面向智能体网络的去中心化身份体系综述

薄兆一[1]，王旭[1]，张栋[1]，李融[1]，卓书果[2]，陈慧慧[2]，任奎[2]
（1. 中央网信办数据与技术保障中心, 北京市 100048；2. 浙江大学 区块链与数据安全全国重点实验室，浙江省 杭州市 310027）

摘要：去中心化标识（DID）作为一种由用户自主控制的新型身份体系，为应对智能体（Agent）大规模、动态化及跨域协同带来的身份管理挑战提供了创新解决方案。本文系统探讨了面向智能体的DID技术，首先分析了传统互联网的通信模式和以中心化认证为核心的标识体系在智能体场景下面临的可扩展性、移动性、跨域互信及隐私保护等瓶颈。其次，本文梳理了数字身份技术从中心化到去中心化的演进脉络，并重点剖析了基于区块链账本的DID架构及其在增强通信安全、实现细粒度动态权限管理以及保障数据主权与隐私合规等方面的核心价值。文章进一步分析了当前DID在性能扩展、跨链互操作性及法律合规治理等方面存在的挑战，并展望了未来在构建智能体信誉体系、实现跨域信任传递及设计面向监管的身份架构等方向的发展路径，旨在为构建安全、可信、自主可控的智能体网络身份基础设施提供理论参考与实践指引。

关键词：分布式数字身份；可验证凭证；智能体；标识体系；区块链
中图分类号：TP309
文献标志码：A
doi：10.11959/j.issn.1000

Decentralized Identity Systems for Agent Networks: A Review

Bo Zhaoyi[1], Wang Xu[1], Zhang Dong[1], Li Rong[1], Zhuo Shuguo[2], Chen Huihui[2], Ren Kui[2]
1. Data and Technical Support Center of Cyberspace Administration of China, Beijing 100048, China
2. The State Key Laboratory of Blockchain and Data Security, Zhejiang University, Hangzhou 310027, China

Abstract: Decentralized Identifiers (DID), as a novel user-controlled identity system, offer an innovative solution to the identity management challenges posed by large-scale, dynamic, and cross-domain collaboration of agents. This paper systematically explores DID technologies for agents. It first analyzes the bottlenecks encountered by the traditional Internet identity paradigm and centralized authentication in agent scenarios, including scalability, mobility, cross-domain mutual trust, and privacy protection. Second, the paper traces the evolution of digital identity technologies from centralization to decentralization, and conducts an in-depth examination of the blockchain ledger-based DID architecture, highlighting its core value in enhancing communication security, enabling fine-grained dynamic permission management, and safeguarding data sovereignty and privacy compliance. The paper further analyzes the current challenges of DID in terms of performance scalability, cross-chain interoperability, and legal compliance governance, and looks forward to future development directions such as building agent reputation systems, realizing cross-domain trust transfer, and designing regulation-oriented identity architectures. This paper aims to provide theoretical references and practical guidance for building secure, trusted, and autonomous network identity infrastructures for Agents.

Key words: Decentralized Identifier, Verifiable Credentials, Agent, Identity System, Blockchain

0 引言

随着人工智能与分布式系统技术的快速发展，智能体（Agent）正从单一应用组件演进为具备自主决策与协同能力的网络主体，逐步广泛应用于自动驾驶、物联网、数字服务及复杂信息系统等场

收稿日期：XXXX–XX–XX；修回日期：XXXX–XX–XX
通信作者：卓书果，shuguozhuo@zju.edu.cn

景。在这一过程中，网络中的参与主体不再仅限于“人—应用”的交互模式，而是逐渐扩展为“人—智能体—智能体”的多主体协同体系。其中，人作为决策与价值的最终承担者，通常通过配置、训练或授权的方式驱动智能体行为，而智能体则在一定范围内具备自主执行任务与参与交互的能力。随着智能体规模与自治程度的不断提升，如何在网络中有效区分不同类型主体（如人、智能体、主机及应用程序），建立清晰、可信的身份边界，实现适应智能体特点的授权委托与信任传递，成为支撑复杂系统运行的基础性问题[1][2]。

在传统互联网体系中，网络架构设计的核心长期聚焦于主机互联与数据传输效率，身份问题并非其首要关切。然而这并不意味着传统互联网完全没有身份机制，而是这些机制以分层、分散的方式嵌入在各个协议层次之中：IP负责网络层寻址，DNS负责名称解析，TLS/X.509负责服务器认证，OAuth/OIDC负责应用授权，IAM负责组织身份治理，SPIFFE/SPIRE等则负责工作负载身份管理。这种分层设计使得各类主体的身份依赖于不同的外围系统进行补充定义，缺乏对"身份"与"标识"的统一原生抽象。在以人为中心、规模相对可控的应用环境中，这种分散的身份机制尚能满足需求；但随着智能体逐渐成为独立参与网络交互的实体，传统互联网中各层身份机制碎片化、缺乏跨层统一协调的设计范式开始显现出明显局限。特别是在跨域协同、自动化决策及高频交互场景下，缺乏统一、可验证且可扩展的原生身份体系，已成为制约系统安全性与协同效率的重要因素[3]。

与此同时，各国相继出台的数据安全与隐私保护法规，对数字身份的安全性、可控性与合规性提出了更高要求。在此背景下，构建面向智能体的可信身份体系，逐渐从“可选能力”转变为“基础设施需求”。去中心化标识（Decentralized Identifier，DID）作为一种新型身份管理范式，通过将标识控制权赋予主体自身，并结合分布式账本与密码学机制，实现身份的自主生成、可信解析与安全验证、授权和审计，为构建跨域互信与隐私保护兼顾的身份体系提供了新的技术路径。

基于上述背景，本文围绕面向智能体的去中心化标识技术展开系统研究。首先，从智能体网络通信特征出发，分析传统互联网体系在缺乏原生身份设计背景下面临的关键挑战；其次，梳理数字身份技术由中心化向分布式演进的发展脉络，并对比不同标识体系的技术特征与适用边界；在此基础上，重点探讨DID的结构模型、解析机制及其主流技术路线，深入分析其在通信安全、访问控制、授权委托与信任传递、数据主权保护等方面所带来的核心价值。同时，本文还对当前DID在性能扩展、跨域互操作及治理合规等方面面临的问题进行讨论，并展望未来在信任体系构建、跨域协同及监管适配等方向的发展趋势，为构建安全、可信且可持续演进的智能体身份基础设施提供参考。

1 智能体网络标识挑战与身份技术演进

1.1 智能体演进与网络通信新需求

智能体系统正呈现出多样化与规模化发展趋势，涵盖软件代理、机器人、无人机、物联网终端以及基于大模型的虚拟代理等多种形态。随着这些实体逐步成为网络中的“主动参与者”，网络通信从传统的人机交互模式演进为以智能体为核心的多主体协同模式，对标识与身份体系提出了系统性的新要求，也对网络体系结构、资源组织形态带来新变化。

从网络体系结构角度来看，传统互联网的设计重点在于数据传输路径与连接管理，而并未对“谁在通信”这一问题提供统一的身份表达机制。主机、应用程序与用户之间的关系通常是临时绑定的：主机作为通信载体，应用程序作为功能实现单元，而用户身份则依赖于上层业务系统进行管理。这种松散耦合的身份模型在智能体环境中逐渐失效——智能体既可能运行于不同主机之上，也可能在多个应用之间迁移或重构，其行为主体性无法通过传统网络结构中的位置或连接关系进行唯一刻画。

在资源组织形态方面，智能体呈现出显著的动态性与弹性特征。软件智能体及容器化服务可按需实例化并快速撤销，其生命周期高度短暂且不可预测。从本质上看，智能体与云原生环境中的动态工作负载（Workload）、服务实例及自动化代理高度相似——它们同样面临身份动态分配、生命周期短暂以及跨环境迁移等挑战。传统互联网中虽已出现专门面向工作负载身份的标准（如SPIFFE/SPIRE），为微服务和容器化服务提供了运行时身份认证能力，但这类机制主要面向静态或半静态的

服务实例，难以覆盖智能体所具备的自主决策、跨域协同及高度动态部署等特性。在这种环境下，传统互联网中依赖连接与位置关系的通信机制难以提供稳定、持续的主体标识能力，从而影响通信双方对彼此身份的一致认知。其次，智能体的跨环境迁移能力进一步加剧了身份不稳定问题。无论是运行于云边协同环境中的软件智能体，还是具备物理移动能力的机器人与无人系统，其运行上下文均可能频繁变化。传统互联网中的连接状态往往与具体网络环境绑定，一旦环境发生变化，通信关系需要重新建立，这不仅增加了系统开销，也使得身份连续性难以维持，进而影响安全机制的有效性。再次，从身份表达与信任机制角度来看，传统互联网缺乏统一的主体标识框架。不同组织通常构建各自独立的身份管理体系，导致身份标准不一致、信任边界割裂。在多组织、多域协同的智能体网络中，这种割裂状态显著增加了认证与授权的复杂度，使得跨域协同成本高昂且难以规模化扩展。此外，智能体通信模式正逐步由连接导向转向消息驱动。大量智能体通过消息队列、中继网络等方式实现解耦交互，通信双方之间不再依赖持续连接关系。在此模式下，安全机制需要从“连接级安全”转向“消息级可信”，即通过对消息本身进行签名、加密与验证来建立信任。这进一步凸显了独立于网络连接之外的身份标识的重要性。与此同时，边缘设备与物联网节点在计算与存储能力方面存在显著约束，使其难以承载复杂的传统身份管理机制。在云计算与多租户环境中，同一服务名称往往对应多个动态变化的智能体实例，传统互联网缺乏跨生命周期的一致标识能力，给授权审计、行为追踪与责任界定带来了现实挑战。

