基于数据对象的跨域情报可信共享

彭泰[1,2]，孙晶[1]，陈旭涧[1,2]，周纤[2]，叶宇铭[2]，白晓颖[2]
1. 北方工业大学信息学院，北京 100144；
2. 中国人民解放军军事科学院军事科学信息研究中心，北京 100142

摘要
情报数据作为一种高价值的数据资产，被存储在不同平台，被不同主体所持有，具有高分散性和低可用性特点。由于结构形态与存储方式不同，多源异构情报数据难以实现高效汇聚共享，多主体间情报信息融合和综合分析利用存在较大困难。因此亟须在跨域情报主体间建立安全可信的共享互操作机制，在满足数据确权、安全审计等管理要求的同时，实现情报信息的深度挖掘。针对跨域情报数据可信共享需求与应用特点，提出基于数据对象的情报管理方法，并采用数字对象体系架构及区块链可信访问控制技术构建跨域情报数据可信共享系统，实现多源异构情报数据的视图统一和跨域可信共享，为情报数据融合汇聚、情报信息智能分析提供技术支撑，充分挖掘情报信息的巨大潜力。

关键词
可信共享；数据对象；情报分析

中图分类号：TP315 文献标志码：A doi: 10.11959/j.issn.2096-0271.2023049

Trusted sharing of cross-domain intelligence based on data objects

PENG Tai[1,2], SUN Jing[1], CHEN Xujian[1,2], ZHOU Xian[2], YE Yuming[2], BAI Xiaoying[2]
1. School of Information, North China University of Technology, Beijing 100144, China
2. Information Research Center of Military Services, PLA Academy of Military Science, Beijing 100142, China

Abstract
Intelligence data, as a high-value data asset, is stored on different platforms and held by different subjects, and has the characteristics of high dispersion and low availability. Due to the different structural forms and storage methods, it is non-trivial to achieve efficient aggregation and sharing of multi-source heterogeneous intelligence data, and there are great difficulties in the fusion and comprehensive analysis and utilization of intelligence information among multiple subjects. Therefore, there is an urgent need to establish a secure and trustworthy sharing and interoperability mechanism among cross-domain intelligence subjects to meet management requirements such as data validation and security auditing, while realizing correlation analysis and cross-validation of intelligence information and deeply mining the intelligence value therein. To address the needs and application characteristics of cross-domain intelligence data trustworthy sharing, this paper proposes a data object-based intelligence management method and adopts the digital object architecture and blockchain trusted access control technology to build a cross-domain intelligence data trustworthy sharing system, realizing the view unification and cross-domain trustworthy sharing of multi-

source heterogeneous intelligence data, providing technical support for intelligence data fusion and convergence and intelligence information intelligent analysis, and fully exploiting the huge potential of intelligence information. This will provide technical support for intelligence data convergence and intelligent analysis, and fully exploit the huge potential of intelligence information.

Key words
trusted sharing, data object, intelligence analysis

0 引言

情报数据是高价值的数据资产。然而，受收集和产生方式的局限，资产所有权往往归属于不同主体机构[1]，数据采用不同的机制和技术平台存储管理，因此情报数据具有高分散性和低可用性[2]的特点。相较于数据总体的高价值，数据被不同主体分散持有后，数据相关性和完整性被隔断，许多关联信息无法被发现[3]。然而跨域多源数据通常比单一来源数据具有更多的维度和广度，可以从不同的视角来观察、比较、理解同一个问题。为充分挖掘情报数据价值，情报研究从文献传递、信息服务、知识服务到智能服务的关键转变在于分布在不同主体机构的信息能够汇聚融合、综合分析实现跨域情报数据的有效利用[2,4]，未来情报服务方向也将走向面向问题的、以用户为中心的协同工作模式。该模式下微观情报工作流程由线性结构转变为开放透明的网络结构，各方主体会以多种角色参与到整个情报分析过程中，情报数据也从单主体内部的线性流动转变为多主体之间的共享与互操作。

在多主体协同的情报研究模式下，多方主体的情报数据类型繁多、结构各异，呈现碎片化多源异构特点。充分利用多来源高质量碎片化信息与关联性价值，集结多方情报研究人员共同分析成果，是提升情报研究效率和提高研究成果质量的有效手段[5]。因此，如何在不改变现有数据所有权和管理模式的前提下，将分布于各数据节点的异构数据统一管理，打通信息系统间的数据壁垒，推进情报数据跨主体间可信共享是推进协同情报研究工作的关键。

一方面，受收集和产生方式的局限，分散在各主体间的情报数据有着各自不同的目录结构以及组织方式，且由于用户需求在系统开发阶段的不完整性、在运行阶段的持续扩展性和演化性，以及信息技术更新换代的阶段性等原因，各主体间形成数以百万计的“信息孤岛”，主体之间的数据往往很难被外界发现、获取和访问，不利于促进跨机构、跨部门、跨系统的数据资源互通和互操作[6-7]。另一方面，针对全域多维信息来源广泛而分散的情况，需要实现跨层级、跨地域、跨系统、跨部门、跨业务的大数据共享与融合，其安全要求高、共享难度大、需要安全、有效的信任体系支撑。因缺乏访问控制及事后监管、审计、溯源、确权等方面要求的可信共享方案，情报数据的各个持有主体对共享数据往往存在着各种顾忌，导致分散的情报数据在使用者和所有者之间难以有效共享利用，并发挥其情报价值，“数字生产力”[8]作用大为受限。

为了实现多主体间协同模式下的情报数据融合共享，无缝嵌入现有碎片化情报研究方法[5]，本文基于对象化数据抽象思想，围绕碎片化情报数据对象化与可信共

享，提出一种数据对象化的抽象与管理方法，利用数字对象体系架构及区块链相关技术，构建基于数字对象体系的跨域情报数据可信共享系统，将碎片化多源异构情报数据从各自所属的信息系统中抽象为数据对象统一管理，对外表现为统一的数据视图；并基于数据对象实现对碎片化情报数据的标识、检索和可信访问控制，对外按需共享并提供数据接口服务，为情报数据融合汇聚、情报信息智能分析提供技术支撑，充分挖掘情报信息的巨大潜力。

1 数据对象管理

1.1 情报数据对象

数据对象是指存储在计算机系统中的数据元素，它是数据处理和分析的基本单元。在信息领域，数据对象具有重要的作用。数据对象的种类多种多样，在不同的领域具有不同的应用。其中，情报领域和非情报领域的数据对象有着显著的区别。非情报领域的数据对象主要涉及商业、科研和日常生活等方面。它们可以来自各种渠道，包括交易记录、库存数据、客户信息等。这些数据对象通常是可公开的，处理和分析的技术比较成熟和通用，主要用于支持业务流程和决策效率。类似地，情报数据对象同样具有来源广泛、形式复杂等特点，除此之外还具有以下3点特殊性。

