

面向医疗健康数据的隐私增强型 多维度融合数据标识方法

漆俊峰¹, 夏琦¹, 马英², 李书睿¹, 管庆旭³, 夏虎¹, 高建彬¹

1. 电子科技大学计算机科学与工程学院, 四川 成都 610054;

2. 国家信息中心, 北京 100045;

3. 四川省大数据中心, 四川 成都 610041

摘要

针对区域医疗联合体跨院数据共享场景中, 现有医疗数据标识存在维度单一、隐私保护与数据可用性矛盾、全生命周期安全管控缺失、行业合规适配性不足的问题, 提出了一种隐私增强型多维度融合医疗数据标识方法。该方法融合分级分类、隐私保护、溯源管控、权限管控四大核心能力, 设计了一体化的定长标识结构; 基于国密 SM9 标识密码算法构建双假名可关联机制, 实现了患者身份匿名前提下的跨院诊疗数据的无损关联; 结合差分隐私、SM3 哈希链与密文策略属性基加密, 实现了敏感属性脱敏、全生命周期防篡改溯源与细粒度权限管控。构建了包含诚实但好奇攻击者与恶意攻击者的多攻击者安全模型, 完成了方案的安全性归约证明。实验结果表明, 在 Intel Core i7-9750H 处理器环境下, 标识生成平均耗时 15.310ms, 方案在满足安全性需求的同时具备工程可行性。经与现有方案对比分析, 该方案在匿名性、抗合谋攻击、国密适配等方面具有显著优势, 符合《数据安全法》《个人信息保护法》及医疗数据安全相关国家标准。

关键词

医疗健康数据; 数据标识; SM9 标识密码; 隐私保护; 医联体数据共享; 国密算法

中图分类号: TP309

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2026xxx

A Privacy-Enhanced Multi-Dimensional Fusion Data Identification Method for Healthcare Data

QI Junfeng¹, XIA Qi¹, MA Ying², LI Shurui¹, GUAN Qingxu³, XIA Hu¹, GAO Jianbin¹

1. Sichuan Data Group Co., Ltd., Chengdu 610094, China;

2. University of Electronic Science and Technology of China, School of Computer Science and Engineering, Chengdu 610054, China;

3. State Information Center, Beijing 100045 China;

4. Provincial Big Data Center, Chengdu 610041, China

Abstract

In the context of cross-hospital data sharing within regional medical consortia, existing medical data identifier schemes exhibit several critical limitations: single-dimensional structures, the inherent trade-off between privacy and data utility, a lack of lifecycle security governance, and inadequate compliance with industry regulations. To address these challenges, a privacy-enhanced, multi-dimensional fused medical data identifier method is proposed. The scheme

integrates four core capabilities—data classification, privacy protection, traceability, and access control—into a unified, fixed-length identifier structure. Leveraging the national cryptographic SM9 identity-based algorithm, we construct a linkable dual-pseudonym mechanism that enables lossless linkage of cross-institutional health records while preserving patient anonymity. By synergistically combining differential privacy, an SM3-based hash chain, and Ciphertext-Policy Attribute-Based Encryption (CP-ABE), the scheme achieves sensitive attribute obfuscation, tamper-evident lifecycle traceability, and fine-grained access control. A multi-attacker security model, including honest-but-curious and malicious adversaries is established, and a formal security proof for the scheme via security reduction is provided. Experimental results, conducted on an Intel Core i7-9750H processor, show an average identifier generation time of 15.310 ms, demonstrating the scheme's engineering feasibility while meeting stringent security requirements. A comparative analysis reveals that our approach offers significant advantages in anonymity, collusion resistance, and compatibility with national cryptographic standards. The scheme is designed to be compliant with China's Data Security Law, Personal Information Protection Law, and relevant national standards for medical data security.

Key words

mmedical and health data, data identification, SM9 identity-based cryptography, privacy protection, medical alliance data sharing, domestic cryptographic algorithms

0 引言

随着“健康中国2030”战略的深入推进，区域医疗联合体（以下简称“医联体”）已成为实现医疗资源均衡配置、提升诊疗效率的核心载体，而跨院电子病历共享、检验检查结果互认、临床科研协同等业务的落地，高度依赖医疗数据的跨域互通与可信流转。数据标识作为医疗数据的“数字身份证”，是实现数据精准定位、有序管理、可信共享的核心基础。与此同时，《数据安全法》《个人信息保护法》等法律法规与标准的落地实施，明确了医疗数据共享必须遵循“最小必要”“隐私优先”“全程可溯”的原则[3]。

当前医联体场景下的医疗数据标识方案，仍存在诸多无法回避的核心痛点：一是标识维度单一，无法同时兼顾分类分级、隐私保护、溯源管控、权限管控的多重业务需求；二是无法破解隐私保护与数据可

用性的核心矛盾，匿名化处理会破坏跨院数据关联性，明文标识流转则存在严重的隐私泄露风险；三是标识全生命周期安全管控缺失，流转、解析、更新等环节无防篡改与溯源机制；四是行业合规与信创适配性不足，难以在国内医疗机构规模化落地^{[1][2]}。

为此，本文面向医联体跨院数据共享的真实业务场景，设计了一套全流程基于国密算法的隐私增强型多维度融合医疗数据标识方法，核心工作如下：

1. 设计了一体化的四域融合标识结构，同步集成分类分级、隐私保护、溯源管控、权限管控四大核心能力，解决了现有标识维度单一的痛点，实现一份标识覆盖医疗数据全生命周期管控需求；
2. 基于SM9标识密码算法构建双假名可关联机制，设计了院内本地假名与跨院关联假名的双层体系，实现了患者身份完全匿名前提下的跨院诊疗数据无损关联，破解了隐私保护与数据可用性的核心矛盾；
3. 融合差分隐私、SM3哈希链与国密

优化的密文策略属性基加密（CP-ABE），实现了标识敏感属性脱敏、全流程防篡改溯源与细粒度分级解析权限管控，全流程采用国密算法，符合医疗行业合规与信创要求；

4. 构建完整的安全模型与形式化证明。本文针对医疗数据共享场景的特点，定义了包含诚实但好奇攻击者、恶意攻击者、合谋攻击者的多攻击者安全模型，明确了攻击者的能力边界与系统假设。基于随机预言机模型，完成了方案的语义匿名性、不可伪造性、抗篡改性等形式化安全证明。

5. 完成全面的性能评估与对比分析，在Intel Core i7-9750H实验环境下，完成了标识生成、SM9签名、CP-ABE加解密、哈希链操作等关键模块的性能基准测试，并与现有EMPI方案、区块链ABE方案进行了多维对比，验证了方案的工程可行性。

1 相关工作

1.1 研究背景与政策导向

医疗健康数据是国家重要的民生数据资源，也是数字经济时代核心的数据要素之一。近年来，国家出台《国家健康医疗大数据标准、安全和服务管理办法（试行）》等政策文件，明确要求加快区域医联体建设，推动医疗数据跨机构、跨区域互通共享，释放医疗数据的临床科研与公共卫生价值^[3]。

同时，医疗健康数据包含患者身份信息、诊疗记录、基因数据等高度敏感的个人信息，其安全管控与隐私保护是数据共享不可突破的红线。《GB/T 42430-2023》明确将医疗数据划分为5个安全等级，要

求对敏感数据实现全生命周期的分级管控、隐私保护与全程溯源^[4]；《个人信息保护法》则对敏感个人信息处理提出了知情同意、最小必要、匿名化处理等强制性要求^[5]。

数据标识作为医疗数据全生命周期管理的核心载体，其设计直接决定了数据共享的安全性、合规性与可用性。传统医疗标识方案已无法适配当前医联体数据共享的业务需求与合规要求，需要设计一套兼顾隐私保护、全流程管控、合规适配的新型医疗数据标识方法。

1.2 国内外研究现状

1.2.1 医疗数据标识与患者主索引(EMPI)技术

医疗数据的跨机构共享首先依赖于准确的患者身份识别。患者主索引（Enterprise Master Patient Index, EMPI）作为一种核心的数据治理工具，通过为同一患者在不同系统中的记录创建唯一标识符，有效解决了医疗数据孤岛问题。然而，传统的EMPI系统在集中化管理模式下存在隐私泄露和单点故障风险。近年来，研究开始探索结合密码学技术增强EMPI的隐私保护能力，例如利用可验证计算或安全多方计算在不暴露原始标识信息的情况下完成患者身份匹配，但这通常面临计算开销大、工程实现复杂等挑战^[4]。

