不更新或更新了错误的根哈希,即便存在一致性的跨域身份解析协议,身份跨域解析服务也难以正常提供服务.此外由于具备跨域访问能力的边缘节点可以自由地加入和退出参与跨域身份解析,因此无法保证参与节点的数量是否能达到跨域解析服务稳定运行的最低节点数量,也难以保证参与节点的可用性.

因此需要一个完善的可用性保障机制将边缘节点组织起来,对恶意、不活跃的边缘节点进行发现和惩罚,对正常提供服务的边缘节点进行激励,保障身份跨域解析服务稳定运行.本节提出的跨域身份解析的可用性保障机制分为2个部分:1)边缘节点解析范围的划分机制和激励机制,此机制为参与解析节点分配解析范围,并根据边缘节点的工作量,奖励参与跨域身份解析的边缘节点,激励边缘节点积极参与到自己权限内的外域身份解析中;2)恶意节点的发现和惩罚机制,此机制用于识别异常的解析行为和及时发现恶意节点,以便采取措施惩罚恶意节点、预防恶意行为.

在跨域解析系统中,每个外域的边缘节点都有一个唯一的DID标识符,这些标识符组成集合$EN=\{EN_1, EN_2, \cdots, EN_n\}$.同时,待解析的外域$q$的DID标识符也构成集合$Q=\{Q_1, Q_2, \cdots, Q_m\}$.通过散列这些DID标识符,可以将它们均匀分布在一个大小为$2^{256}$的整数空间中.如算法2所示,解析范围划分机制采用一致性哈希算法,将$EN_i$散列后的值$Hash(EN_i)$映射到一个环状的整数空间上.一致性哈希具备优越的负载均衡能力和扩展性,便于节点的动态加入和移除,在这些操作发生时,只需进行最小的范围调整,从而减少维护成本.在这个哈希环上,如果$Hash(EN_i)$与$Hash(EN_j)$相邻,那么$\forall Q_k \in [Hash(EN_i), Hash(EN_j))$,都应由边缘节点$EN_i$来解析.

为了确保解析过程的可用性和安全性,系统在算法2中设置冗余度参数$R$,为每个解析区间指派了多个边缘节点.用户可以通过比较这些节点返回的结果来验证解析的正确性.此外,为了防止节点在负责特定区间解析任务的长期过程中出现安全性下降或受到攻击的风险,系统会定期更新节点负责的解析范围.这一更新过程在算法2中表现为使用当前区块号*BlockNumber*和$EN_i$作为哈希函数的输入,以定期重新分配边缘节点在哈希环上的位置.

在此基础上,区块链根据节点所负责的解析范围的大小来分配奖励.负责更大解析范围的节点因处理更多解析请求而获得的奖励比例也更高,这进一步激励了节点的参与和性能优化.这种机制不仅提高了系统的效率,也增强了网络的整体安全性和可靠性.

**算法2.** 解析范围划分和激励机制.

输入:参与者节点列表*DIDs*,冗余度$R$,当前块号*Bn*,激励总额$C$;

输出:缓存范围边界划分*Bs*,每个范围负责的节点列表*RDIDs*,每个节点获得的激励*Cs*.

① set $Bs=\varnothing$, $RDIDs=\varnothing$, $Cs=\varnothing$;
② for each *DID* in *DIDs*
③ $b$ =hash(*DID* | *Bn*);
④ put $b$ into *Bs*;
⑤ end for
⑥ sort *Bs*;
⑦ for $i$ from 1 to $R$
⑧ for each *DID* in *DIDs*
⑨ $b$ =hash(*DID* | $i$ | *Bn*); /*找到*DID*负责的区间,并根据区间大小为此*DID*提供激励*/
⑩ find interval $j$ in *Bs* include $b$;
⑪ $Cs[DID]+=C/R/(Bs[j+1]-Bs[j])/2^{256}$;
⑫ put *DID* into $RDIDs[i][j]$;
⑬ end for
⑭ end for
⑮ return *Bs*, *RDIDs*, *Cs*.

跨域身份解析协议中的恶意节点存在2种可行的攻击模式:一是通过上传错误的域外区块链根哈希值,使用户认为错误的解析结果在区块链上存在;二是在当域外身份信息已经更新时,仍返回旧的身份信息生成的解析结果.

对于第一种攻击,本系统通过在根哈希值的更新中引入投票机制来防止此类攻击.参与的各个边缘节点都有权限更新外域区块链的根哈希值,并获取收益,如果有恶意节点更新错误的根哈希值则会出现多个候选的根哈希值,此时参与节点开始进行投票选择,找到正确的根哈希值,并惩罚恶意节点.当参与的边缘节点中恶意节点小于1/2时,本机制能保证外域区块链根哈希值的正确性和实时性.

当用户发起解析请求时,为降低请求到恶意节点的概率,本机制将采取交叉验证的方式提升安全性.如算法3所示,用户首先将随机向负责待解析的DID标识符所在范围的边缘节点$EN_{in}$发送解析请求,然后用户将会随机选择一个不在此解析范围的边缘节点$EN_{out}$(避免恶意节点通过注册多个DID标识符,提升其被分配到某个解析范围的概率)发送解析请

求，$EN_{out}$ 会向外域请求解析，获得解析结果. 用户可以通过比对 $EN_{in}$ 的解析结果和 $EN_{out}$ 获取到的解析结果，验证解析结果是否正确. 如果 $EN_{in}$ 是恶意节点则将带有 $EN_{in}$ 签名的解析结果发送到区块链上，证明其是恶意节点，区块链会奖励用户，并对恶意节点 $EN_{in}$ 进行惩罚. 同时，参与解析的边缘节点也可以伪装成用户，通过上述方式对其他参与解析的边缘节点的解析结果进行验证，频繁地递归查询和验证，能够快速地发现、惩罚、驱除恶意节点.

**算法 3.** 安全解析机制.

输入：待解析的 DID 标识符 *DID*，参与解析外域 *r* 的节点列表 *L*；

输出：解析结果 *DIDDoc*，欺诈证明 *FraudProof*.