- 高敏数据机密安全难。由于领域的特殊性，情报数据对象往往是高度敏感和机密的，数据的保密和安全工作对于情报数据归属主体十分重要。因此，情报数据对象对安全性的要求也更为严格，需要进行严格的访问控制并设置保密措施，以确保数据的安全性和保密性。
- 数据跨域可信共享难。情报数据对象可以来自各种渠道，包括开源情报、人员情报、信号情报等，通过多元化手段进行收集和处理的情报数据对象也具有非常复杂的形式，包括文本、音频、图像、视频等，再加上情报数据对象高安全性和机密性的要求，导致情报数据对象难以进行跨域的可信共享。
- 多主体间协同处理难。情报中蕴含的深层次有价值的信息往往需要情报分析人员利用专业技能和工具收集、整理并加以综合分析才能获发掘出来，复杂情报研究场景下还会出现多主体共同参与分析与研判的情况，以实现更加准确和全面的情报分析。因此还需要实现多主体间情报数据的获取和使用，这无疑增加了情报数据对象可信共享的难度。

综上所述，情报分析领域亟须探索统一抽象的数据对象来规范情报数据的可控访问使用，使不同来源的情报数据可以进行有效整合和可信共享，同时在确保情报数据安全性和保密性的前提下提高情报分析的效率和准确性。

1.2 数据抽象与对象化

在信息系统中，处理业务的逻辑与处理数据的算法相较于数据本身往往更受关注。数据的表现形式和存储结构被设计为系统更加便捷地取用、更加快速地处理[9-10]。大数据时代，数据的价值逐渐被认识和发掘，使数据的资源属性也变得越发重要。数据对象是针对数据在资源层面的抽象，是数据作为可获取的独立资源而存在的表现形式。借由数字对象这层抽象表达，数据得以脱离其具体存在情景，拥有统一的视图并作为资源被调配与使用。数据对象化是指将数据从信息系统中“独立”出来，屏蔽其复杂各异结构、管理与存

储方式等细节，强调其作为可获取的独立的资源属性的过程。

多源异构数据在对象化之后便具有了统一的数据视图，作为数据资源而存在。由于数据本身广泛的来源和复杂多样的结构，数据对象应具有足够强大的包容性从而可以容纳所有的数据，但数据最终还是需要依托现实的物理存储而存在，其生成与来源也与具体的生成环境相关，因此数据对象应当有一个可自定义的部分用于存储数据本身。其内部结构可能十分简单，也可能比较复杂，但相对于外界而言这一部分应当类似于“资源黑盒”，且需要针对“黑盒”属性进行描述的部分。数据对象总体而言应包含两个部分，数据本体以及对数据的描述，如图1所示。

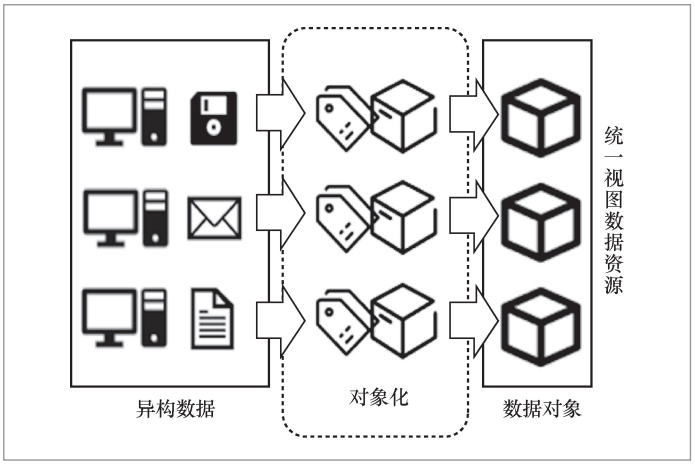

图 1 数据对象抽象

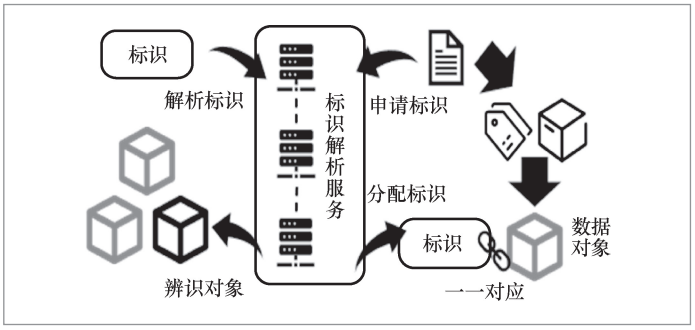

图 2 数据对象标识

1.3 数据对象唯一标识

对数据进行标识是进行数据管理的常用手段，在系统内给予数据一个标识，用于对数据进行辨识和区分。分散于各个信息系统中的数据在其各自系统内的标识方式不一致，这导致在研究过程中无法有效辨识和调度资源。数据的统一标识也是跨系统、跨主体之间数据互操作需要解决的根本问题之一[11]，数据对象作为多源异构数据的进一步抽象分装，更需要有将其分辨与识别的手段。

数据对象标识作为一项永久属性存在于数据对象中属性与描述的部分中，是辨识和管理数据对象的关键，如图2所示。一方面，数据对象之间需要一种相互分辨和区别的凭据，用于在相同视图的资源环境中准确地辨识和代表每一个对象；另一方面，在被调用、共享和流动的过程中，同一数据对象应能被准确地识别到。为了保证标识的永久性，应尽量避免标识与数据本身的内容或信息相关。为了保证标识的唯一性，在资源网络中应设有分配标识的服务，所有数据对象的标识应由且只能由标识服务分配而来。在进行数据抽象化形成数据对象的同时，应同步向标识服务申请唯一标识作为其数据对象身份的证明，只有拥有标识数据对象才能参与进一步的流动和共享等活动。标识服务则负责对所有的数据标识进行分配、登记与管理，维护整个系统中的标识体系，以保证后续的其他工作顺利进行。