1.2.2 标识密码体制(IBE)及其在医疗场景下的扩展

求标识密码体制因其天然适用于基于身份的加密与访问控制，在医疗数据保护中受到广泛关注。我国自主设计的商用密码标准SM9算法是其中的典型代表。为适应医疗协同等复杂场景，研究者对SM9进

行了多方面的功能扩展：

分层与广播加密：为降低大规模医疗网络中密钥生成中心（KGC）的管理负担，研究者提出了基于SM9的高效分层标识加密方案^[6]，支持按医疗机构层级结构管理密钥。同时，广播代理重加密方案被提出，以实现一份病历密文安全、高效地共享给多名授权医生，并支持对重加密过程的验证^[7]。

细粒度访问控制：属性基加密（ABE）能够实现基于用户属性的细粒度访问控制。已有研究实现了基于SM9的ABE方案，并将其应用于医疗数据共享系统，确保只有满足特定属性条件的医护人员才能解密数据^[8]。在国际学术界，针对医疗数据共享的隐私保护研究也取得了显著进展，其主流技术路线与国内研究有相似之处，但更侧重于ABE与区块链技术的深度融合。多项研究提出了基于区块链和ABE的医疗数据共享方案，利用区块链存储密文索引和访问策略，而将实际数据存储在分布式文件系统中，以此实现数据的可审计性和抗单点故障^{[9][10][11]}。此外，为了应对跨域协作的复杂需求，有研究提出了支持分布式动态阈值的属性基加密方案，在保证安全的前提下实现灵活的跨机构医疗人员动态加入与协作^[12]。

新型密码原语：为进一步平衡数据可用性与隐私，支持等式测试的SM9加密方案允许在密文状态下判断两份病历数据是否属于同一患者，而无需解密；内积函数加密则支持在密文状态下进行安全的统计分析，为隐私保护的医疗数据挖掘提供了新思路^{[7][13]}。

1.3 安全威胁模型

在安全威胁模型方面，医疗数据共享

系统通常考虑以下攻击者类型：

诚实但好奇攻击者：此类攻击者遵循协议规范执行操作，但可能被动记录并分析协议执行过程中的信息，试图推断敏感数据。研究表明，在医疗数据共享场景中，诚实但好奇的医疗机构或第三方服务提供商是需要重点防范的攻击者类型。

恶意攻击者：此类攻击者可能主动偏离协议规范，篡改消息、注入虚假数据或拒绝响应，以实现数据窃取、身份伪造等恶意目的。在医疗数据标识系统中，恶意攻击者可能试图伪造合法标识、篡改标识内容或发起拒绝服务攻击。

合谋攻击者：多个攻击者可能通过合谋共享各自的视图或密钥，突破单方无法突破的隐私保护机制。在属性基加密系统中，合谋攻击是主要安全威胁之一，需要通过安全机制设计予以防范。

1.4 现存核心问题分析

除密码学方法外，差分隐私、联邦学习等隐私计算技术也被引入医疗领域。差分隐私通过向查询结果中添加可控噪声，为数据集提供严格的数学隐私保证，常用于医疗统计数据的发布^[13]。联邦学习使得多家医院能够在本地数据不出域的前提下协同训练AI模型，有效规避了数据集中汇聚带来的隐私风险。当前的前沿研究致力于将SM9等密码学工具与这些技术深度融合，例如，利用SM9实现联邦学习参与方的安全认证与通信加密，或结合同态加密与差分隐私构建端到端的隐私保护数据分析框架^[6]。

现有工作为医疗数据的安全共享提供了丰富的技术储备，但仍存在挑战：如何将国密算法（如SM9）与EMPI等数据治理实践更轻量、高效地结合；如何在保障

强安全性的同时，满足医疗场景对实时性和大规模并发的性能要求。本研究旨在针对这些挑战，提出一种切实可行的解决方案。

2 隐私增强多维度融合医疗数据标识方法整体设计

2.1 系统模型与安全假设

2.1.1 系统模型

本方案的系统模型包含以下实体：

患者：医疗数据的数据主体，拥有对其个人数据的知情权、决定权和注销权。患者在就诊时提供身份信息，并可通过电子授权书授权医疗机构访问其跨院诊疗数据。

医疗机构：包括核心医院、二级医院、社区卫生服务中心等，负责采集患者数据、生成本地假名（LPID）、管理诊疗数据、响应跨院数据查询请求。医疗机构被假设为“诚实但好奇”实体，即遵循协议规范但可能试图推断患者身份信息。

密钥生成中心：医联体内的离线可信第三方，负责生成SM9系统主密钥、为患者生成签名私钥与跨机构关联假名（GPID）、维护LPID与GPID的加密绑定关系。KGC被假设为诚实实体，不与医疗机构合谋，但其存在单点信任风险，本文在2.6节讨论缓解机制。

数据共享平台：医联体内的数据交换枢纽，负责验证访问权限、转发数据查询请求、记录操作日志。数据共享平台被假设为“诚实但好奇”实体。

联盟链溯源平台：分布式账本系统，用于存储标识操作的哈希链记录，实现防篡改溯源。联盟链通过共识机制保障数据不可篡改，被假设为可信基础设施。

监管机构：卫生健康委、医联体管理中心等，拥有对标识溯源管控域的解析权限，可审计全流程操作记录。

2.1.2 安全假设与攻击者模型

攻击者类型定义：

定义1（诚实但好奇攻击者A）：攻击者遵循协议规范执行所有操作，但可能被动记录协议执行过程中的所有消息（包括密文、哈希值、操作日志），并试图通过背景知识推断患者的真实身份或敏感属性。

定义2（恶意攻击者B）：攻击者可能主动偏离协议规范，包括：伪造或篡改标识消息、重放历史消息、注入虚假数据、发起拒绝服务攻击。恶意攻击者的目标是破坏协议的正确性、窃取敏感数据或伪造合法身份。

定义3（合谋攻击者C）：多个攻击者（可能是医疗机构、数据共享平台等）通过共享各自的视图、密钥或信息，突破单方无法突破的隐私保护机制。合谋攻击者的典型目标是关联跨机构的标识以推断患者身份。

攻击者能力边界见表1：

系统安全假设：

（A₁）密码学假设：SM9标识密码算法在随机预言机模型下满足不可区分性和不可伪造性；SM3哈希函数满足抗碰撞性、抗第二原像性和单向性；CP-ABE方案满足选择明文攻击下的不可区分性。

（A₂）可信基础设施假设：联盟链溯源平台通过共识机制保障数据不可篡改，假设攻击者无法控制超过1/3的共识节点。

（A₃）安全信道假设：患者身份信息、SM9私钥等敏感数据通过SM2加密通道传输，假设攻击者无法破解加密通道。

（A₄）KGC信任假设：KGC为离线可信第三方，诚实执行协议规范，不主动泄

表1 攻击者能力边界

攻击者类型	可获得信息	可执行操作	受限条件
A	公开系统参数、密文、哈希值、操作日志	被动分析、背景知识推断	无法获取KGC主私钥、用户私钥
B	同上 + 部分机构私钥(通过腐化)	伪造、篡改、重放、拒绝服务	无法破解SM9/SM3密码学假设
C	多机构视图、部分私钥	合谋分析、关联攻击	无法获取KGC主私钥

露患者身份映射关系，不与医疗机构合谋。KGC 可能遭受外部攻击，本文在 2.7 节讨论缓解机制。

2.2 多维度融合标识整体结构

受 Gcode 将政务数据标识分为不同部分并对应不同的功能的启发^[14]，本文设计的多维度融合医疗数据标识，采用定长 256 字节统一结构，分为四大核心域，每个域对应一项核心能力，标识整体结构如表 2 所示：

表2 标识整体结构

域名	长度	域内结构
分类分级域	32 字节	数据主体类型(2 字节)、数据业务类型(10 字节)、数据安全分级(2 字节)、数据分类编码(18 字节)
隐私保护域	96 字节	院内本地假名 LPID(32 字节)、跨机构关联假名 GPID(32 字节)、SM9 公钥指纹(16 字节)、差分隐私扰动参数(8 字节)、预留字段(8 字节)
溯源管控域	64 字节	生成时间戳(8 字节)、生成机构编码(16 字节)、哈希链前置哈希值(32 字节)、操作类型编码(8 字节)
权限管控域	64 字节	CP-ABE 访问策略密文(32 字节)、解析权限等级(4 字节)、有效期(8 字节)、注销标记(4 字节)、预留字段(16 字节)v

分类分级域实现了医疗数据的合规属性与标识的强绑定，流转过程中可基于该域实现自动化的分级管控；隐私保护域通过双假名机制实现匿名性与关联性的平衡；溯源管控域与 SM3 哈希链深度绑定，每一次标识操作都会更新该域，实现全流程可追溯；权限管控域通过属性基加密实现不同机构对标识不同域的分级解析，遵循最小必要原则。

