基于区块链的噪声化数据分享控制协议

谢晴晴[1]，杨念民[1]，冯霞[2]
（1. 江苏大学计算机科学与通信工程学院，江苏 镇江 212013；2. 江苏大学汽车与交通工程学院，江苏 镇江 212013）

摘 要：区块链提供了一种在不可信网络中进行信息分享的可信通道。然而区块链账本是公开透明的，敏感数据直接明文存储在区块链中会泄露用户隐私。考虑到隐私保护，用户更倾向于提供噪声化数据。另外，面对不同的应用场景，用户需要提供的数据噪声化程度也不同。为此，基于区块链和智能合约技术，采用密文策略属性基加密（CP-ABE）算法设计了一套安全、高效且支持安全搜索的噪声化数据分享控制协议。首先，采用可外包的密文策略属性基加密算法，减少用户端的计算负担。其次，利用智能合约来实现密文之上的数据搜索，预防恶意服务器的恶意操作。再次，所提协议能将数据更新复杂度降低至 $O(1)$，对数据实时更新的应用场景友好。最后，安全性分析和仿真实验证明了所提协议的安全性和可行性。

关键词：区块链；噪声化数据分享；访问控制；可搜索加密
中图分类号：TP309
文献标志码：A
DOI: 10.11959/j.issn.1000−436x.2023177

Blockchain-based noisy data sharing control protocol

XIE Qingqing[1], YANG Nianmin[1], FENG Xia[2]
1. School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China
2. School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013, China

Abstract: Blockchain provides a trusted channel for information sharing in an untrusted network. However, blockchain ledgers are public and transparent. Hence, the sensitive data stored in plaintext on the blockchain will compromise user privacy. Considering privacy preservation, users are more inclined to provide noisy data. And in different application scenarios, the level of data noise that users need to provide also varies. To this end, a secure, efficient, and searchable noisy data sharing control protocol was proposed based on the blockchain and smart contract technology with the ciphertext policy attribute-based encryption (CP-ABE) algorithm. First, an outsourced CP-ABE algorithm was introduced to reduce the computing burden on the user side. Secondly, smart contract was implemented to perform data search on ciphertext, which could prevent malicious operations by malicious servers. Third, the proposed protocol reduced the data update complexity to $O(1)$, which is friendly to the scenarios with real-time data update. Finally, security analysis and simulation experiments demonstrate the security and feasibility of the proposed protocol.

Keywords: blockchain, noisy data sharing, access control, searchable encryption

0 引言

区块链技术具有难以篡改、公开透明、去中心化等优点，是数字加密货币的底层核心技术[1]。目前已有许多学者基于区块链技术来实现数据的可信分享。以金融供应链为例，传统金融体系存在着信息不对称、信息孤岛、结算不能自动完成等缺点。以区块链为底层技术的金融供应链可以将银行、核

收稿日期：2023−05−11；**修回日期：**2023−09−04
基金项目：国家自然科学基金资助项目（No.62002139, No.62272203）；江苏省自然科学基金资助项目（No.BK20200886）；中国博士后科学基金资助项目（No.2019M651738）
Foundation Items: The National Natural Science Foundation of China (No.62002139, No.62272203), The Natural Science Foundation of Jiangsu Province (No.BK20200886), China Postdoctoral Science Foundation (No.2019M651738)

心企业、二三级供应商和其他金融机构上链，支持资金流、信息流、信任流同时传递，并通过嵌入智能合约实现协议的自动执行，提高信息共享的效率，降低信任和资金传递成本[2]。然而，区块链的开放特性使所有节点都可以复制和共享区块链上的数据，查看所有交易历史。这给敏感数据的隐私保护带来了挑战[3]。

为了解决隐私泄露的问题，许多学者从密码学的角度出发来设计解决方法。牛淑芬等[4]提出了一种基于联盟链的可搜索加密方案。该方案将区块链技术和可搜索加密算法相结合，通过代理重加密技术保障数据在不同医院之间进行传输时不会泄露患者的敏感信息，从而解决了病例数据在不同医院中共享的问题。薛腾飞等[5]建立了基于区块链的医疗数据共享模型，采用代理重加密方案解决了访问控制问题。Chen 等[6]提出了一种基于区块链的物联网跨域数据共享方法，使用门限代理重加密的方法对密文进行处理，避免恶意的代理机构与访问者合谋，保证数据在跨域存储、共享过程的安全性。

但是目前已有的基于区块链的数据分享协议皆不适用于数据的噪声化分享场景。例如，导航服务应用程序需要精确的位置数据，外卖服务中顾客可以仅提供楼栋号而非具体的房间号以保护隐私。在医疗健康数据的共享中，医院为了保护病人的隐私，对一些敏感数据进行不同程度的噪声化处理，然后根据数据用户的不同需求来提供不同精度的数据。在社交平台上，用户出于人身安全的考虑，可以先将自己的位置信息进行零度、轻度和重度的噪声化处理后，再分别发布给家人、普通朋友和应用程序。上述情况皆可以归结为数据的噪声化分享控制问题，即通过控制用户访问到的数据精度来实现数据的噪声化分享控制。

为此，本文将智能合约和密文策略属性基加密（CP-ABE, ciphertext policy attribute-based encryption）算法相结合，设计了一种基于区块链的噪声化数据分享控制协议。首先采用支持外包的 CP-ABE 算法来减少用户端的加解密计算负担。其次，利用智能合约来实现密文之上的数据搜索，预防了恶意服务器的非法操作，实现了高效、安全且可搜索的数据分享功能。具体来说，本文协议的主要贡献如下。

1) 本文将智能合约、云计算技术和 CP-ABE 算法相结合，实现了数据的噪声化细粒度分享控制，不但保护了数据隐私，而且能够提供可信的数据搜索结果。

2) 本文协议可以支持密文同态运算，使数据更新的计算复杂度为 $O(1)$。因此本文协议适用于数据实时更新的应用场景。

3) 本文协议的安全性和性能评估分别得到了充分的分析和验证，表明本文协议具有安全性和可行性。

1　相关工作

本节主要介绍访问控制技术和基于区块链的数据分享两方面的相关工作。

1.1　访问控制技术

基于属性的加密（ABE, attribute-based encryption）技术使数据所有者能够根据用户的属性对敏感数据进行细粒度的访问控制。ABE 的概念最早由 Sahai 等[7]提出，其原型来源于基于身份的加密。根据解密策略的位置不同，ABE 可以分为密钥策略属性基加密[8]（KP-ABE, key policy attribute-based encryption）和密文策略属性基加密[9]（CP-ABE, ciphertext policy attribute-based encryption）。在 KP-ABE 中，密钥关联于一个访问控制策略，密文关联于数据属性集合；在 CP-ABE 中，密文关联于一个访问控制策略，密钥关联于用户属性集合。无论是 KP-ABE 还是 CP-ABE，只有属性集和访问控制策略完全匹配时，用户才能解密该密文。与传统加密方法相比，ABE 实现了一对多加密，降低了密钥管理开销。

ABE 技术是最适用于云存储的加密方法之一，已被广泛研究。随着区块链技术的发展，分布式存储技术得到广泛关注，迎来了 ABE 与区块链相结合的热潮。汪玉江等[10]提出基于 ABE 和区块链的个人隐私数据保护方案，实现个人数据的一对多的安全传输和数据的细粒度访问控制。牛淑芬等[11]针对中心化云存储带来的数据安全和隐私保护问题，提出了一种区块链上基于云辅助的密文策略属性基数据共享加密方案，确保了数据的安全存储。Zhang 等[12]结合区块链技术、密文策略分级属性加密技术和星际文件系统设计了一种语音数据加密的分布式存储方案，确保了语音数据的安全性、难以篡改性和可扩展性。Wu 等[13]提出一种基于区块链的隐藏策略和属性访问控制方案，实现属性和策

略的私密性，增强数据的安全性。Yin 等[14]针对区块链物联网中指定接收者的敏感数据安全共享问题，提出一种基于去中心化密钥生成和可编程密文的私有数据共享方案。该方案设计了一种新的密文策略去中心化密钥属性基加密算法，以防止敏感数据泄露。Zhang 等[15]针对车联网面临的敏感数据暴露、数据易受非法访问和篡改、云服务器单点故障等问题，提出一种基于属性加密的区块链数据访问方法。为了加强隐私保护，属性被隐藏，所有生成的交易都记录在区块链上以供审计。Yu 等[16]提出了一种基于区块链的具有属性撤销的细粒度访问控制算法，并应用于物联网应用系统中。Li 等[17]针对医疗行业的安全存储、可靠共享和隐私保护问题，提出了一种基于属性和同态密码系统的区块链电子医疗系统。系统中应用的属性基加密算法实现了基于部分密文的半策略隐藏和动态权限更改的功能。

1.2 基于区块链的数据分享

随着区块链技术的发展，去中心化存储模式进入大众视野。去中心化存储方式可以解决传统云存储系统的单点故障问题，与中心化存储相比具有价格低、安全性高等诸多优势。此外，区块链具有难以篡改、多方可信等特征，可以有效解决数据篡改和服务器不可信等问题。为此相关学者基于区块链技术陆续提出了一系列的数据安全分享方案。