① $EN_{in}$←random chose a node in *L*;

② $EN_{out}$← random chose a node in *EN*, but out of *L*;

③ {*DID*, *DIDDoc*, *Sig*, $EN_{in}$} ←resolve *DID* by $EN_{in}$;

④ {*DID*, *DIDDoc*, *Sig*, $EN_{out}$} ←resolve *DID* by $EN_{out}$; /*验证 2 次解析结果是否一致*/

⑤ if *DID* and *DIDDoc* from $EN_{in}$ and $EN_{out}$ not equal and *DID* not match with *DIDDoc* from $EN_{in}$ or $EN_{out}$ /*如果 $EN_{in}$ 和 $EN_{out}$ 返回的 DID 和 DID 文档不相等，或返回的 DID 与 DID 并不匹配*/

⑥ *FraudProof*←{{*DID*, *DIDDoc*, *Sig*, $EN_{in}$}, {*DID*, *DIDDoc*, *Sig*, $EN_{out}$}};

⑦ return *nil*, *FraudProof*;

⑧ else

⑨ return {*DID*, *DIDDoc*}, *nil*;

⑩ end if

## 5 支持跨域的数据交易系统设计

支持跨域的数据交易系统需要特别关注交易的原子性和细粒度的监管[42]. 通过实现身份跨域解析机制，可以有效地支持这 2 个关键特性. 首先，交易原子性可以通过该解析机制确保跨域交易合约状态的准确传递，从而保证交易的服务可用性和状态一致性. 此外，细粒度监管也得以实现，因为一旦身份跨域解析完成，就可以利用 DID 文档和可验证凭证传递的身份属性来执行基于身份属性粒度的监管[43]. 基于身份跨域解析机制实现的数据交易系统，可以有效地保证跨域数据交易的原子性，也使得系统能够适应多样化的监管要求.

### 5.1 数据交易层的模块分解

本节首先定义数据层支持跨域数据交易需要设计的模块. 数据交易层由区块链实现，因此其中的模块应通过智能合约实现. 图 5 中定义的模块都通过智能合约实现，部署到区块链上运行.

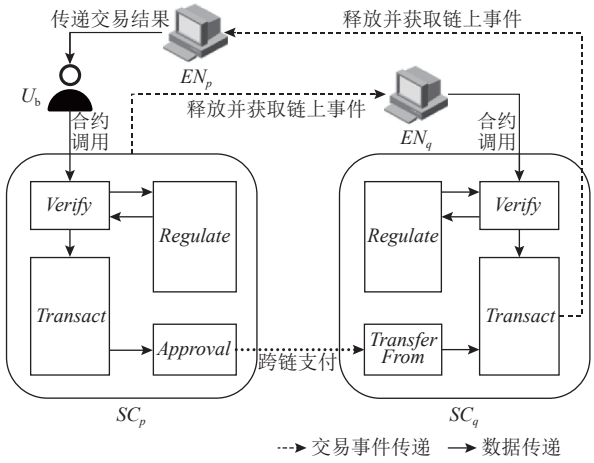



Fig. 5 Cross-domain data trading smart contract structure

图 5 跨域数据交易智能合约结构

1）跨链信息校验模块. 跨链信息校验模块用于校验外域传递来的信息，例如跨域解析的身份、外域区块链上的某些状态等. 此模块使用第 4 节中设计的方案保证传递数据的完整性和实时性. 本模块功能可定义为：

$$SC_p.Verify(BC_q, rootHash, Key, Value) \rightarrow Bool.$$

2）监管模块. 监管模块应根据数据交易系统所在地的法律法规编写，对交易双方身份和被交易的数据资产进行监管. 监管模块可以通过对数据资产的 DID 标识符解析到的 DID 文档中的信息对出入域的数据资产进行细粒度的监管；也可以指定身份属性，要求交易双方用户提交对应可验证凭证，进行基于属性的监管. 本模块可定义为：

$$SC_p.Regulate(DID, []Attr, statement) \rightarrow Bool.$$

3）交易验证模块. 交易验证模块用于执行交易逻辑，包括校验买方支付的交易金额是否满足卖方定义的价格、交易的附加条款、交易双方的某些特殊约定等. 本模块没有返回值，但会释放一组合约事件，将交易执行的结果固定在区块链上. 本模块可定义为：

$$SC_p.Transact([]param, statement) \Rightarrow Event(BuySuccess, []data).$$

4）跨链支付模块. 当前存在多种跨链支付的方案，本文不深入跨链支付的技术细节. 跨链支付功能可以抽象为，从一条区块链上的账户将一笔资金转移到另一条区块链上的账户中. 因此本模块参考

ERC20 标准[44]，定义为：

$$SC_p.Approve(DID_b,DID_s,num)\Rightarrow Event(TransferSuccess),$$

$$SC_q.TransferFrom(DID_b,DID_s,num).$$

### 5.2 跨域数据交易机制

跨域数据资产交易流程如图 5 所示，可以分成 2 个阶段，各自运行在买方域 $D_p$ 和卖方域 $D_q$ 上的区块链智能合约上.

1)针对于买方合约

①当用户 $U_b$ 通过身份跨域机制，充分了解域外的数据资产信息，并决定购买此数据后，用户 $U_b$ 调用跨域信息校验模块，校验待购买资产和卖方身份信息.

$$SC_p.Verify(BC_q,rootHash,DID_s,DIDDoc_s)\&\ SC_p.Verify(BC_q,rootHash,DID_b,DIDDoc_b).$$

②当身份信息校验通过后，开始调用监管模块，对本次交易进行监管. 属性信息可以通过解析 DID 文档获取，也可以通过主动提供可验证凭证获取.

$$SC_p.Regulate(DID_s,\{DIDDoc_s,VC_1,\cdots,VC_n\},statement)\&\ SC_p.Regulate(DID_b,\{DIDDoc_b,VC_1,\cdots,VC_n\},statement)\&SC_p.Regulate(DID_b,\{DIDDoc_b\},statement).$$

③当通过监管后，调用交易验证模块，传入交易必要的参数、待购买数据的价格、自己支付的金额等.

$$SC_p.Transact(\{DID_b,DIDDoc_b,token,number,\cdots\},statement).$$

④在交易验证模块执通过后，意味着买方系统批准交易，调用 $SC_p.Approve(DID_b, DID_s, num)$，进行跨链的转账.