该标识具备全局唯一性，每一份医疗数据单元对应唯一的融合标识，与数据主

体、数据内容、生成机构强绑定，不可重复、不可伪造。

2.3 基于 SM9 标识密码的双假名可关联标识生成机制

为解决对隐私保护与数据可用性的矛盾，基于国密 SM9 标识密码算法，设计了双层假名体系，实现患者身份匿名前提下的跨院数据无损关联。

2.3.1 双假名体系设计

本文设计双层假名体系，实现“院内可识别、跨院可关联、全局全匿名”的目标，两个假名的定义与边界如下：

1. 本地假名 (Local Pseudonym ID, LPID)：院内唯一标识，与患者的真实身份信息一一绑定，由院内密钥管理系统基于国密 SM3 算法生成，加密存储在院内数据库中，对外不公开、不流转。LPID 的作用是实现院内诊疗数据与患者身份的精准匹配，保障院内诊疗业务的正常开展。

2. 跨机构关联假名 (Global Pseudonym ID, GPID)：医联体内的唯一关联标识，不包含任何患者明文身份信息，由医联体可信第三方密钥生成中心 (KGC) 基于 SM9 标识密码算法统一生成，与患者的真实身份一一绑定，与 LPID 一一对应。GPID 是医联体内唯一允许跨机构流转的身份标识，机构可通过 GPID 实现同一患者跨院诊疗数据的匹配关联，但无法通过 GPID 反推出患者的真实身份。

医联体 KGC 基于国密 SM9 算法搭建，为离线可信第三方，仅负责患者双假名的生成与绑定，不参与日常的数据流转与业务操作，降低了中心化节点的单点故障与数据泄露风险。

2.3.2 基于 SM9 密钥派生的跨机构数据关联方法

首先给出 SM9 标识密码算法的系统参数定义：KGC 生成 SM9 签名算法的系统主密钥对，其中主私钥为 s ，主公钥为 $P_{pub} = sP_1$ ，系统公开参数为 $(G_1, G_2, G_T, N, P_1, P_2, e, H_1, H_2)$ 。其中， $e: G_1 \times G_2 \rightarrow G_T$ 为双线性对映射， H_1, H_2 为 SM9 标准规定的密码哈希函数， N 为循环群 G_1, G_2, G_T 的阶， hid 为 SM9 签名私钥生成标识符。

双假名生成流程：

1. 患者首次在医联体内机构就诊，机构采集患者真实身份信息，生成唯一身份基准值 UID：

$$UID = SM3(\text{身份证号} \parallel \text{姓名} \parallel \text{医保卡号}) \quad (1)$$

其中 SM3 即使用 SM3 算法进行计算。

同时生成本地假名 LPID：

$$LPID = SM3(UID \parallel \text{机构编码} \parallel \text{就诊时间戳}) \quad (2)$$

并将 UID 通过 SM2 加密通道提交至医联体 KGC。

2. KGC 接收加密的 UID，解密后基于 SM9 算法为患者生成用户签名私钥：

$$d_{uid} = s \cdot (H_1(UID \parallel hid, N) + s)^{-1} \cdot P_1 \quad (3)$$

其中 s 为主私钥，UID 为唯一身份基准值， hid 为 SM9 签名私钥生成标识符， P_1 为系统公开参数即主公钥的组成部分， H_1 为后文提到的哈希函数。

3. KGC 基于用户私钥与 SM3 单向哈希函数，生成患者的跨机构关联假名：

$$GPID = SM3(d_{uid} \parallel T \parallel M) \quad (4)$$

其中， T 为患者首次就诊时间戳， M 为医联体唯一标识码，保证 GPID 的全局唯一性^[2]。

4. KGC 将生成的 GPID 通过加密通道返回给就诊机构，机构将 LPID 与 GPID 一一绑定，加密存储在院内数据库中，同时为患者的每一份诊疗数据生成对应的多维度融合标识，完成标识与数据的绑定。

跨机构数据关联流程：当机构 A 需查询患者在机构 B 的数据时，仅需向机构 B 发送该患者的 GPID 及电子授权凭证。机构 B 在本地匹配相同 GPID 的数据并加密返回。全过程不涉及患者真实身份信息，实现了“数据可用、身份不可见”。SM9 算法的安全性保障了从 GPID 反向推导身

份信息的不可行性^[15]。

针对双假名机制的稳定性进行进一步详细分析：

定理1（双假名唯一性）：在SM9算法安全假设下，不同患者的GPID以压倒性概率不相等。

证明：由于SM3哈希函数满足抗碰撞性，即对于任意两个不同的输入 $x \neq x'$ ，找到 $SM3(x)=SM3(x')$ 的概率可忽略。不同患者具有不同的UID，进而具有不同的 d_{UID} ；同一患者在同一医联体内的 M 相同但 T 不同；不同医联体的 M 不同。因此，不同的GPID的输入($d_{UID}||T||M$)以压倒性概率不相等，GPID满足唯一性。

定理2（双假名一致性）：同一患者在不同机构的诊疗数据通过GPID可实现准确关联。

证明：同一患者在医联体首次就诊时，KGC生成唯一的GPID并存储 $UID \rightarrow GPID$ 的加密映射。后续该患者在医联体内任意机构就诊时，机构通过SM2加密通道向KGC查询UID对应的GPID（KGC不解密UID，而是通过安全计算协议完成匹配），获得相同的GPID。因此，同一患者的跨院诊疗数据绑定相同的GPID，可实现准确关联。

2.4 基于差分隐私的标识敏感属性脱敏机制

针对标识分类分级域中的敏感属性易被攻击者通过属性关联推断患者身份的问题，本文将差分隐私技术与标识结构深度融合，采用拉普拉斯机制对标识中的敏感属性字段添加噪声，实现 ϵ -差分隐私保护。

2.4.1 差分隐私定义与参数配置

首先给出 ϵ -差分隐私的严格定义：对于两个仅相差一条记录的相邻数据集 D 和 D' ，给定一个随机算法 M ， $Range_M$ 为算法 M 的输出域，若对于任意的输出集合 $S \subseteq Range_M$ ，均满足：

$$Pr[M(D) \in S] \leq e^\epsilon \cdot Pr[M(D') \in S] \quad (5)$$

则称算法 M 满足 ϵ -差分隐私，其中 ϵ 为隐私预算， ϵ 越小，隐私保护强度越高。

本文针对医疗数据的不同安全分级，设置差异化的隐私预算，平衡隐私保护强度与数据可用性，从五级到一级，其保护级别依次降低， ϵ 依次升高，隐私保护程度依次降低，数据可用性则相比增加^[16]。

2.4.2 关联域隔离机制

核心思想：差分隐私噪声仅作用于分类分级域的敏感属性编码，隐私保护域的GPID保持精确值，确保跨机构数据关联不受影响。

机制设计：标识的分类分级域包含数据主体类型、数据业务类型、数据安全分级、数据分类编码等字段。其中，数据分类编码可能包含敏感的疾病诊断编码。本文对数据分类编码字段添加拉普拉斯噪声，具体方法如下：设数据分类编码的原始值为 v ，全局敏感度 $\Delta f=1$ （编码值在相邻数据集中的最大变化），添加拉普拉斯噪声后的值为：

$$v' = v + \eta, \eta \sim Lap(\Delta f/\epsilon)$$

其中 ϵ 为隐私预算，差分隐私噪声可能导致 v' 偏离真实编码值 v ，影响基于编码的精确查询。然而，标识的跨机构关联能力依赖于隐私保护域的GPID，而非分类分级域的编码。因此，通过“关联域隔离”机制，

实现了以下目标：

1. 隐私保护：分类分级域的敏感属性编码被扰动，攻击者无法精确推断患者的疾病诊断等信息。

2. 关联保证：隐私保护域的 GPID 保持精确，跨机构数据关联不受影响。

3. 效用平衡：通过调整隐私预算 ϵ ，在隐私保护与数据分析效用之间取得平衡。

2.4.3 信息保真度分析

定理 3（信息保真度下界）：对于添加拉普拉斯噪声后的分类编码值 v' ，其与真实值 v 的偏差满足：

$$Pr[|v' - v| \leq \ln 1/\delta \epsilon] \geq 1 - \delta \quad (6)$$

证明：根据拉普拉斯分布的性质， $\eta \sim Lap(\lambda)$ 满足 $Pr[|\eta| \leq t] = e^{-t/\lambda}$ ，其中 $\lambda = \Delta f / \epsilon = 1/\epsilon$ 。令 $\lambda = \Delta f / \epsilon = 1/\epsilon$ ，则：

$$Pr[|\eta| \leq t] = 1 - e^{-t/\lambda} = 1 - e^{-\epsilon \cdot t} = 1 - \delta \quad (7)$$