Liu 等[18]针对医疗数据泄露问题，将区块链技术和基于属性的可搜索加密技术结合，实现医疗数据的安全共享。Long 等[19]针对工业区块链隐私泄露问题，结合对称加密和同态加密技术，提出数据隐私保护方法。Huang 等[20]将联盟链与云计算相结合提出了一套分布式医疗数据隐私保护方案。该方案通过云服务器为区块链节点提供服务并设计了基于区块链的智能医疗分布式数据管理架构。Regueiro 等[21]利用区块链技术和 Paillier 加密算法提出了一种用于健康数据统计分析的隐私保护方法。该方法能够在保护患者隐私的前提下提高数据分析的准确性。Gochhayat 等[22]针对物联网和边缘设备的快速增长所带来海量数据的安全存储问题，提出了一种新颖的基于区块链技术的轻量级去中心化加密云存储架构。Agyekum 等[23]提出了一种基于区块链和代理重加密算法的物联网数据安全共享方案，实现数据的机密性、完整性和安全性。Zhang 等[24]基于 CP-ABE 和区块链提出了一套轻量级、去中心化且多权限的访问控制方案，用以保证车联网用户的信息安全和隐私保护。Liu 等[25]提出了一种基于 CP-ABE 的分层物流数据访问控制方案，以超级账本中的成员身份代替传统的密钥分发中心来保证用户私钥的安全性。

但是以上工作仅能实现部分的如下功能：1)隐私保护的数据搜索；2)安全可信的数据搜索；3)针对搜索权限的细粒度控制。

2 预备知识

本节主要介绍本文使用的密码学知识，包括双线性配对和 CP-ABE 方案。

2.1 双线性配对

设 G_0 和 G_T 均为 q 阶大素数循环群，称映射 $e: G_0 \times G_0 \to G_T$ 为双线性配对，e 满足下列条件。

1) 双线性：对于任意 $x, y \in G_0$ 和任意 $a, b \in Z_q$，都有 $e(x^a, y^b) = e(x, y)^{ab}$。

2) 非退化：若 g 是 G_0 的生成员，则 $e(g, g) \neq 1_{G_T}$。

3) 可计算：对于任意 $x, y \in G_0$，$e(x, y) \in G_T$ 可被有效计算。

2.2 密文策略属性基加密（CP-ABE）方案

CP-ABE 将访问策略嵌入密文中，将用户的身份属性集合嵌入用户的身份私钥中。只有当用户的身份属性集合满足数据的访问策略时，该用户才能成功解密获得明文。然而，在传统的 CP-ABE 系统中，加解密需要花费较大的计算开销，不但导致数据分享延迟，而且不适用于资源受限的物联网场景。为此，很多研究学者提出了支持加解密外包的 CP-ABE 方案[26]，基本框架如下。

1) $\text{Setup}(1^k, L) \to (\text{PK}, \text{MSK})$。输入安全参数 1^k 和属性空间 L，输出公钥 PK 和系统主密钥 MSK。该算法首先选择一个生成元为 g、阶为素数 p 的双线性群 G_0，任意选取 $h \in G_0$ 和 $\alpha, \beta \in Z_p$；然后，为属性空间 L 中的每个元素 a_i 随机选择 $v_i \in Z_p$，并计算 $\text{pk}_i = g^{v_i}$；最后，输出系统公钥 $\text{PK} = \left\{G_0, g, h, g^{\beta}, \{\text{pk}_i = g^{v_i} \mid a_i \in L\}\right\}$ 和系统主密钥 $\text{MSK} = \left\{\beta, g^{\alpha}, \{v_i \mid a_i \in L\}\right\}$。

2) $\text{KeyGen}(\text{MSK}, A^u) \to \text{ASK}_u$。输入主密钥 MSK 和用户属性集 A^u，输出用户属性私钥

$\mathrm{ASK}_u=(\mathrm{AK}_1,\mathrm{AK}_2)$。该算法随机选择$r,\varepsilon\in Z_p$，计算 $\mathrm{AK}_1=\left\{D_1=g^{\alpha r}h^{\varepsilon},D_2=g^{\varepsilon},D_j=g^{\alpha r v_j^{-1}}\,\middle|\,a_j\in A^u\right\}$，$\mathrm{AK}_2=g^{\beta+\alpha r}$，$\mathrm{ASK}_u=\langle\mathrm{AK}_1,\mathrm{AK}_2\rangle$。

3) $\mathrm{Encrypt}_1(\mathrm{PK},T)\to\mathrm{CT}'$。输入公钥 PK 和访问策略树 T，输出中间密文 CT′。该算法的计算复杂度与访问策略树 T 有关，即 $O(\mathrm{NumLeaf})$，其中 NumLeaf 是访问策略树 T 的叶子节点数。从访问策略树 T 的根节点开始为 T 的每个节点 x 赋予一个阶为 d_x 的多项式 q_x，令 k_x 为节点 x 的门限值，则 $d_x=k_x-1$。对于 T 的根节点 r，随机选择 $s\in Z_p$，令 $q_r(0)=s$，并选择 d_r 个随机整数来确定多项式 q_r。对于其他节点 x，令 $q_x(0)=q_{\mathrm{parent}(x)}(\mathrm{index}(x))$，选择 d_x 个随机整数来确定多项式 q_x，其中 $\mathrm{parent}(x)$ 是节点 x 的父节点，当前节点是 x 的第 $\mathrm{index}(x)$ 个子节点。令 X 表示 T 中所有叶子节点构成的集合，$\mathrm{att}(x)$ 表示节点 x 对应的属性，该算法输出 $\mathrm{CT}'=\left\{T,C_3'=g^{s_1},C_4'=h^{s_1},\left\{C_j=g^{v_jq_x(0)}\,|\,a_j=\mathrm{att}(x),x\in X\right\}\right\}$。

4) $\mathrm{Encrypt}_2(\mathrm{PK},m,\mathrm{CT}')\to\mathrm{CT}$。输入公钥 PK、明文 m 和中间密文 CT′，输出密文 CT。该算法的计算复杂度为 $O(1)$。该算法选择一个随机数 $s_2\in Z_p$，计算 $C_1=me(g,g)^{\beta s_2},C_2=g^{s_2},C_3=g^{s_1+s_2},C_4=h^{s_1+s_2}$，输出密文 $\mathrm{CT}=\left\{T,C_1,C_2,C_3,C_4,\{C_j=g^{v_jq_x(0)}\,|\,a_j=\mathrm{att}(x)\}\right\}$。

5) $\mathrm{Decrypt}_1(\mathrm{PK},\mathrm{CT},\mathrm{ASK}_u.\mathrm{AK}_1)\to C'$。输入公钥 PK、密文 CT、用户属性私钥分量 $\mathrm{ASK}_u.\mathrm{AK}_1$，输出外包解密结果 C'。该算法的计算复杂度与访问策略树 T 有关，即 $O(k+2|S|)$，其中，$k=\left|A^u\right|$ 表示用户属性个数，S 表示满足访问策略树最小的内部节点集合。该算法首先引用一个递归算法 decryptNode，其具体定义参考文献[9]。假设 x 为访问策略树的任一节点，若 S 能够满足访问策略树 T，则 $\mathrm{decryptNode}(\mathrm{CT},\mathrm{ASK}_u.\mathrm{AK}_1,x)=e(g,g)^{\alpha r q_x(0)}$。然后，该算法对根节点 r 执行 decryptNode 算法，记作 $F_r=\mathrm{decryptNode}(\mathrm{CT},\mathrm{ASK}_u.\mathrm{AK}_1,r)=e(g,g)^{\alpha r q_r(0)}=e(g,g)^{\alpha r s_1}$，并计算 $B=\dfrac{e(D_1,C_3)}{e(D_2,C_4)F_r}=e(g,g)^{\alpha r s_2}$。最终，该算法输出 $C'=\{C_1,C_2,B\}$，其中 $C_1,C_2\in\mathrm{CT}$。

6) $\mathrm{Decrypt}_2(C',\mathrm{ASK}_u.\mathrm{AK}_2)\to m$。输入中间密文 $C'=\{C_1,C_2,B\}$ 和用户属性私钥分量 $\mathrm{ASK}_u.\mathrm{AK}_2$，输出明文 m。该算法的计算复杂度为 $O(1)$。该算法计算 $m=\dfrac{C_1B}{e(\mathrm{ASK}_u.\mathrm{AK}_2,C_2)}=\dfrac{me(g,g)^{\beta s_2}e(g,g)^{\alpha r s_2}}{e(g^{\beta+\alpha r},g^{s_2})}$。

3　模型设计

本节主要介绍系统模型和安全模型。

3.1　系统模型

基于区块链的噪声化数据分享控制系统模型如图 1 所示，主要包括四类主体：①可信机构（TA, trusted authority）、②数据主（DO, data owner）、③数据用户（DU, data user）、④云服务器（CS, cloud server）。