⑤本域的边缘节点 $EN_b$ 收集交易验证模块释放事件，其中包含批准交易的证明 *BuySuccess* 和跨链支付模块释放的事件、支付成功证明 *TransferSuccess*. $EN_b$ 的构建并发送交易到区块链 $BC_q$ 上、$\{BuySuccess, TransferSuccess, DID_b, DID_d, Sig_b\}$调用卖方合约 $SC_q$.

2)针对于卖方合约

①跨域信息校验模块在校验域外身份信息之外，还应当校验交易是否批准、跨链支付是否成功.

$$SC_q.Verify(BC_p,rootHash,tx,BuySuccess)\&\ SC_q.Verify(BC_p,rootHash,tx,TransferSuccess).$$

②监管模块和买方合约基本一致，但 *statement* 应当根据当前系统遵守法律法规进行调整.

③交易验证模块应当验证数字资产价格和跨链支付的金额是否匹配，并根据系统的交易规定进行其他的验证.

④在交易验证模块执行通过后，意味着卖方系统批准交易，调用 $SC_q.TransferFrom(DID_b, DID_s, num)$ 进行跨链的转账.

⑤域 $BC_p$ 的边缘节点 $EN_p$ 发起域外请求 $EN_p.request(BC_q,\{GetSellState\})$，获取域 $BC_q$ 交易验证模块释放的事件；响应结果包含批准交易的证明、数据块根哈希值、数据资产所在托管节点的节点列表等$\{SellSuccess, rootHash_d,\{DN_d\}\}$和跨链支付模块释放的事件. $EN_p$ 将跨域请求结果发送给用户，完成跨链交易，之后用户可以凭借证明发起域外数据传输请求.

在本机制中，跨域数据交易的可监管性可通过智能合约的监管模块和身份层获得支持. 身份层中可以解析 DID 标识符得到数据资产和交易双方的公开属性，此外，交易双方还可以通过可验证凭证向区块链系统提供自己更详细的属性. 监管模块可以实现基于属性的智能合约调用机制，避免跨域数据交易脱离监管执行.

由于智能合约内部执行的安全性通过区块链网络保证，数据安全和资金安全的关键问题在于买卖双方合约的调用是否安全. 买方合约的输入是对数字资产和卖方的身份的解析结果，卖方合约的输入是买方合约的批准交易的证明和跨链支付证明支付. 因此资金安全和数据安全建立在可信的跨链信息传递机制的基础上. 上述需要可信跨域的信息可以看作是 $SC_p$ 和 $SC_q$ 合约的身份，第 3 节和第 4 节设计的身份跨域机制可以保证跨域数据交易状态信息跨域的可信性和安全性，保证跨域数据交易机制的数据安全和资金安全.

### 5.3 跨域数据传输机制

在数据交易被双方数据交易系统批准，且资金已经跨域支付后，就进入了数据的跨域传输阶段. 在这个阶段中涉及到数据托管者、买卖双方系统的边缘节点、买方 3 个角色. 在此阶段应当保证跨域传输数据的完整性和传输效率.

如图 6 所示，跨域数据传输过程为：1)数据托管节点到边缘节点的传输；2)两域边缘节点之间的传输；3)边缘节点到买方的传输. 此流程共包含 2 次域内的数据传输和 1 次域外的数据传输. 当数据量很大时，会带来很大的开销，因此本文将使用去中心化存储技术[45-46]，使用并行传输的方式提高数据传输效率.

首先，卖方将数据托管到数据托管节点后，数据被拆分成大小一致的数据块存储. 之后，数据托管节点通过第 6 节的解析范围划分算法，根据数据块哈希值存储到参与托管的数据托管节点中. 当展开数据传输时，买方根据跨域交易执行过程中的各个事

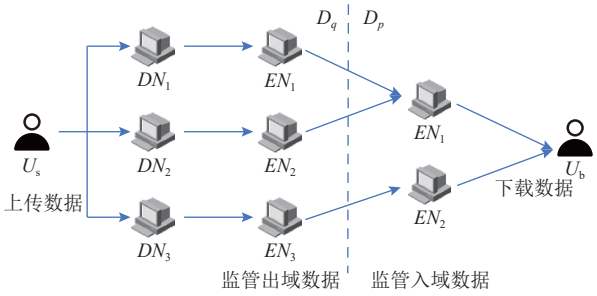



Fig. 6 Cross-domain data transmission process

图 6 跨域数据传输过程

件，包含 *SellSuccess*、*BuySuccess*、$rootHash_d$ 等信息作为凭证，展开对所有数据块的请求，并在传输完成后构建出完整的数据资产内容.

在这个传输过程中，两域的边缘节点应当承担起审查跨域传输数据块内容的职责. 包括审查数据块是否为被交易的数据资产的一部分、数据块包含的内容是否符合出入域规则等.

在跨域数据传输阶段，资金已经支付到卖方，买方也已经收到数据传输凭证，因此本机制中并不存在数据安全风险和资金安全风险. 本系统中数据被分割并冗余存储到数据托管节点中，这种去中心化的存储方式保证了即便存在拒绝提供服务的恶意数据托管节点，买方依然能获取到全部的数据资产内容，同时买方可以通过并行传输保证了买方获取数据的效率. 买方在交易确认后也已经将跨域交易资金打入到跨链的代理账户中，如果卖方将数据的传输凭证已经交付给买方，则卖方可以直接从代理账户中提取资金，否则资金会在一定的锁定期限后自动返还给买方. 上述策略保护了买卖双方的资金和数据安全.