1.4 数据对象发现与存储

被外界发现，是数据资源发挥其价值的前提，标识仅仅作为一个永久标识符与数据对象绑定，而不带有对数据本身内容的描述。数据对象的发现需要依托数据对

象中对数据本身的描述与属性部分。元数据是一种结构化的数据，用于记录数据的内容、环境、结构、相互关系和溯源等信息，更容易被管理、检索、传播和使用[12]。元数据是数据对象得以被发现的关键，它能够提供给访问者数据对象的描述信息，使数据对象更容易被检索与发现。元数据主要包括资源的使用方式、类别、领域、标签、关键词等信息。在数据对象管理中应有一项发现服务，集中存储与管理数据对象的元数据，以便外部访问或需求能够适时地获知数据对象的存在。数据对象分散存储在各对象仓库中，发现服务汇聚其元数据，对外集中提供元数据检索与发现的入口。在发现服务中，用户得知数据对象的存在之后便可以利用标识在仓库中寻找并获取到目标数据对象，如图3所示。

数据对象的存储服务需要支持数据对象的访问管理和持久化服务。考虑到数据对象统一的形式以及各数据所在节点的实际存储方式差异，数据对象存储服务可以作为实际数据存储系统的上层结构。所有数据对象存储服务仅需要保持对外暴露的数据对象操作接口的一致性，接收标准消息并返回结果，而不必关心数据对象操作的具体实现，各式数据存储系统能够根据情况自行选择实现方式。对外而言，各数据对象存储服务能够拥有一致的操作接口与消息标准，进而促进数据对象的流动与系统间的互操作。

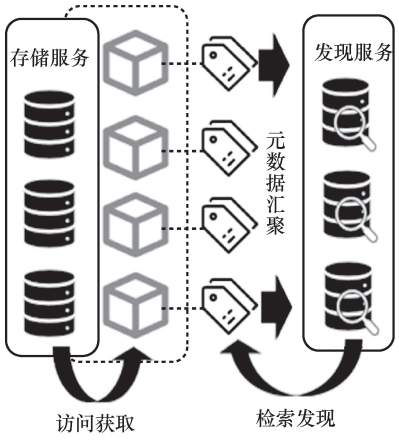

图 3 数据对象发现

2 基于数字对象体系的跨域情报数据管理系统

跨域情报数据管理系统以情报数据对象为基础，以情报数据对象服务网络为核心，以标识解析网络联通各主体间的分布式情报数据仓库，并建立情报数据多元情报数据发现服务，支撑跨域情报数据的共享和各主体间的互操作。

2.1 数字对象体系

数字对象体系结构（digital object architecture，DOA）由图灵奖获得者Rober E. Kahn提出。DOA以数字对象的形式对当前互联网中的数据进行统一抽象以提高数据的交互能力[13]。DOA除了将数据资源抽象为数字对象外，还涉及数字对象的存储、访问和管理等方面，其构成组件包含了负责对数字对象（DO）进行标识和解析标识的标识与解析系统（identifier/resolution system，IRS）、负责DO存储和访问的数字对象仓库（repository）以及负责管理DO元数据以提供对外发现的元数据注册表（registry）3个部分。此外，DOA还制定了标识与解析协议（identifier/resolution protocol，IRP）和数字对象接口协议（digital object interface protocol，DOIP），用于规范DOA与外部的交互以及3个组件内部的交互[14]，总体结构如图4所示。当前DOA的实现有由DONA基金会管理的DOA/Handle[15-18]和CNRI推出的Cordra[19]。DOA的典型应用——

数字对象标识符（DOI）已被广泛应用于ACM和IEEE等数字图书馆。

2.2 基于DOA的情报数据对象服务网络

碎片化情报分析是一种针对情报研究课题需求，基于内容相关的碎片化信息，结合大数据与人工智能技术，嵌入情报业务环节，通过流程、数据、技术标准规范，充分利用高质量碎片化信息与开发关联性价值，提升情报研究效率与成果质量的情报研究方法[5]。本文使用情报数据对象代替碎片化情报分析中的碎片化情报数据，帮助碎片化情报分析在多主体间协同进行。

本文基于标识解析系统（IRS）搭建情报数据对象管理中的标识服务，基于注册表系统搭建数据对象中的发现服务，基于仓库系统搭建数据对象存储服务，使用DOIP和IRP在各服务间建立通信，共同组建情报数据对象服务网络，实现情报数据对象的跨主题间共享，如图5所示。

碎片化情报数据是对情报数据的进一步解构与拆分，其本质是将收集而来的情报数据进行拆解，进而形成多条碎片化的情报数据。每一条碎片化情报数据都是原情报数据的一个部分，代表着原情报数据中的某一个观点或者某一个信息片段。碎片化情报数据作为情报数据的拆分与凝练结果，具有结构粒度较小、数量多、内容多样、形式多样、蕴含信息相对独立且发散等特点，因此本文将碎片化情报数据抽象为情报数据对象。

情报数字对象包含标识、元数据、状态数据以及实体4个部分，如图6右侧所示。对每一个情报数字对象，关联一个唯一的“持久性标识符”，标识永久且唯一地指向一个情报数字对象，是情报数字对象永久不变的属性；元数据是对情报数字对象的描述信息，用于情报数字对象的检索和发现；状态数据包含情报数字对象当前所处位置、访问入口、访问方式等信息，用于情报数字对象的定位和访问；实体则是情报数据包含的实际内容，可以是任何比特序列或一系列比特序列的集合。采用标识和状态数据分离的方式使情报数字对象的标识与其访问入口的耦合性减弱，因此无论是否在互联网环境中，情报数字对象都可以被共享、访问。

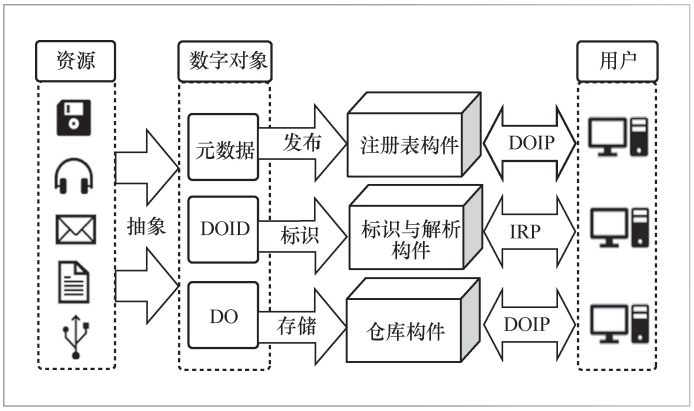

图 4 数字对象体系架构

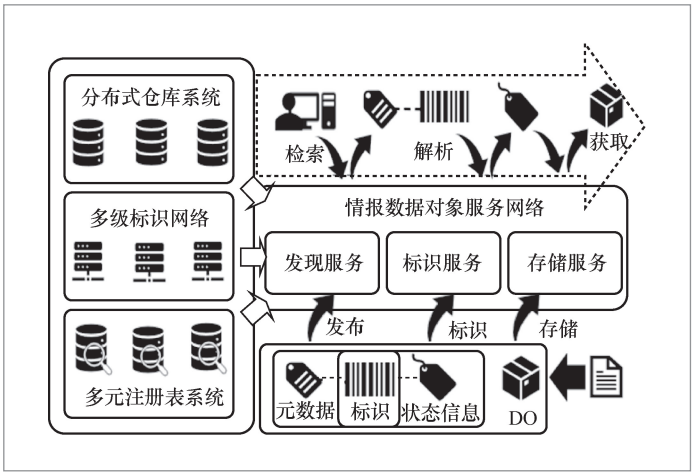

图 5 基于 DOA 的情报数据对象服务

情报数据的抽象分为4个过程，如图6左侧所示。首先对情报数据进行自动/半自动化的状态信息和元数据抽取，再将情报数据本体进行序列化，之后两者组合为未标识的对象，并交于进行标识分配与解析