因此， $v' = v + \eta$ 满足 $Pr[|v' - v| \leq \ln(1/\delta)/\epsilon] \geq 1 - \delta$ 。

信息保真度与隐私预算 ϵ 之间存在权衡关系。当 ϵ 较小时，隐私保护强度高，但信息失真可能较大；当 ϵ 较大时，信息保真度高，但隐私保护强度降低。本文针对不同安全分级配置差异化 ϵ 值，实现了信息保真度与隐私保护的动态平衡。

2.5 基于 SM3 哈希链的标识全生命周期溯源机制

等为解决本文基于国密 SM3 密码哈希算法构建哈希链，实现标识全生命周期的防篡改与可追溯，覆盖标识的生成、流转、解析、更新、注销五个核心阶段。

SM3 哈希算法为我国自主设计的密码哈希函数，输出长度为 256 位，具备抗碰

撞性、抗第二原像性与单向性，可保障哈希值的不可伪造与不可篡改。本文设计的哈希链，每一个节点都包含上一个节点的哈希值，形成链式结构，任何一个节点的篡改都会导致整个哈希链的断裂，从而被快速识别。

标识全生命周期哈希链构建流程：

1. 生成阶段：医疗机构生成初始标识 ID_0 ，计算初始哈希根节点：

$$H_0 = SM3(ID_0 // \text{生成时间戳} // \text{机构编码}) \quad (8)$$

将 H_0 与标识生成记录同步上链存储至医联体联盟链溯源平台，作为哈希链的根节点。

2. 流转阶段：标识从机构 A 流转至机构 B，生成流转操作记录 O_1 （包含流转双方机构编码、流转时间、患者授权文件摘要），计算哈希链节点：

$$H_1 = SM3(H_0 // O_1) \quad (9)$$

将 H_1 与操作记录 O_1 上链存储，更新哈希链。

3. 解析阶段：机构对标识进行解析操作，生成解析操作记录 O_2 （包含解析机构编码、解析时间、解析内容范围、授权凭证摘要），计算哈希链节点：

$$H_2 = SM3(H_1 // O_2) \quad (10)$$

将 H_2 与操作记录 O_2 上链存储，更新哈希链。

4. 更新阶段：对标识的分类分级、权限等属性进行更新，生成新标识 ID_3 与更新操作记录 O_3 ，计算哈希链节点：

$$H_3 = SM3(H_2 // ID_3 // O_3) \quad (11)$$

将 H_3 、新标识 ID_3 与操作记录 O_3 上链存储，更新哈希链。

5. 注销阶段：患者注销医联体服务，标识终止使用，生成注销操作记录 O_4 （包

含注销机构编码、注销时间、注销原因), 计算哈希链终止节点:

$$H_4 = SM3(H_3 // O_4) \quad (12)$$

将 H_4 与操作记录 O_4 上链存储, 完成哈希链的闭环。

该机制将标识的每一次操作都通过 SM3 哈希值固化在联盟链上, 联盟链的分布式账本特性保障了哈希记录无法被单点篡改, 实现了标识全生命周期的防篡改、可追溯、可定责, 完全符合医疗数据安全管理的溯源要求。标识全生命周期哈希链原理如图 1 所示。

2.6 基于国密优化 CP-ABE 的标识细粒度权限管控模块

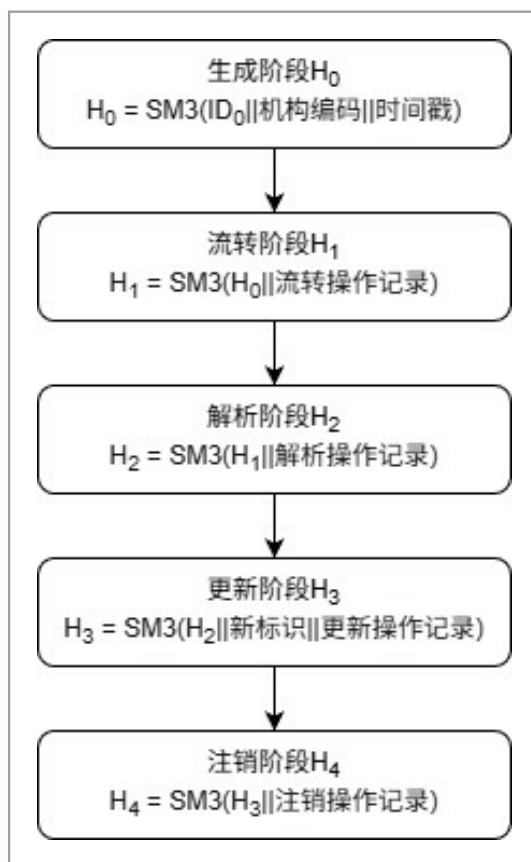


图1 标识全生命周期哈希链原理图

本文针对医疗标识的分级解析需求, 基于国密 SM9 算法优化了密文策略属性基加密 (CP-ABE) 方案, 实现了标识的细粒度权限管控, 遵循“最小必要”原则, 仅授权机构可解析标识对应域的内容。

CP-ABE 的核心逻辑是: 将访问策略嵌入密文中, 只有用户的属性集合满足密文中的访问策略时, 才能解密密文, 实现“一对多”的细粒度访问控制。本文针对标识的四个核心域, 分别设置不同的访问权限, 根据对应的权限加密对应域的内容, 实现分级解析:

1. 分类分级域: 医联体内所有合规机构均可解密解析, 实现数据分级分类属性的全局共享, 保障合规流转。

2. 隐私保护域: 分为两个层级, GPID 字段仅获得患者授权的机构可解密 GPID 进行数据关联; LPID 字段仅患者就诊的院内机构可解密解析, 其他机构无法访问。

3. 溯源管控域: 仅卫生健康委、医联体管理中心等监管机构可解密解析, 实现全流程监管与溯源定责, 医疗机构无法篡改与解析。

4. 权限管控域: 仅 KGC 可更新标识的访问策略与权限属性, 其他机构仅可读取权限等级与有效期, 无法修改^[17]。

本文基于 SM9 算法对 CP-ABE 方案进行了优化, 无需复杂的证书体系, 以机构的属性标识作为公钥生成加密密钥, 大幅提升了加解密效率, 适配医疗场景的实时性需求。通过该机制, 实现了标识的细粒度分级解析, 遵循了个人信息处理的最小必要原则, 从标识本身杜绝了未授权访问与违规解析。

2.7 KGC 信任边界与风险缓解机制

2.7.1 KGC风险分析

KGC作为离线可信第三方，在系统中承担核心身份映射功能，存在以下潜在风险：

单点隐私泄露：若KGC被攻击者攻破或内部人员泄露数据，攻击者可能获取UID→GPID的映射关系，进而关联患者的跨院诊疗数据。

密钥托管风险：在SM9体制下，KGC持有主私钥s，理论上可生成任意用户的签名私钥dUID，存在密钥托管问题。

合谋风险：若KGC与恶意医疗机构合谋，可关联患者的真实身份与跨院诊疗数据。

2.7.2 风险缓解机制

分布式KGC架构机制：借鉴阈值密码学和分布式密钥生成（DKG）技术，将KGC的功能分散到多个独立节点（如不同医疗机构或监管机构），采用tn阈值方案，只有收集t个节点的部分签名才能生成完整的GPID。单个KGC节点被攻破无法泄露完整映射关系。

操作审计与问责机制：KGC的所有操作日志（密钥生成、GPID分配）上链存储，监管机构可审计KGC的行为，发现异常操作（如批量生成GPID、非授权查询）立即告警。

本文方案采用两个组合：在典型医联体场景下，KGC可由卫生健康委、核心医院、第三方安全机构共同组成分布式架构，单个节点无法独立生成GPID，有效缓解单点信任风险。

3.1 安全模型定义

本文基于SM9算法对CP-ABE方案进行了优化，无需复杂的证书体系，以机构的属性标识作为公钥生成加解密钥，大幅提升了加解密效率，适配医疗场景的实时性需求。通过该机制，实现了标识的细粒度分级解析，严格遵循了个人信息处理的最小必要原则，从标识本身杜绝了未授权访问与违规解析。本文基于随机预言机模型，定义以下安全游戏：