本系统可以通过激励机制促使更多的边缘节点和数据托管节点参与数据的传输和存储，以便可以使用并行传输的方式加快数据的传输效率. 在此过程中当参与数据托管节点、两域的边缘节点比例，与相互之间链接的传输带宽的比例相同时，可以避免信道堵塞和空闲，达到最优的传输效率. 由于参与的节点数量不定，将数据划分成较小的数据块，然后通过动态变换链接方式进行传输和分别进行去重，也可以接近最优的传输效率.

## 6 实现与评估

### 6.1 系统实现

为了实现本文提出的数据交易系统，我们配置了以下实验环境. 本文的实验环境为 20 台 2.4 GHz 8 核 CPU，8 GB 内存，50 GB SSD 硬盘，操作系统为 CentOS 7 的服务器组成. 为验证跨域数据交易的场景，服务器分成 2 个域，域内的服务器通过高速局域网进行连接，两域之间通过设置不同的网段和防火墙保证域内的安全性. 服务器上安装在每个设备上，安装由 Go 语言编写的以太坊 Geth 客户端[47]、每个域内的区块链节点通过共识算法构建一个区块链.

如图 7 所示，区块链上的 DID 标识符注册和解析合约、买方合约、卖方合约通过 Solidity[48] 实现. 域内用户和数据托管节点通过 Web3.py 库[49] 与区块链节点进行交互，实现在区块链上执行身份的解析和注册、数据的购买和支付、数据的发布等操作.

```
function processTransaction(
    IdentityAttributes memory buyer,
    IdentityAttributes memory seller,
    IdentityAttributes memory product,
    string memory buyerMessage,
    bytes memory buyerSignature,
    string memory sellerMessage,
    bytes memory sellerSignature,
    string memory productMessage,
    bytes memory productSignature
) public returns (bool) {
    // 验证所有参与方的签名
    require(verify(buyer.issuerAddress, buyerMessage, buyerSignature),
        "Buyer identity invalid");
    require(verify(seller.issuerAddress, sellerMessage, sellerSignature),
        "Seller identity invalid");
    require(verify(product.issuerAddress, productMessage, productSignature),
        "Product identity invalid");

    // 监管身份信息
    require(regulate1(buyer), "Buyer identity not approved");
    require(regulate2(seller), "Seller identity not approved");
    require(regulate3(product), "Product identity not approved");

    // 进行交易
    return transact(buyer.walletAddress, seller.walletAddress, product.value);
}
```

Fig. 7 Implementation of partial cross-domain data smart contract

图 7 部分跨域数据智能合约实现

### 6.2 系统分析

本节将围绕设计目标展开分析，验证身份跨域解析机制和数据交易系统设计是否满足了设计目标.

1）身份跨域解析结果一致性. 本文的身份跨域解析协议可以根据本域中区块链上的域外区块根哈希验证解析结果的正确性，并可以通过解析结果的区块号，判断是否为最新的解析结果. 如果不是缓存的解析结果，或不是最新的解析结果，可以通过交叉验证的方式获取正确的、最新的解析结果. 此外，对于缓存机制可能造成过时的解析结果的问题，本文通过缓存的被动更新机制，在不影响解析效率的前提下，保证了解析结果一致性.

2）身份跨域解析服务可用性. 假设在区块 $t$，跨域身份解析服务参与节点数量为 $n$，本区块提供的激励为 $s$，则本区块中平均每个区块获得的激励为 $s/n$.

若当参与节点不足以满足当前用户的跨域解析请求，可以通过在区块链中提高激励 $s$，节点能获得的平均激励激励会增加，促使更多的边缘节点参与，提高身份跨域解析服务能力. 如果其中出现恶意节点，可以通过发现机制和惩罚机制，针对恶意行为生成欺诈证明，对恶意节点进行惩罚或驱除，保证服务的可用性. 通过对发现恶意节点的用户进行奖励，可以激励边缘节点积极验证其他节点的解析结果，让破坏服务可用性的节点尽可能快地被发现.

3)跨域数据交易原子性. 跨域数据交易的数据购买可分成 2 个部分，智能合约的执行和数据交易状态的传递. 智能合约的执行是原子性的，$SC_p$ 和 $SC_q$ 的状态信息通过跨域身份解析机制的方式进行跨域. 跨域身份解析机制可以保证一致性和可用性，即便 2 个域中存在恶意节点，状态信息也能实时和正确地传递，保证跨域数据交易的执行不会因异常中止. 同时跨域数据交易存在回滚机制，如果因为身份、监管或其他因素导致交易失败，边缘节点能主动地发起交易，撤销已经变更的中间状态，返还支付的资金. 在数据交付部分，由于数据托管的去中心化特性，因此即便存在恶意的数据托管节点，也不会影响数据正常的跨域传输.

4)细粒度的跨域数据交易监管. 如图 7 所示的合约示例，跨域数据交易合约 $SC_p$ 可以传入买卖双方 DID 标识符、DID 文档、可验证凭证和数据的 DID 标识符、DID 文档作为参数. $SC_p$ 可以身份验证模块验证 DID 标识符和 DID 文档的对应关系，并通过验证可验证凭证的签名验证其完整性. 验证通过后，DID 标识符可以进行基于标识符监管、DID 文档的内容和可验证凭证的内容支撑基于属性的监管.

### 6.3 性能评估

对本系统的性能评估的目的是为了证明利用本系统实现跨域数据交易的可行性. 因此本节首先将根据系统的分层架构对每一层分别进行测试；然后对一组完整的跨域数据交易进行测试，分析系统性能瓶颈，探讨可行的系统优化方案.