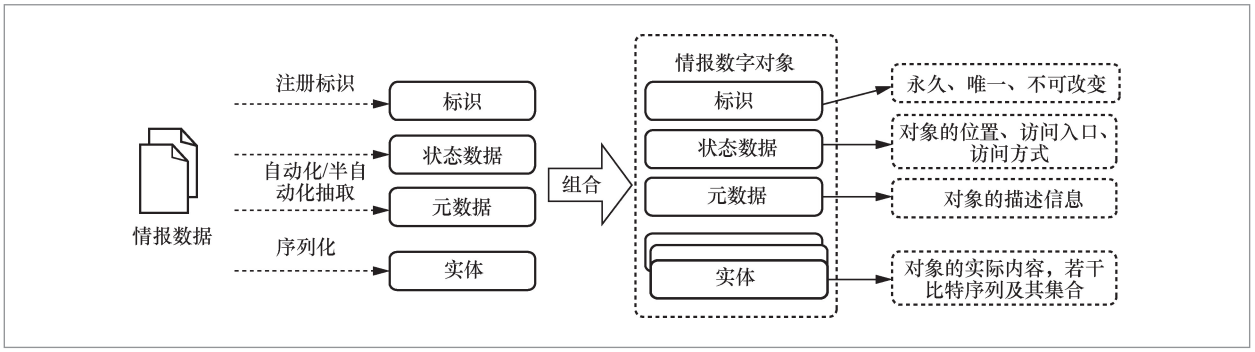

图 6 情报数据对象结构与抽象过程

的服务进行标识的分配注册，最后将所有部分组合为完整的情报数字对象。

在情报数据的整个生命周期中，存在被多个系统使用和修改，在多个机构或部门之间流动的情况，需要合理地设立标识注册与解析节点。本文暂且考虑在情报研究领域，针对某一情报研究体系或者多个情报研究主体合作情况下的IRS设立。对于由多个情报机构组成的情报研究体系，每一个情报机构下属存在多个部门，同时考虑情报数据的整个生命周期。本文采用按机构部门和功能划分的多级IRS设立方法，如图7所示。

首先设立一个全局根节点，其代表整个领域下的最高级别IRS节点，由此往下，每一个机构设立一个IRS节点，以区分各情报数字对象资源所属的主体。往下每一机构下属部门通常拥有各自不同的职能以及相对应的情报数据处理系统。通常情况下，各个部门系统之间负责的情报数据处理职能与情报数据生命周期是大致吻合的，例如负责情报收集的部门、负责情报融合的部门、负责情报分析的部门等，通过对各部门设立IRS节点，也能在一定程度上反映情报数据在其生命周期的位置。

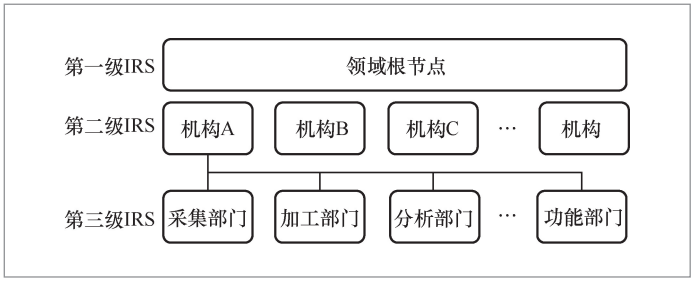

图 7 多级 IRS 结构

在涉及多个数据仓库以及部门的情况下，DOIP以及对象仓库本身的机制可以确保数字对象的安全性。DOIP是一种用于分布式系统中的对象标识和验证的协议，它能够确保数字对象在不同仓库之间的唯一性和完整性。同时，仓库本身也提供了一系列安全机制，如访问控制、权限管理和数据加密等，以确保数字对象在存储和传输过程中的安全性。DOIP和仓库的安全机制相结合，可以有效地保护数字对象免受篡改、丢失或未经授权访问的风险。这样的综合保护措施可以提高数字对象的安全性，并为多仓库环境下的数据管理提供可靠的保障。

考虑到情报数字对象可能来自不同主体，同时又需要被多方主体发现和使用，因此情报数字对象元数据应具有统一的元数据标准，对DO进行规范化描述。情报数字对象元数据规范应具有一个较宽的覆盖范围，并且支持灵活扩展，既需要能够描述情报数据原始信息的部分，也需要能够包含唯一标识符、数据资源当前位置、发布者、发布时间、授权信息等描述资源状态的部分。此

外从情报分析人员检索的角度，还可以增添元数据扩展部分，记录数据被检索的频次与后续评价等信息，从而在提升情报数据检索体验的同时，进一步发掘情报数据的潜藏价值。

本文将情报数字对象的元数据划分为3个部分：数据本体描述性部分、状态信息部分以及可扩展部分。描述性部分参考机读目录（machine readable catalogue，MARC）和都柏林核心元数据（dublin core，DC），主要用于对情报数字对象本身信息进行描述，可以包含情报数据本身的信息片段的概括关键词、拆分来源、类型、名称、标题等，是用于检索的主要部分；状态信息部分包含情报数字对象的唯一标识符、所在仓库标识、获取方式、授权信息、所有者等信息，因此需要与DO标识关联且主要用于管理使用，一般由IRS存储与管理；可拓展部分可以根据领域、业务等需要进行设计，辅助描述性部分更好地对情报数字对象进行全面且进一步的描述。描述性部分与可扩展部分通常发布于元数据注册表，供外部检索与发现，如图8所示。