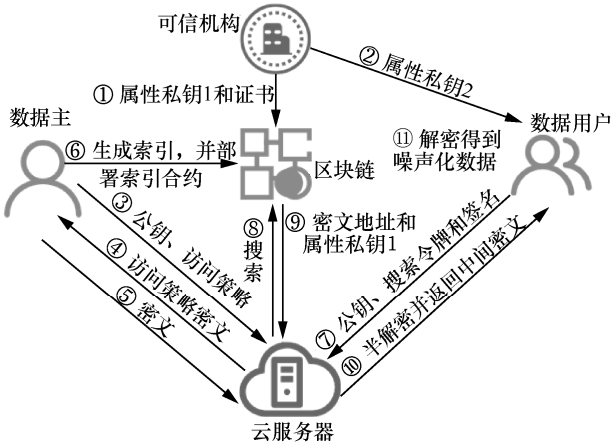

图 1　基于区块链的噪声化数据分享控制系统模型

1) TA。TA 生成系统参数和系统主密钥，为用户分发属性私钥，维持属性列表。

2) DO。DO 首先为待分享的敏感数据设置访问策略，从数据文件中提取关键字，然后加密数据文件，将密文存储到云服务器中并获取存储地址，最后在区块链中生成关键字索引列表，并部署索引合约。

3) DU。DU 首先计算搜索令牌，将公钥和搜索令牌发送给云服务器，然后云服务器对数据密文进行半解密，并将半解密结果返回给 DU，最后 DU 在本地计算相应的噪声化数据。

4) CS。CS 负责存储 DO 上传的数据密文，并提供外包加解密服务。

智能合约包括属性合约、查询合约和索引合约。其中，属性合约维护属性列表；查询合约根据 DU 给出的搜索令牌来寻找相应的数据密文存储地址；索引合约建立关键字搜索令牌和密文地址的映

射关系。

各个主体之间的交互流程如下。TA 为 DU 计算属性私钥 1、属性私钥 2 和证书，并将属性私钥 1 和证书上传到区块链，属性私钥 2 交给 DU 保存；DO 将公钥和访问策略发送给 CS；CS 执行外包 CP-ABE 的加密算法，并将访问策略密文发送给 DO；DO 加密消息和噪声并将完整密文上传到 CS；DO 根据关键词生成索引，部署并调用索引合约；DU 将搜索请求<公钥，搜索令牌，签名>发送给云 CS；CS 将搜索请求转发给查询合约；区块链运行查询合约，并将查询到的密文地址和属性私钥 1 返回给 CS；CS 执行外包 CP-ABE 的解密操作，将得到的中间密文并返回给 DU；DU 进行本地解密，计算得到噪声化数据。

3.2 安全模型

考虑到本文所提方案是通过加密来实现数据的隐私保护和密文搜索的，因此本节讨论的安全模式针对 2 种典型的安全标准：选择明文攻击下的不可区分性和适应性选择关键词语义安全。

1) 选择明文攻击下的不可区分性

定义 1 为证明选择明文攻击下的不可区分性，本文定义了敌手和挑战者之间的交互性游戏。

初始化。挑战者 $\mathcal{C}$ 运行初始化程序并且向敌手 $\mathcal{A}$ 发送公钥 PK。

阶段 1。敌手 $\mathcal{A}$ 提交属性集合 $A^1, A^2, \cdots, A^n$ 进行解密密钥查询。作为响应，挑战者 $\mathcal{C}$ 运行属性密钥生成算法生成相应的属性密钥 ASK_j，其中 $j \in [1, n_1]$，这些属性私钥发送给敌手 $\mathcal{A}$。

挑战。敌手 $\mathcal{A}$ 向挑战者 $\mathcal{C}$ 传递 2 个明文集 $\left\{m_0, \mathrm{NoiS}_0 = \{\tau_i^0\}_{i\in[1,N]}\right\}$ 和 $\left\{m_1, \mathrm{NoiS}_1 = \{\tau_i^1\}_{i\in[1,N]}\right\}$，2 个集合中对应的元素长度相同，并设置相关的访问策略树为 $\mathrm{TS}^* = \{T_i\}_{i\in[1,N]}$ 和 $T_\forall^*$。注意，阶段 1 中敌手查询属性私钥的任一属性集合皆不满足任一访问策略树 $\mathrm{TS}^* = \{T_i\}_{i\in[1,N]}$ 和 $T_\forall^*$。在这里如果任何授权用户的属性满足 $T_i, i \in [1, N]$，那么它也一定满足 $T_\forall^*$。挑战者随机抽取一个比特 $\mu \in [0,1]$，对明文集 $\left\{m_u, \mathrm{NoiS}_u = \{\tau_i^u\}_{i\in[1,N]}\right\}$ 的元素加密，生成密文集合 $\left\{\mathrm{CT}_{m_u}^*, \mathrm{CT}_{\mathrm{NOiS}_u}^*\right\}$。随机生成一个会话密钥 k，计算相应的密文 $\varepsilon_{T_\forall^*}(k)$。然后把生成的密文发给敌手 $\mathcal{A}$。在实际的加密过程中，中间密文并不会发送出去，而是将最终生成的密文发送给敌手 $\mathcal{A}$。所以在安全游戏中无须考虑中间密文。

阶段 2。同阶段 1，敌手 $\mathcal{A}$ 提交属性集合 $A^{n_1+1}, A^{n_2+1}, \cdots, A^{n_2}$ 进行解密密钥查询，挑战者返回相应的属性私钥。注意，此处 $A^{n_1+1}, A^{n_2+1}, \cdots, A^{n_2}$ 要求均不能满足挑战阶段的任何访问策略树。

猜测。敌手 $\mathcal{A}$ 输出猜测 $\mu' \in [0,1]$。

本文定义敌手 $\mathcal{A}$ 的优势在这个游戏中为 $\mathrm{Adv}_{\mathcal{A}} = \Pr[\mu' = \mu] - \frac{1}{2}$。

定理 1 如果任意多项式时间的敌手 $\mathcal{A}$ 在以上的安全游戏中的优势 $\mathrm{Adv}_{\mathcal{A}}$ 是可忽略的，那么本文协议是 CPA 安全的。

2) 适应性选择关键词语义安全

本文在有状态的模拟器 $\mathcal{S}$ 和敌手 $\mathcal{A}$ 之间采用基于模拟的游戏，允许泄露访问模式和搜索模式来证明安全。信息泄露情况采用 2 个泄露函数进行描述，即 $L_1(D) = \left\{|D|, n, \{|D_i|, (D_i)\}_{i\in[1,n]}\right\}$，输入数据文件集合 D，输出数据文件集合的大小、数据文件数量、每个数据文件的大小和标识符；L_2 定义为 $L_2(D, w) = (\mathrm{AP}(w), \mathrm{Tok}_w)$，输入数据文件集合和查询关键词 w，输出关键词的访问模式 $\mathrm{AP}(w)$ 和搜索令牌。在挑战者 $\mathcal{C}$、敌手 $\mathcal{A}$ 以及模拟器 $\mathcal{S}$ 之间进行的游戏定义如下。

$\mathrm{Real}_{\mathcal{A}}(\lambda)$。挑战者 $\mathcal{C}$ 根据安全参数 λ 初始化系统、敌手 $\mathcal{A}$ 给挑战者文件集合 D，挑战者根据索引生成算法和数据加密算法生成安全索引 I 和加密文档 D 并给敌手。敌手向挑战者发起多项式次查询，对于每次查询 q，查询的关键词产生搜索令牌并发送给 $\mathcal{A}$，最后 $\mathcal{A}$ 输出一个比特位作为游戏的输出。

$\mathrm{Ideal}_{\mathcal{A}}(\lambda)$。数据文件集合 D，给定泄露函数 L_1 和 L_2，模拟器 $\mathcal{S}$ 产生并发送 (I^*, C^*) 给 $\mathcal{A}$。然后敌手 $\mathcal{A}$ 重复向模拟器 $\mathcal{S}$ 发起多项式次查询，对于每次查询 q，模拟器 $\mathcal{S}$ 根据泄露函数 L_2 返回相应的搜索令牌 Tok_w^*，最后敌手 $\mathcal{A}$ 输出一个比特位作为该概率实验的结果。

4 交易数据隐私保护方案

本节介绍协议的具体构造，假设本文涉及的所有成员都有自己的公私钥对，记为 $\mathrm{pk}_{\mathrm{entity}}/\mathrm{sk}_{\mathrm{entity}}$，其中 entity 指代实体名称。本文协议的主要标记符说明如表 1 所示。