为验证跨域身份解析的一致性，本文在两域同时发起多次跨域身份解析，验证解析结果是否具备一致性. 由于不同的缓存命中率对一致性的影响不同，过高的缓存命中率可能导致更多的跨域身份解析结果不一致. 因此本文设计在不同的缓存命中率下进行跨域身份解析结果的一致性对比. 如图 8 所示，在无缓存的跨域身份解析中，可以保证解析结果的一致性. 当缓存命中率在 50% 后，已经存在 14% 的初次解析结果产生了不一致，需要使用二次解析来保证一致性，缓存的实际命中率提高了 72%. 随着缓存命中率的提升，由于缓存的被动更新机制，产生不一致的初次解析结果比率仅是略微上升到 19%，且都通过二次解析机制保证了一致性. 可见本方案能有效地保证系统的一致性. 身份解析仅需要在链下应用或通信中执行 1 次，即便需要二次解析，延迟相较于通信中的其他操作的延迟，如页面渲染、数据传输等占比很低，因此延迟对现实的应用并不会产生明显影响.

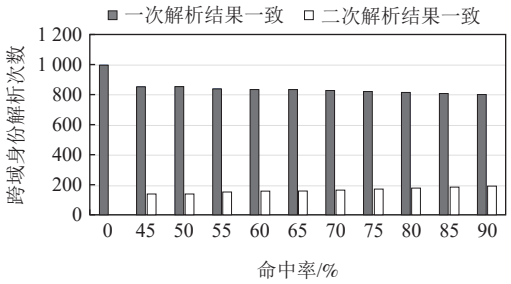



Fig. 8 Cross-domain identity resolution consistency

图 8 跨域身份解析一致性

在身份层中，本文使用 ERC-1056[50] 实现的去中心化身份管理合约在区块链上写入了 100 000 条 DID 标识符和其相应的属性信息，然后测试跨域身份解析的平均延迟. 如图 9 所示，由于每次跨域身份解析需要 2 次域内通信(用户和边缘节点通信)和 2 次域间通信(边缘节点之间通信)，因此平均延迟较高，平均在 180 ms. 由于身份解析具有局部性，因此可以通过缓存的方式加速身份解析速度. 通过 LRU 缓存机制将跨域解析的缓存命中率提升到 70% 之后，能有效地将跨域身份解析的平均延迟降到 82 ms.

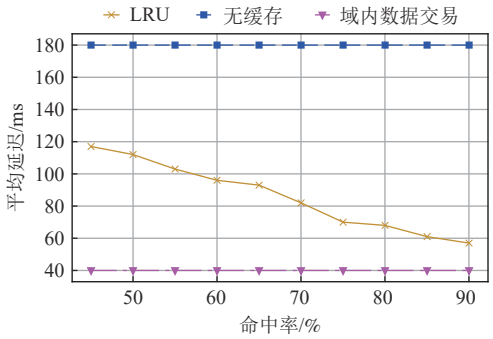



Fig. 9 Cross-domain identity resolution operation delay

图 9 跨域身份解析操作延迟

在交易层中，本文使用 Solidity 实现卖方合约和买方合约，使用 Web3.py 实现用户客户端中与域外或域内的区块链交互的功能. 在一次成功的跨域数据

交易中，至少需要 2 次 EVM 执行智能合约和 4 次用户客户端与区块链的交互(买方调用买方合约，边缘节点调用卖方合约，边缘节点之间传递交易成功事件，边缘节点向用户传递交易成功事件)．在交易层中，主要的性能瓶颈在于智能合约的执行．由于区块链上的智能合约调用交易需要被打包进块后才能执行，因此本层中的智能合约执行的时间并不确定，在区块链上交易量过大时，有可能需要等待多个区块才能被执行．区块链设置出块时间为 1 s，平均每个区块可以打包 200 个交易．交易被打包进块的时间与交易设置的 gas 费用密切相关．如图 10 所示，在以 500 个/s 的速率发送数据交易的区块链上交易时，会造成区块链网络的拥堵，导致每个交易平均需要等待 2.3 个区块才能被打包进块，2 次串行的智能合约执行需要等待 5 s 左右．在交易不拥堵的情况下，跨域交易相较于域内交易的延迟会增加一倍，在拥堵情况下，域内交易的延迟一般也要在 1 个区块以上，跨域交易并不会影响用户的使用体验．同时，随着交易速率的提升，交易成功的延迟会进一步增大．在这个过程中，客户端和区块链的交互所使用的时间可以忽略不计，因此如果想要提升交易层的效率，应当着重提升区块链的吞吐量．如果用户想要降低自己的数据交易的延迟，可以适当提升数据交易的 gas 费用，使得交易能尽快上链执行．

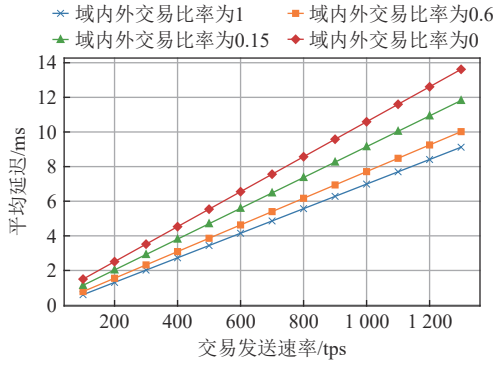



Fig. 10 Cross-domain data trading contract execution delay

图 10 跨域数据交易合约执行延迟

在具体的应用场景中，本系统的效率表现为执行数据交易操作的吞吐量．本节批量构建了大量域内数据交易和域外数据交易，以不同的域内域外数据交易的比例进行数据交易．实验环境中有 2 个域，每个域中存在 8 个区块链节点，其中 4 个区块链节点同时为边缘节点．区块链的吞吐量在区块 gas 上限为 10 000 000 时可以达到 200 tps 左右．但数据交易只是区块链上交易中的一部分，而且其吞吐量会随着域外数据交易比例的提升随之下降．如图 11 所示，当数据交易都为跨域数据交易时，吞吐量将稳定在 90 tps 左右．虽然跨域数据交易每一域中仅需发送 1 次区块链上交易，但由于需要处理另一域的跨域数据交易请求，导致本域区块中的半数区块链上交易被另一域中的跨域请求占用．此外跨域数据交易中需要执行额外的跨域数据监管规则，并需要释放额外的跨域数据交易事件，导致需要其单个交易相较于域内数据交易消耗更多的 gas，也降低了单个区块中的区块上限．如果想要提高本系统的吞吐量，应当从提升区块容纳交易的上限(*Gaslimit*)和减少出块时间两方面入手．例如，Layer2[51] 技术可以有效地降低交易的 gas 费用，让区块中可以打包更多的交易；有向无环图(DAG)技术可以同时打包多个区块，减少平均出块时间．类似的技术还有分片、并行执行等，利用上述区块链前沿技术可以进一步提高系统对跨域数据交易的吞吐量．