在整个系统中，机构与机构之间、机构下部门与部门之间都有情报数据共享与互操作的需求，因此在每一个机构标识解析节点下需设立元数据中心，以满足不同机构针对数据对象不同的共享需求。机构可以根据共享的不同数据类型、不同作用或者不同的目标使用者，设立多个元数据中心以满足多元情报数据对象发现的需求。此外由于通常数据不在该级标识解析节点下而是在部门级标识解析节点下存储，故该级标识解析节点下可以不设直接数据对象仓库。在部门级别标识解析节点下，至少需设立一个注册表以供不同部门之间的数据发现共享需求。除此之外还需设立数据对象仓库，以存储具体的情报数据对象。存储在各仓库中的情报数据对象可以根据需求将其元数据放在各注册表中，以供检索与发现。

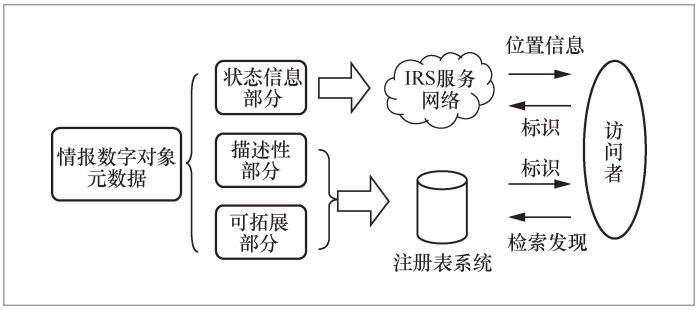

图 8 情报数字对象元数据

情报数据对象元数据作为抽象描述情报数据的数据，其本身的描述性概要（包括来源、设计主题、领域方向及内容概要）甚至其本身具有相当的敏感性。参考现代保密制度中分级保密的思想，本文中访问控制的标的不局限于情报数据对象本身，还包括情报数据对象的元数据，当且仅当情报数据对象访问者密级高于情报数据对象密级，且因业务范围、特定事务等原因存在访问需求时，访问者才可获取情报数据对象的元数据信息。综上所述，情报数据对象的元数据仅对特定范围的访问者开放，以保护参与可信共享机制的情报数据安全可控。

2.3 情报数据对象服务网络下的情报检索发现与关联分析

在情报数据对象服务网络中，情报数据被赋予全局唯一的标识，并且以统一的数据对象形式存储在拥有统一接口的分布式数据对象仓库中，其状态信息和标识交于标识解析系统管理，其元数据交于元数据发布中心进行公开发布。当有新的主体加入协同情报研究时，情报提供可以选择直接将数据交给现有的情报数据对象仓库，仓库会将数据转化为数据对象形式，并且向标识系统通信请求为新的数据对象分配唯一标识，之后再将返回的标识与数据资源组成完整已标识的数据对象并存入仓

库中；此外，主体也可选择改造自己现有的存储系统或数据库，使其对外暴露标准接口接受标准请求消息，并实现数据对象基础操作，最后将其在标识系统中进行注册和标识分配以便仓库能够解析到。

多主体间的情报研究人员在进行协同情报研究时，根据研究课题在发现服务中利用关键词检索等方法筛选出所需的元数据信息，通过元数据对情报数据进行进一步的了解，确认后将元数据中包含的数据对象标识发送给标识服务进行解析，从而获得数据本身所属主体及其仓库访问地址，最后通过存储服务向仓库请求获取数据。对于情报研究人员而言，情报数据对象服务网络是一个集情报信息发现与获取于一体的平台，汇聚了大量分散、自治的情报数据，屏蔽了各种异构问题，为情报研究人员提供统一的数据视图，且可基于规范访问数据。如此，情报研究人员便能够聚焦于数据和研究本身，从而更好地发掘情报数据本身的价值，提高研究效率。

3 基于区块链的多主体情报可信访问控制

3.1 多主体情报数据可信共享

访问控制是多主体间数据共享及互操作安全的一道重要防线。随着当今时代数据内容价值的提高，为了保护共享数据资产的安全和保密，一直以来在计算机系统中有许多机制用于防止在数据共享时发生数据泄露的问题[20]。随着各种第三方共享接口的出现，情报大数据的可信共享需要高效、灵活、动态的可信授权访问控制服务系统。传统的权限管理和访问控制框架在应用于多主体间情报大数据可信共享时在灵活性、扩展性、可信性等方面存在缺陷和不足。

当前已存在的主流权限管理和访问控制服务框架，如Oauth 2.0 权限管理和访问控制框架，大多有相同的缺陷，即都依赖一个中心化的授权认证服务器负责权限的管理和授权的认证。在实现情报数据共享时，中心化的服务框架天然存在以下无法回避的问题，难以达到可信共享的要求。首先，中心化的权限管理和访问控制服务器被恶意攻击者攻击，可能会导致大规模的敏感数据外泄；其次，中心化的权限管理和访问控制服务器的具体计算过程和计算环境难以对参与本体系的全部主体做到完全的公开透明。

分布式账本技术的出现让去中心化公开可溯源计算系统成为现实，基于分布式账本技术的密码学分布式记账系统采用P2P网络以及共识算法，能让网络中的节点共同参与到交易的确认、区块的认证以及区块链账本的存储和更新维护中，使参与各方主体可以不依赖传统的中心化权限管理和访问控制计算服务器，在一个去中心化的权限管理和访问控制计算系统中进行计算，保障了权限管理和访问控制计算过程的安全、可信、可溯源[20-21]。

3.2 基于分布式账本的权限管理和访问控制机制

情报数据对象具有敏感性高与机密性高的特点，在开展跨部门、体系间情报数据对象访问时，本文使用了分布式账本通过访问与授权流程，保证权限管理和访问控制计算过程的安全、可信、可溯源。

本文基于分布式账本提出了去中心化权限管理和访问控制机制，将权限管理和访问控制的授权计算过程使用分布式账本去中心化地执行，借助分布式账本去中心化、高安全性的特点实现可信共享中的可

信访问控制计算。利用分布式账本难以篡改、无法抵赖的特性实现授权记录的公开可追溯。通过采用分布式账本技术，基于联盟链的思想，在共享数据的多个部门、系统之间，通过共识机制，建立数据互操作的权限管理和访问控制机制：采用智能合约，基于数据资源目录，定义多个主体对不同数据的操作权限，并在数据互操作过程中进行授权和监管；采用分布式账本、存储管理权限定义的智能合约以及数据互操作历史记录的方式建立情报大数据可信共享机制，从而打消各情报主体对参与共享的各种顾忌，将情报数据在使用者和所有者之间有效共享，发挥其价值，释放其“数字生产力”。