综上所述，传统互联网以连接与位置为中心的通信模型难以适应智能体网络所呈现的动态性、自治性与跨域协同特征。随着智能体逐步成为网络中的基本单元，构建具备全局唯一性、可验证性与去中心化特征的新型标识体系，已成为支撑智能体互联网发展的迫切需求，也是网络体系结构的必然演进方向。

1.2　数字身份标识技术的发展脉络

数字身份管理技术经历了由中心化向分布式演进的多个阶段。早期互联网环境中，身份主要依赖设备层标识（如MAC地址、IP地址）以及基于用户名/密码的人机交互式认证机制，此阶段的身份信息通常由各类应用系统分别维护，呈现出明显的“孤岛化”特征，在安全性、可扩展性及跨域互认方面均存在不足。20世纪90年代以来，随着网络安全需求的提升，公共密钥基础设施（PKI）及X.509证书体系逐步成为主流，通过引入证书颁发机构（CA）实现公钥的可信分发，从而为网络通信提供了基础的安全保障。尽管PKI通过信任链机制增强了身份验证能力，但其本质上仍依赖中心化信任根，存在单点故障及信任集中等问题。

随着互联网服务形态的复杂化及跨组织业务协同需求的增强，联邦身份管理体系逐渐兴起，以SAML、OAuth及OpenID Connect（OIDC）等为代表的标准通过引入第三方身份提供者（IdP），实现了跨域单点登录与统一认证，有效提升了用户体验和系统互操作能力。然而，该类模式依然建立在中心化身份管理机构之上，用户身份数据的控制权仍掌握在平台或服务提供方手中，在隐私保护及数据主权方面存在一定局限。

随着云计算与微服务架构的普及，面向服务实例与自动化进程的工作负载身份（Workload Identity）逐渐成为身份技术体系中不可忽视的重要层次。以SPIFFE（面向所有人的安全生产身份框架）及其参考实现SPIRE为代表的工作负载身份标准，为容器化服务、微服务及自动化代理提供了基于密码学的运行时身份认证机制，使得服务间通信无需依赖静态凭据即可完成相互认证。这一技术方向与智能体的需求高度契合——智能体本质上是一类高度动态的工作负载，具备自主执行任务、动态实例化及跨环境迁移等特征。然而，现有工作负载身份标准主要面向相对可预测的微服务场景，在支持智能体的自主决策、多域信任建立及隐私保护等方面仍存在不足，难以完整覆盖智能体复杂的协同与交互需求。

近年来，区块链技术的发展推动了数字身份模型的进一步演进。以比特币、以太坊为代表的区块链系统中，基于公私钥生成的链上地址可作为去中心化环境下的身份标识，具备一定的抗篡改与不可伪造特性。然而，此类标识主要侧重于账户层面的唯一性表达，缺乏对身份属性、语义信息及生命周期管理（如更新与撤销）的有效支持。针对上述不足，自我主权身份（Self-Sovereign Identity, SSI）

理念逐渐提出，其核心在于将身份控制权从中心化机构转移至身份主体自身，实现身份数据的自主生成、自主管理与自主披露[3]。在此背景下，W3C发布了去中心化标识符（Decentralized Identifiers, DID）规范，定义了一种无需依赖中心化注册机构、由主体自主生成并可通过密码学机制证明控制权的新型标识体系。同时，可验证凭证（Verifiable Credentials, VC）为身份属性的表达与交换提供了标准化数据结构，使得各参与方在无需建立全局信任前提下即可完成身份验证与数据共享。目前，围绕DID与VC已形成多种技术实现路径，包括基于Hyperledger Indy的Sovrin网络、Veres One身份网络以及基于以太坊的did:ethr方法等[4]。此外，相关国际标准组织亦在持续推进该领域的规范化进程，例如IEEE P2958正在制定面向物联网场景的去中心化身份与访问管理框架，ISO亦在推进基于分布式账本的身份系统参考架构研究[5][6][7][8][9]。

总体来看，数字身份技术正由以中心化信任为核心的传统模型，逐步演进至强调分布式架构、工作负载身份、主体可控的新型模式。不同阶段技术在安全性、互操作性与自主性方面各有侧重：PKI体系奠定了可信通信基础但依赖中心化信任根；联邦身份体系提升了跨域互通能力但受制于身份提供者；区块链地址实现了一定程度的去中心化标识但缺乏语义表达能力；而DID/VC模型则在此基础上进一步融合可信性与自主性，也解决了语义表达、授权等工作负载的载体问题，为构建面向智能体的下一代标识体系提供了重要技术支撑。

2 现有标识机制局限与智能体标识需求分析

2.1 现有标识机制比较

现有传统标识机制种类繁多，其技术特征与信任锚点各异，适用于不同的网络层级与应用场景。首先，基于“IP地址+端口号（Port）”的组合标识是当前网络通信中最基础的寻址与逻辑识别方式。这种基于套接字（Socket）的标识机制具有全球寻址能力与自动化分配的优势，能够精确区分同一物理宿主机上的不同软件实例，适用于基于端到端的直接通信与静态路由。然而，该标识本质上属于位置标识而非身份标识，本质是应用层与传输层间的一个抽象层，无法承载主体的语义属性，且在NAT（网络地址转换）穿透、动态IP分配以及微服务高频扩缩容环境下，IP+Port的映射关系极易发生偏移，难以维持长效稳定的身份映射与安全认证。用户名/密码认证是最早的应用层身份方案，部署简单、易于理解，广泛用于Web登录和内部系统。然而，它依赖弱认证凭据，难以抵御钓鱼和暴力破解攻击，对智能体场景来说强度不足。X.509/PKI技术方案通过数字证书和证书颁发机构（CA）提供了加密级的身份验证，为HTTPS等通信提供安全信任基础。PKI模式为Web和物联网设备通信提供了可靠的加密机制，但由于CA中心化，其存

表1 现有标识机制的技术特性和优劣

标识机制	技术特征	适用场景	优势	缺点
IP+Port	传输层套接字标识，支持逻辑进程区分；在公网环境下具备寻址能力	端到端直接通信、静态服务寻址、网络层级流量转发	标准化程度高，实现寻址与逻辑识别的统一	语义贫乏，难以应对动态扩缩容，缺乏原生安全验证机制
用户名/密码	应用层凭证，基于共享密码	传统网络服务登录	部署简单，用户熟悉	安全性弱、跨域互认能力差、难以支持自动化主体大规模管理
X.509/PKI	数字证书与CA体系，基于公钥基础设施	Web/物联网安全通信	加密传输，认证级别高，标准成熟	CA中心化单点信任；证书管理复杂
OAuth/OIDC/SAML	联合身份协议，引入授权服务器与Token机制	Web单点登录、企业联邦身份、移动应用	跨域单点登录便捷	依赖IdP，隐私与控制权受限，协议复杂
区块链地址	匿名公钥地址，去中心化账本托管	加密货币、去中心化应用	去中心化，防篡改	无身份属性；难以撤销
硬件序列号	嵌入式设备唯一ID	设备防伪	固有唯一性，与设备物理绑定	可被复制、模拟或中继攻击，存在软硬件层面的仿冒风险

在信任单点和证书管理负担，且对资源受限的设备和动态环境支持不佳[10]。OAuth/OIDC/SAML 等联邦身份协议引入了第三方身份提供者(IdP)，使用户能够跨服务单点登录，提升了跨域互操作性和用户体验。这类方案适用于社交登录、企业单点登录等场景，但其身份管理依赖大型平台或组织，缺乏用户身份数据自主权，不适合高度自主或多租户的智能体环境。区块链地址（如比特币、以太坊地址）是去中心化的标识符，无需中央机构便可验证归属，且天生具有防篡改特性。它们在数字货币和区块链应用中广泛使用，但主要表达账户唯一性而非完整身份信息，不包含姓名或角色权限等属性，同时缺乏内置的撤销与更新机制。硬件 ID（如设备序列号）提供了设备级的唯一性，对于设备注册和防盗有用，但容易被复制或以软件手段伪造，且通常未与加密机制直接关联。表 1 对现有主要标识机制的技术特性和优劣进行了比较。