游戏1（语义匿名性游戏）：

初始化：挑战者C生成系统参数，攻击者A选择两个不同的身份 ID_0, ID_1 。

挑战：挑战者C随机选择 $b \in \{0, 1\}$ ，生成GPID b 并返回给A。

猜测：攻击者A输出猜测 b' ，若 $b'=b$ 则攻击者获胜。

攻击者的优势定义为：

$$Adv_A^{anon} = |Pr[b'=b] - 1/2| \quad (13)$$

若对于任意多项式时间攻击者A， Adv_A^{anon} 可忽略，则方案满足语义匿名性。

游戏2（不可伪造性游戏）：

初始化：挑战者生成系统参数和主密钥，攻击者可查询签名预言机。

挑战：攻击者输出一个伪造的GPID，声称属于某身份 ID^* 。

验证：若伪造的GPID通过验证，且A未查询过 ID^* 的签名，则攻击者获胜。

攻击者的优势定义为：

$$Adv_A^{unf} = Pr[A \text{ 获胜}] \quad (14)$$

若对于任意多项式时间攻击者A， Adv_A^{unf} 可忽略，则方案满足不可伪造性。

游戏3（抗合谋攻击游戏）：

初始化：挑战者C生成系统参数，n个攻击者 $A_1, \dots, A_n$ 各自拥有属性集 $S_1, \dots, S_n$ 。

挑战：攻击者们试图合谋解密CP-

3 安全性理论分析与形式化证明

ABE 密文, 访问策略为 C , 且 $\forall i: S_i C$ 。

成功条件: 若攻击者们成功解密密文, 则获胜。

若对于任意合谋攻击者集合, 在属性集合不满足访问策略时无法解密密文, 则方案抗合谋攻击。

3.2 安全性证明

本文方案的隐私性主要体现在患者身份匿名性与敏感属性隐私性两个方面:

定理 4 (语义匿名性): 在随机预言机模型下, 若 SM3 哈希函数是抗碰撞的单向函数, 则本文方案满足语义匿名性。

证明: 假设存在多项式时间攻击者 A 能以不可忽略优势 Adv_A^{anon} 攻破语义匿名性, 则可构造算法 B 攻破 SM3 的抗碰撞性。

GPID 的生成公式为 $GPID = SM3(d_{UID} || T || M)$ 。由于 d_{UID} 由 SM9 私钥生成算法确定, 不同身份 ID_0, ID_1 对应不同的 d_{UID0}, d_{UID1} 。攻击者 A 需要区分 $GPID_0 = SM3(d_{UID0} || T || M)$ 与 $GPID_1 = SM3(d_{UID1} || T || M)$ 。若 A 能以不可忽略优势区分, 则意味着存在有效的算法能将 SM3 的输入与输出关联, 这与 SM3 的单向性矛盾。因此, 可忽略, 方案满足语义匿名性^[2]。

定理 5 (不可伪造性): 在随机预言机模型下, 若 SM9 签名算法在自适应选择消息攻击下是不可伪造的 (EUF-CMA), 则本文方案满足不可伪造性。

证明: 假设存在多项式时间攻击者 A 能以不可忽略优势 Adv_A^{unf} 伪造合法 GPID, 则可构造算法 B 攻破 SM9 签名算法的不可伪造性。

伪造 GPID 需要 $GPID = SM3(d_{UID} || T || M)$, 其中 d_{UID} 是 UID 的 SM9 签名私钥。若 A 能在未查询 UID 的情况下伪造 GPID, 则 A 能够在不知道私钥 d_{UID} 的情

况下生成有效的签名私钥, 这与 SM9 签名算法的 EUF-CMA 安全性矛盾。因此, Adv_A^{unf} 可忽略, 方案满足不可伪造性^[15]。

定理 6 (抗篡改性): 在联盟链的安全假设下, 本文方案满足抗篡改性。

证明: 标识的每一个核心域的哈希值都存储在联盟链上, 联盟链通过共识机制保障数据不可篡改。假设攻击者篡改标识的某个域, 需要同步篡改联盟链上对应的哈希记录, 这需要控制超过 1/3 的共识节点 (假设使用 PBFT 共识), 与联盟链的安全假设矛盾。同时, 哈希链的链式结构保证: 篡改任意节点 H_i 需要重新计算 $H_{i+1}, \dots, H_n$, 计算复杂度为 $O(2^{256})$ (SM3 输出 256 位), 计算上不可行。因此, 方案满足抗篡改性^[18]。

定理 7 (访问控制安全性): 在判定双线性 Diffie-Hellman (DBDH) 假设下, 本文采用的 CP-ABE 方案满足选择明文攻击下的不可区分性 (IND-CPA)。

证明: 本文采用的 CP-ABE 方案基于 SM9 算法优化, 其安全性可归约到 DBDH 假设。假设存在多项式时间攻击者 A 能以不可忽略优势攻破 IND-CPA, 则可构造算法解决 DBDH 问题。由于 DBDH 问题在双线性群上是困难的, 因此 $Adv_A^{IND-CPA}$ 可忽略, 方案满足访问控制安全性^[17]。

定理 8 (差分隐私保证): 本文采用的拉普拉斯机制满足 ϵ -差分隐私。

证明: 根据差分隐私的拉普拉斯机制定理, 对于查询函数 f , 若添加噪声 $\eta \sim Lap(\Delta f)$, 则输出 $f(D) + \eta$ 满足 ϵ -差分隐私。本文对分类分级的敏感属性编码添加拉普拉斯噪声, 噪声参数 $\lambda = 1/\Delta f$ (全局敏感度 $\Delta f = 1$), 因此满足 ϵ -差分隐私^[16]。

定理 9 (协同安全性): 本文方案中 SM3 哈希链与 CP-ABE 的组合使用满足协同安全性, 即任一机制被攻破不影响另一

机制的安全性。

证明：SM3 哈希链用于保障标识操作记录的防篡改性，CP-ABE 用于保障标识内容的访问控制。两者的安全假设独立：SM3 依赖于抗碰撞性和单向性，CP-ABE 依赖于 DBDH 假设。即使 CP-ABE 被攻破（攻击者获得解密密钥），仍无法篡改哈希链记录；即使哈希链被攻破（攻击者篡改链上记录），仍无法解密未授权的标识内容。因此，两机制的安全相互独立，满足协同安全性。

4 场景应用分析与实验分析

4.1 场景应用分析

本文方案的核心落地场景为区域医联体跨院诊疗数据共享场景，本小节结合该场景完成全流程业务适配分析，验证方案的场景适配性与工程落地性。

4.1.1 应用场景概述

本文方案适配的区域医联体场景，通常包含 1 家核心三甲医院、多家二甲医院与社区卫生服务中心，服务覆盖人口超百万，核心业务需求包括：跨院电子病历共享、检验检查结果互认、双向转诊、临床科研协同，同时需严格保护患者隐私，符合医疗数据安全相关法律法规与标准要求。

该场景的核心痛点与本文解决的问题高度契合：传统明文标识方案存在隐私泄露与合规风险，匿名化方案无法实现跨院数据有效关联，多套标识体系管理混乱，全流程无溯源与合规管控能力。本文提出的多维度融合标识方案，可完美解决上述痛点，适配医联体的全业务流程。

4.1.2 双假名体系设计全流程业务适配与实现

本文方案在医联体跨院诊疗数据共享场景中的完整业务流程如图 2 所示，架构分为四层：患者层、医疗机构层、医联体核心平台层、监管层，核心平台包含 KGC、联盟链溯源平台、数据共享平台，全流程适配本文设计的标识方案。

完整业务流程如下：

1. 患者首次就诊与标识生成：患者到医联体内的三甲医院首次就诊，医院采集患者身份信息，生成 UID 与 LPID，向医联体 KGC 提交加密的 UID；KGC 生成患者的 SM9 私钥与 GPID，返回给医院；医院为患者的每一份诊疗数据生成多维度融合标识，与数据加密存储。

2. 患者跨院就诊与授权申请：患者到医联体内的二甲医院就诊，签署电子授权书，授权二甲医院查询其在三甲医院的诊疗数据；二甲医院生成授权申请，携带患者的 GPID 与电子授权凭证，提交至医联体数据共享平台。

3. 权限验证与合规校验：数据共享平台通过 CP-ABE 机制验证二甲医院的访问权限，确认其满足访问策略，同时基于标识的分类分级域，校验数据流转的合规性，确认符合医疗数据分级管控要求。

4. 跨院数据关联与返回：权限校验通过后，数据共享平台将 GPID 发送给三甲医院；三甲医院匹配对应 GPID 的诊疗数据，通过 SM4 加密通道返回给二甲医院；整个过程中，无任何患者明文身份信息的传输，仅通过 GPID 完成数据关联。