情报数据提供方对外提供数据时需要提供三部分功能：用于外部访问者获取数据相关信息的数据发现服务、用于访问控制计算的授权计算合约以及数据仓库服务和访问验证合约。数据访问者可以通过数据发现服务获取情报数据的描述性信息及其MD5摘要，之后通过分布式账本网络向授权计算合约发送访问授权申请，获得合约许可后凭合约授权向数据仓库服务请求获取相应的数据资源。数据仓库服务向数据访问者传输加密数据，获得访问者确认回执后，将解密密钥和签名回执上传至智能合约完成密钥验证过程，实现对数据访问过程及其内容的存证。

3.3 情报数据授权访问流程

情报数据可信共享授权过程如**图9**所示。数据访问者试图获取具体数据资产时，应首先通过调用授权计算智能合约申请访问授权，该过程中访问者将其主客体信息、被访问数据内容及MD5摘要等相关信息通过其公钥签名后打包为一次符合该合约要求的调用，并在分布式账本网络中

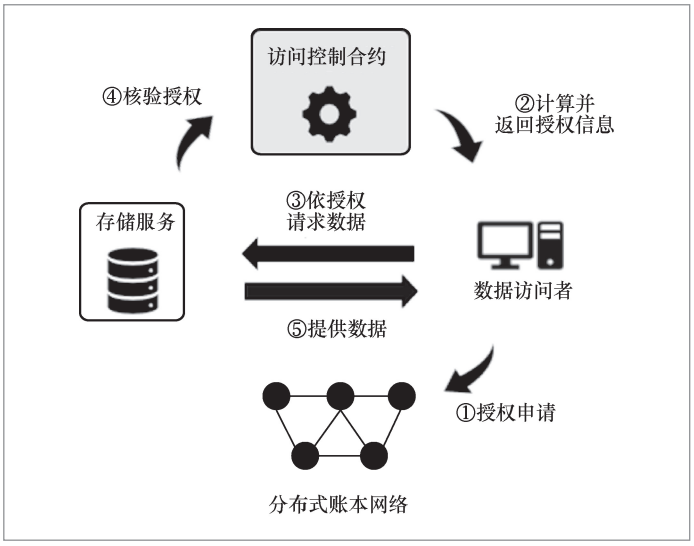

图 9 授权访问流程

以该次调用生成，并广播该交易事务。智能合约授权验证相关内容，包括授权策略等可被数据访问者调用的功能接口，在收到该授权申请事务后，对该次调用进行访问控制授权计算，完成授权计算后，合约将计算结果打包生成另一笔目的地为申请访问者（也即该次执行调用者地址）的事务记录。该事务记录不包含价值转移，但蕴含了对访问操作的授权信息，故可认为它是一条经过分布式账本网络存档确认的授权记录。

数据访问者通过分布式账本网络获取到智能合约的授权（事务）记录后，即可访问数据仓库服务并请求相应数据资源，访问者请求资源时，将请求操作同智能合约授权记录所在区块编号及其所在事务哈希值等信息一并发送至数据仓库服务。数据仓库收到访问请求后，根据访问者提供的授权记录所在区块等信息查验其授权记录。数据仓库在验证其授权记录之后，即可向数据访问者提供数据，并向智能合约发送该次授权已被执行的信息，合约收到授权被执行的信息后，生成一条目的地为数据访问者的账户地址，内容记载着授权记录与该次授权记录相应的执行信息的执

行记录，事后通过验证记载于区块内的事务记录和相应的执行记录，即可验证该次共享中曾进行过的情报数据访问控制授权记录。

4 系统实验与分析

4.1 情报数据检索实验分析

在情报对象管理系统中，情报数据对象的检索发现是系统的核心功能之一。其中数据对象的检索耗时是保证情报数据对象管理和共享的关键步骤，也是制约情报共享效率和可靠性的重要因素。随着情报数据对象的数量不断增加，数据对象的检索耗时成为制约情报对象管理系统效率和可靠性的重要因素。因此，情报对象管理系统中对象检索的低耗时可以提高情报共享系统的效率和准确性，从而促进实现情报共享的可信性和有效性。

由于单次检索的响应耗时较低不便于观察与统计，因此本文以1 000次检索的总耗时为基准，记录其随系统中DO数量规模的变化情况，实验结果如图10所示。从实验结果中可见，在基于DOA的情报对象管理系统中，考虑到网络因素的影响，情报数据对象的检索耗时总体在可接受范围内，随系统中对象规模的增大而变长，但增长幅度仍在可接受范围内。考虑到情报对象管理系统的结构以及网络因素的影响，笔者认为该系统能够满足后续情报数据对象共享的基础需求。

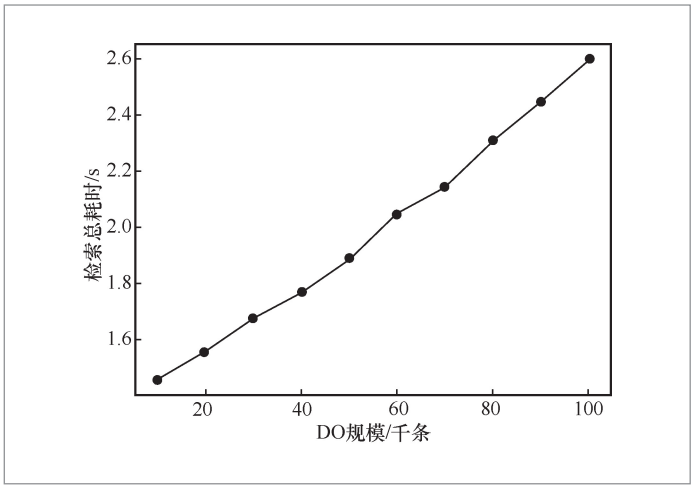

图 10 检索响应实验结果

4.2 情报数据访问请求实验分析

在情报对象管理系统中，每秒可处理的请求数量是衡量系统性能的重要指标之一。在协同研究工作的模式，随着用户数量的增多，系统需要能够及时响应每一位用户的请求以保证用户体验，进而保证情报对象的共享效率。因此，系统需要能够高效地处理和响应这些情报对象请求。情报对象管理系统的高并发处理能力和高效率能够促进实现更加高效、稳定、可靠的情报数据对象共享。

随着系统内DO规模的增大，系统内部针对特定DO的查询耗时会随之增加，并且延长情报对象的获取耗时，进而影响多位用户同时获取情报对象时的效率。本文选择每秒可处理的DO获取请求数量为基准，记录其随系统中DO数量规模的变化情况，实验结果如图11所示。从实验结果可知，系统每秒可处理的最大DO访问请求数量随系统中DO数量规模增大而减小，在最初的DO数量规模变化中，每秒可处理的访问请求数量降低的幅度相较于后期更大。当DO规模达到5万条时，其下降幅度逐步减小，全体的每秒可处理最大访问请求数量在可接受范围内，能够保证多用户的使用体验和访问效率。