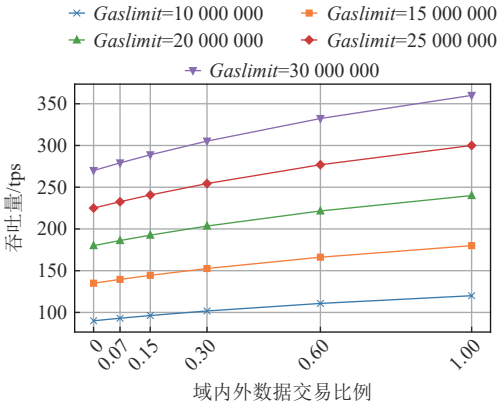



Fig. 11 Cross-domain data trading throughput

图 11 跨域数据交易吞吐量

## 7 总 结

本文设计并实现了一种基于去中心化身份的跨域数据交易系统．本文设计的身份跨域解析机制和跨域数据交易系统模型，可以安全、可监管、高效地支撑更大规模的数据要素市场的构建．

本文的创新点在于让数据交易系统充分应用了去中心化技术．通过去中心化身份技术，将多个数据交易系统以去中心化的形式形成了一个统一的身份层，解决了在跨域数据交易中面临的身份解析结果一致性和身份解析服务可用性问题．在身份跨域解析机制和去中心化身份的基础上，使用区块链、智能合约和去中心化存储技术解决了跨域数据交易的安全和合规问题．

本文所提出的方案能兼容多种区块链系统,现有的中心化和去中心化交易系统都可以通过兼容去中心化身份标准加入本文提出的身份跨域解析机制.此外,本文提出的跨域身份解析机制可以支持多种跨域数据交易场景,保障数据交易的安全性和可监管性.本文方案通过智能合约实现数据交易逻辑,在保证数据交易安全的同时,提高了系统效率,具有很强的灵活性,适用于多种数据要素市场的应用场景.本文方案在数据交易之外,也能应用于其他的需要高安全性保证的跨域场景中,例如物联网设备身份互通、政府业务的跨机构和跨地域的办理等.以物联网为例,物联网中的设备往往被分割到多个互不连通的网络中,这使得跨域的操作难以执行.本文的身份方案可以用作多种场景的身份底座,支撑丰富的业务逻辑.

本文的工作局限于当前区块链、去中心化身份等去中心化技术的发展,未来工作可以应用前沿的区块链可扩展性技术,例如分片、Layer2、并行执行等,进一步优化跨域数据交易的效率,并讨论如何在更大规模的数据交易系统中实现低成本、高可用的跨域数据交易.分片技术可以在链内划分出存储和交易子集,可以通过并行处理更多的交易来减少状态存储的冗余度以存储更多的状态数据;Layer2 方案可以将链上的执行和存储转移到链下,提升链上交易的吞吐量.在系统的效率优化之外,区块链上可以通过零知识证明技术,在链上执行交易,但隐藏交易细节;使用环签名技术,可以实现对交易双方身份的隐藏,提升系统的对用户隐私的保护.

**作者贡献声明**:郝嘉琨完成了系统设计以及实验部分,并撰写论文;向鹏和何逸飞对系统设计提出建议并修改论文;高健博提供了论文思路和技术方案指导并修改论文;关志、谢安明和陈钟提出指导意见并修改论文.

## 参 考 文 献

[1] Dong Xiangqian, Guo Bing, Shen Yan, et al. An efficient and secure decentrailizing data sharing model[J]. Chinese Journal of Computers, 2018, 41(5): 1021−1036 (in Chinese)
(董祥千, 郭兵, 沈艳, 等. 一种高效安全的去中心化数据共享模型[J]. 计算机学报, 2018, 41(5): 1021−1036)

[2] Chen L, Koutris P, Kumar A. Towards model-based pricing for machine learning in a data marketplace[C]//Proc of the 2019 Int Conf on Management of Data. New York: ACM, 2019: 1535−1552

[3] Wang Yujue, Wu Qianhong, Qin Bo, et al. Identity-based data outsourcing with comprehensive auditing in clouds[J]. IEEE Transactions on Information Forensics and Security, 2016, 12(4): 940−952

[4] Jing Dong, Yuan Ye, Zhang Xiaowei, et al. Survey on data pricing and trading research[J]. Journal of Software, 2023, 34(3): 1396−1424 (in Chinese)
(江东, 袁野, 张小伟, 等. 数据定价与交易研究综述[J]. 软件学报, 2023, 34(3): 1396−1424)

[5] Zhao Yanqi, Yu Yong, Li Yannan, et al. Machine learning based privacy-preserving fair data trading in big data market[J]. Information Sciences, 2019, 478: 449−460

[6] Bajoudah S, Dong C, Missier P. Toward a decentralized, trust-less marketplace for brokered IoT data trading using blockchain[C]//Proc of 2019 IEEE Int Conf on Blockchain (Blockchain). Piscataway, NJ: IEEE, 2019: 339−346

[7] Shen Meng, Liu Huisen, Zhu Liehuang, et al. Blockchain-assisted secure device authentication for cross-domain industrial IoT[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(5): 942−954

[8] Su Liya, Shen Xinyue, Du Xiangyu, et al. Evil under the sun: Understanding and discovering attacks on Ethereum decentralized applications[C]//Proc of the 30th USENIX Security Symp (USENIX Security 21). Berkeley, CA: USENIX Association, 2021: 1307−1324