2.2　智能体标识需求特征

面向智能体的标识体系需满足区别于传统互联网主体（用户或服务器）的多维需求特征。从规模与生命周期角度看，智能体系统呈现出大规模部署、按需生成与快速销毁的典型特征，这要求标识体系具备良好的可扩展性与弹性能力，能够支持海量实例的快速注册、解析与注销过程，并尽可能降低对人工授权配置及中心化管理机制的依赖。从运行环境角度看，智能体普遍具备跨网络迁移能力，尤其是在无人机、移动机器人及边缘计算等场景中，其网络接入位置频繁变化，因而标识体系需实现与网络位置的解耦，以支持跨域、跨网络环境下身份的持续保持与安全通信[11]。从协同与信任机制角度分析，多智能体系统通常运行于多组织、多域的复杂环境中，即要区分出人、智能体、资源库的区别，又要明确之间的继承式关系，还要符合可验证、可传递、可控制的互操作信任框架。标识体系需提供统一的身份语义表达能力及跨域信任建立机制，从而降低认证与授权过程的复杂性。同时，随着智能体通信模式由传统的同步连接逐步向异步、消息驱动模式演进，标识体系还需支持面向消息级的身份认证与数据保护机制，以适应中继转发、离线交互等复杂通信场景。在隐私与数据保护方面，智能体在执行任务过程中往往涉及敏感数据交换，因此标识体系应支持最小化披露与可验证声明机制，以避免主体属性信息的过度暴露。此外，智能体运行环境呈现出显著的异构性特征，大量边缘设备与物联网终端在计算与存储资源方面存在约束，这要求标识体系在安全机制设计上具备轻量化特性，以降低计算与通信开销。在系统治理层面，随着智能体规模的持续增长及行为复杂性的提升，标识体系还需具备全生命周期的可追溯性与可审计能力，以满足监管合规及责任界定等现实需求[12]。

进一步而言，智能体虽然与微服务、容器等传统工作负载同样具有弹性部署、短生命周期和运行时身份需求，但二者在身份语义和信任边界上存在本质差异。传统工作负载通常由单一组织编排系统统一创建、调度和销毁，其身份多服务于服务间认证、集群内访问控制和运维审计，标识含义主要绑定于服务名称、命名空间或运行环境；智能体则可能代表人、组织或其他智能体执行具有目标导向的任务，具备自主决策、跨平台迁移、调用外部工具、递归委托以及生成内容和交易行为等能力。因此，智能体标识不仅要证明“该进程或实例是谁”，还需要表达“其代表谁、被授权做什么、能力边界是什么、委托链来自何处、行为责任如何追溯”等更丰富的语义。与 SPIFFE/SPIRE 等工作负载身份主要解决集群内服务身份不同，面向智能体的标识体系还需支持跨组织可验证凭证、权限衰减式委托、会话级或成对 DID、可撤销授权、最小披露以及与声誉和合规治理机制的结合。这种“主体身份—授权上下文—行为责任”一体化表达，构成了智能体标识区别于传统工作负载身份的独特性。

然而，现有传统互联网标识体系在智能体互联网关键需求方面普遍存在不足。具体而言，基于 IP 地址+Port 的网络层标识仅具备定位属性，缺乏身份语义表达与跨域能力；基于用户名/密码的认证机制在安全性与可扩展性方面存在明显缺陷；基于 X.509 证书的 PKI 体系在证书管理复杂性及中心化信任依赖方面存在瓶颈；基于 OAuth、OIDC 及 SAML 的联邦身份体系依赖中心化身份提供者，难以满足去中心化与自主控制需求；而区块链地址虽具备一定的去中心化特征，但在身份属性表达与生命周期管理方面仍存在不足。上述问题共同导致传统标识体系难以支撑智能体系统的复杂运行需求。为直观刻画相关需求与不足之间的对应关系，表 2

表2 智能体标识需求

智能体标识需求	传统互联网标识体系缺陷	潜在的智能体互联网 DID/VC 支撑机制
大规模扩展性	静态注册限制数量；CA 证书管理难以快速增量扩展	支持自动注册、分布式目录或区块链记录身份，以应对海量代理规模
动态寻址	缺乏即时生成唯一标识的机制(IP/用户名静态分配);NAT 问题复杂	每个实例可本地生成 DID，并视需要锚定至账本或通过轻量级目录发布
跨网络移动性	基于 IP 的会话在切换网络时中断；基于中心化 IdP 的认证不适用漫游场景	DID 允许主体跨网络保留相同标识；采用多路径密钥管理维护连接安全
跨域信任	不同域无共享身份根；缺少跨组织信任协议	去中心化标识和可验证凭证构建公开信任根，实现跨域身份互认
端到端消息级安全	TLS 依赖对等连接，不支持异步代理之间的逐条消息加密	DIDComm 等协议提供消息级别加密，确保任意中继信道中的数据安全
隐私与最小披露	认证时需暴露全部身份标识(如邮箱、手机号);联合登录追踪性强	可验证凭证支持选择性披露身份属性，保护隐私；DID 允许使用多个别名
资源受限适配	证书和握手开销大，不适合微控制器；用户名密码易被窃取	轻量级密钥对和简化协议(无证书签名)适配低功耗设备；对称加密替代复杂公钥
可审计与可追溯性	日志与应用日志分散；无统一身份标识跨体系关联	DID 可为实体提供可追踪的标识基础，便于关联日志与行为审计

主要对传统互联网标识体系与智能体互联网 DID 技术进行了系统性对比分析。

综上所述，智能体标识体系需要在可扩展性、可移动性、去中心化信任、隐私保护及轻量化等方面实现综合提升。传统标识机制由于在标识生成、信任建立以及身份管理模式上的结构性约束，难以从根本上满足智能体系统对“动态主体+跨域协同+自主控制”的统一需求。近年来提出的去中心化标识符（DID）通过将标识控制权赋予主体自身，并结合分布式账本与密码学机制，实现了标识生成、解析与验证过程的去中心化与可验证化。在该模型中，智能体无需依赖中心化注册机构即可自主生成全局唯一标识，并通过可验证凭证机制实现身份属性的安全表达与选择性披露，从而在保障隐私的同时支持跨域信任建立。此外，DID 体系通过将标识与网络位置解耦，使其天然适用于动态迁移与多网络环境，有效契合智能体的运行特征。因此，从体系结构与能力匹配角度来看，DID 不仅能够弥补传统标识体系的不足，还为构建面向智能体的统一身份基础设施提供了关键技术支撑，逐渐成为该领域的重要研究方向与发展趋势。

2.3 DID 标识结构与解析机制

去中心化标识符（DID）采用统一的 URI 语法结构，表示为"did:<method>: <methodSpecificId>"[8]。其中，<method> 为 DID 方法名称（DID Method），用于指示标识符所依赖的具体技术规范（如 indy、ethr、web 等）；<methodSpecificId>则为该方法体系下生成的全局唯一标识字符串，通常与密钥或分布式账本记录相关联。此外，DID 规范还定义了 DID URL 的扩展形式，允许在 DID 基础上附加路径（path）、查询参数（query）或片段（fragment），以引用 DID 文档中的特定资源（如某一验证方法或服务端点）。在主体关系方面，DID Subject 字段描述 DID 所标识的实体（可以是人、组织、设备或智能体等），而 DID Controller 字段描述被授权对该 DID 文档执行创建、更新与停用操作的主体，二者可以相同，也可以分离，以支持代理控制等委托场景。

每一个 DID 均对应一个结构化的 DID 文档（DID Document），用于描述与该标识符关联的公开验证材料和交互信息。根据 W3C DID 规范[8]，DID 文档可包含以下类型的验证关系（Verification Relationship），分别服务于不同的密码学用途：认证（Authentication），用于证明 DID Subject 对标识符的控制权；断言方法（Assertion Method），用于发布可验证凭证或声明；密钥协商（Key Agreement），用于建立加密通信信道；能力调用（Capability Invocation），用于授权特定操作的执行；能力委托（Capability Delegation），用于将操作权限转授予其他主体。每种验证关系可引用 DID 文档中定义的验证方法（Verification Method），其中包含密钥类型、公钥及控制器信息等。服务端点（Ser-

vice Endpoint）字段则用于描述与该DID主体进行交互的网络接口，例如消息服务、凭证存储或智能体通信端点等。需要特别指出的是，DID文档本身不应直接承载姓名、地址等敏感身份属性，此类信息通常通过可验证凭证（Verifiable Credentials，VC）在独立数据结构中加以表达，以实现身份属性的最小化披露与隐私保护。DID文档的表达格式以JSON-LD为主流，但规范并不强制要求唯一使用该格式，也支持纯JSON或CBOR等其他序列化形式。

如图1所示，在解析机制方面，DID的解析过程依赖于与特定DID方法对应的解析器（Resolver），其标准化输入为一个DID字符串，输出结果包含三个组成部分：DID文档本身（DID Document）、解析元数据（Resolution Metadata）以及文档元数据（Document Metadata）。解析元数据描述解析过程的状态信息，如错误类型、内容类型等；文档元数据则记录DID文档的创建时间、更新时间及停用状态等生命周期信息。解析过程首先根据DID中的<method>字段确定解析路径，继而从对应数据源中获取DID文档。对于基于区块链或有向无环图（DAG）结构的DID方法（如Hyperledger Indy、IOTA等），解析器通常通过访问相应的分布式账本网络，查询与目标DID相关联的链上记录，并从中提取和验证DID文档；对于基于Web基础设施的DID方法（如did:web），则通过标准HTTPS协议访问预定义路径以获取相应文档。在多方法共存的环境下，通用解析器（Universal Resolver）可对不同类型的DID进行统一解析并输出标准化结果，从而提升系统的互操作性与可扩展能力。