在情报对象管理系统中除了大量存在的文本对象外，还可能包含其他形式的数据形成的数字对象。针对这些不同类型的

数据及其数字对象，本文选取了文本、图片、音频3种类型的DO进行访问响应实验，实验结果如**图12**所示。由实验结果可知，图片和音频类型的DO的请求耗时相较于文本类型明显更长，这是由图片与音频数据的大小导致的，通常情况下文本信息占用的存储空间是最小的。此外3种类型DO的请求耗时都会随着仓库内DO的规模而增加，这与前文的实验结果相符。

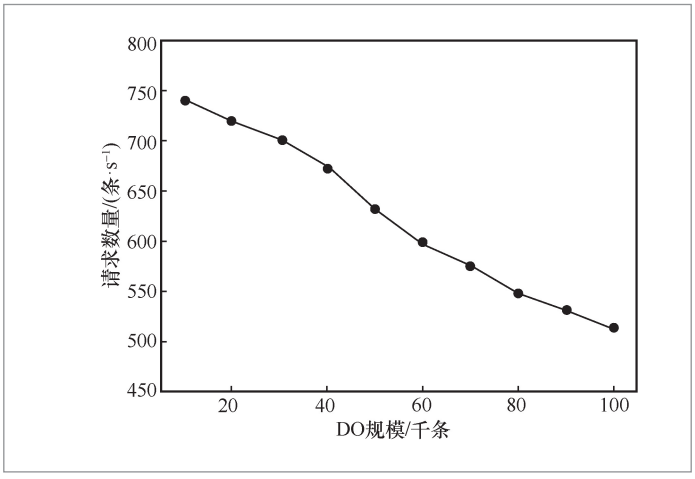

图 11 系统吞吐量实验结果

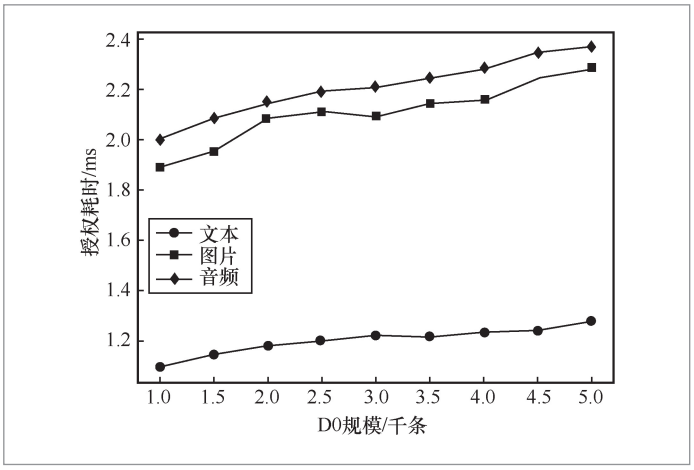

图 12 不同种类对象请求耗时

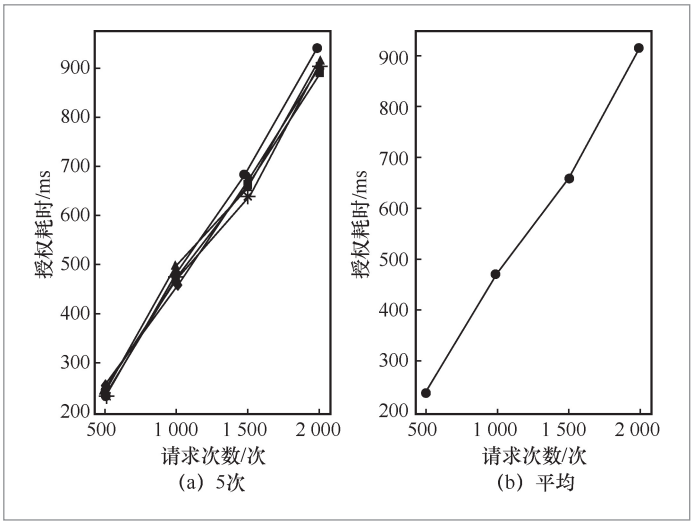

图 13 授权响应耗时实验结果

4.3 授权请求响应实验分析

系统中情报数据对象的授权耗时是指用户请求访问某一情报数据对象时，系统进行访问授权所需的时间。这个授权时间的长短，对情报数据对象的共享具有非常重要的意义。授权时间过长会严重影响情报数据对象的共享和流通效率。因此，在情报对象管理系统中，授权时间的缩短是一项重要的改善目标。通过记录并分析系统中授权耗时的变化，可以为系统性能的优化提供重要的指导。同时，对于情报对象共享，授权时间的缩短也能够提高共享效率，促进情报对象的快速可信流通，进而推动情报研究工作的进展。

本文分5次进行情报对象的授权请求实验，分别记录在不同授权请求次数下的授权耗时，实验结果如**图13**所示。由实验结果可知授权请求的总耗时随着请求数量的增加呈线性增加，总体保持稳定增长态势。此结果说明系统内的授权操作有一个较为恒定的处理耗时，且能够平稳地处理授权操作，不受请求数量的影响。这与情报对象授权操作的性质保持一致，所有情报对象的授权操作进行的步骤一样，其消耗的时间也大致相同。这也意味着可以通过增加服务器的硬件配置等方法对外置条件进行改善，进而缩短授权时间，提高授权效率。

在考虑到授权请求的耗时后，整个数字对象的访问请求耗时定会有所增加，其总耗时应为访问请求耗时与授权请求耗时之和。本文针对添加访问控制后的DO访问耗时进行实验，分5次进行情报对象的请求总耗时实验，分别记录在不同请求次数下的访问总耗时，实验结果如**图14**所示。由实验结果可知，加入访问控制之后的DO访问总时长符合预期，与上述实验对比可知，其大致为两者请求耗时的总和。

5 结束语

情报数据具有流通性、多源数据融合和多方参与主体的特性[4]，在数据采集生产、汇聚融合、挖掘利用的过程中，数据生产者、使用者、监管者等各参与主体需要在兼顾各方权利、责任和利益的前提下充分发挥数据价值。本文探讨了多主体情报研究协同环境下数据可信共享技术，构建的原型系统可为数据接入、交互、管理、和综合利用提供基础服务支撑。在不改变数据所有权和管理模式的前提下，在分布的数据网络节点之间，建立按需服务的数据共享模式，打破信息孤岛的壁垒，融合汇聚发现有价值的信息并形成可以指导行动的知识，最终进入数据驱动的决策应用场景。