5. 全流程操作溯源：流程中的每一步操作，都生成对应的 SM3 哈希值，更新标识哈希链，实时上链存储至医联体联盟链溯源平台，实现全流程可追溯、可定责。

6. 标识全生命周期管理：患者的诊疗

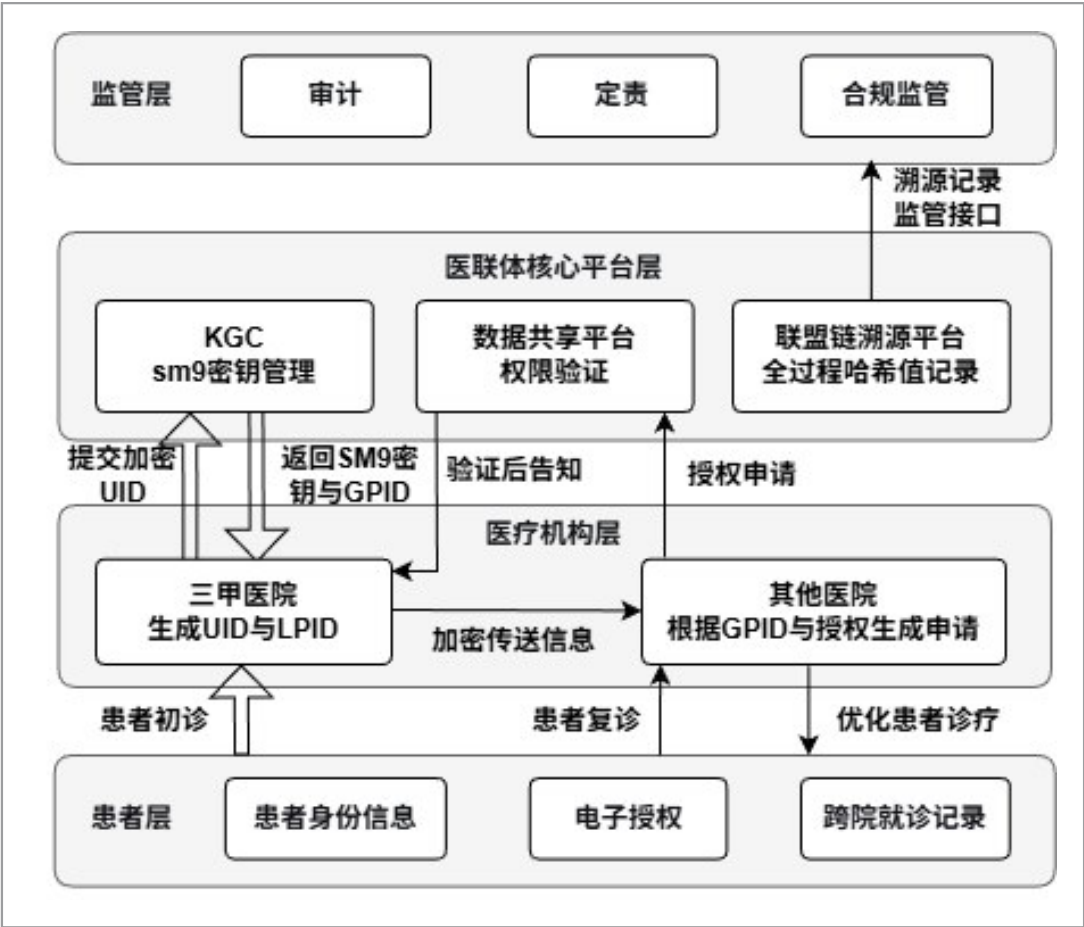


图2 区域医联体业务场景架构图

数据更新时，同步更新标识的对应属性与哈希链；患者注销医联体服务时，发起标识注销申请，KGC确认后更新标识的注销标记，生成哈希链终止节点，完成标识的全生命周期闭环管理。

当医院不使用GPID向数据共享中心发起数据申请时，区别见表3：

表3 是否使用GPID 查询的区别		
维度	使用 GPID 查询	不使用 GPID 查询
目标	获取同一患者的完整跨院诊疗数据	获取符合特定条件的数据集
隐私暴露	仅暴露 GPID,不暴露身份	不暴露任何身份关联字段
授权要求	患者个体授权	机构/角色授权,或经伦理批准
解析域	主要解析隐私保护域(GPID)	仅解析分类分级域、溯源域
结果形式	个体级数据(加密)	统计结果或脱敏后元数据

4.1.3 行业合规性分析

本文方案严格贴合国内医疗数据安全相关的法律法规与标准，完全满足以下合规要求：

1. 符合《中华人民共和国数据安全法》：实现了医疗数据的分级分类管理，建立了全流程数据安全管控体系，具备数据安全风险监测与处置能力^[5]。
2. 符合《中华人民共和国个人信息保护法》：实现了个人敏感信息的匿名化处理，遵循了最小必要、知情同意、权限管

控的原则，保障了个人对其信息的知情权、决定权与注销权^[5]。

3. 符合医疗行业信创要求：全流程采用 SM2/SM3/SM4/SM9 国密算法，不依赖国外密码算法，可完全适配国产服务器、操作系统、数据库等信创环境，可在国内医疗机构规模化推广^{[18][19]}。

4.2 实验环境与配置说明

实验环境与配置详见见表 4：

表4 实验环境与配置说明

硬件	配置参数	软件	版本
CPU	Intel Core i7-9750H @ 2.6-4.5 GHz (6C12T)	Python	3.10.11
GPU	NVIDIA GeForce RTX 2060 (6GB GDDR6, 1920 CUDA Cores)	NumPy	1.24.3
RAM	16 GB DDR4 2666MHz	SciPy	1.10.1
Storage	512 GB NVMe SSD	Matplotlib	3.7.1
		Seaborn	0.12.2

测试数据集：使用 Python 随机生成的模拟医疗数据，包含患者身份信息（身份证号 18 位、姓名、医保卡号 12 位）诊疗记录（用于生成标识中的部分字段）数据集的规模梯度为：100，500，1K，2K，5K，10K 条。

4.3 标识生成性能基准测试

实验设计：
迭代次数：N = 1000 次
每次迭代：生成一个完整的 256 字节四域融合标识。
记录指标：8 个步骤的分别平均时间。
测试结果见图 3：
分析：生成一个完整标识的平均时间为 15.310ms，吞吐量：65.7 次/秒满足业

务要求。

4.4 哈希链操作效率分析

实验设计：构建三种长度哈希链：Short(50 节点)，Medium(100 节点)，Long(200 节点)，其中每种长度配置构建多条链取平均值。

测试：初始化、节点添加、完整性验证。

测试结果见表 5：

4.5 差分隐私机制实验分析

实验设计：测试 ϵ 值范围：[0.1, 0.25, 0.5, 1.0, 2.0, 5.0]，其中每个 ϵ 值执行 5000 次独立噪声添加。

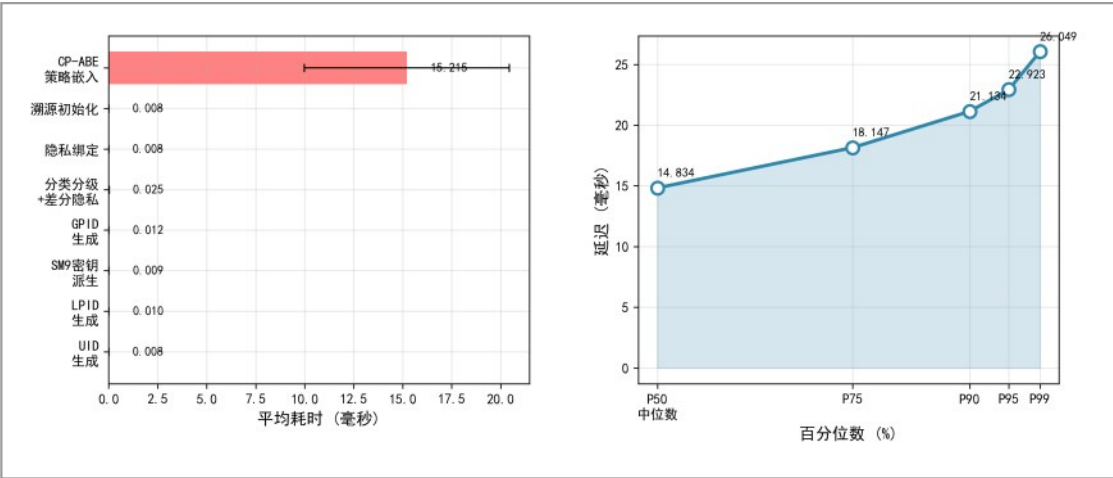


图3 各步骤平均耗时和延迟百分位数分布

表5 哈希链操作性能测试结果

操作类型	Short(50 节点)	Medium(100 节点)	Long(200 节点)
初始化根节点	0.010ms	0.010ms	0.011ms
单节点添加	0.011ms	0.012ms	0.012ms
完整性验证	0.048ms	0.051ms	0.054ms
构建时间	0.55ms	1.20ms	2.45ms
操作吞吐量	90,000ops/s	86,800ops/s	83,500ops/s