[9] Lu Qinghua, Xu Xiwei, Liu Yue, et al. uBaaS: A unified blockchain as a service platform[J]. Future Generation Computer Systems, 2019, 101: 564−575

[10] Han Rong, Yan Zheng, Liang Xueqin, et al. How can incentive mechanisms and blockchain benefit with each other? a survey[J]. ACM Computing Surveys, 2022, 55(7): 1−38

[11] Islam M M, Islam M K, Shahjalal M, et al. A low-cost cross-border payment system based on auditable cryptocurrency with consortium blockchain: Joint digital currency[J]. IEEE Transactions on Services Computing, 2022, 16(3): 1616−1629

[12] Wu Zhenhao, Gao Jianbo, Li Qingshan, et al. Research on security technologies in data security governance[J]. Journal of Information Security Research, 2021, 7(10): 907−914 (in Chinese)
(吴振豪, 高健博, 李青山, 等. 数据安全治理中的安全技术研究[J]. 信息安全研究, 2021, 7(10): 907−914)

[13] Hu Zhenyu, Deng Lunzhi, Gao Yan, et al. A designated tester identity-based searchable authenticated encryption scheme[J]. Journal of Guizhou Normal University (Natural Sciences), 2022, 40(3): 56−63, 101 (in Chinese)
(胡震宇, 邓伦治, 高岩, 等. 一个指定验证者的基于身份可搜索认证加密方案[J]. 贵州师范大学学报: 自然科学版, 2022, 40(3): 56−63, 101)

[14] Huang Yicai, Yu Bin. Dynamic searchable symmetric encryption scheme for conjunctive queries basedon SHVE[J]. Journal of Computer Research and Development, 2024, 61(6): 1545−1558 (in Chinese)
(黄一才, 郁滨. 一种基于 SHVE 的连接查询动态对称可搜索加密方案[J]. 计算机研究与发展, 2024, 61(6): 1545−1558)

[15] Feng Shuai, Deng Lunzhi. Identity-based data auditing scheme with privacy protection[J]. Journal of Guizhou Normal University (Natural Sciences), 2023, 41(2): 105−112 (in Chinese)

(冯帅, 邓伦治. 基于身份的隐私保护数据审计方案[J]. 贵州师范大学学报: 自然科学版, 2023, 41(2): 105−112)

[16] Shen Yuan, Song Wei, Zhao Changsheng, et al. A cross-domain ciphertext sharing scheme supporting access behavior identity tracing[J]. Journal of Computer Research and Development, 2024, 61(7): 1611−1628 (in Chinese)
(申远, 宋伟, 赵常胜, 等. 支持访问行为身份追踪的跨域密文共享方案[J]. 计算机研究与发展, 2024, 61(7): 1611−1628)

[17] Feng Chaosheng, Liu Bin, Guo Zhen, et al. Blockchain-based cross-domain authentication for intelligent 5G-enabled Internet of drones[J]. IEEE Internet of Things Journal, 2021, 9(8): 6224−6238

[18] Chen Dajiang, Zhang Ning, Cheng Nan, et al. Physical layer based message authentication with secure channel codes[J]. IEEE Transactions on Dependable and Secure Computing, 2018, 17(5): 1079−1093

[19] Chen Dajiang, Jiang Shaoquan, Zhang Ning, et al. On message authentication channel capacity over a wiretap channel[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 3107−3122

[20] Li Ming, Weng Jian, Yang Anjia. CrowdBC: A blockchain-based decentralized framework for crowdsourcing[J]. IEEE transactions on parallel and distributed systems, 2018, 30(6): 1251−1266

[21] Qin Zhiguang, Xiong Hu, Wu Shikun, et al. A survey of proxy re-encryption for secure data sharing in cloud computing[J]. IEEE Transactions on Services Computing, 2016. Doi:10.1109/TSC.2016.2551238

[22] Lai Junzuo, Deng R H, Guan Chaowen, et al. Attribute-based encryption with verifiable outsourced decryption[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(8): 1343−1354

[23] Lai Junzuo, Deng R H, Guan Chaowen, et al. Cryptographic solutions for cloud storage: Challenges and research opportunities[J]. IEEE Transactions on Services Computing, 2019, 15(1): 567−587

[24] Wan Zhiguo, Liu Wei, Cui Hui. HIBEChain: A hierarchical identity-based blockchain system for large-scale IoT[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 20(2): 1286−1301

[25] Yang Xiaohui, Li Wenjie. A zero-knowledge-proof-based digital identity management scheme in blockchain[J]. Computers & Security, 2020, 99: 102050

[26] Xu Jie, Xue Kaiping, Tian Hangyu, et al. An identity management and authentication scheme based on redactable blockchain for mobile networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(6): 6688−6698

[27] Koutsos V, Papadopoulos D, Chatzopoulos D, et al. Agora: A Privacy-Aware Data Marketplace[J/OL]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(6): 3728−3740

[28] Li Chunlin, Liang Songyu, Zhang Jing, et al. Blockchain-based data trading in edge-cloud computing environment[J/OL]. Information Processing & Management, 2022, 59(1): 102786

[29] Gupta P, Dedeoglu V, Kanhere S S, et al. TrailChain: Traceability of data ownership across blockchain-enabled multiple marketplaces[J]. Journal of Network and Computer Applications, 2022, 203: 103389

[30] Dixit A, Singh A, Rahulamathavan Y, et al. FAST DATA: A fair, secure, and trusted decentralized IIoT data marketplace enabled by blockchain[J]. IEEE Internet of Things Journal, 2023, 10(4): 2934−2944

[31] Lin C H V, Huang C C J, Yuan Yanghao, et al. A fully decentralized infrastructure for subscription-based IoT data trading[C/OL]//Proc of 2020 IEEE Int Conf on Blockchain (Blockchain). Piscataway, NJ: IEEE, 2020: 162−169

[32] Ou W, Huang S, Zheng J, et al. An overview on cross-chain: Mechanism, platforms, challenges and advances[J]. Computer Networks, 2022, 218: 109378