表 1　　**主要标识符说明**

参数	含义
$\mathrm{PK},\mathrm{MSK},\mathrm{SK}_S$	系统公钥、系统主密钥和搜索密钥
A^u	用户 u 的身份属性集合
$\mathrm{ASK}_u=\langle \mathrm{AK}_1,\mathrm{AK}_2\rangle$	用户 u 的属性私钥，是二元组
$\mathrm{Cert}(\mathrm{sk}_{\mathrm{TA}},\mathrm{pk}_u)$	TA 所签发的用户 u 证书，其中 $\mathrm{sk}_{\mathrm{TA}}$ 是 TA 的私钥， pk_u 是用户 u 的公钥
$X.Y$	取 X 元组中的 Y 分量值
$E_k(\cdot)$	对称加密算法，其中 k 是对称密钥
$\mathrm{NoiS}=\{\tau_i\}_{i\in[1,N]}$	噪声集合，其中 $\tau_i\in G_T$
$\mathrm{noi}(m,\tau_i)$	噪声化函数，输出对原数据 m 添加噪声值 τ_i 之后的噪声化数据，其中 $\tau_i\in\mathrm{NoiS}$
$\{T_i\}_{i\in[1,N]}$	访问策略树集合，其中 T_i 规定了只有满足该访问策略树的用户才能获得噪声化数据 $\mathrm{noi}(m,\tau_i)$
$T_\triangledown$	所有 $T_i,i\in[1,N]$ 析取得到的访问策略树，其定义为用户属性集满足 $T_\triangledown$ 当且仅当其满足任一访问树 $T_i\in\mathrm{TS}$
CT_m	原数据 m 的密文共享组件
$\mathrm{CT}_{\mathrm{NoiS}}$	噪声集合 NoiS 的密文
$\varepsilon_{T_\triangledown}(k)$	会话密钥 k 的密文
$\mathrm{KWS}=\{w_1,w_2,\cdots,w_v\}$	关键词空间，包含 v 个关键词
Tok_{w_i}	关键词 w_i 搜索令牌

4.1　系统初始化

TA 首先设置用户属性空间 $L=\{a_i\}_{i\in[1,n]}$，调用 $\mathrm{Setup}(1^k,L)\to(\mathrm{PK},\mathrm{MSK})$ 产生系统公钥 PK 和主密钥 MSK，然后定义公开的伪随机函数 $F:\{0,1\}^k\times\{0,1\}^*\to\{0,1\}^l$，并生成随机搜索密钥 $\mathrm{SK}_S\leftarrow\{0,1\}^k$。

4.2　用户注册

用户注册过程是 DU、TA 和区块链之间的交互，如图 2 所示，主要步骤如下。

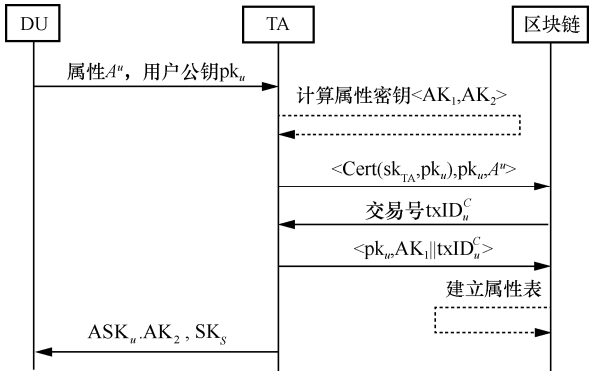

图 2　用户注册时序

步骤 1　用户 u 向 TA 发送二元组 $\langle A^u,\mathrm{pk}_u\rangle$ 以请求计算属性私钥 ASK_u，其中 A^u 是该用户的身份属性集合，pk_u 是该用户的公钥。

步骤 2　TA 首先调用 $\mathrm{KeyGen}(\mathrm{MSK},A^u)\to\mathrm{ASK}_u$ 算法为用户生成属性私钥 $\mathrm{ASK}_u=\langle \mathrm{AK}_1,\mathrm{AK}_2\rangle$ 并为用户分发证书 $\mathrm{Cert}(\mathrm{sk}_{\mathrm{TA}},\mathrm{pk}_u)$；然后将 $\langle \mathrm{Cert}(\mathrm{sk}_{\mathrm{TA}},\mathrm{pk}_u),\mathrm{pk}_u,A^u\rangle$ 上传到区块链中，返回交易号 txID_u^C；再以 pk_u 为键，$\mathrm{AK}_1\|\mathrm{txID}_u^C$ 为值调用属性合约，属性合约的执行过程如算法 1 所示。若 TA 账户没有足够的余额来支付矿工收取的燃料费，则系统回滚；最后将属性私钥分量 $\mathrm{ASK}_u.\mathrm{AK}_2$ 和搜索密钥 SK_S 发回给 DU。

算法 1　属性合约

输入　键值对 $\left\langle \mathrm{PK}_u,\mathrm{AK}_1\|\mathrm{txID}_u^C\right\rangle$

输出　属性表 ASKList

1) 判断属性表 ASKList 是否存在，若存在则跳至步骤 3)，若不存在则执行步骤 2)；

2) 定义并令属性表 ASKList 为空；

3) 在 ASKList 中新增键值对 $\left\langle \mathrm{PK}_u, \mathrm{AK}_1 \| \mathrm{txID}_u^C \right\rangle$；

4) 返回 ASKList；

4.3 数据加密

DO 首先设置噪声集合 $\mathrm{NoiS}=\{\tau_i\}_{i\in[1,N]}$ 和相应的访问策略树 $\mathrm{TS}=\{T_i\}_{i\in[1,N]}$，其中 T_i 规定了能获得噪声化数据 $\mathrm{noi}(m,\tau_i)$ 的用户类型，τ_i 随着 i 的增加而变大。在 TS 的基础上，DO 设计一个新的访问策略树 $T_\forall$，其定义为用户属性集满足 $T_\forall$ 当且仅当其满足任一访问树 $T_i \in \mathrm{TS}$。另外，DO 还需要产生一个对称的会话密钥 k，用于加密原始数据的相关信息，支持数据的高效更新。

然后，DO 和 CS 计算 NoiS 的密文 $\mathrm{CT}_{\mathrm{NoiS}}=\left\{\varepsilon_{T_i}(A_{\tau_i})_{i\in[1,N]}\right\}$ 和会话密钥 k 的密文 $\varepsilon_{T_\forall}(k)$，计算过程如下。

步骤 1 DO 将 PK和$T_\forall$ 发送给 CS，CS 执行外包属性基加密算法，计算密文 $\mathrm{CT}'_\forall=\mathrm{Encrypt}_1(\mathrm{PK},T_\forall)$，并将 $\mathrm{CT}'_\forall$ 返回给 DO。

步骤 2 DO 加密会话密钥 k 为 $\varepsilon_{T_\forall}(k)=\mathrm{Encrypt}_2(\mathrm{PK},k,\mathrm{CT}'_\forall)$。

步骤 3 DO 随机选择 $\tilde{s},\hat{s}\in Z_q$，计算噪声集合密文 $\mathrm{CT}_{\mathrm{NoiS}}$。对于 $\forall \tau_i \in \mathrm{NoiS}$，执行以下步骤。

1) DO 将 PK和T_i 发送给 CS，CS 计算 $\mathrm{CT}'_i=\mathrm{Encrypt}_1(\mathrm{PK},T_i)$ 并发送给 DO。

2) DO 计算 $A_{\tau_i}=(\tau_i)^{-1}e(g,g)^{-\alpha(\tilde{s}-\hat{s})}$ 和 $\varepsilon_{T_i}(A_{\tau_i})=\mathrm{Encrypt}_2(\mathrm{PK},A_{\tau_i},\mathrm{CT}'_i)$。

步骤 4 针对待分享的原数据 m，DO 计算 $C_m=me(g,g)^{\alpha\tilde{s}}, p_{\mathrm{up}}=e(g,g)^{\alpha\hat{s}}$ 和 $\mathrm{CT}_m=\{C_m, E_k(P_{\mathrm{up}})\}$，其中 $E_k(\cdot)$ 是对称加密算法。

至此，数据加密阶段结束，DO 得到噪声集合 NoiS 密文 $\mathrm{CT}_{\mathrm{NoiS}}=\{\varepsilon_{T_i}(A_{\tau_i})\}_{i\in[1,N]}$、会话密钥 k 密文 $\varepsilon_{T_\forall}(k)$ 和原数据 m 的密文共享组件 $\mathrm{CT}_m=\{C_m, E_k(P_{\mathrm{up}})\}$。值得注意的是，$\mathrm{CT}_{\mathrm{NoiS}}$ 和 $\varepsilon_{T_\forall}(k)$ 的计算与原数据 m 无关，因此在数据 m 持续更新的场景中，DO 仅需更新数据的密文共享组件 CT_m 即可。

4.4 生成索引

假设数据 m 的关键词集为 KW_m，其索引的计算步骤如下。

步骤 1 DO 向 TA 申请搜索密钥 SK_S，TA 使用 DO 的公钥 $\mathrm{pk}_{\mathrm{DO}}$ 来加密搜索密钥 $\mathrm{Enc}(\mathrm{SK}_S,\mathrm{pk}_{\mathrm{DO}})$ 并传送给 DO。