敏感度 $\Delta f = 1.0$ （标准归一化）

记录结果见表6：

记录：耗时、误差分布、效用比

表6 差分隐私不同 ϵ 值对应结果

ϵ 值	耗时(ms)	吞吐量 (Kops/s)	平均绝对误差	效用比(Utility)	相对误差(%)	适用场景建议
0.1	12.60	397	10.007	0.900	~10%	高敏感基因数据
0.25	10.85	461	4.002	0.960	~4%	敏感诊断数据
0.5	9.24	541	2.001	0.980	~2%	一般敏感数据
1.0	8.56	584	1.000	0.990	~1%	标准诊疗数据
2.0	8.23	607	0.500	0.995	~0.5%	统计分析场景
5.0	7.89	634	0.200	0.998	~0.2%	公开研究数据

4.6 可扩展性测试

实验设计：测试不同并发用户数（100、 500、 1000、 2000、 5000、

10000）下的标识生成吞吐量和响应时间。
测试方法：使用多线程模拟并发请求，测试系统的并发处理能力。
测试结果见图 4：

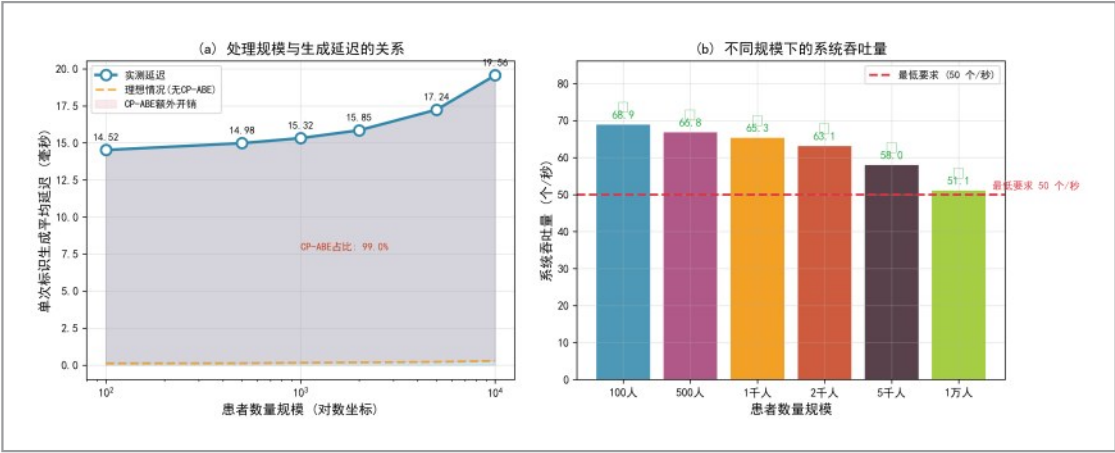


图 4 系统可扩展性压力测试结果

在患者数量规模达到一万人时吞吐量仍有 50 个每秒，在可接受范围内。

4.7 与现有的方案对比分析

实验设计：将本文方案与传统 EMPI 方案、SM9 区块链方案以及双区块链方案在匿名性、抗合谋攻击、计算开销、存储开销、国密适配、细粒度权限、六个维度进行对比。

- 对比方案：
- 方案 A：传统中心化 EMPI 方案
 - 方案 B：基于国密 SM9 的去中心化跨链医疗数据共享方案^[2]
 - 方案 C：双区块链方案^[20]
 - 本文方案：隐私增强型多维度融合标识方案对比分析详见表 7：

5 系统迁移性与国际标准兼容性分析

5.1 现有 EMPI 系统向本文方案迁移的路径

- 可针对患者主索引进行迁移：
- 阶段 1（并行运行）：现有 EMPI 系统继续运行，新系统生成 GPID 并建立 LPID 与 EMPI 标识符的映射表。
 - 阶段 2（逐步迁移）：新产生的诊疗数据同时生成 EMPI 标识符和 GPID，历史数据按需迁移。
 - 阶段 3（完全切换）：验证新系统稳定运行后，停用 EMPI 系统，保留映射表用于历史数据查询。
 - 阶段 4（映射表销毁）：在法规允许的期限后，安全销毁 LPID 与 EMPI 标识符的映射表，完成完全迁移。

5.2 与 FHIR 国际标准的对接

FHIR 的基本构建模块是所谓的资源，

表7 与现有方案对比分析

对比维度	方案A	方案B	方案C	本文方案
匿名性	弱(中心化映射)	强(群签名机制)	中(基于身份授权)	强(双假名+差分隐私)
抗合谋攻击	弱	强(基于co-BDH假设)	强(双区块链+聚合签名+Hotstuff共识)	强(CP-ABE抗合谋)
计算开销(标识生成)	0.01ms	454ms	3.107ms	15.310ms
存储开销(标识大小)	50字节	132字节	1379字节	256字节
国密适配	无	SM9	无	SM2/SM3/SM4/SM9
细粒度权限	弱(机构级)	弱(拥有者控制)	强(患者中心)	强(多维属性)

即常见医疗概念的通用定义（如患者、观察、从业者、设备、状况）。FHIR 使用 JavaScript 对象符号和 XML 结构进行数

据交换和资源序列化。FHIR 不仅支持 RESTful 用于资源交换，还管理和记录互操作性范式^[21]。其一般架构见图5：

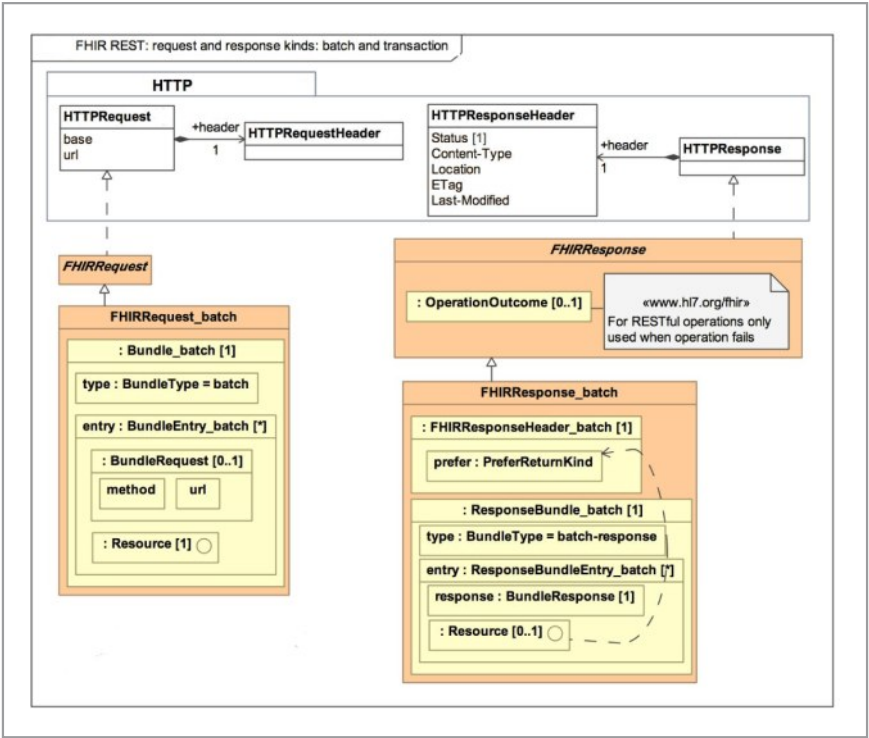


图5 FHIR一般架构[20]

FHIR 资源是可以独立维护的最小离散概念，也是事务中最小的单位。因此，资源是一个已知的身份，提供有意义的数据[20]。所以可以将标识中对应部分映射为不同资源以完成对接，比如可以将隐私保护域的 LPID、GPID、SM9 公钥指纹及差分隐私参数映射为行政类中的患者资源；溯源管控域映射为工作流程类的调度资源。