图1　DID文档解析

从身份验证模式来看，DID体系将身份信任根锚定于主体自身所控制的密钥及其对应的分布式记录之上，使主体能够在无需依赖中心化机构的前提下完成身份声明与控制权证明。DID文档中公开的验证方法与服务描述在功能上可支持类似于PKI的加密通信与身份验证，但其信任建立过程不再依赖单一中心节点，而是通过分布式机制实现。此外，DID体系通常结合消息级安全协议（如DIDComm）实现应用层的端到端加密与身份认证，在异步通信与多主体交互场景中表现出更强的适应能力[13]。相较于基于令牌的联邦身份模型，DID在架构上更加去中心化，更有利于支撑面向智能体的开放式网络生态。

3　基于DID的智能体标识架构与关键优势

3.1　面向智能体标识的DID主流技术路线

去中心化标识（DID）为智能体（自治软件或物联网设备）提供了无需中心化机构即可管理身份的机制。本节总结了面向智能体标识的DID主流技术路线，包括基于区块链账本、基于密钥/对等网络、基于域名等不同类型，对其原理、关键组件、代表实现、优缺点和典型场景进行比较，并讨论各路线在互操作性、标准化、安全隐私及可扩展性方面的挑战与机遇。

传统身份管理依赖中心化机构，难以满足智能体生态中自治、隐私和跨域互信的需求。W3C发布的DID标准将DID定义为“新型标识符”，可将标识主体与DID文档关联，使标识控制者无需任何中心化机构即可证明对身份的控制权。DID与可验证凭证（VC）结合，可支持加密安全、隐私保护的身份信息表达、授权委托与信任传递。在智能体网络中，DID可为每个实体分配全局唯一且自主可控的标识；智能体间通过DIDComm协议进行异步、端到端加密通信。DIDComm协议定义了面向去中心化标识的安全通信机制，通过端到端加密、消息认证及完整性保护等手段，实现了通信过程中的机密性与可信性。同时，该协议支持基于DID的伪匿名通信与选择性身份披露，在降低对中心化中介依赖的前提下，为智能体之间的可验证凭证与授权数据交换提供了安全传输基础。

3.1.1　基于区块链账本的DID

基于区块链账本的去中心化标识是当前自主管理身份体系中最具代表性的实现路径之一，其核心思想是利用区块链的分布式账本特性，为标识的生成、解析与状态管理提供可信基础设施。在该技术路线下，DID作为全局唯一标识符，由用户或智能体本地生成，并通过链上注册或锚定的方式完成标识与公钥的绑定，从而实现“标识—密钥—控制权”的统一。

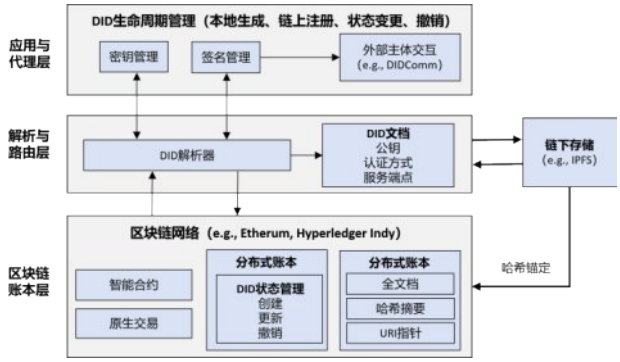

图2　基于区块链账本的DID架构

如图2所示，从技术实现角度来看，基于区块链账本的 DID 体系通常包含链上与链下协同的分层架构。底层为区块链网络（如 Ethereum、Hyperledger Indy 等），主要承担 DID 注册、更新与撤销等状态操作的持久化记录功能[14]。链上通过智能合约或原生交易结构维护 DID 文档的索引信息，典型做法包括：将 DID 文档全文存储于链上，或仅存储其哈希摘要与指针（如 URI），以降低存储开销并提升扩展性。中间层为 DID 解析与路由机制，即通过 DID Resolver 根据 DID 方法规范（DID Method）解析对应的链上记录，并获取最新的 DID 文档。该文档通常采用 JSON-LD 格式，包含公钥、认证方式、服务端点等关键元数据，用于支持身份认证与安全通信。上层则为应用与代理层（Agent），负责密钥管理、签名验证以及与其他主体之间的交互逻辑，例如结合 DIDComm 实现端到端加密通信，或基于 VC 完成凭证的签发与验证[15]。

在具体运行流程上，DID 的生命周期管理体现了区块链驱动的技术特点。首先，用户在本地生成公私钥对，并构造 DID 文档初始版本；随后将 DID 标识及其关联信息注册至分布式账本区块链，完成“创建（Create）”操作。当密钥轮换或服务端点发生变化时，可通过提交更新交易实现“更新（Update）”；若标识不再使用，则通过链上标记完成“撤销（Deactivate）”。整个过程遵循不可篡改与可追溯原则，任何状态变更均以时间序列形式记录在链上，从而形成完整的身份演化历史。此外，为提升性能与隐私性，部分实现采用链下存储与链上锚定相结合的方式，即将敏感数据保存在去中心化存储系统（如 IPFS），仅将其哈希值写入区块链，实现数据完整性校验与访问控制的分离[16][17]。

在技术特征方面，基于区块链账本的 DID 体系具有去中心化信任、抗篡改和高可验证性等显著优势。其一，系统不依赖单一身份提供者，用户对智能体的标识及其密钥拥有完全控制权；其二，区块链的共识机制保证了 DID 状态的全局一致性与不可篡改性；其三，通过密码学机制与链上记录的结合，可实现对身份声明与凭证的高效验证。同时，该技术路径也面临一定挑战，如链上存储成本较高、交易延迟影响实时性，以及隐私保护与公开可验证之间的权衡问题。总体而言，基于区块链账本的 DID 为构建可信、可验证且去中心化的智能体标识体系提供了重要支撑，是当前分布式数字身份研究与应用的重要方向之一。

3.1.2　基于P2P对等网络的DID

基于 P2P 对等网络的 DID 也不依赖区块链账本，而是通过点对点通信与分布式存储机制实现标识管理与解析。该方法以去中心化网络为基础，将标识的控制权与数据存储权进一步下沉至网络边缘节点（如用户代理或智能体），通过直接交互完成身份的创建、解析与验证，从而形成轻量化、高实时性的DID体系。

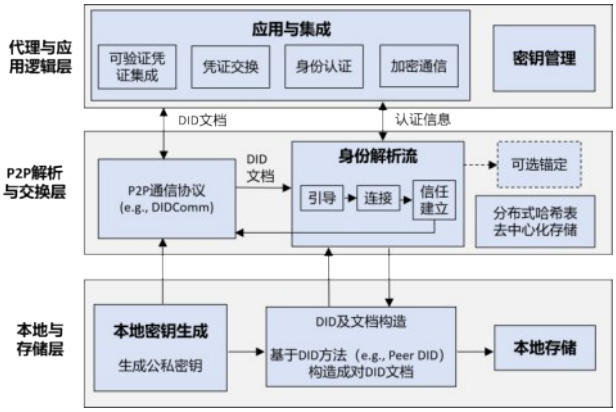

图3　基于P2P对等网络的DID架构

如图3所示，基于P2P网络的DID体系通常采用“本地生成+对等解析+可选锚定”的分层架构。首先，在标识生成阶段，用户或智能体在本地环境中生成公私钥对，并基于特定 DID 方法（如 peer DID）构造DID标识符及其对应的DID文档。该文档同样采用结构化格式（如 JSON 或 JSON-LD），包含公钥、认证方式及服务端点等信息，但与区块链方案不同，其默认不写入全局共享账本，而是保

存在本地或通过点对点方式分发给通信对手。其次，在标识解析与交换阶段，系统依赖P2P通信协议（如DIDComm）实现DID文档的直接传输与动态解析。当两个主体建立连接时，通常通过“引导（invitation）—连接（connection）—信任建立（trust establishment）”流程交换各自的DID及其文档，从而完成身份绑定。对于需要跨会话或多方复用的场景，可结合分布式哈希表（Distributed Hash Table，DHT）或去中心化存储网络实现轻量级索引，使得DID文档在一定范围内可发现与可检索。最上层则为代理与应用逻辑层，该层负责密钥管理、加密通信及与可验证凭证系统（VC）的集成，从而支持凭证交换、身份认证等功能。

在运行机制上，该体系强调“连接驱动”的身份使用模式。DID通常以成对（pairwise）形式存在，即每一对通信主体之间生成唯一的DID标识，从而避免全局标识带来的可关联性风险。在具体流程中，一方通过生成一次性邀请信息（包含临时DID及通信端点）发起连接请求，另一方接收后建立安全信道，并在后续交互中逐步交换长期DID及相关凭证信息。由于整个过程基于端到端加密和本地密钥控制，身份数据无需经过中心化服务器或全局账本即可完成验证与使用。此外，为提升可用性，部分实现允许将关键状态（如公钥哈希）选择性锚定至外部系统（如区块链或公共注册表），以增强持久性与可审计性。