步骤 2 DO 将 CT_m 和 $\varepsilon_{T_\forall}(k)\|\mathrm{CT}_{\mathrm{NoiS}}\|\mathrm{TS}$ 分别存放到 CS 中，并返回存储地址 addr_m 和 addr_τ，令 $\mathrm{addr}_{m\tau}=\{\mathrm{addr}_m,\mathrm{addr}_\tau\}$。

步骤 3 对 $\forall w_j \in \mathrm{KW}_m$，DO 计算搜索令牌 $\mathrm{Tk}=\{\mathrm{tok}_{w_i}=F(\mathrm{SK}_S,w_i)\}_{w_i\in\mathrm{KW}_m}$，将 $\langle \mathrm{Tok},\mathrm{addr}_{m\tau}\rangle$ 通过交易发送给索引合约，调用索引合约更新索引 I，如算法 2 所示。若 DO 账户没有足够的余额来支付燃料费，则系统回滚。假设只有 2 个数据 m_1 和 m_2 被分享，其中数据 m_1 有关键词 w_1 和 w_2，密文存储地址为 $\mathrm{addr}_{m_1\tau}$；数据 m_2 有关键词 w_1,w_2,w_3，密文存储地址为 $\mathrm{addr}_{m_2\tau}$，则生成的索引如表 2 所示。索引合约的执行过程如算法 2 所示。

表 2 生成的索引

键	值
Tok_{w_1}	$\mathrm{value}_1=\{\mathrm{addr}_{m_1\tau},\mathrm{addr}_{m_2\tau}\}$
Tok_{w_2}	$\mathrm{value}_2=\{\mathrm{addr}_{m_1\tau},\mathrm{addr}_{m_2\tau}\}$
Tok_{w_3}	$\mathrm{value}_3=\{\mathrm{addr}_{m_2\tau}\}$
Tok_{w_4}	$\mathrm{value}_4=\varnothing$
⋮	⋮

算法 2 索引合约

输入 搜索令牌 TK，密文地址 $\mathrm{addr}_{m\tau}$

输出 索引表 I

1) 判断索引表 I 是否存在，若存在跳至步骤 3)，若不存在执行步骤 2)；

2) 定义并初始化映射 I，使 I 中存储的键与关键词空间的关键词一一对应，每个键值对的值为空集；

3) for $\mathrm{tok}_{w_i}\in \mathrm{TK}$ do

4) 检索 I，得到键值对 $\langle \mathrm{key}_i,\mathrm{value}_i\rangle$；

5) if $\mathrm{key}_i == \mathrm{tok}_{w_i}$

6) $\mathrm{value}_i \leftarrow \mathrm{value}_i \bigcup \mathrm{addr}_{m\tau}$；

7) 更新索引表 I；

8) end if

9) end for

10) 返回 I；

上述数据加密和生成索引流程如图 3 所示。

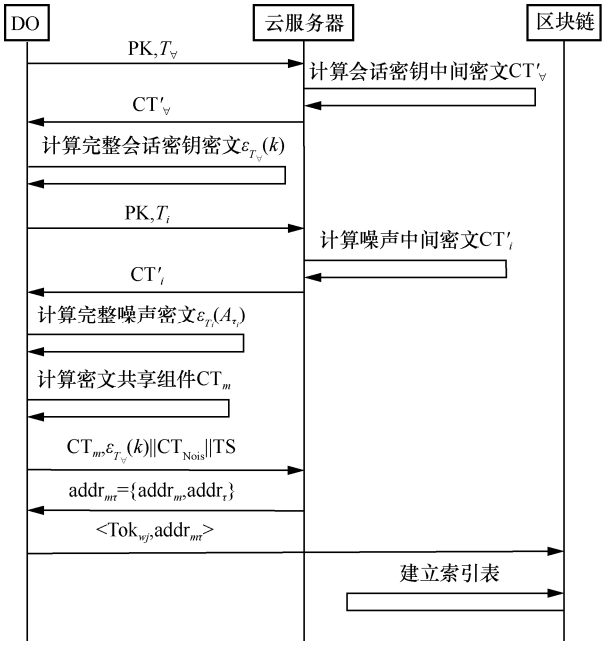

图 3　数据加密和生成索引流程

4.5　访问数据

假设用户 u 的搜索关键词集合为 KW_u，初始化查询结果 R_u 为空集。访问数据过程如下。

步骤 1　用户 u 首先计算搜索令牌集 $\mathrm{TK}_u=\{\mathrm{Tok}_{W_i}=F(\mathrm{SK}_S,w_i)\,|\,w_i\in\mathrm{KW}_u\}$，然后签名搜索令牌集得到 $\mathrm{Sig}(\mathrm{sk}_u,\mathrm{TK}_u)$，并形成搜索请求 $\langle\mathrm{TK}_u,\mathrm{pk}_u,\mathrm{Sig}(\mathrm{sk}_u,\mathrm{TK}_u)\rangle$ 发送给 CS。

步骤 2　CS 将搜索请求 $\langle\mathrm{TK}_u,\mathrm{pk}_u,\mathrm{Sig}(\mathrm{sk}_u,\mathrm{TK}_u)\rangle$ 转发给查询合约。该查询合约根据搜索请求执行搜索，其执行过程如算法 3 所示。

算法 3　查询合约

输入　搜索请求 $\langle\mathrm{TK}_u,\mathrm{pk}_u,\mathrm{Sig}(\mathrm{sk}_u,\mathrm{TK}_u)\rangle$

输出　密文地址集合 R_u，属性私钥 AK_1，用户身份属性集合 A^u

1) 判断 $\mathrm{TK}_u=\mathrm{verify}(\mathrm{pk}_u,\mathrm{Sig}(\mathrm{sk}_u,\mathrm{TK}_u))$，若成立则进行下一步，若不成立则返回；

2) 根据 pk_u 在 ASKList 中检索得到 $\mathrm{AK}_1\|\mathrm{txID}_u^c$；

3) 根据 txID_u^c 取得用户身份属性集合 A^u；

4) 初始化 R_u；

5) for $\mathrm{tok}_{w_i}\in\mathrm{TK}_u$ do

6) 检索 I，得到满足 $\mathrm{key}_i==\mathrm{tok}_{w_i}$ 的键值对 $<\mathrm{key}_i,\mathrm{value}_i>$，并更新 $R_u\leftarrow R_u\bigcup\mathrm{value}_i$；

7) end for

8) 返回 $R_u\|\mathrm{AK}_1\|A_u$；

步骤 3　针对 $\forall\mathrm{addr}_{m\tau}\in R_u$，CS 执行操作如下。

1) 从 $\mathrm{addr}_{m\tau}.\mathrm{addr}_\tau$ 中获取 $\varepsilon_{T_\forall}(k)\|\mathrm{CT}_{\mathrm{NoiS}}\|\mathrm{TS}$。

2) 对于 $\forall T_i\in\mathrm{TS}$，一旦找到数据用户 u 的身份属性集合 A^u 满足的访问策略树 T_i，则从 $\mathrm{addr}_{m\tau}.\mathrm{addr}_m$ 中获取数据密文共享组件 $\mathrm{CT}_m=\{C_m,E_k(P_{\mathrm{up}})\}$，并跳到第 3)步；否则结束该进程，继而处理下一个数据密文地址 $\mathrm{addr}_{m\tau}\in R_u$。

3) CS 使用属性私钥 $\mathrm{ASK}_u.\mathrm{AK}_1$ 外包解密，调用 $\mathrm{Decrypt}_1(\mathrm{PK},\varepsilon_{T_i}(A_{\tau_i}),\mathrm{ASK}_u.\mathrm{AK}_1)$ 和 $\mathrm{Decrypt}_1(\mathrm{PK},\varepsilon_{T_\forall}(k),\mathrm{ASK}_u.\mathrm{AK}_1)$，并将结果 $\left\{\mathrm{CT}_m,C'_{A_{\tau_i}},C_k\right\}$ 返回给 u。

4) u 首先对收到的密文做第二次解密，执行式(1)和式(2)。

$$\mathrm{Decrypt}_2\left(\mathrm{PK},C'_{A_{\tau_i}},\mathrm{ASK}_u.\mathrm{AK}_2\right)=A_{\tau_i}\tag{1}$$

$$\mathrm{Decrypt}_2(\mathrm{PK},C'_k,\mathrm{ASK}_u.\mathrm{AK}_2)=k\tag{2}$$

然后对称解密 $D_k(E_k(P_{\mathrm{up}}))$ 得到 P_{up}，最后可得噪声化数据 $\mathrm{noi}(m,\tau_i)=C_m\dfrac{A_{\tau_i}}{P_{\mathrm{up}}}$。