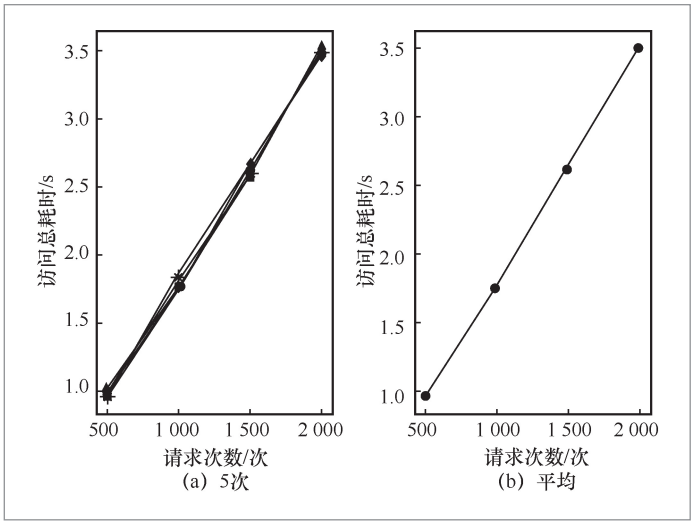

图 14 访问 DO 总耗时

参考文献:

[1] XIA X Y, CHEN F F, HE Q, et al. Online collaborative data caching in edge computing[J]. IEEE Transactions on Parallel and Distributed Systems, 2021, 32(2): 281–294.

[2] 祝振媛, 李广建. 多视角下的情报分析模型研究综述[J]. 图书情报工作, 2019, 63(19): 136–147.
ZHU Z Y, LI G J. Review on the research of information analysis models under multiple perspectives[J]. Library and Information Service, 2019, 63(19): 136–147.

[3] QI Y N, QIN X M, SUN D H, et al. Cloud-blockchain fusion system for secure and trusted sharing of domain. data[J]. Journal of Communication University of China(Science and Technology), 2022, 29(2): 9–18.

[4] 戴国强. 推进竞跑阶段的创新情报研究[J]. 情报学报, 2019, 38(8): 771–777.
DAI G Q. Intelligence studies for innovation in the new era[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(8): 771–777.

[5] 罗冲凌, 韩妍娜. 基于碎片化分析工具的开源情报研究新范式[J]. 国防科技, 2022, 43(4): 24–29.
LUO C L, HAN Y N. A new paradigm for open-source intelligence research based on an analytical tool for fragmented information[J]. National Defense Science & Technology, 2022, 43(4): 24–29.

[6] XIA X Y, CHEN F F, HE Q, et al. Online collaborative data caching in edge computing[J]. IEEE Transactions on Parallel and Distributed Systems, 2021,

32(2): 281–294.

[7] 金澈清, 钱卫宁, 周敏奇, 等. 数据管理系统评测基准: 从传统数据库到新兴大数据[J]. 计算机学报, 2015, 38(1): 18–34.
JIN C Q, QIAN W N, ZHOU M Q, et al. Benchmarking data management systems: from traditional database to emergent big data[J]. Chinese Journal of Computers, 2015, 38(1): 18–34.

[8] GUO H, ZHANG Z F, XU J, et al. Accountable proxy re-encryption for secure data sharing[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 18(1): 145–159.

[9] NAIMI A I, WESTREICH D J. Big data: a revolution that will transform how we live, work, and think[J]. American Journal of Epidemiology, 2014, 179(9): 1143–1144.

[10] 付熙雯, 郑磊. 开放政府数据的价值: 研究进展与展望[J]. 图书情报工作, 2020, 64(9): 122–132.
FU X W, ZHENG L. The value of open government data: insights from literature and a research agenda[J]. Library and Information Service, 2020, 64(9): 122–132.

[11] JÖRG B, HÖLLRIGL T, SICILIA M Á. Entities and identities in research information systems[C]//Proceedings of 11th International Conference on Current Research Information Systems. [S.l.:s.n.], 2012.

[12] 李善青, 郑彦宁, 赵辉, 等. 大数据背景下科学元数据的重要问题研究[J]. 科技管理研究, 2019, 39(18): 184–188.
LI S Q, ZHENG Y N, ZHAO H, et al. Study on key problems of scientific metadata under the background of big data[J]. Science and Technology Management Research, 2019, 39(18): 184–188.

[13] KAHN R, WILENSKY R. A framework for distributed digital object services[J]. International Journal on Digital Libraries, 2006, 6(2): 115–123.

[14] SHARP C. Overview of the digital object architecture(DOA)[Z]. 2016.

[15] Cross-Industry Working Team. Managing access to digital information: an approach based on digital objects and stated operations[Z]. 1997.

[16] The dona foundation[Z]. 2018.

[17] SUN S X. Internationalization of the handle system – a persistent global name service[Z]. 1998.

[18] KOPONEN T, CHAWLA M, CHUN B G, et al. A data-oriented (and beyond) network architecture[J]. ACM SIGCOMM Computer Communication Review, 2007, 37(4): 181–192.

[19] REILLY S, TUPELO-SCHNECK R. Digital object repository server: a component of the digital object architecture [J]. D-Lib Magazine, 2010, 16(1/2).

[20] 范吉立, 李晓华, 聂铁铮, 等. 区块链系统中智能合约技术综述[J]. 计算机科学, 2019, 46(11): 1–10.
FAN J L, LI X H, NIE T Z, et al. Survey on smart contract based on blockchain system[J]. Computer Science, 2019, 46(11): 1–10.

[21] WU K D, MA Y, HUANG G, et al. A first look at blockchain-based decentralized applications[J]. Software: Practice and Experience, 2021, 51(10): 2033–2050.

作者简介

彭泰(1998–), 男, 北方工业大学信息学院硕士生, 主要研究方向为数字对象体系。

孙晶(1968–),女,北方工业大学信息学院副教授,中国计算机学会(CCF)会员,主要研究方向为软件体系结构。

陈旭涧(1998–),男,北方工业大学信息学院硕士生,主要研究方向为数据分析和区块链技术。

周纤(1995–),女,博士,军事科学院军事科学信息研究中心助理研究员,主要研究方向为数据处理和分析。

叶宇铭(1990–),男,军事科学院军事科学信息研究中心工程师,主要研究方向为大数据处理技术、信息服务。

白晓颖(1973–),女,博士,军事科学院军事科学信息研究中心研究员、博士生导师,CCF高级会员,主要研究方向为计算机软件。

收稿日期: 2023-02-28
通信作者: 周纤, zhouxian@alumni.sjtu.edu.cn