在技术特征方面，基于P2P网络的DID体系具有低延迟、部署灵活及隐私保护能力强等优点，特别适用于智能体之间的动态交互与临时信任建立场景。然而，其局限性也较为明显，例如缺乏全局一致的解析机制、标识发现能力较弱，以及在大规模网络中的互操作性与可管理性仍有待提升。

3.1.3　基于域名/web网络的DID

基于域名/Web网络的DID是一类依托现有互联网基础设施（尤其是DNS体系与HTTPS协议）实现的DID技术路线，其核心思想是在不引入区块链或复杂P2P网络的前提下，利用域名所有权作为信任锚点，将标识解析与验证过程映射到成熟的Web架构之上。该方法通过将DID与域名空间进行绑定，使标识控制权与域名控制权保持一致，从而在兼容现有互联网体系的同时，实现一定程度的去中心化身份管理。

如图4所示，基于Web网络的DID体系通常采用“域名映射+标准解析+安全传输”的分层架构。底层依托域名系统（DNS）与传输层安全协议（HTTPS），作为标识解析与数据获取的基础通道。中间层为DID方法规范与解析机制，其中最典型的是基于Web的DID方法（如did:web）。在该方法下，DID标识符通常直接由域名派生，例如将域名转换为DID格式，并通过路径规则定位对应的DID文档资源。具体而言，解析器（Resolver）在接收到DID后，会将其转换为标准HTTPS请求路径，并通过安全连接获取目标服务器上托管的DID文档。该文档一般采用JSON-LD格式，包含公钥、认证方法、服务端点等信息，用于支持身份认证与后续交互。最上层则为应用与代理层，负责密钥管理、签名验证以及与其他主体之间的安全通信，例如结合DIDComm实现基于DID的加密消息交换，或集成VC支持凭证的签发与验证流程。

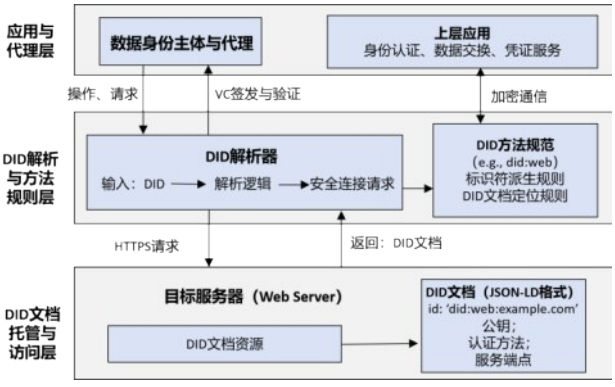

图4　基于域名/web网络的DID架构

在具体运行机制上，该体系强调“域名控制即身份控制”的映射关系。实体首先需拥有某一域名的管理权限，并在其Web服务器上部署对应的DID文档文件；随后，通过HTTPS服务对外提供该文档的访问接口。验证方在解析DID时，通过标准的TLS证书校验机制确认服务器身份，从而间接验证DID文档的可信来源。为增强安全性，部分实现还可结合DNSSEC或证书透明性机制，对域名解析与证书链进行进一步验证。此外，在密钥更新或服务端点变更时，仅需更新服务器上的DID文档即可完成状态变更，无需提交链上交易或广播至P2P网络，从而具备较高的灵活性与实时性。

在技术特征方面，基于域名/Web网络的DID体系具有实现简单、兼容性强及易于部署等优点，

能够充分复用现有互联网基础设施，降低系统接入门槛。然而，其去中心化程度相对有限，仍依赖DNS 体系及证书颁发机构等传统信任基础设施；同时，域名本身的注册与管理也依旧存在中心化控制的风险。表 3 对上述提到的三种 DID 主要实现路线进行了系统性地总结和对比。

3.2　DID 赋能智能体系统的关键价值

在众多的去中心化标识技术路线中，基于区块链账本的 DID 方案凭借其独有的分布式共识机制与全局一致性特性，已经成为了当前支撑智能体系统较具代表性的技术路线之一。相比于依赖现有域名体系的 did:web（存在中心化根域名风险）或纯点对点的 did:peer（缺乏全局可发现性与持久性锚点），基于区块链的 DID 能够提供跨信任域的全局一致性与身份存续性，确保智能体在复杂的开放环境交互中具备高度的自主性与可审计性。因此，本文重点聚焦于基于区块链账本的技术路线，深入剖析其架构优势及对智能体生态的赋能逻辑。

3.2.1　增强智能体网络通信安全与抗攻击能力

基于区块链账本的 DID 体系通过构建去信任化的公钥基础设施，为智能体间通信提供了高强度的安全保障。如图 5 所示，在该体系下，每个智能体均可持有唯一的链上 DID 标识，并与其对应的公私钥对绑定。智能体间的安全通信过程可描述为以下步骤：首先，智能体 A 生成或持有一个 DID，其密钥与 DID 文档锚定于分布式账本；其次，智能体 B 通过 DID 解析器获取 A 的 DID Document，提取其公钥及密钥协商信息。随后，双方基于所获公钥通过 DIDComm 协议完成连接握手，握手阶段依次交换例如 connection-request 与 connection-response 两类消息，其消息体字段中包含发起方 DID 标识、所支持的加密算法及密钥协商参数。连接建立后，后续业务消息采用加密消息封装格式传输，消息中包含消息标识符、发送方与接收方 DID、消息类型、时间戳等元数据字段；消息负载一般采用对称密钥加密，而会话密钥再通过接收方公钥进行封装，从而实现端到端机密性保护。同时，系统还可结合发送方密钥认证或数字签名机制，实现消息来源验证与完整性校验。该机制无需依赖中心化身份认证机构，即可在多智能体环境中实现可信、安全的点对点通信。在此基础上，可验证凭证（VC）进一步用于证明通信方的组织归属、能力声明或访问权限，并结合状态列表或撤销注册表对凭证的当前有效性进行核验。值得强调的是，DID 本身仅保证对方控制着与该标识对应的密钥，并不直接赋予对方可信性；通信主体是否可信，取决于 VC 的内容与签发机构、所处信任框架、声誉系统以及相应治理规则的综合约束。

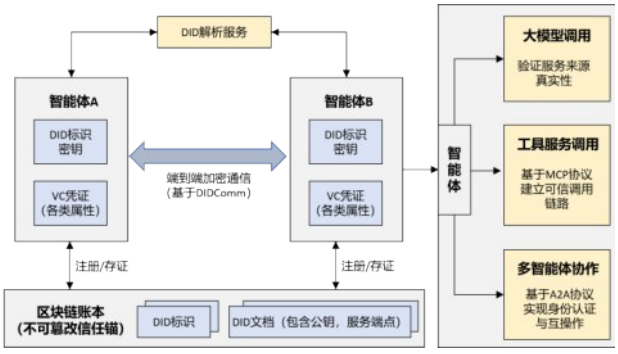

图 5　基于区块链账本的 DID 与智能体间交互流程示意

在当前主流智能体通信协议不断演进的背景下，DID 可作为底层身份与信任锚嵌入其中。例如，在基于模型上下文协议（MCP）的工具调用框架中[34]，智能体可通过 DID 对工具提供方进行身份解析与公钥获取，从而建立可信调用链路，DID 所承载的可验证公钥与服务端点信息，为此类动态信任建立提供了去中心化的身份锚点。在多智能体交互协议（A2A）中，不同智能体之间的协作、任务分发与结果回传同样依赖身份认证与消息

表 3　　DID 主要实现路线对比

路线	信任锚	优点	缺点	性能对比	适合智能体场景
基于区块链账本型	区块链/分布式账本	公共可验证、可审计	成本高、有一定延迟	延迟通常为秒级；吞吐量一般为 10^2 – 10^3TPS	跨组织协作、公共身份、监管审计
基于 P2P 对等型	双方交换的密钥	隐私好、低延迟	不易全局发现	时延通常低于 50 ms；理论吞吐量可达 10^4TPS 以上	临时会话、点对点智能体通信
基于 Web 型	DNS/HTTPS/WebPKI	易部署、兼容互联网	去中心化程度有限	延迟通常为 10 – 100 ms；可支持 10^3 – 10^5级并发请求	企业智能体、服务提供方、工具服务

完整性校验，DID 可作为统一的跨平台身份标识，实现跨系统的安全互操作[35]。由此，DID 不仅支撑“智能体—工具服务”，还扩展至“智能体—智能体”通信、“智能体—数据接口”等多元交互场景。与此同时，智能体往往依赖其源大模型（如 LLM）进行推理与决策，并通过插件或 API 调用外部工具完成任务，此时 DID 可用于标识模型服务提供方及工具服务节点，使智能体在调用大模型推理服务或外部能力时，能够验证服务来源的真实性与完整性，防止模型接口被劫持或工具返回结果被篡改，从而构建覆盖“智能体—模型—工具”的端到端可信调用链[19]。