4.6　更新数据

在数据噪声化分享场景中，一般只更新原数据 m，而无须更新噪声。具体的更新过程如下。

假设 DO 欲将原始数据 m 更新为 m'。DO 需进行以下步骤，随机选择 $\Delta s\in Z_q$，计算 $\Delta P_{\mathrm{up}}=e(g,g)^{\alpha\Delta s}$，$\Delta C_{m'}=\dfrac{m'}{m}e(g,g)^{\alpha\Delta s}$，$P_{\mathrm{up}}{}'=P_{\mathrm{up}}\Delta P_{\mathrm{up}}=e(g,g)^{\alpha(\hat{s}+\Delta s)}$，并加密 $P_{\mathrm{up}}{}'$ 为 $E_k(P_{\mathrm{up}}{}')$。DO 将原数据的地址 addr_m 和 $\{\Delta C_{m'},E_k(P_{\mathrm{up}}{}')\}$ 发送给 CS。CS 根据 addr_m 获得原数据密文共享组件 C_m，计算 $C_{m'}=C_m\Delta C_{m'}=m'e(g,g)^{\alpha(\hat{s}+\Delta s)}$，最后，用新的密文共享组件 $\mathrm{CT}_{m'}=\{C_{m'},E_k(P_{\mathrm{up}}{}')\}$ 替换原数据的密文共享组件 $\mathrm{CT}_m=\{C_m,E_k(P_{\mathrm{up}})\}$。

5　安全性证明

本文协议通过加密实现了数据的隐私保护和密文搜索，因此本节将证明本文协议具有选择明文攻击下的不可区分性和适应性选择关键词语义安全。

5.1　选择明文攻击下的不可区分性证明

证明　假设一个概率多项式时间敌手 $\mathcal{A}$ 的优

势在之前的安全模型的定义中是 $\mathrm{Adv}_{\mathcal{A}}$，然后证明可以基于 $\mathcal{A}$ 构建敌手 $\mathcal{A}'$，使 $\mathcal{A}'$ 能够破解属性基加密算法难以区分的多重加密，在接下来的安全博弈中具有与 $\mathrm{Adv}_{\mathcal{A}}$ 相同的优势。

初始化。$\mathcal{A}$ 获取属性基加密算法的公钥，然后发送公钥给敌手 $\mathcal{A}$。相应的主密钥只有属性基加密算法的挑战者 $\mathcal{C}$ 知道。

阶段 1。$\mathcal{A}$ 提交属性集合 $A^1, A^2, A^3, \cdots, A^{n_1}$。为了产生相应的属性密钥，$\mathcal{A}'$ 对于每个属性集 $A^l, l \in [1, n_1]$ 向挑战者 $\mathcal{C}$ 生成一个属性密钥查询。然后 $\mathcal{A}'$ 将生成的私钥 $\mathrm{ASK}_l = \{\mathrm{AK}_{1,l}, \mathrm{AK}_{2,l}\}_{l\in[1,n_1]}$ 发送给敌手 $\mathcal{A}$。

挑战。敌手 $\mathcal{A}$ 给 $\mathcal{A}'$ 一个挑战访问策略树 $\langle TS^* = \{T_i\}_{i\in[1,N]}, T_\forall^* \rangle$ 和 2 个对应部分长度相同的明文组 $\{m_0, \mathrm{NoiS}_0 = \{\tau^0\}_{i\in[1,N]}\}$ 和 $\{m_1, \mathrm{NoiS}_1 = \{\tau^1\}_{i\in[1,N]}\}$，满足 $\left|\tau_i^0\right| = \left|\tau_i^1\right|_{\forall i\in[1,N]}, |m_0| = |m_1|$。要求查询阶段所查询的属性集合 $A^1, A^2, A^3, \cdots, A^{n_1}$ 均不满足任何访问策略树。$\mathcal{A}'$ 按照以下 2 个步骤对挑战明文进行加密并返回给敌手 $\mathcal{A}$。

步骤 1 $\mathcal{A}'$ 随机选择一个秘密的会话密钥 k 并且在相应的访问策略树 $T_\forall^*$ 加密 k。相应的密文计算过程为

$$\varepsilon_{T_\forall^*}(k) = \mathrm{CPABE.Encrypt}(T_\forall^*, k, \mathrm{PK}) \tag{3}$$

步骤 2 $\mathcal{A}'$ 随机选择 $\hat{s}^* \in Z_q$，然后生成两组明文列表 $\left\{m_0, \frac{m_0}{\tau_1^0} e(g,g)^{\alpha\hat{s}^*}, \cdots, \frac{m_0}{\tau_N^0} e(g,g)^{\alpha\hat{s}^*}\right\}$，$\left\{m_1, \frac{m_1}{\tau_1^1} e(g,g)^{\alpha\hat{s}^*}, \cdots, \frac{m_1}{\tau_N^1} e(g,g)^{\alpha\hat{s}^*}\right\}$。然后 $\mathcal{A}'$ 发送这 2 个列表消息和相应的树集合 $\{T_x, T_1, \cdots, T_N\}$ 给挑战者 C，其中 $T_x \in T^*$ 是从 T^* 中任意选择的一个访问策略树。挑战者 C 第一次随机选择一个值 $\mu \in [0,1]$，然后使用属性基加密算法在访问策略树 T_x 下加密 m_μ，结果如式(4)所示。类似地，对于每个 $i \in [1, N]$，C 在访问策略树 T_i 下加密 $\frac{m_\mu}{\tau_N^\mu} e(g,g)^{\alpha\tilde{s}^*}$，如式(5)所示。密文发送给敌手 $\mathcal{A}'$，对于每个 $y \in Y_i$, $i \in [1, N]$，计算 $\tilde{C}_i^* = \frac{C_i'}{C_{m_\mu}} = (\tau_i^u)^{-1} e(g,g)^{-\alpha(\tilde{s}^* - \hat{s}^* - s^*)}$ 设置 $C_{m_\mu}^* = C_{m_\mu}, \hat{C}_i^* = \hat{C}_i, C_i'^* = C_i', C_i''^* = C_i'', C_i^{j*} = C_i^j$，$\varepsilon_{T_i}\left(A_{\tau_i^*}^*\right) = \left\langle T_i, \tilde{C}_i^*, \hat{C}_i^*, C_i'^*, C_i''^*, \{C_i^j, \forall a_j = \mathrm{att}(y) \in Y_i\}\right\rangle$。除此以外，$\mathcal{A}'$ 设置 $P_{\mathrm{up}}^* = e(g,g)^{\alpha\hat{s}^*}$，使用对称加密密钥加密 P_{up}^* 生成密文 $E_k^*(P_{up}^*)$。现在 $\mathcal{A}'$ 获得密文 $\mathrm{CT}_{m_\mu} = \left\{C_{m_\mu}^*, E_k^*\left(P_{\mathrm{up}}^*\right)\right\}$, $\mathrm{CT}_{\mathrm{NoiS}} = \left\{\varepsilon_{T_i}(A_{\tau_i^*})\right\}_{i\in[1,N]}$。

最终，$\mathcal{A}'$ 将挑战密文 $\varepsilon_{T_\forall^*(k)}, \mathrm{CT}_{m_\mu}$ 和 $\mathrm{CT}_{N_\mu^*}$ 返回给 $\mathcal{A}$。

$$\begin{aligned}&\varepsilon_{T_x}(m_\mu) = \mathrm{CPABE.Encrypt}(\mathrm{PK}, m_u, T_x) = \\ &\left\{T_x, C_{m_\mu} = m_\mu e(g,g)^{\alpha\cdot\tilde{s}^*}, \hat{C} = h^{\tilde{s}^*}, C_x' = g^{\tilde{s}_1^*} g^{\tilde{s}^*},\right. \\ &\left. C_x'' = h^{\tilde{s}_1^*} h^{\tilde{s}^*}, \left\{C_x^j = g^{v_j q_y(0)} \middle| a_j = \mathrm{att}(y) \in Y_x\right\}\right\}\end{aligned} \tag{4}$$

$$\begin{aligned}&\varepsilon_{T_x}\left(\frac{m_\mu}{\tau_i^\mu} e(g,g)^{\alpha\hat{s}^*}\right) = \\ &\mathrm{CPABE.Encrypt}\left(PK, \frac{m_\mu}{\tau_i^\mu} e(g,g)^{\alpha\hat{s}^*}, T_i\right) = \\ &\left\{T_i, C_i' = \frac{m_\mu}{\tau_i^\mu} e(g,g)^{\alpha(\hat{s}^* + s_i^*)}, \hat{C} = h^{\hat{s}_i^*}, C_i' = g^{\hat{s}1_i^*} g^{\hat{s}_i^*},\right. \\ &\left. C_i'' = h^{\hat{s}_{1_i}^*} h^{\hat{s}_i^*}, \left\{C_i^j = g^{v_j q_y(0)} \middle| a_j = \mathrm{att}(y) \in Y_i\right\}\right\}\end{aligned} \tag{5}$$

阶段 2。在没有一组属性可以满足挑战中的任何访问策略树的条件下重复阶段 1。

猜测。敌手 $\mathcal{A}$ 从 $\mu \in [0,1]$ 输出猜测 u'。然后，敌手 $\mathcal{A}'$ 输出 u' 结束游戏。

根据之前定义的安全模型，敌手 $\mathcal{A}'$ 的优势攻破具有多重加密的属性基加密算法的概率是 $\mathrm{Adv}_{\mathcal{A}} = \Pr[\mu' = \mu] - \frac{1}{2} = \mathrm{Adv}_{\mathcal{A}'}$。上述等式表明，如果在本文的安全模型中 $\mathrm{Adv}_{\mathcal{A}}$ 是不可忽略的，那么对于敌手也具有不可忽略的优势 $\mathrm{Adv}_{\mathcal{A}} = \mathrm{Adv}_{\mathcal{A}'}$ 来打破具有多重加密的属性基加密算法的安全性。证毕。

5.2 适应性选择关键词语义安全

证明 本文假设存在一个外部敌手 $\mathcal{A}$ 和一个模拟器 $\mathcal{S}$，若对于外部敌手 $\mathcal{A}$ 和模拟器 $\mathcal{S}$ 能够满足 $\left|\Pr[\mathrm{Real}_{\mathcal{A}}(\lambda)] - \Pr[\mathrm{Ideal}_{\mathcal{A},\mathcal{S}}(\lambda)]\right| \leqslant \mathrm{negl}(\lambda)$（$\mathrm{negl}(\lambda)$ 是可忽略函数）。那么，可以认为本文协议满足适应性选择关键词语义安全。通过文献[21]的等价性证明可知，在模拟中，敌手需区分模拟器产生的密文、索引才能在游戏中取得胜利。所以，需证明

$\left|\Pr[\mathrm{Ind}_{\mathcal{A}}]=1\right| \leqslant \frac{1}{2}+\mathrm{neg}(\lambda)$ 才能证明本文协议是安全的。

模拟器生成模拟数据密文 C^* 和模拟索引 I^* 的过程如下。

阶段 1。生成模拟数据密文 C^*。模拟器通过泄露函数 $L_1(D)=\left\{|D|,n,\left|D_i\right|_{i\in[1,n]}\right\}$，随机生成 n 个长度为 $\left|D_i\right|_{i\in[1,n]}$ 的数据密文文件 $C^*=\{C_1,C_2,\cdots,C_n\}$。由于本文使用的文件加密算法是安全的属性基加密算法，即在模拟中的 $\mathrm{Real}_{\mathcal{A}}(\lambda)$ 的 C 和 $\mathrm{Ideal}_{\mathcal{A},S}$ 的 C^* 在计算中是不可区分的，故：$\left|\Pr[\mathrm{FileEnc}[m,\mathrm{PK}]\rightarrow C]-\Pr[\mathrm{Random}\rightarrow C^*]\right| \leqslant \mathrm{negl}_1(\lambda)$。

阶段 2。生成模拟安全索引 I^*。模拟器在 $\{0,1\}^l$ 随机选择 n 个元素作为模拟搜索令牌，并生成模拟安全索引 I^*。在 $\mathrm{Real}_{\mathcal{A}}(\lambda)$ 中采用伪随机方法 F 构造索引，而模拟安全索引 I^* 由模拟器随机生成的字符串代替。根据伪随机函数的安全性，在搜索密钥 SK_s 未知的情况下，敌手 $\mathcal{A}$ 在计算中无法区分随机字符串和伪随机函数 F 的输出结果。因此，敌手 $\mathcal{A}$ 在计算中无法区分 I 和 I^*，故 $\left|\Pr[\mathrm{IndexGen}(\mathrm{KWS},\mathrm{SK}_S)\rightarrow I]-\Pr[\mathrm{Random}\rightarrow I^*]\right| \leqslant \mathrm{negl}_2(\lambda)$，其中 IndexGen 是索引生成算法。

本文用 $\mathrm{Advan}(\mathcal{A}(C))$ 来表示对手 $\mathcal{A}$ 能辨别真实的密文和随机密文的优势，用 $\mathrm{Advan}(\mathcal{A}(I))$ 来表示对手 $\mathcal{A}$ 辨别真实的加密索引和随机字符串的优势，则

$$\Pr[\mathrm{Ind}_{\mathcal{A}}(\lambda)=1]=\frac{1}{2}+\mathrm{Advan}(\mathcal{A}(C))+$$

$$\mathrm{Advan}(\mathcal{A}(I))=\frac{1}{2}+\left|\Pr[\mathrm{FileEnc}(m,\mathrm{PK})\rightarrow C]-\right.$$

$$\left.\Pr[\mathrm{Random}\rightarrow C^*]\right|+\left|\Pr[\mathrm{IndexGen}(\mathrm{KWS},\mathrm{SK}_S)\rightarrow I)]-\right.$$

$$\left.\Pr[\mathrm{Random}\rightarrow I^*]\right| \leqslant \frac{1}{2}+\mathrm{negl}_1(\lambda)+\mathrm{negl}_2(\lambda)$$

最后，本文令 $\mathrm{negl}_1(\lambda)+\mathrm{negl}_2(\lambda)=\mathrm{negl}(\lambda)$ 可以得到 $\Pr[\mathrm{Ind}_{\mathcal{A}}(\lambda)=1] \leqslant \frac{1}{2}+\mathrm{negl}(\lambda)$。

根据上述证明，对于在任意概率多项式时间的外部敌手 $\mathcal{A}$，$\mathrm{Real}_{\mathcal{A}}(\lambda)$ 和 $\mathrm{Ideal}_{\mathcal{A},\mathcal{S}}$ 的输出是不可区分的，因此，本文协议满足适应性选择关键词语义安全。证毕。

6　性能分析

本节将介绍所提出的数据噪声化分享协议在功能和性能方面与其他现有工作的对比。

6.1　功能对比

文献[10-11,27]与本文协议类似，因此将本文协议与其进行功能对比，对比结果如表 3 所示。文献[10]方案实现了解密计算的外包，但是用户端的加密计算开销与访问策略树属性数量成正比，而且数据关键字以明文的形式出现在区块链。文献[11]方案将可搜索加密和区块链技术相结合实现数据隐私保护，但不支持高效的数据更新。文献[27]方案依赖半可信服务器，而且未考虑密文之上的数据搜索。

6.2　性能对比

表 4 展示了本文协议与文献[10-11,27]在数据主端的加密计算、数据用户端的解密计算、搜索令牌计算、数据搜索计算和数据更新开销的计算代价进行对比。其中，T_{e} 表示指数运算；T_{m} 表示乘法运算；T_{p} 表示双线性对操作；T_{D} 表示一次对称加解密；H 表示散列运算；F 表示伪随机函数；$|Y|$ 表示访问策略树中的叶子节点个数；k 表示用户属性数量；n_k 表示搜索关键词的数量；符号—表示不适用；n_z 表示噪声集合中噪声因子的数量；S 表示满足访问

表 3　　**功能对比**

方案	抵抗恶意服务器	细粒度访问控制	密文之上的数据搜索	数据更新	外包加解密
文献[10]方案	√	√	×	×	外包解密
文献[11]方案	√	√	√	×	×
文献[27]方案	×	√	×	√	×
本文协议	√	√	√	√	√

表 4 性能对比

方案	数据主端的加密计算	数据用户端的解密计算	搜索令牌计算	数据搜索计算	数据更新开销
文献[10]方案	$T_{\mathrm{m}}+(\lvert Y\rvert+2)T_{\mathrm{e}}$	$S(T_{\mathrm{p}}+T_{\mathrm{e}})+2T_{\mathrm{m}}$	0	n_k	—
文献[11]方案	$2T_{\mathrm{p}}+(2\lvert Y\rvert+3)T_{\mathrm{e}}+$ $(\lvert Y\rvert+3)T_{\mathrm{m}}+(1+\lvert Y\rvert)H+T_{\mathrm{D}}$	$T_{\mathrm{p}}+T_{\mathrm{m}}+T_{\mathrm{D}}$	$n_k[(2k+1)T_{\mathrm{e}}+$ $kT_{\mathrm{m}}+(1+k)H]$	$n_k[2kT_{\mathrm{p}}+(2k+2)T_{\mathrm{e}}+$ $(2k-1)T_{\mathrm{m}}]$	—
文献[27]方案	$(2n_{\mathrm{z}}+2)T_{\mathrm{m}}+(3n_{\mathrm{z}}+2n_{\mathrm{z}}\lvert Y\rvert+$ $2\lvert Y\rvert+4)T_{\mathrm{e}}+T_{\mathrm{D}}+(n_{\mathrm{z}}+1)\lvert Y\rvert H$	$(4S+2)T_{\mathrm{p}}+(2S+1)T_{\mathrm{m}}+$ $2ST_{\mathrm{e}}+T_{\mathrm{D}}$	—	—	$3T_{\mathrm{m}}+T_{\mathrm{e}}+T_{\mathrm{D}}$
本文协议	$(4n_{\mathrm{z}}+4)T_{\mathrm{m}}+(4n_{\mathrm{z}}+5)T_{\mathrm{e}}+T_{\mathrm{D}}$	$2T_{\mathrm{P}}+3T_{\mathrm{m}}+T_{\mathrm{D}}$	n_kF	n_k	$3T_{\mathrm{m}}+T_{\mathrm{e}}+T_{\mathrm{D}}$