[33] Gilad-Bachrach R, Laine K, Lauter K, et al. Secure Data Exchange: A marketplace in the cloud[C/OL]//Proc of the 2019 ACM SIGSAC Conf on Cloud Computing Security Workshop. New York: ACM, 2019: 117−128

[34] Hu Donghui, Li Yifan, Pan Lixuan, et al. A blockchain-based trading system for big data[J]. Computer Networks, 2021, 191: 107994

[35] Manu S, Amy G, Markus S, et al. Decentralized Identifiers (DIDs) v1.0[EB/OL]. [2024-05-15]. https://www.w3.org/TR/did-core/

[36] Manu S, Grant N, Dave L, et al. Verifiable Credentials Data Model v1.1[EB/OL]. [2024-05-15]. https://www.w3.org/TR/vc-data-model/

[37] Manu S, Kyle D H, Michael P. DID Specification Registries[EB/OL]. [2024-07-16]. https://www.w3.org/TR/did-spec-registries/

[38] Decentralized Identity Foundation. decentralized-identity/universal-resolver: Universal Resolver implementation and drivers[CP/OL]. [2024-05-10]. https://github.com/decentralized-identity/universal-resolver

[39] Decentralized identity foundation. decentralized-identity/universal-registrar[CP/OL]. [2024-05-10]. https://github.com/decentralized-identity/universal-registrar

[40] Ren Yuzheng, Xie Renchao, Yu F R, et al. Potential identity resolution systems for the industrial Internet of Things: A survey[J]. IEEE Communications Surveys & Tutorials, 2020, 23(1): 391−430

[41] Inci M S, Gulmezoglu B, Irazoqui G, et al. Cache attacks enable bulk key recovery on the cloud[C]//Proc of the18th Int Conf on Cryptographic Hardware and Embedded Systems (CHES 2016). Berlin: Springer, 2016: 368−388

[42] Han Xuan, Yuan Yong, Wang Feiyue. Security problems on blockchain: The state of the art and future trends[J]. Acta Automatica Sinica, 2019, 45(01): 206−225 (in Chinese)
(韩璇, 袁勇, 王飞跃. 区块链安全问题: 研究现状与展望[J]. 自动化学报, 2019, 45(01): 206−225)

[43] Gao Jianbo, Zhang Jiashuo, Li Qingshan, et al. RegLang: A smart contract programming language for regulation[J]. Computer Science, 2022, 49(S1): 462−468 (in Chinese)
(高健博, 张家硕, 李青山, 等. RegLang: 一种面向监管的智能合约编程语言[J]. 计算机科学, 2022, 49(S1): 462−468)

[44] ERC−20: Token standard[EB/OL]//Ethereum Improvement Proposals. [2024−05−16]. https://eips.ethereum.org/EIPS/eip-20

[45] Jayabalan J, Jeyanthi N. Scalable blockchain model using off-chain IPFS storage for healthcare data security and privacy[J]. Journal of Parallel and Distributed Computing, 2022, 164: 152−167

[46] Kostamis P, Sendros A, Efraimidis P S. Data management in Ethereum DApps: A cost and performance analysis[J]. Future Generation Computer Systems, 2024, 153: 193−205

[47] Wood G. Ethereum: A secure decentralised generalised transaction ledger[J]. Ethereum Project Yellow Paper, 2014, 151(2014): 1−32

[48] DANNEN C. Introducing Ethereum and Solidity[M]. Berlin: Springer, 2017

[49] Ethereum. ethereum/Web3. py[CP/OL]. [2024-05-15]. https://github.com/ethereum/web3.py

[50] Ethereum. ERC−1056: Ethereum lightweight identity[EB/OL]// Ethereum Improvement Proposals. [2024-05-10]. https://eips.ethereum.org/EIPS/eip-1056

[51] Li Bei, Wu Hao, He Xiaowei, et al. Survey of storage scalability in blockchain systems[J]. Computer Science, 2023, 50(1): 318−333 (in Chinese)
(李贝, 吴昊, 贺小伟, 等. 区块链系统的存储可扩展性综述[J]. 计算机科学, 2023, 50(1): 318−333)



**Hao Jiakun**, born in 1998. PhD condidate. His main research interests include blockchain, distributed system, and cryptography.
**郝嘉琨**, 1998 年生. 博士研究生. 主要研究方向为区块链、分布式系统、密码学.



**Xiang Peng**, born in 1994. PhD condidate. His main research interests include blockchain and network security.
**向 鹏**, 1994 年生. 博士研究生. 主要研究方向为区块链、网络安全.(xiangp@pku.edu.cn)



**He Yifei**, born in 1998. PhD candidate. His main research interests include blockchain technology and cryptography.
**何逸飞**, 1998 年生. 博士研究生. 主要研究方向为区块链技术、密码学.(heyifei@pku.edu.cn)



**Gao Jianbo**, born in 1994. PhD, associate professor. Member of CCF. His main research interests include blockchain, digital identity, and domain-specific software engineering.
**高健博**, 1994 年生. 博士, 副教授. CCF 会员. 主要研究方向为区块链、数字身份、特定领域软件工程.(gao@bjtu.edu.cn)



**Guan Zhi**, born in 1980. PhD, associate professor. Member of CCF. His main research interests include blockchain and cryptography.
**关 志**, 1980 年生. 博士, 副研究员. CCF 会员. 主要研究方向为区块链、密码学.(guan@pku.edu.cn)



**Xie Anming**, born in 1977. PhD. His main research interests include big data security, blockchain security, and cryptography application.
**谢安明**, 1977 年生. 博士. 主要研究方向为大数据安全、区块链安全、密码学应用.



**Chen Zhong**, born in 1963. PhD, professor, PhD supervisor. Fellow of CCF. His main research interests include domain-specific software engineering, network and information security, and system software.
**陈 钟**, 1963 年生. 博士, 教授, 博士生导师. CCF 会士. 主要研究方向为特定领域软件工程、网络与信息安全、系统软件.(zhongchen@pku.edu.cn)