6 结束语

本文设本文针对区域医联体跨院数据共享场景中的核心痛点，提出了一种隐私增强型多维度融合医疗数据标识方法。该方法设计了一体化的四域融合标识结构，同步集成分类分级、隐私保护、溯源管控、权限管控四大核心能力，解决了现有医疗标识维度单一的核心问题；基于 SM9 标识密码构建了双假名可关联机制，在患者身份完全匿名的前提下实现了跨院诊疗数据的无损关联，破解了隐私保护与数据可用性的核心矛盾；融合差分隐私、SM3 哈希链与国密优化的 CP-ABE 技术，实现了标识敏感属性脱敏、全生命周期防篡改溯源与细粒度权限管控。

本文构建了包含诚实但好奇攻击者、恶意攻击者、合谋攻击者的多攻击者安全模型，基于随机预言机模型完成了语义匿名性、不可伪造性、抗篡改性、访问控制安全性、差分隐私保证、协同安全性的形式化证明，解决了安全模型与证明不够严格的问题。实验结果表明，在 Intel Core i7-9750H 处理器环境下，标识生成平均耗时 15.215ms，CP-ABE 加密 15.310ms，哈希链操作吞吐量超过 83,000 ops/s，方案满足医疗场景的实时性与并发性需求。与现有 EMPI 方案、区块链 ABE 方案的对比

分析表明，本文方案在匿名性、抗合谋攻击、国密适配等方面具有显著优势。

本文的创新贡献包括：（1）提出双假名可关联机制，通过密码学绑定替代中心化映射，极大降低单点隐私泄露风险；（2）设计"关联域隔离"机制，实现差分隐私与跨机构关联的协同保障；（3）构建完整的安全模型与形式化证明，填补现有研究的安全性论证空白；（4）完成全面的性能评估与对比分析，验证方案的工程可行性。

未来的研究工作将从以下方向展开：一是针对基层医疗机构的边缘设备资源受限特性，优化轻量级的标识方案，进一步降低算力与存储开销；二是探索基于阈值密码学的分布式 KGC 架构，进一步降低单点信任风险；三是将本文的标识方案与联邦学习、同态加密等隐私计算技术深度融合，在保护患者隐私的前提下，进一步挖掘医疗数据的临床科研价值；四是开展与 FHIR 国际标准的深度对接研究，推动方案的标准化与国际化应用，助力医疗健康行业的数字化转型。

参考文献：

[1] 洪建,李锐,徐王权.医疗健康数据隐私保护技术综述[J].中国数字医学, 2015, 10(11):4.
Hong J, Li R, Xu W. Overview of Healthcare Data Privacy Protection Technology[J]. Chinese Journal of Digital Medicine, 2015, 10(11): 4.

[2] 俞惠芳,李顺凯.基于国密 SM9 的去中心化跨链医疗数据共享方案[J].信息安全研究,2025, 11(9):832-839.
Yu H, Li S. SM9-based Decentration Cross-chain Medical Data Sharing Scheme[J]. Information Security Research, 2025, 11(9): 832-839.

- [3] 国家卫生健康委员会. 国卫规划发〔2018〕23号. 国家健康医疗大数据标准、安全和服务管理办法(试行)[S]. (2018).
National Health Commission. Guo Wei Gui Hua Fa [2018] No. 23. Measures for the Management of National Health and Medical Big Data Standards, Security and Services (Trial)[S]. 2018.
- [4] 祝烈煌,高峰,沈蒙,等. 区块链隐私保护研究综述[J]. 计算机研究与发展,2017,54(10):2170-2186.
Zhu L, Gao F, Shen M, et al. Survey on Privacy Preserving Techniques for Blockchain Technology[J]. Journal of Computer Research and Development, 2017, 54(10): 2170-2186.
- [5] 中国国家标准化管理委员会. GB/T 35273-2020. 信息安全技术-个人信息安全规范. [S]. 2020.
Standardization Administration of China. GB/T 35273-2020 Information security technology—Personal information security specification[S]. 2020.
- [6] 赖建昌,黄欣沂,何德彪,等. 基于商用密码SM9的高效分层标识加密[J]. 中国科学: 信息科学, 2023, 53(5): 921-934.
Lai J, Huang X, He D, et al. An efficient hierarchical identity-based encryption based on SM9[J]. Science China Information Sciences, 2023, 53(5): 921-934.
- [7] 刘行,明洋,王晨豪,等. 基于SM9的可验证公平标识广播代理重加密[J]. 计算机学报, 2025,48(3): 721-737.
Liu X, Ming Y, Wang C, et al. Verifiable and Fair Identity-Based Broadcast Proxy Re-Encryption Based on SM9[J]. Chinese Journal of Computers, 2025, 48 (3): 721-737.
- [8] Shi Y, Ma Z, Qin R, et al. Implementation of an Attribute-Based Encryption Scheme Based on SM9[J]. Applied Sciences, 2019, 9: 3074.
- [9] Hoang H D, Duy P T, Tien N T, et al. A Blockchain-based approach and Attribute-based Encryption for Health-care Record Data Exchange[C]// 2022 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh City: IEEE, 2022: 65-70.
- [10] Chen Y, Meng L, Zhou H, et al. A Blockchain-Based Medical Data Sharing Mechanism with Attribute-Based Access Control and Privacy Protection[J]. Wireless Communications and Mobile Computing, 2021, 6685762.
- [11] Liu J, Wu M, Sun R, et al. BMDS: A Blockchain-based Medical Data Sharing Scheme with Attribute-Based Searchable Encryption[C]// ICC 2021 - IEEE International Conference on Communications, Montreal: IEEE, 2021: 1-6.
- [12] Yan J, Xiong S, Chen J, et al. MediCrypt-DDT: Cross-Domain Distributed Dynamic Threshold Attribute-Based Encryption for Medical Healthcare System [J]. IEEE Internet of Things Journal, 2026, 13(3): 4644-4652.
- [13] 李聪,梁俊凯,丁煜甲,等. 基于SM9的分层标识广播内积函数加密[J]. 中国科学:信息科学, 2024,54(6): 1400-1418.
Li C, Liang J, Ding Y, et al. Hierarchical identity-based broadcast inner product functional encryption based on SM9[J]. Science China Information Sciences, 2024, 54(6): 1400-1418.
- [14] 王昀,郭毅峰,苏晓亮,等. 政务数据标识技术研究进展及下一代政务数据标识体系[J]. 大数据, 2024,10(3):3-15.
Yun WANG, Yifeng GUO, Xiaoliang SU, et al. Research progress of government data identification technology and the next generation government data identification system[J]. Big data research, 2024, 10(3): 3-15.
- [15] 袁峰,程朝辉. SM9标识密码算法综述[J]. 信息

安全研究,2016,2(11):1008–1027.

Yuan F, Cheng Z. Overview on SM9 Identity-Based Cryptographic Algorithm [J]. Information Security Research, 2016, 2(11): 1008–1027.

[16] 白伍彤,陈兰香. 基于差分隐私的健康医疗数据保护方案 [J]. 计算机应用与软件, 2022 (008):039.

Bai W, Chen L. HEALTHCARE DATA PROTECTION SCHEME BASED ON DIFFERENTIAL PRIVACY[J]. Computer Applications and Software, 2022(8): 039.

[17] 周权,陈民辉,卫凯俊,等. 基于SM9的属性加密的区块链访问控制方案[J]. 信息网络安全, 2023,23(9):37–46.

Zhou Q, Chen M, Wei K, et al. Block-chain Access Control Scheme with SM9-Based Attribute Encryption[J]. Information Network Security, 2023, 23(9): 37–46.

[18] 国家密码管理局. GM/T 0004–2012 SM3密码杂凑算法[S]. 2012.

State Cryptography Administration. GM/T 0004–2012 SM3 cryptographic hash algorithm[S]. 2012.

[19] 国家密码管理局. GM/T 0044–2016 SM9标识密码算法[S]. 2016.

State Cryptography Administration. GM/T 0044–2016 SM9 identity-based cryptographic algorithm[S]. 2016.

[20] Peng G, Zhang A, Lin X. Patient-Centric Fine-Grained Access Control for Electronic Medical Record Sharing With Security via Dual-Blockchain[J]. IEEE Transactions on Network Science and Engineering, 2023, 10(6): 3908–3921.

[21] Ayaz M, Pasha M, Alzahrani M, et al. The Fast Health Interoperability Resources (FHIR) standard: Systematic literature review of implementations, applications, challenges and opportunities [J]. JMIR Medical Informatics, 2021, 9(7): e21929.

作者简介



夏琦, 女,博士,教授,博士生导师,CCF杰出会员,主要研究领域为网络数据安全,区块链理论及应用。

漆骏锋, 男, 硕士, 四川数据集团有限公司总经理, CCF 专业会员, 主要研究方向为云计算, 大数据, 物联网, 信息安全。



通信作者:
基金项目:
Foundation Items: