基于本地化差分隐私和属性基可搜索加密的区块链数据共享方案

冯涛，陈李秋，方君丽，石建明
（兰州理工大学计算机与通信学院，甘肃 兰州 730050）

摘 要：针对传统基于云的数据共享方案依赖可信第三方、只关注数据隐私保护或访问控制问题，提出一种基于本地化差分隐私和属性基可搜索加密的区块链数据共享方案。将区块链和云服务器结合，链上链下协同存储数据，提供高效可靠防篡改的数据共享。首先，引入本地化差分隐私对共享数据进行预处理，保证数据拥有者身份隐私的同时抵御不可信第三方攻击；其次，将可搜索加密技术和属性基加密结合，支持密文检索实现数据隐私保护、为共享数据提供细粒度访问控制；最后，通过安全性、正确性证明及实验分析证明所提方案满足安全目标。

关键词：区块链；本地化差分隐私；数据共享；属性基可搜索加密；隐私保护
中图分类号：TP309
文献标志码：A
DOI: 10.11959/j.issn.1000−436x.2023103

Blockchain data sharing scheme based on localized difference privacy and attribute-based searchable encryption

FENG Tao, CHEN Liqiu, FANG Junli, SHI Jianming
School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

Abstract: Aiming at the problem that traditional cloud-based data sharing schemes rely on trusted third parties and only focus on data privacy protection or access control, a blockchain data sharing scheme based on localized difference privacy and attribute-based searchable encryption was proposed. The blockchain and cloud server were combined to store data chain by chain and provide efficient, reliable and tamper-proof data sharing. Firstly, the localization difference privacy was introduced to preprocess the shared data to protect the privacy of the data owner and resist the attack of the untrusted third party. Secondly, the searchable encryption technology and attribute-based encryption were combined to realize data privacy protection, support ciphertext retrieval, and provide fine-grained access control for shared data. Finally, the safety, the correctness proof, and the experimental analysis proves that the proposed scheme meets the safety objectives.

Keywords: block chain, localized differential privacy, data sharing, attribute-based searchable encryption, privacy protection

0 引言

近年来，云计算技术不断发展，并且对数据共享的要求越来越严格，目前的数据共享方案要么依赖可信第三方，要么只关注数据隐私保护或访问控制。传统基于云的数据共享方案通常依赖于可信第三方，这些可信第三方可能会存在单点故障、不可追溯、数据维护预算高等问题。

现有研究表明[1]，借助区块链去中心化、数据防篡改以及可追溯的特性，可以为数据共享提供新途径。Fan 等[2]提出了一个基于区块链的信息管理系统 MedBlock 来处理患者的信息，通过区块链实

收稿日期：2022−11−28；**修回日期：**2023−04−21
基金项目：国家自然科学基金资助项目（No.62162039，No.61762060）；甘肃省科技厅重点研发计划基金资助项目（No.20YF3GA016）
Foundation Items: The National Natural Science Foundation of China (No.62162039, No.61762060), The Key Research and Development Program of Gansu Provincial Science and Technology Department (No.20YF3GA016)

现高效安全的医疗数据共享，但该系统的安全性完全依赖于区块链技术，不能抵御不可信第三方攻击。Xia 等[3]提出了一个基于区块链为大数据实体之间云存储库中的共享医疗数据提供数据来源、审计和控制，从而解决医疗大数据管理在无信任环境中共享医疗数据问题的系统，但该系统在共享数据的过程中没有很好地考虑到数据拥有者（Do, data owner）的身份隐私保护。杜瑞忠等[4]针对数据共享方案中的可信第三方问题，构建了区块链环境下的公钥可搜索加密方案，解决了私有云环境中一对多的数据分享问题，但该方案应用场景单一，不具有可扩展性。

在数据共享方案中，为了保护需要共享的数据，首先要解决的是隐私保护问题。在以往的研究中，研究者提出了 K-anonymity[5-6]、L-diversity[7]、T-closeness[8]、M-invariance[9]的隐私保护模型。由于这些模型不能抵御攻击者的背景知识，于是出现了差分隐私（DP, differential privacy）技术。Hassan 等[10]对区块链中应用差分隐私技术进行了全面调查，并讨论了将差分隐私应用于区块链的各种研究方向，验证了在区块链中引入差分隐私是目前解决身份隐私问题的最佳方案。但传统差分隐私技术通常假设第三方服务器是可信的，不会窃取或者泄露数据拥有者的个人敏感信息。但在实际应用中，真正可信的第三方不成立，本地化差分隐私（LDP, localized differential privacy）技术作为传统差分隐私技术的一种改进技术[11-12]，可使每个数据拥有者直接对数据进行扰动，将数据隐私化处理掌握在自己手中。由于上传的共享数据不是自己的原始数据，而是对原始数据进行扰动后的数据，因此即使第三方机构不可信，获得的数据也不会泄露数据拥有者的身份隐私，直接杜绝了不可信第三方的攻击，从而在根源上保证了对数据拥有者身份隐私的保护。

对于身份隐私保护，若敏感数据泄露，则会暴露数据拥有者的身份，使数据拥有者不愿意共享数据。Liu 等[13]模拟政府部门间共享统计数据的场景, 在区块链环境下提出了一种基于本地化差分隐私的政务数据共享方法，这种方法适用于匿名投票、民意调查等，可以很好地保护数据拥有者的身份信息，但该方法没有提供细粒度访问控制。Sun 等[14]提出了一种基于区块链的两级隐私保护机制，在第一个阶段，利用本地化差分隐私干扰轨迹的位置信息，以此保护身份隐私；在第二个阶段，所有的传感数据通过边缘节点上传到区块链，由边缘云处理并反馈给请求者，以此保护数据隐私，该方案安全性得到了保证，但验证效率降低了。

对于数据隐私保护，研究者提出了可搜索加密（SE, searchable encryption）技术[15]对关键字进行加密和搜索，以此保证数据隐私及数据可用性。Tang 等[16]使用智能合约为区块链上的多个数据拥有者构建全局加密搜索索引和验证索引，提供了功能齐全的访问控制，以确保患者和医疗机构可以控制对数据的访问，但该方案无法保障数据拥有者的身份信息。牛淑芬等[17]利用 SE 技术和属性基加密（ABE, attribute-based encryption）技术，设计了一个密文策略的属性基 SE 方案，实现了对加密数据的有效搜索和细粒度访问授权，但容易产生单点故障。