值得注意的是，DID 技术及其安全性已在现实中得到初步验证。例如，在欧盟《电子身份识别与信任服务条例》（eIDAS 2.0）框架下推进的“欧洲数字身份钱包（EUDI Wallet）”，兼容以去中心化标识（DID）与可验证凭证（VC）为核心的身份管理模式[36]。该体系允许用户在本地钱包中持有由政府或受信任机构签发的数字凭证（如身份信息、驾驶执照、学历证明及支付相关资质等），并在跨境政务服务、银行开户、支付认证等高安全等级场景中，通过选择性披露与加密验证机制完成身份核验与授权过程。与此同时，更贴近智能体场景的实践探索已开始出现。例如，Garzon 等提出将去中心化标识与可验证凭证嵌入 AI Agent 体系[18]，使每个智能体拥有可解析的 DID、可轮换的密钥材料以及由可信主体签发的能力凭证。在智能体请求调用工具、数据接口或其他智能体时，验证方可先解析其 DID 文档获取公钥，再校验其 VC 中声明的组织归属、模型能力、工具访问范围和有效期，从而在不依赖单一身份提供方的情况下完成身份认证与权限核验。

3.2.2　细粒度访问控制与动态权限管理

在面向智能体系统的新兴通信架构中，访问控制对象已不再局限于单一智能体节点，而是扩展至大模型服务、工具插件及数据资源等多类主体，这要求引入更灵活、可验证的授权机制。这主要体现在智能体系统相关的授权委托与信任传递等方面上。具体来说，在智能体网络中，“授权”是指赋予智能体相关的行动权限，“委托”允许智能体将子任务及相应权限转交其他智能体，“信任传递”确保多跳委托链中权限从原始用户到末端智能体的可验证延续，“撤销”旨在终止智能体的相关行动权限，而“下架”则是对智能体身份及其所有关联权限的永久性移除。在这一体系中，人作为决策与价值的最终承担者，通常通过配置或授权的方式驱动智能体行为，而智能体则在一定范围内具备自主执行任务与参与交互的能力。

W3C 可验证凭证（VC）规范为此提供了“发行者—持有者—验证者”三方标准模型：凭证由发行者签名生成，由持有者本地安全存储，并在交互时提交给验证者进行真实性与有效性校验。在此基础上，将 VC 机制与基于区块链账本的 DID 体系相结合，可支撑多种主流访问控制模型：在基于属性的访问控制（ABAC）模型中，智能体携带含属性字段的 VC，由策略引擎依据属性匹配结果执行授权决策；在基于角色的访问控制（RBAC）模型中，凭证签发方在 VC 中声明持有者的角色信息，服务方据此映射相应权限集合；在基于能力的访问控制（Capability-based Access Control）模型中，VC 本身作为可验证的能力令牌，持有者凭其可直接访问目标资源。这种融合架构能够实现权限的动态管理与自动更新，一旦凭证过期或被撤销，其访问权限将随之失效，无需依赖中心化权限管理服务器，从而显著提升系统的实时性与自治性[20]。在实际交互中，智能体可依据最小披露原则仅出示相应 VC 以证明其具备特定属性或权限，而无需暴露多余身份信息，兼顾安全性与隐私保护。例如，在基于模型上下文协议（MCP）的工具调用过程中，工具服务可要求调用方智能体出示特定 VC（如“具备某类任务执行权限”或“来自可信组织”），并结合其 DID 验证后方可开放接口访问；在多智能体交互协议（A2A）的协作场景中，任务发起方可基于对方智能体的 VC 属性（如能力标签、信誉等级）动态分配任务与权限，从而实现面向能力与信任的自适应协作[21][22]。需要注意的是，在智能体递归委托场景中，委托链末端的资源服务器须能通过加密方式验证回溯至原始授权用户的完整委托路径，而非仅验证最终子智能体身份；权限的逐级衰减机制是确保多跳委托链安全性与最小权限原则落地的关键技术保障。

由此，智能体能够以 DID 为核心标识，自主定义资源访问策略，并对访问主体范围进行精细化控制，从而实现面向多主体协同环境的高粒度权限治

理。在实际应用中，相关技术路径已在企业级身份与数据共享场景中得到探索与验证。例如，Microsoft 推出的去中心化身份解决方案 Microsoft Entra Verified ID，基于 W3C DID 与可验证凭证（VC）标准构建，形成了以DID为信任锚的身份验证体系[37]。在具体业务流程中，用户（或智能体）在访问企业服务或进行在线业务办理时，可通过数字身份钱包向服务提供方提交VC；验证方则通过解析用户的DID文档获取公钥信息，对凭证的签名及其状态进行校验，从而在无需与发行方建立实时连接的情况下完成可信认证与授权过程。另一方面，随着智能体规模的增长，逐操作交互式同意模式将引发不可管理的"同意疲劳"问题，因此面向智能体的访问控制研究需从传统交互式同意向策略驱动的预授权模型及基于风险等级的动态授权机制演进，以实现人类监督与系统自主性的有效平衡。

3.2.3 *数据主权保护与隐私合规*

基于区块链账本的DID体系将身份控制权从中心化机构转移至智能体（或其对应用户），契合数据主权与自我主权身份的核心理念。在该模式下，智能体通过持有与DID绑定的公私钥对，实现对自身身份标识及相关凭证的自主生成、管理与使用，相关身份数据不再集中存储于单一数据库中，从根本上降低了集中式身份库泄露所带来的系统性风险。智能体能够依据具体应用场景及监管要求，自主决定数据的访问范围与披露粒度，实现“最小必要披露”的数据共享模式。

在隐私保护机制方面，基于区块链账本的DID体系可结合多种密码学技术实现差异化的最小化披露：选择性披露允许凭证持有者仅出示VC中满足验证需求的部分属性字段；零知识证明（Zero-Knowledge Proof，ZKP）支持在不暴露原始属性内容的前提下证明某一命题成立；匿名凭证则可在验证过程中完全隐去持有者的身份标识[23][24]。上述机制的核心价值在于将"出示凭证"与"暴露身份"相解耦，使智能体能够在满足验证需求的同时将信息披露控制在最小必要范围内。例如，在多智能体协同场景中，参与方仅需证明其持有某类权威机构颁发的合规凭证，而无需披露具体机构信息或其他敏感数据。以自动驾驶与车路协同场景为例，车辆智能体在接入道路基础设施或参与协同决策时，仅需向路侧单元证明其"具备合法上路资质"或"通过安全认证"，而无需披露车辆唯一标识、所属车主或注册机构信息。在数据市场场景中，数据提供方智能体可在不暴露数据来源或主体身份的情况下，仅证明其数据“来源合规且经过授权”或“满足特定质量标准”，从而支持安全的数据交易与共享。通过上述机制，在满足协同验证与访问控制需求的同时，有效降低了敏感信息泄露与跨场景关联风险，从而在保证可验证性的同时强化隐私保护。需要指出的是，上述隐私保护能力并非引入DID后自动获得。链上的DID标识、服务端点、交易时间戳及哈希记录等元数据均可能成为行为关联与身份推断的信息来源，需在系统设计阶段主动选用适配的隐私增强机制加以规避。在此基础上，如何将智能体身份与其所执行的操作及生成内容进行不可否认的绑定，是建立完整问责链条的深层挑战。

此外，区块链账本通常仅记录DID文档、凭证哈希摘要及撤销状态等非敏感信息，具体身份数据与凭证内容则可存储于链下分布式存储系统（如IPFS）中，并通过加密与哈希锚定机制与链上数据建立关联。这种“链上锚定、链下存储”的架构有效避免了敏感信息直接上链所带来的隐私暴露风险[25]。

综上，表4从通信安全、访问控制与数据隐私三个维度，对比总结了DID体系相比传统IP标识

表4 传统标识体系与DID赋能智能体的能力对比

能力维度	传统互联网标识体系	DID赋能智能体系统能力
通信安全与抗攻击	依赖中心化身份提供方与静态密钥体系，易形成单点失效与集中式攻击目标；通信认证机制薄弱。	支持端到端加密与签名认证，结合不可篡改账本与共识机制，显著提升抗攻击与通信安全能力
访问控制与动态权限管理	权限管理依赖中心化IdP或静态ACL模型，更新滞后且难以跨域协同；缺乏灵活的属性与角色级控制能力。	实现基于属性/角色的细粒度访问控制；支持权限动态更新、链上校验与自动化授权决策
数据主权保护与隐私合规	身份与数据集中存储于中心化机构，存在隐私泄露与滥用风险；跨系统数据共享缺乏可验证隐私机制	用户自主控制身份与凭证；结合VC与零知识证明实现最小化披露；采用“链上锚定+链下存储”提升隐私与合规能力。