对于细粒度访问控制，ABE 技术[18]被认为是提供细粒度访问控制最有效的解决方案之一。Ding 等[19]提出了一种在区块链模型下可以进行细粒度访问控制的属性基加密安全访问控制模型，并通过智能合约将所提模型应用于区块链，实现了对区块链用户的访问控制，但不能抵抗不可信第三方攻击。

近年来，有研究者开始将数据共享服务中的隐私保护及访问控制相结合考虑。徐红[20]针对隐私保护及访问控制问题设计了一个基于差分隐私和属性基加密的云数据安全共享机制，避免了未授权用户访问存储在第三方的共享数据，保障了共享数据的安全，但单点故障及不可溯源问题未得到解决。Chen 等[21]提出了一种基于区块链并且具有基于属性的访问控制和隐私保护的医疗数据共享机制，该机制利用 K-anonymity 和可搜索加密相结合，实现了身份及数据的隐私保护，并可应用在区块链上，利用智能合约实现基于身份的访问控制，但该机制不能抵御不可信第三方攻击。

综上所述，虽然目前已有许多研究者提出基于区块链的多种数据共享方案，但这些方案大多只考虑共享数据过程中的隐私保护或访问控制问题，很少有方案将这 2 个问题一起结合考虑。本文在此研究基础上，将隐私保护细分为身份隐私保护和数据隐私保护，构建了区块链环境下的本地化差分隐私和属性基可搜索加密的数据共享方案，并通过理论安全性、正确性证明及实验分析证明了本文方案满足安全目标。

针对数据共享过程中依赖可信第三方的隐私保护和访问控制问题，本文构建了区块链环境下的本地化差分隐私和属性基可搜索加密的数据共享方案，主要研究工作如下。

1) 提出了基于本地化差分隐私和属性基加密的区块链数据共享方案，并将隐私保护细分为身份隐私及数据隐私。首先，利用 LDP 技术中的 RAPPOR 方法对数据进行预处理，以模糊可能反映数据拥有者（DO）身份的敏感数据，确保 DO 身份隐私保护并抵御不可信第三方攻击；然后，利用 SE 技术，实现对加密关键字的搜索，提供数据隐私保护。

2) 考虑到访问控制问题，本文利用 ABE 技术为共享数据提供灵活的细粒度访问控制。数据访问者（DV, data visitor）可根据感兴趣关键字产生搜索陷门，将相关陷门上传至区块链并进行搜索等操作。DO 有权决定数据能否被访问，可以有效保证共享数据的安全。

3) 利用区块链和云服务器链上链下协同存储数据，对包含关键字的数据进行索引，将索引和搜索工作放到区块链，扰动后的敏感数据及数据密文存储在云服务器上，解决数据共享依赖可信第三方的问题，保证搜索结果的不可篡改性及溯源性。

1 预备知识

1.1 双线性映射

1) 双线性。对任意的 $x,y\in G_1$ 和 $a,b\in G_T$，有 $e(x^a,y^b)=e(x^b,y^a)=e(x,y)^{ab}$。

2) 非退化性。存在 $g\in G_1$，使 $e(g,g)\neq 1$。

3) 可计算性。对所有的 $x,y\in G_1$，存在有效算法计算 $e(x,y)$。

1.2 判定性双线性 Diffie-Hellman 假设

设 G_1、G_2 是阶为素数 p 的循环群，双线性映射 $G_1\times G_1\to G_2$，g 是群 G_1 的生成元，随机生成 g^a、g^b、g^c 且 $(a,b,c)\leftarrow Z_q^*$，对于四元组 (g,g^a,g^b,g^c)，不存在概率多项式时间敌手以不可忽略的优势区分 $e(g,g)^{abc}\in G_2$。

1.3 访问结构

令 $P=\{P_1,P_2,\cdots,P_n\}$ 表示基于属性加密方案的实体集，若存在访问结构 $A\subseteq 2^P$，$\forall B,C$，若 $B\in T$ 且 $B\subseteq C$，则有 $C\in T$，称 T 是单调的访问结构。访问结构 A 是 $P=\{P_1,P_2,\cdots,P_n\}$ 的非空子集。在访问结构 T 中的集合为授权访问集合，不在访问结构 T 中的集合为非授权访问集合。

1.4 线性秘密共享

定义实体集 P 上的秘密共享方案在 Z_p 上是线性的，且满足以下条件。

1) 各方的秘密组成域 Z_p 上的矩阵。定义 P 为 $(\boldsymbol{A},\ \rho)$，其中，$\boldsymbol{A}$ 表示一个 $l\times n$ 矩阵，ρ 表示 $\boldsymbol{A}$ 的每一行映射到相应属性的映射函数。对于所有的 $1\leqslant i\leqslant l$，有 $\boldsymbol{A}_i$ 为 $\boldsymbol{A}$ 的第 i 行向量，ρ_i 为 $\boldsymbol{A}$ 的第 i 行参与方标识。随机选取一个向量 $\boldsymbol{v}=(s,y_2,\cdots,y_n)$，其中，$s$ 为共享的秘密值，$y_2,\cdots,y_n$ 为随机值，则 $\boldsymbol{Av}$ 为利用秘密共享得到的关于秘密值 s 的 l 个共享子秘密，$\lambda_i=\boldsymbol{A}_i\boldsymbol{v}$ 属于 ρ_i 且表示共享子秘密。

2) 假设 $p'\in P$ 是任意的授权子集，定义 $I\subset\{1,2,\cdots,l\}$ 且 $I=\{ip_i\in p'\}$，则一定存在常量 $\{\omega_i\in Z_P\}_{i\in I}$，对于任意秘密值 $\{\lambda_i\}_{i\in I}$，有

$$\sum_{i\in I}\omega_i\lambda_i=s$$

1.5 本地化差分隐私

假设有 n 个用户，每个用户对应一条记录，给定一个隐私算法 M、定义域 $D(M)$及值域 $R(M)$，若算法 M 在任意 2 条记录 t 和 $t'(t,t'\in D(M))$ 上得到相同的输出结果 $t^*(t^*\in R(M))$ 且满足下列不等式，则 M 满足 ε-本地化差分隐私。

$$\Pr[M(t)=t^*]\leqslant \mathrm{e}^{\varepsilon}\Pr[M(t')=t^*]$$

由此可知，隐私算法 M 对任意一条记录进行扰动后，得到输出集的概率分布变化相对较小，其概率比值不超过 e^{ε}，这意味着攻击者就算知道输出结果，也无法推理出输入数据为哪一条记录，从而保证数据可以抵御不可信第三方的攻击。

2 方案描述

2.1 方案简介

本文设计的基于本地化差分隐私和属性基可搜索加密的区块链数据共享方案模型如图 1 所示，主要包括 5 类参与实体：属性授权中心（AAS, attribute authorization center）、数据拥有者（DO）、云服务器（CS, cloud server）、区块链（B, block-chain）、数据访问者（DV）。

1) 属性授权中心。AAS 是完全可信的，主要负责系统初始化生成公共参数及密钥分发。

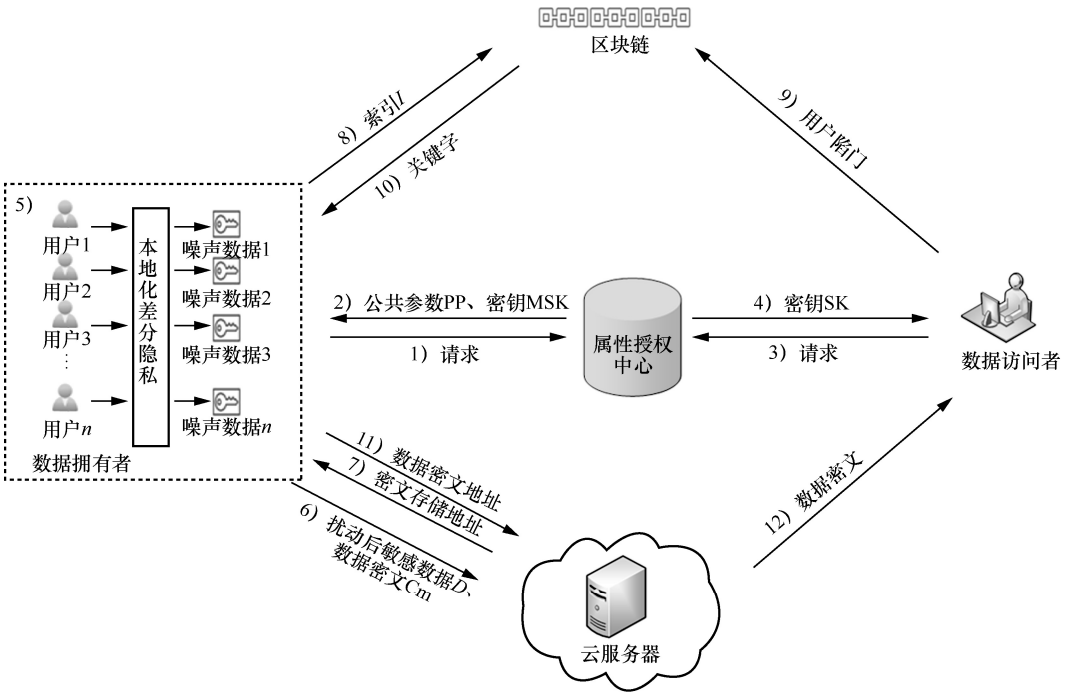

图 1　基于本地化差分隐私和属性基可搜索加密的区块链数据共享方案模型

2) 数据拥有者。每个 DO 拥有原始数据，并使用 LDP 技术中的 RAPPOR 方法对数据进行预处理，使 DO 将数据隐私化，保护 DO 身份隐私。

3) 云服务器。CS 存储数据密文及扰动后的敏感数据，并将其密文存储地址返回给 DO；当搜索成功后，区块链会返回关键字给 DO，DO 根据索引关系找到 CS 上存储密文地址；CS 接收 DO 发起的数据密文地址，查找数据密文并将数据密文返回给 DV。

4) 区块链。区块链的区块有容量限制，主要负责存储由 DO 上传的包含关键字的数据索引，接收由 DV 上传的陷门并进行搜索匹配服务。若搜索成功，区块链上的节点将返回关键字给 DO，反之则失败。

5) 数据访问者。DV 可根据感兴趣关键字和私钥产生搜索陷门，并将陷门上传至区块链，由区块链节点进行搜索操作。若搜索成功，最终 DV 会收到 CS 返回的数据密文，然后解密其密文；若搜索失败，DV 无法访问获取共享数据。

2.2　方案概述

首先，AAS 初始化 DO 的公共参数 PP 和主密钥 MSK 及 DV 的密钥 SK。其次，每个 DO 使用 LDP 技术中的 RAPPOR 方法对数据进行预处理，将扰动后的敏感数据 D 上传到 CS；利用可搜索加密对数据集中包含关键字的数据创建索引 I，得到数据密文 Cm 并上传至 CS；收到密文后的 CS 将其存储地址返回给 DO；DO 将索引广播存储至区块链的新区块中。最后，DV 根据感兴趣关键字和私钥产生搜索陷门，并将陷门上传至区块链，由区块链节点进行搜索操作。若搜索成功，区块链返回关键字给 DO，DO 根据索引关系找到 CS 上存储的密文地址，最终 DV 会收到 CS 返回的数据密文，并解密其密文。方案概述时序图如图 2 所示。

2.3　安全模型

1) 本文方案预处理阶段是满足 ε- 本地化差分隐私的，对数据拥有者具有身份隐私保护功能且能抵御不可信第三方攻击。

2) 通过概率多项式时间敌手 A 和挑战者 B 的游戏来定义方案抵御关键字攻击。

初始阶段　B 运行系统建立算法并将公共参数输出给 A 。

阶段 1　A 向 B 查询关键字 $w_a,\cdots,w_x$ 的密文索引。

挑战　A 向 B 提交 2 个挑战关键字 w_0 和 w_1，然后 B 随机选择参数 $u\in\{0,1\}$，并将关键字索引密文发送给 A。

阶段 2　A 重复阶段 1 继续查询关键字 w_i 的密文索引，其中，$w_i\neq w_0,w_1$。

猜测　最后 A 输出值 $u\in\{0,1\}$ 作为对 μ 的猜测，若 $\mu'=\mu$，则 A 攻击成功。A 攻击成功的优势被定义为

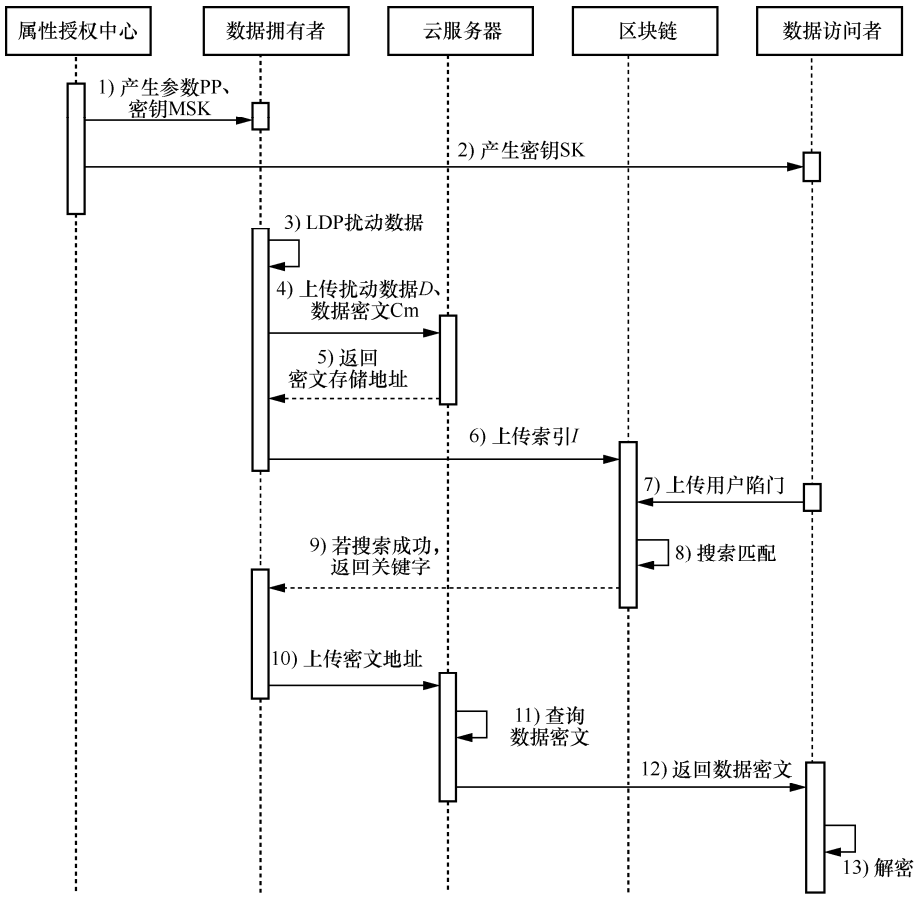

图 2　方案概述时序图

$$\left|\Pr[\mu'=\mu]-\frac{1}{2}\right|$$

若不存在概率多项式时间敌手能以不可忽略的优势赢得上述安全游戏，则方案满足关键字语义安全性。

3　方案流程

3.1　数据预处理

本文使用本地化差分隐私技术中的 RAPPOR 方法对原始数据进行预处理，得到 ε-本地化差分隐私保护后的数据 D 以模糊 DO 身份信息，共包含 4 个步骤：布隆过滤、永久随机响应（PRR）、瞬时随机响应（IRR）、聚合。数据预处理如图 3 所示。

1) 布隆过滤。初始状态时，对于字符串长度为 k 位的数组，将它的 h 个映射函数在位图中置为 h 个 1，由布隆过滤技术将真实数据 w 的值表示为长度为 h 的向量 $\boldsymbol{M}=(0,1)^h$，并记录布隆串与字符串的映射关系。

2) 永久随机响应。对于真实数据 w，对向量 $\boldsymbol{M}$ 的每一位 $i\left(0\leqslant i\leqslant k\right)$ 进行扰动生成一个新向量 $\boldsymbol{M}'$，$f\in[0,1]$ 表示概率取值，扰动的方式满足

$$p(M_i^{'}=w)=\begin{cases}0.5f, & w=1\\ 0.5f, & w=0\\ 1-f, & w=M_i\end{cases} \tag{1}$$

3) 瞬时随机响应。对向量 $\boldsymbol{M}'$ 的每一位 i 进行第二次扰动，生成一个 k 位的二进制串结果集 D，分别表示 M_i' 取值为 1 和 0 时置为 1 的概率。

$$P(D_i=1)=\begin{cases}p, M_i'=1\\ q, M_i'=0\end{cases} \tag{2}$$

因为瞬时随机响应是第二次的扰动过程，上传数据中的 1 可由原始数据中的 0 转变而来，也可由原始数据中的 1 转变而来。若由原始数据中的 1 转变而来，则将这种转变的概率记为 p*。这种转变可由 2 种变化组成：由 1 变为 0 再变为 1 或由 1 变为 1 再变为 1。那么

$$p^*=P(D_i=1\cdot m_i=1)=\frac{1}{2}f(p+q)+(1-f)p \tag{3}$$

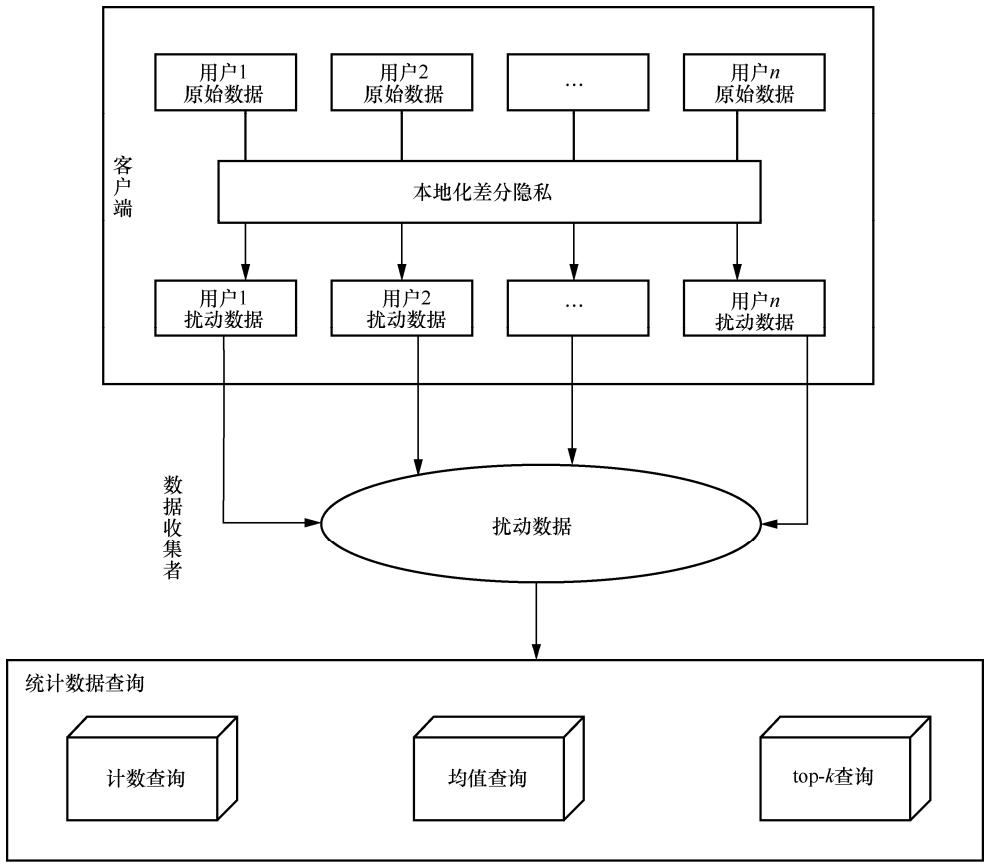

图 3　数据预处理

若由原始数据中的 0 转变而来，则将这种转变概率记为 q^*。这种转变也可由 2 种变化组成：由 0 变为 0 再变为 1 或由 0 变为 1 再变为 1。那么

$$q^* = P(D_i = 1 \cdot m_i = 0) = \frac{1}{2} f(p+q) + (1-f)q \tag{4}$$

4) 聚合。将瞬时随机响应操作过后产生的敏感数据密文 D 上传到 CS。

3.2　系统建立

在 ABE 算法中，每个用户的属性集合都是不同的，数据拥有者在上传用户的共享数据时，为了实现对不同用户的权限控制，根据用户的身份属性集合制定一个访问策略，只有满足访问策略的用户才可以访问数据。其中，访问策略有布尔表达式、线性秘密共享矩阵和访问树，本文选择的访问策略为线性秘密共享矩阵。除此以外，本文将 SE 技术和 ABE 技术相结合，使数据隐私保护和访问控制问题得到了解决。

$\text{Setup}(\lambda) \to (\text{PP}, \text{MSK})$：系统初始化。该算法由 AAS 执行，输入参数 λ，产生双线性映射 $e: G \times G \to G_T$，其中，G 和 G_T 是阶为素数 p 的循环群，g 是 G 的生成元。定义哈希函数 $H:\{0,1\}^* \to G$，选取随机数 $\alpha, \beta \in Z_p$，计算系统公共参数 PP 和主密钥 MSK。

$$\text{PP} = (g, g^{\beta}, e(g,g)^{\alpha}, H) \tag{5}$$

$$\text{MSK} = g^{\alpha} \tag{6}$$

$\text{KeyGen}(\text{PP}, \text{MSK}, S) \to \text{SK}$：密钥生成。该算法由 AAS 执行，输入主密钥 MSK、公共参数 PP 和用户属性集合 S，对于每一个属性 $x \in S$，算法随机选择参数 $t \in Z_p$，计算 $k_x = H(x)^t$。生成用户解密密钥 SK 为

$$\text{SK} = (K = g^{\alpha} g^{\beta t}, L = g^t, \{K_x = H(x)^t\}_{x \in S}) \tag{7}$$

3.3　加密与上传

$\text{Encrypt}(\text{PP}, (\boldsymbol{A}, \rho), M) \to (\text{Cm}, I_x)$：数据加密。该算法由 DO 执行，输入公共参数 PP、明文 m 和访问策略 $(\boldsymbol{A}, \rho)$，其中，$\boldsymbol{A}$ 表示一个 $l \times n$ 矩阵，ρ 表示 $\boldsymbol{A}$ 的每一行映射到相应属性的映射函数。算法随机选择向量 $\boldsymbol{v} = (s, y_2, \cdots, y_n) \in Z_p$，对于 $\boldsymbol{A}$ 的每一行 $\boldsymbol{A}_i$，计算共享子秘密 $\lambda_i = \boldsymbol{A}_i \boldsymbol{v}$，随机选择参数 $r_t \in Z_p$。计算数据密文 Cm 为

$$\begin{aligned}&\mathrm{Cm}=(C=me(g,g)^{\alpha s},C'=g^{s},\\&\{C_t=\mathrm{g}^{\beta\lambda_i}H_{\rho(t)}{}^{-rt},D_t=g^{r_t}\}_{t=1}^{l})\end{aligned}\tag{8}$$

对于关键字集合 $w=\{w_x\}$，将包含关键字 w_x 的数据生成索引 I_x。随机选取参数 $\theta\in Z_p$，计算关键字索引 I_x 为

$$I_x=(E_x=(e(g,g)^{\alpha\theta}e(g,H(w_x))^{\theta},F_x=g^{\theta})\tag{9}$$

DO 将数据密文 Cm 上传至 CS，然后 CS 将其存储地址 Address 返回给 DO；索引 I_x 广播存储至区块链的新区块中。

3.4 数据查询与匹配

$\mathrm{Trapdoor}(\mathrm{PP},\mathrm{SK},w')\rightarrow T_{w'}$：陷门生成。该算法由 DV 执行，输入公共参数 PP、个人私钥 SK 和搜索的关键字 w'，算法随机选择参数 $u\in Z_p$。生成陷门 $T_{w'}$ 并发送给区块链

$$T_{w'}=(G_{w'}=H(w')g^{\alpha}g^{\beta u},J_{w'}=g^{\beta u})\tag{10}$$

$\mathrm{Search}(\mathrm{PP},T_{w'},I_x)\rightarrow\mathrm{Address}$：搜索匹配。该算法由区块链节点执行，输入公共参数 PP、关键字索引 I_x 和用户陷门 $T_{w'}$，若搜索关键字和索引中包含的关键字一样，则满足式(11)，表明搜索匹配成功，区块链将 Address 返回给 DO。

$$E_xe(F_x,J_{w'})=e(F_x,G_{w'})\tag{11}$$

3.5 下载解密

$\mathrm{Decrypt}(\mathrm{PP},\mathrm{Cm},\mathrm{SK})\rightarrow m$：密文解密。该算法由 DV 执行，输入公共参数 PP、访问策略 $(\boldsymbol{A},\rho)$ 下的密文和属性集合 S 下的解密密钥分别为

$$\mathrm{Cm}=\left(C,C',\{C_t,D_t\}_{t=1}^{l}\right)\tag{12}$$

$$\mathrm{SK}=(K,L,\{K_x\}_{x\in S})\tag{13}$$

若属性集合 S 满足访问策略 $(\boldsymbol{A},\rho)$，那么计算数值 $w_t=Z_p$，满足

$$\sum_{\rho(t)\in S}w_t\boldsymbol{A}_i=s\tag{14}$$

解密算法为

$$\frac{e(C',K)}{\prod_{\rho(t)\in S}(e(C_t,L)e(D_t,K_{\rho(t)}))^{w_i}}=e(g,g)^{\alpha s}\tag{15}$$

最后，算法计算 $\dfrac{C}{e(g,g)^{\alpha s}}$ 得到明文 m。

4 安全性证明及正确性证明

4.1 安全性证明

1) 身份隐私保护及抵御不可信第三方攻击

DO 首先输入原始数据 w_1 和 w_2，然后对 w_1 和 w_2 进行编码，得到 m_1 和 m_2，最后对 m_1 和 m_2 进行噪声扰动，得到扰动值 d_1 和 d_2，其中，编码过程和扰动过程是相互独立的，D 为所有扰动数据所构成的集合，有 $\forall d\in D$。根据本地化差分隐私定义可得

$$\begin{aligned}&\frac{\Pr(d_1\in D\mid w_1)}{\Pr(d_2\in D\mid w_2)}=\frac{\Pr(d_1\in D\mid w_1,m_1)}{\Pr(d_2\in D\mid w_2,m_2)}=\\&\frac{\Pr(d_1\in D\mid w_1)\Pr(w_1\mid m_1)}{\Pr(d_2\in D\mid w_2)\Pr(w_2\mid m_2)}=\frac{\Pr(d_1\in D\mid m_1)}{\Pr(d_2\in D\mid m_2)}=\\&\left[\frac{p^*(1-q^*)}{q^*(1-p^*)}\right]^h=\left(\mathrm{e}^{\frac{\varepsilon}{h}}\right)^h=\mathrm{e}^{\varepsilon}\end{aligned}\tag{16}$$

因此，满足 ε-本地化差分隐私。

2) 关键字语义安全性

证明 存在敌手 A 以不可忽略的优势赢得关键字攻击游戏，则挑战者 B 能以不可忽略的优势解决双线性 Diffie-Hellman（BDH）困难假设问题。

初始化 为了计算 $e(g,g)^{\alpha\beta\gamma}\in G_T$，给定 BDH 参数为 $u_1=g^{\alpha},u_2=g^{\beta},u_3=g^{\gamma}\in G$，B 选择散列函数 $H:\{0,1\}^*\rightarrow G$，执行初始化，得到公共参数 PP 并发送给 A。

阶段 1 敌手 A 询问随机预言机 H_1，B 开始计算初始化为空的 H_1 列表 (w_i,h_i,e_i,c_i)，A 询问 H_1 的任何关键字 $w_i\in\{0,1\}$，A 做出以下应答。

①若 w_i 已经在 H_1 列表中，B 将 $H_1(w_i)=h_i\in G$ 发送给 A。

②否则，B 选择随机数 $c_i\in\{0,1\},e_i\in Z_p$。若 $c_i=0$，B 计算 $h_i=u_2{}^{e_i}=g^{\beta e_i}\in G$；若 $c_i=1$，B 计算 $h_i=g^{e_i}\in G$，并将数组 (w_i,h_i,e_i,c_i) 添加到 H_1 列表中，同时将 $H_1(w_i)=h_i$ 返回给 A。

挑战 A 选择 2 个关键字 w_0 和 w_1 给 B，B 从 H_1 列表中执行两次询问，可得 $H_1(w_0)=h_0$ 和 $H_1(w_1)=h_1$。若 c_0=0 且 c_1=1，B 宣布失败并且终止；否则，B 选择 $\mu\in(0,1)$，使 $c_\mu=0$，并任取参数 $\gamma\in Z_p$ 及关键字索引 $I_x^*=(E_x^*,F_x^*)$，其中，E_x^*和F_x^*分别为

$$E_x^* = (e(g,g)^{a_\mu \gamma} e(g,H(w_\mu)^\gamma) = (e(g,g)^{a_\mu \gamma} e(g,u_2^{a_\mu})^\gamma) = (e(g,g)^{a_\mu \gamma} e(g,g^{\beta a_\mu})^\gamma) = e(g,g)^{a_\mu (1+\beta)\gamma} \in G_T \tag{17}$$

$$F_x^* = g^\gamma \in G \tag{18}$$

因此，I_x^*是合法的关键字密文索引。

阶段 2　A 重复阶段 1 并适应性地选择关键字 w_i 的密文索引，其中，$w_i \neq w_0, w_1$。

猜测　A 输出 w_i 关键字的猜测值 μ'。若 $\mu' = \mu$，则 A 获胜，表示 A 能以可忽略的优势输出 $e(g,g)^{a_\mu (1+\beta)\gamma} \in G_T$；否则，A 在游戏中失败。因此，敌手 A 在游戏中没有任何优势获胜。

由此可得，在 BDH 假设成立的前提下，A 只能以可忽略的优势赢得关键字攻击游戏。

4.2　正确性证明

1) 搜索匹配正确性证明

当搜索的关键字 $w' = w$ 时，有

$$e(F_x, G_{w'}) = e(g^\theta,\ H(w')g^\alpha g^{\beta u}) = e(g,H(w'))^\theta e(g,g)^{(\alpha+\beta u)\theta} \tag{19}$$

$$e(F_x, J_{w'}) = e(g^\theta, g^{\beta u}) = e(g,g)^{\beta\theta u} \tag{20}$$

$$E_x = e(g,g)^{\alpha\theta} e(g,H(w_x))^\theta \tag{21}$$

2) 授权访问用户解密数据正确性证明

$$\prod_{\rho(t)\in S} (e(C_t,L)e(D_t,K_{\rho(t)}))^{w_t} = \prod_{\rho(t)\in S} (e(g^{\beta\lambda_i} H_{\rho(t)}{}^{-r_t}, g^t) e(g^{r_t}, H(x)^t))^{w_t} = \prod_{\rho(t)\in S} (e(g^{\beta\lambda_i} H_{\rho(t)}{}^{-r_t}, g^t) e(g^{r_t}, H(x)^t))^{w_t} = \prod_{\rho(t)\in S} (e(g^{\beta\lambda_i}, g^t))^{w_t} = (e(g^{\beta\lambda_i}, g^t))^{\sum w_t} = (e(g^\beta, g^t))^{\sum \lambda w_t} = (e(g^\beta, g^t))^S \tag{22}$$

$$e(C^{'}, K) = e(g^S, g^\alpha g^{\beta t}) = e(g^S, g^\alpha) e(g^S, g^{\beta t}) = e(g,g)^{\alpha S} e(g,g)^{\beta t S} \tag{23}$$

因此，式(15)是正确的。

5　方案对比与实验分析

5.1　方案对比

本节将本文方案与文献[20]方案和文献[21]方案的功能特性进行对比，如表 1 所示。

从表 1 中可以看出，文献[20]在传统云环境下，通过本地化差分隐私中的 RAPPOR 方法和属性基加密，考虑数据共享过程中的隐私泄露及访问控制问题，避免非法访问者访问存储在第三方的共享数据，保障拥有者共享数据时的安全，虽然不需要可信第三方，但不具有可追溯性和安全搜索。文献[21]在区块链环境下，通过 K-anonymity 方法及可搜索加密确保数据在不泄露隐私的情况下共享，由于区块链的区块大小限制及透明共享问题，完全依赖区块会使数据效率随数据的增加而降低，并且 K-anonymity 不能抵御不可信第三方攻击。本文将区块链和云服务器结合，链上链下协同存储数据，提供防篡改可追溯的数据共享；利用本地化差分隐私中的 RAPPOR 方法对数据进行预处理，式(16)证明了本文方案满足 ε-本地化差分隐私，实现拥有者的身份隐私保护，能抵御不可信第三方攻击；可搜索加密技术和属性基加密的结合，搜索匹配正确性证明了 $E_x = e(g,g)^{\alpha\theta} e(g,H(w_x))^\theta$，实现数据隐私保护、支持密文安全检索，$\dfrac{e(C',K)}{\prod_{\rho(t)\in S} (e(C_t,L)e(D_t,K_{\rho(t)}))^{w_t}} = e(g,g)^{\alpha s}$ 成立为共享数据提供细粒度访问控制。

5.2　实验分析

本文采用实验环境为 64 位的 Windows 操作系统，Intel 酷睿 i5-6300HQ CPU 3.20 GHz、内存 8 GB，通过本地虚拟机 VMare 加载开源项目 OpenStack 来进行性能测试，使用 Python 语言，加密函数由 PBC 函数库提供。通过实验结果，本节将本文方案与文献[20]方案和文献[21]方案分别在引入区块链前后关键字检索时间上进行对比分析，如图 4 所示，其中，TPS（transaction per second）为每秒处理关键字交易数。

表 1　功能特性对比

方案	云存储	区块链	可追溯性	抵御不可信第三方攻击	隐私性	安全搜索
文献[20]方案	√	×	×	√	√	×
文献[21]方案	×	√	√	×	√	√
本文方案	√	√	√	√	√	√

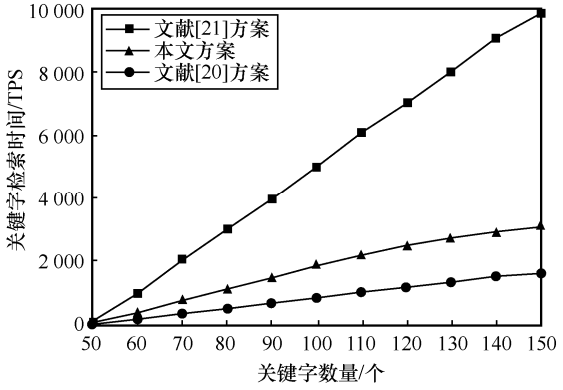

图 4 引入区块链前后关键字检索时间

由图4可知，将关键字数量分别设置为50、100、150时，本文方案相较于文献[20]方案引入区块链后的关键字检索时间增加，但安全性、隐私性得到保证，随着关键字数量的增多，引入区块链后的关键字检索时间增长量逐渐减少；本文方案与文献[21]方案均引入区块链，但文献[21]关键字检索随操作请求频率变化，即 DV 每进行一次数据访问，就处理一次关键字，使区块链上的操作计算成本与访问率成线性增长关系，而本文方案只将包含关键字的数据索引 I_x 上传到区块链，进行关键字检索时，对应的陷门 $T_{w'}$ 发送给区块链，若搜索关键字和索引 I_x 中包含的关键字一样，则满足 $E_x e(F_x, J_{w'}) = e(F_x, G_{w'})$，区块链直接将 Address 返回给 DO，使在保证检索效率的情况下限制不可信第三方的恶意行为。从与文献[20]方案和文献[21]方案的对比中可以看出，本文方案具有较高的效率及安全性。

6 结束语

目前，数据共享方案大多依赖于可信第三方，只关注数据隐私保护或访问控制问题，由此本文提出一种基于本地化差分隐私和属性基可搜索加密的区块链数据共享方案。该方案利用 LDP 技术中的 RAPPOR 方法对数据进行预处理以模糊身份信息，切实保障 DO 的身份隐私问题；若 CS 非法获得数据，则获得的数据也不是原始数据，因此从根本上抵御了不可信第三方攻击；将 SE 技术和 ABE 技术相结合，使 DO 实现数据隐私保护以及细粒度访问控制；将数据密文及关键字密文上传到 CS，对包含关键字的索引广播至区块链的新区块中，搜索工作在区块链中进行，保证数据共享的安全。如何使基于区块链的本地化差分隐私数据共享应用于各种实际场景中，解决现存的重要问题，是下一步的研究重点。

参考文献：

[1] JACYNYCZ V, CALVO A, HASSAN S, et al. Betfunding: a distributed bounty-based crowdfunding platform over Ethereum[C]//Proceedings of the 13th International Conference on Distributed Computing and Artificial Intelligence. Berlin: Springer, 2016: 403-411.

[2] FAN K, WANG S, REN Y, et al. MedBlock: efficient and secure medical data sharing via blockchain[J]. Journal of Medical Systems, 2018, 42(8): 1-11.

[3] XIA Q, SIFAH E B, ASAMOAH K O, et al. MeDShare: trust-less medical data sharing among cloud service providers via blockchain[J]. IEEE Access, 2017, 5: 14757-14767.

[4] 杜瑞忠, 谭艾伦, 田俊峰. 基于区块链的公钥可搜索加密方案[J]. 通信学报, 2020, 41(4): 114-122.
DU R Z, TAN A L, TIAN J F. Public key searchable encryption scheme based on blockchain[J]. Journal on Communications, 2020, 41(4): 114-122.

[5] SWEENEY L. K-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.

[6] SWEENEY L. Achieving K-anonymity privacy protection using generalization and suppression[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 571-588.

[7] MACHANAVAJJHALA A, KIFER D, GEHRKE J, et al. L-diversity: privacy beyond K-anonymity[J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1): 3.

[8] LI N H, LI T C, VENKATASUBRAMANIAN S. T-closeness: privacy beyond K-anonymity and L-diversity[C]//Proceedings of 2007 IEEE 23rd International Conference on Data Engineering. Piscataway: IEEE Press, 2007: 106-115.

[9] XIAO X K, TAO Y F. M-invariance: towards privacy preserving re-publication of dynamic datasets[C]//Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2007: 689-700.

[10] MUH A, MHR B, JC A. Differential privacy in blockchain technology: a futuristic approach[J]. Journal of Parallel and Distributed Computing, 2020, 145: 50-74.

[11] 叶青青, 孟小峰, 朱敏杰, 等. 本地化差分隐私研究综述[J]. 软件学报, 2018, 29(7): 1981-2005.
YE Q Q, MENG X F, ZHU M J, et al. Survey on local differential privacy[J]. Journal of Software, 2018, 29(7): 1981-2005.

[12] 李杨, 温雯, 谢光强. 差分隐私保护研究综述[J]. 计算机应用研究, 2012, 29(9): 3201-3205, 3211.
LI Y, WEN W, XIE G Q. Survey of research on differential privacy[J]. Application Research of Computers, 2012, 29(9): 3201-3205, 3211.

[13] LIU L P, PIAO C H, JIANG X H, et al. Research on governmental data sharing based on local differential privacy approach[C]//Proceedings of 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE). Piscataway: IEEE Press, 2018:

39-45.

[14] SUN Z C, WANG Y J, CAI Z P, et al. A two-stage privacy protection mechanism based on blockchain in mobile crowdsourcing[J]. International Journal of Intelligent Systems, 2021, 36(5): 2058-2080.

[15] 秦志光, 徐骏, 聂旭云, 等. 公钥可搜索加密体制综述[J]. 信息安全学报, 2017, 2(3): 1-12.
QIN Z G, XU J, NIE X Y, et al. Overview of public key searchable encryption systems[J] Journal of Cyber Security Information Security, 2017, 2(3): 1-12.

[16] TANG X, GUO C, CHOO K K R, et al. A secure and trustworthy medical record sharing scheme based on searchable encryption and blockchain[J]. Computer Networks, 2021, 200: 108540.

[17] 牛淑芬, 谢亚亚, 杨平平, 等. 区块链上基于云辅助的属性基可搜索加密方案[J]. 计算机研究与发展, 2021, 58(4): 811-821.
NIU S F, XIE Y Y, YANG P P, et al. Cloud-assisted attribute-based searchable encryption scheme on blockchain[J] Computer Research and Development, 2021, 58(4): 811-821.

[18] BETHENCOURT J, SAHAI A, WATERS B. Ciphertext-policy attribute-based encryption[C]//Proceedings of 2007 IEEE Symposium on Security and Privacy. Piscataway: IEEE Press, 2007: 321-334.

[19] DING X W, YANG J M. An access control model and its application in blockchain[C]//Proceedings of 2019 International Conference on Communications, Information System and Computer Engineering (CISCE). Piscataway: IEEE Press, 2019: 1-5.

[20] 徐红. 基于差分隐私和属性基的云数据安全共享机制[D]. 桂林: 桂林电子科技大学, 2020.
XU H. Secure sharing mechanism of cloud data based on differential privacy and attribute basis[D]. Guilin: Guilin University of Electronic Technology, 2020.

[21] CHEN Y W, MENG L H, 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, 2021: 1-12.

[作者简介]

冯涛（1970– ），男，甘肃临洮人，博士，兰州理工大学研究员、博士生导师，主要研究方向为网络与信息安全、区块链、工业互联网安全。

陈李秋（1998– ），女，四川内江人，兰州理工大学硕士生，主要研究方向为网络与信息安全、区块链、隐私保护、访问控制。

方君丽（1985– ），女，甘肃天水人，兰州理工大学讲师，主要研究方向为隐私保护、区块链、工业互联网安全等。

石建明（1994– ），男，甘肃张掖人，兰州理工大学博士生，主要研究方向为工业物联网安全、网络与信息安全。