体系 在安全性、灵活性及用户数据主权保护方面的显著优势。此外，DID 还可用于数据集合标识，为数据流通交易也提供了新思路，应是一个重要的研究方向。

3.3　当前面临的主要挑战

尽管基于区块链账本的去中心化标识（DID）在智能体体系中展现出构建去中心化信任的显著优势，但其广泛应用仍面临多重挑战。在技术层面，链上性能与扩展性的局限导致高并发场景下的标识注册与验证效率低下；跨链互操作性不足与标准化缺失则加剧了多链环境下的互联互通困境。在治理层面，去中心化架构导致法律责任主体模糊，链上数据的不可篡改性与隐私法规中关于数据修正及删除的权利相冲突，同时合规性与伦理风险亦日益凸显。本节将分小节深入剖析上述问题的成因及其对智能体系统带来的影响，需在工程实践中进一步探索和完善。

3.3.1　系统性能与可扩展性瓶颈

区块链账本在吞吐量与延迟方面的固有限制，使去中心化标识（DID）在效率与规模化部署上面临显著瓶颈。在智能体系统中，大量身份注册、凭证颁发与验证请求若依赖链上处理，易引发拥堵、延迟甚至失败；以以太坊等公链为例，其较低 TPS 与共识延迟难以支撑大规模实时认证需求。该瓶颈源于链上链下协同与架构设计：一方面，DID 文档或索引需锚定链上，系统性能受限于区块容量与写入吞吐，多节点共识进一步限制并发能力；另一方面，为保护隐私而采用“链下加密存储+链上存证”的模式，引入额外计算与数据检索开销，对链下访问控制提出更高要求。上述问题综合影响系统实时性、经济性与可部署性，可能会导致认证延迟、通信中断等问题，还因频繁链上操作带来高昂成本，限制资源受限设备接入，整体对智能体系统的规模化扩展与复杂场景落地带来一定挑战。

3.3.2　跨域互操作性与标准化挑战

当前，不同区块链或联盟链间的去中心化标识（DID）体系因缺乏统一互操作标准，成为影响智能体跨网络交互的重要因素。各平台采用不兼容的 DID 方法（如 did:ethr、did:sov、did:ont 等），在标识解析、凭证格式及通信协议等方面存在差异，增加了跨链身份验证与信任建立的复杂性[26][27]。其根源在于各类 DID 方案多围绕本链需求设计，对 W3C 等通用标准的协同遵循尚不充分，同时链间数据结构与接口规范的异构性进一步提升了互联成本。尽管部分跨链机制可实现凭证传输，但多依赖定制化桥接协议，通用性仍有提升空间。互操作能力不足易导致智能体网络出现“信息孤岛”，降低系统协同效率，在跨域认证中往往需借助中介或进行多重注册，从而增加通信开销与系统复杂度，并带来一定安全隐患[28]。在多组织协同与跨域服务场景中，这一问题更为突出，例如以太坊链上的智能体在调用 Hyperledger Indy 链服务时，若缺乏统一信任层或协议适配机制，将难以直接完成身份识别与验证，对业务流程的顺畅衔接带来一定影响。

3.3.3　治理机制、法律责任风险

区块链去中心化与不可篡改特性在提升可信性的同时，也与现行法律法规产生一定张力。在面向智能体的去中心化标识（DID）体系中，由于缺乏集中监管与明确责任主体，一旦出现数据泄露或身份纠纷，责任追溯与认定较为复杂。其成因在于分布式多节点共同维护的数据结构（网络中的多个节点共同存储、验证和维护）使法律责任主体难以清晰界定，在欧盟《通用数据保护条例》（GDPR）、《个人信息保护法》等框架下，涉及个人信息的删除与更新需求难以直接实现，彻底删除或保留痕迹，在不同的场景下也各有利弊，需进一步探讨。这些因素使 DID 在智能体系统中的合规应用面临一定复杂性，尤其在跨国场景下，不同法域的数据治理要求差异带来不确定性；责任界定不清也可能影响系统信任基础；在涉及健康、金融等敏感数据时，还需同时满足多重监管要求，从而提升系统设计与实现难度。

另一方面，可将去中心化标识符（DID）与安全护栏机制有机结合，以在智能体网络中实现多层次风险缓解与治理能力增强。这里所指的安全护栏，主要包括基于策略规则的输入输出过滤机制、权限边界强制执行模块以及运行时行为监控组件等，其能够对智能体的感知输入、决策推理与动作输出实施全流程约束。在此基础上，DID 及其关联的可验证凭证（VC）可进一步作为权限策略与行为规范的可信绑定载体：当智能体发起交互请求时，安全护栏可动态解析凭证中预设的能力范围、身份属性与合规标签，并据此自动激活相应的过滤规则与访问控制策略，从而实现身份认证、权限校

验与行为约束的运行时协同。通过该机制，一方面可有效降低智能体在交互过程中发生敏感信息泄露的风险，主动识别并屏蔽个人可识别信息、财务数据等高风险内容；另一方面，也能够限制智能体执行超出授权范围的异常操作，避免因资源过度调用而导致的系统负载增加与成本损耗。此外，安全护栏与DID的协同机制还有助于保障智能体行为持续满足监管合规要求，并在决策与内容生成过程中维持必要的伦理一致性，减少有害、偏见或歧视性输出的产生。由此，DID不仅能够为智能体提供可验证、可自主控制的身份锚点，还可与安全护栏共同构建面向智能体系统安全运行与可信治理的基础支撑体系。

4 未来研究方向与展望

4.1 面向智能体经济交互的身份与支付协同机制

随着自主智能体逐步具备参与经济活动的能力，其在访问付费资源、发起购买及编排服务等场景中的身份可信性与授权可验证性问题日益突出。传统电子商务与API安全模型建立在人类实时介入与显式同意的范式之上，难以适配智能体自主驱动的高频异步商业交易。为此，未来研究有必要探索将DID体系与金融级API安全规范深度融合的路径，通过发送方受限访问令牌、强客户端认证等机制，为不可逆金融操作提供事务完整性与不可否认性保障。当前，Google提出的智能体支付协议（AP2）以经加密签名的"授权书（Mandate）"为核心，建立了"意图授权书—购物车授权书"两阶段审计链，并结合可验证凭证将授权绑定至用户身份，为智能体商业交易提供了可审计的意图证明机制。然而，该类协议在与W3C DID及VC规范的标准对接、多跳委托场景下的授权链完整性保障，以及智能体自主执行效率与用户可控性之间的动态平衡等方面，仍存在较大研究空间，值得持续深入探索[3][29][30]。

4.2 面向智能体的新型身份与访问管理授权范式

随着自主智能体在企业系统与跨域网络中的规模化部署，传统以人类交互为中心的身份与访问管理（Identity and Access Management，IAM）框架正面临深层挑战。在高速执行与异步长任务场景中，逐操作同意模型既无法应对智能体毫秒级的大量API调用，也难以为跨越数天的持久委托任务提供合理的身份生命周期管理；在跨域协作与递归委托场景中，如何在多跳委托链中安全传递原始授权上下文、并逐级收窄权限范围，现有协议体系尚缺乏成熟的标准化方案。上述困境表明，智能体亟需被纳入IAM体系的设计考量之中。为此，未来研究可重点探索面向智能体的预授权与策略驱动访问控制模型，研究基于DID与可验证凭证的持久委托身份方案，实现跨系统的动态授权与即时撤销。同时，推动OAuth 2.0令牌交换与身份断言等标准在多跳智能体场景下的协议扩展，以及具备范围衰减能力的递归委托凭证设计。上述方向的协同推进，将为构建兼具自主性、可审计性与合规性的智能体身份管理新范式奠定基础[3]。

4.3 DID与智能体通信协议的融合标准化

随着模型上下文协议（MCP）与智能体间交互协议（A2A）成为多智能体系统的核心通信基础设施，其在身份认证、凭证传递与信任建立等环节尚缺乏原生的去中心化支撑机制，DID与VC标准的引入为其提供了重要补充方向。未来研究可重点探索DID体系与上述协议的深度融合路径：在MCP工具调用场景中，工具服务可要求调用方提交基于DID的身份声明及相应VC，以实现可验证的访问授权；在A2A协作场景中，任务发起方可依据对方智能体的DID文档与凭证属性动态评估其可信度，支持能力感知的任务分配。然而，现有协议在DID解析接口、VC传输格式及身份上下文传递等方面尚未形成统一约定，跨平台适配仍依赖定制化实现。为此，有必要在W3C现有标准基础上推动形成面向智能体通信场景的协议扩展规范，明确DID在智能体发现、连接建立与委托授权等关键流程中的标准语义与接口要求，并构建参考实现与测试套件，推动跨平台互信机制的工程落地[31]。

4.4 面向智能体网络的DID统一身份架构

随着智能体生态向多主体、多层次协作网络演进，现有身份体系在主体类型覆盖、标识粒度管理及治理责任归属等方面难以满足统一建模及管理需求。为此，未来研究可着力构建面向智能体网络的DID统一身份架构模型，设计为涵盖主体、标识、凭证与治理四个层次的分层体系。主体层统一纳管人类用户、组织机构、智能体、模型服务、工具服务、数据资源及监管审计机构等多元实体；标识层针对不同主体的持久性与隐私需求，支持长期

DID、会话 DID、Pairwise DID 等多种标识形态的灵活绑定；凭证层基于 W3C VC 规范，对所属组织、授权范围、能力声明、合规资质、工具访问权限及数据使用许可等属性进行结构化表达，支持选择性披露与跨域验证；治理层则通过凭证签发、撤销、审计、责任追溯、声誉评估及监管披露等模块，构建覆盖身份全生命周期的链上链下协同治理机制。该架构模型可同时作为理论分析框架与工程参考基准，推动去中心化身份技术在多主体协同网络中的系统化落地[32][33]。

5　结束语

随着智能体系统向规模化、动态化与跨域协同方向不断演进，传统互联网通信机制与中心化身份管理为核心的标识体系已难以支撑其复杂运行需求。去中心化标识（DID）通过引入分布式信任机制与自我主权身份理念，在标识生成、身份认证、权限管理及隐私保护等方面提供了系统性创新，为构建面向智能体的统一身份基础设施奠定了关键技术基础。本文从智能体标识需求出发，系统梳理了数字身份技术的发展脉络，分析了现有标识体系的局限，并重点探讨了基于区块链账本的 DID 在安全通信、动态授权与数据主权等方面的核心优势。同时，也指出了其在性能扩展、跨域互操作及治理合规等方面仍面临的现实挑战。未来，随着标准体系的逐步完善与关键技术的持续突破，DID 有望与多智能体系统深度融合，推动形成兼具安全性、灵活性与可监管性的下一代网络身份体系，从而为智能体驱动的数字经济社会提供坚实支撑。

参考文献：

[1] 陈健, 蔡智明, 齐佳音, 方滨兴. 面向跨境的去中心分布式数字身份框架设计[J]. 中国工程科学, 2025, 27(1): 88-97.

[2] 焦志伟, 吴正豪, 徐亦佳, 等. 基于隐私保护的分布式数字身份认证技术研究及实践探索[J]. 信息通信技术与政策, 2024, 50(1): 59-66.

[3] South T, Nagabhushanaradhya S, Dissanayaka A, et al. Identity management for agentic AI: The new frontier of authorization, authentication, and security for an AI agent world[R/OL]. OpenID Foundation, 2025.

[4] Soltani R, Nguyen U T, An A. A survey of self-sovereign identity ecosystem[J]. Security and Communication Networks, 2021, 2021: 8873429.

[5] Schardong F, Custódio R. Self-sovereign identity: A systematic review, mapping and taxonomy[J]. Sensors, 2022, 22(15): 5641.

[6] Ahmed M R, Islam A K M M, Shatabda S, et al. Blockchain-based identity management system and self-sovereign identity ecosystem: A comprehensive survey[J]. IEEE Access, 2022, 10: 113436-113481.

[7] Tobin A, Reed D. The inevitable rise of self-sovereign identity[R]. The Sovrin Foundation, 2017.

[8] W3C Recommendation. Decentralized identifiers (DIDs) v1.0[S/OL]. W3C Recommendation, 2022. https://www.w3.org/TR/did-core/.

[9] W3C Recommendation. Verifiable credentials data model v1.1[S/OL]. W3C Recommendation, 2022. https://www.w3.org/TR/vc-data-model/.

[10] Sadique K M, Rahmani R, Johannesson P. DIdM-EIoTD: Distributed identity management for edge internet of things (IoT) devices[J]. Sensors, 2023, 23(8): 4046.

[11] Kim T, Seo D, Kim S H, et al. A comprehensive approach to user delegation and anonymity within decentralized identifiers for IoT[J]. Sensors, 2024, 24(7): 2215.

[12] Alizadeh M, Andersson K, Schelen O. Comparative analysis of decentralized identity approaches[J]. IEEE Access, 2022, 10: 92273-92283.

[13] Decentralized Identity Foundation. DIDComm messaging specification v2.0[S/OL]. Decentralized Identity Foundation, 2022. https://identity.foundation/didcomm-messaging/spec/v2.0/.

[14] Schardong F, Custódio R. Self-sovereign identity: a systematic review, mapping and taxonomy[J]. Sensors, 2022, 22(15): 5641.

[15] Pino A, Margaria D, Vesco A. Combining decentralized IDentifiers with proof of membership to enable trust in IoT networks[C]//Proceedings of 33rd International Telecommunication Networks and Applications Conference (ITNAC). Melbourne: IEEE, 2023: 310-317.

[16] He J, Ma X, Zhang D, et al. Supervised and revocable decentralized identity privacy protection scheme[J]. Security and Safety, 2024, 3: 2024013.

[17] Yin J, Xiao Y, Feng J, et al. Didtrust: Privacy-preserving trust management for decentralized identity[J]. IEEE Transactions on Dependable and Secure Computing, 2025 Jan 1.

[18] Garzon S R, Vaziry A, Kuzu E M, Gehrmann D E, Varkan B, Gaballa A, Küpper A. AI agents with decentralized identifiers and verifiable credentials[J]. arXiv preprint arXiv:2511.02841, 2025.

[19] Sedlmeir J, Smethurst R, Rieger A, et al. Digital identities and verifiable credentials[J]. Business & Information Systems Engineering, 2021, 63(5): 603-613.

[20] Yu L, He M, Liang H, et al. A blockchain-based authentication and authorization scheme for distributed mobile cloud computing services[J]. Sensors, 2023, 23(3): 1264.

[21] Bai Y, Liu Z, Liu X, Lei H. A cross-chain identity authentication scheme based on DID[C]//Proceedings of 2023 IEEE International Conference on Blockchain. IEEE, 2023: 172-179.

[22] Hoops F, Mühle A, Matthes F, Meinel C. A taxonomy of decentralized identifier methods for practitioners[C]//Proceedings of the 2023 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS). IEEE, 2023: 57-65.

[23] Di Francesco Maesa D, Lisi A, Mori P, et al. Self sovereign and blockchain based access control: supporting attributes privacy with zero knowledge[J]. Journal of Network and Computer Applications, 2023, 212: 103577.

[24] Liu J, Liang Z, Lyu Q. Empowering privacy through peer-supervised self-sovereign identity: Integrating zero-knowledge proofs, blockchain oversight, and peer review mechanism[J]. Sensors, 2024, 24(24): 8136.

[25] Goodell G, Aste T. A decentralised digital identity architecture[J]. Frontiers in Blockchain, 2019, 2: 17.

[26] Zhao G, Di B, He H. A novel decentralized cross-domain identity authentication protocol based on blockchain[J]. Transactions on Emerging Telecommunications Technologies, 2022.

[27] Chowdhury M J M, Ferdous M S, Biswas K, et al. A comparative analysis of distributed ledger technology platforms[J]. IEEE Access, 2019, 7: 167930-167943.

[28] Dunphy P, Petitcolas F A P. A first look at identity management schemes on the blockchain[J]. IEEE Security & Privacy, 2018, 16(4): 20-29.

[29] Wang T, Lin Z, Zhang S, et al. Linking souls to humans: Blockchain accounts with credible anonymity for Web 3.0 decentralized identity [C]//Proceedings of the ACM Web Conference 2025 (WWW '25. ACM, 2025. .

[30] Ferdous M S, Chowdhury F, Alassafi M O. In search of self-sovereign identity leveraging blockchain technology[J]. IEEE Access, 2019, 7: 103059-103079.

[31] Vaziry A, Garzon S, Küpper A. Towards multi-agent economies: Enhancing the A2A protocol with ledger-anchored identities and x402 micropayments for AI agents[J]. arXiv preprint arXiv:2507.19550, 2025.

[32] Mazzocca C, Acar A, Uluagac S, et al. A survey on decentralized identifiers and verifiable credentials[J]. IEEE Communications Surveys & Tutorials, 2025.

[33] Čučko Š, Bečirović Š, Kamišalić A, et al. Towards the classification of self-sovereign identity properties[J]. IEEE Access, 2022, 10: 88306-88329.

[34] Anthropic. Model context protocol specification (2025-11-25)[S/OL]. Anthropic, 2025. https://modelcontextprotocol. io/specification/2025-11-25.

[35] Linux Foundation. Agent2Agent (A2A) protocol specification[S/OL]. Linux Foundation, 2025. https://a2a-protocol.org/latest/specification/.

[36] European Parliament and Council. Regulation (EU) 2024/1183 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework[S/OL]. Official Journal of the European Union, 2024. https://eur-lex.europa.eu/eli/reg/2024/1183/oj/.

[37] Microsoft. Introduction to Microsoft Entra Verified ID[EB/OL]. Microsoft, 2024. https://learn. microsoft. com/en-us/entra/verified-id/decentralized-identifier-overview.

薄兆一（1971− ），男，本科，中央网信办数据与技术保障中心高级工程师，主要研究方向为电子政务、标识与密码、数据与技术治理。

王旭（1978− ），男，硕士，中央网信办数据与技术保障中心高级工程师，主要研究方向为数据治理、信息化应用。

陈慧慧（1984− ），女，博士，浙江大学博士后，主要研究方向为数据流通利用安全、网络与信息安全。

任奎（1978– ），男，博士，浙江大学教授，主要研究方向为人工智能安全、数据要素安全、计算机应用。