结构的最小属性集合数量。

1) 数据主端的加密计算。文献[10-11]方案没有实现数据的噪声化处理，所以数据主端的加密开销与访问策略树的叶子节点数量成正比。文献[27]方案的数据主端加密开销与叶子节点数量和噪声因子数量成正比。本文协议提供不同程度的噪声化数据，所以数据主端的加密开销与噪声因子数量成正比；又由于将访问策略计算的加密部分外包给了 CS，所以数据主端的加密开销与叶子节点数量无关。此外，本文协议只在初始化时需要计算噪声因子密文，后续对数据进行噪声化加密时只需采用数据更新算法即可，其复杂度与噪声因子数量无关。而文献[10-11]方案若要提供不同程度的噪声化数据，需要对数据执行 n_{z} 次加密，显然本文协议具有显著优势。

2) 数据用户端的解密计算。文献[10]方案将解密操作的部分加密外包给了 CS，但是数据用户端的解密开销与用户的属性数量成正比。文献[11]方案的数据用户端解密开销与访问控制策略和用户属性数量皆无关，仅需要一次双线性对运算，一次乘法运算和一次对称加密计算。文献[27]方案的数据用户端的解密开销与访问策略树的复杂程度和用户的属性数量有关，需要 $4S+2$ 次双线性对运算，$2S+1$ 次乘法运算，$2S$ 次指数运算，一次对称加密运算。本文协议将与访问控制策略相关的解密运算外包给了 CS，所以数据用户端的解密计算需要 2 次双线性对运算，3 次乘法运算，一次对称加密运算。

3) 搜索令牌计算。文献[27]方案不支持对密文进行检索，所以搜索令牌计算开销不存在。文献[10]方案采用明文关键字进行搜索，所以搜索令牌计算开销记为 0。文献[11]方案将属性基加密与可搜索加密结合，所以计算搜索陷门时需对用户的每个属性做运算，故计算搜索开销与用户的属性数量成正比；计算单个关键字陷门需要 $2k+1$ 次指数运算、k 次乘法运算，$1+k$ 次哈希运算。本文协议需对每个搜索关键词计算一次伪随机函数，因此计算代价为 n_kF 。

4) 数据搜索计算。本文协议和文献[10]方案在构建索引时皆采用基于键值对的查找表方式，因此数据搜索计算的代价皆为常数级。而文献[11]方案将访问控制策略应用于关键字搜索上，当用户的搜索陷门与索引陷门相匹配且用户的属性满足访问控制策略时，搜索匹配才能成功，所以文献[11]方案搜索开销与用户的属性数量成正比。

5) 数据更新开销。文献[27]方案和本文数据更新需要 3 次乘法运算，一次指数计算和一次对称加密计算，与访问策略和噪声因子数量无关，更新开销较少。

6.3 实验分析

为了更准确地评估协议性能，本文在系统初始化、密钥生成、用户加密、索引生成、搜索令牌生成、搜索、用户解密和用户更新密文的时间方面进行仿真测试。本节实验选取 JPBC（java pairing-based cryptography library）库中提供的 A 类椭圆曲线，伪随机函数采用 HMAC-SHA256，智能合约采用 solidity 语言，运行在以太坊虚拟机中，使用 web3.js 库对部署的 3 个智能合约进行调用和测试。本文实验的硬件环境为 Intel(R) Core(TM) i5-4210M CPU @ 2.60 GHz，RAM 为 8 GB。

数据加解密的运行开销如图 4 所示。图 4(a)描述了在不同数量的噪声因子下，数据主端的加密时间。数据主端加密开销主要包含三部分：数据明文的加密即共享组件的计算、会话组件的加密以及噪声集合的加密。共享组件计算和会话组件加密的时间复杂度为 $O(1)$，而噪声集合密文的计算开销与噪声因子数量成正比。由于使用了具有外包加解密的 CP-ABE 算

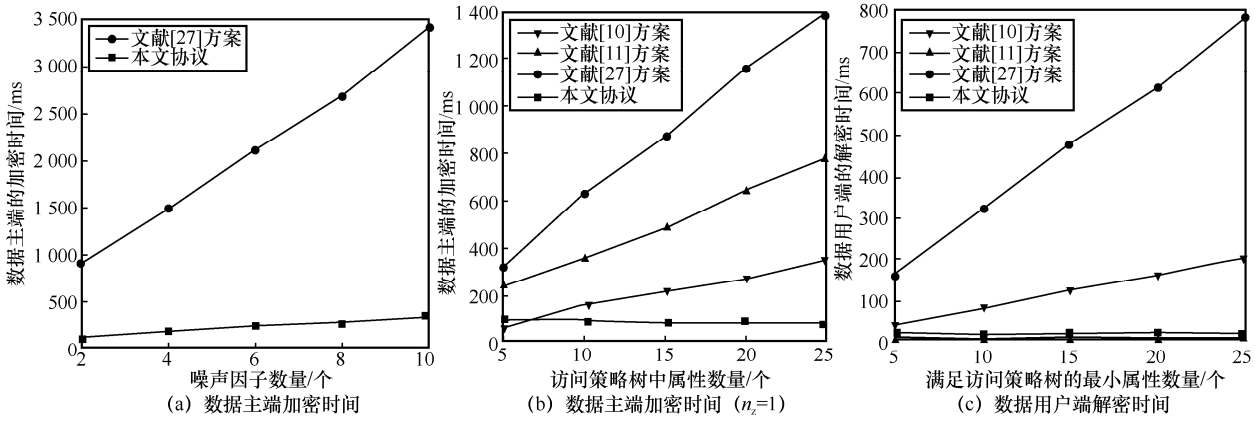

图 4　数据加解密的运行开销

法，有关访问策略的计算部分都外包给了 CS，用户端属性基加密的开销与属性数量无关，这很大程度减少了用户端的计算开销。此外，数据主仅在初始化时需计算噪声集合密文，在后续数据更新时不需要重新计算噪声密文。而文献[27]方案数据主端的加密时间包含了有关访问策略的计算部分，开销大于本文协议。图 4(b)描述了当噪声因子数量为 1 时，各方案的数据主端加密时间。由于使用了外包加解密的 CP-ABE 算法，本文协议的数据主端加密时间几乎不受属性数量影响，而文献[10-11,27]方案的数据主端加密时间与属性数量近似成正比。图 4(c)描述了在不同的满足访问策略树的最小属性数量下，数据用户端的解密时间。本文协议的数据用户端解密开销主要包含三部分：会话组件密文解密运算、噪声密文解密运算、加噪数据的运算。尽管数据主设置了多个噪声密文集，但是数据用户在解密时只需解密与其属性集合匹配的噪声密文，所以用户解密时间与噪声因子数量无关。同样，由于本文协议使用了外包加解密的 CP-ABE 算法，用户端的解密时间与属性数量也无关，解密开销较低。文献[11]方案中的数据用户端解密不包含与访问策略相关的运算，所以解密时间与属性数量无关。而文献[10-27]方案中的数据用户端解密包含与访问树相关的运算，所以解密时间与属性数量成正比。综上，本文协议在解密时间上略高于文献[11]方案，但远低于[10,27]方案。

图 5(a)描述了当用户属性数量为 10 时，各方案搜索令牌的生成时间和搜索时间。本文采用 HMAC-SHA256 算法生成搜索令牌，在 Java 语言环境下经 1 000 次独立实验测得搜索密钥初始化时间平均为 102 ms。当搜索关键词数量为 10 时，在搜索密钥已初始化的情况下，搜索令牌生成时间为 4 ms。当搜索关键词数量为 20 时，在搜索密钥已初始化的情况下，搜索令牌生成时间为 6 ms。搜索关键词数量越多，生成搜索令牌的花费开销平均到计算单个关键词搜索令牌的开销上越小，因此具有较高的效率。而文献[11]方案中搜索令牌的计算步骤包括指数运算、乘法运算和哈希运算，搜索令牌的生成时间与搜索关键词数量成正比。图 5(b)描述了当搜索关键词数量为 10 时，各方案的搜索令牌生成时间。本文协议搜索令牌生成时间与数据用户属性数无关。而文献[11]方案中搜索令牌生成时间与用户属性数量成正比。

图 5(c)和图 5(d)展示了各方案执行查询合约的搜索时间。假设索引表中存储 50 个键值对，每个键值对存储 10 个密文存储地址。图 5(c)描述了当用户属性数量为 10 时，各方案搜索时间与搜索关键词数量之间的关系。本文协议和文献[10]方案在构建索引时皆采用基于键值对的查找表方式，而文献[11]方案采用基于属性的可搜索加密的方式，其需与关键词和访问策略进行匹配，所以本文协议在搜索效率上高于文献[11]方案。图 5(d)描述了当搜索关键词数量为 10 时，各方案搜索时间与用户属性数量的关系。本文协议和文献[10]方案在构建索引时皆采用基于键值对的查找表方式，所以搜索计算开销与用户属性数无关。而文献[11]方案中的搜索计算开销与用户属性数量成正比。

图 6(a)和图 6(b)分别描述了数据主端数据更新时间与用户属性数量和噪声因子数量之间的关系。数据主端更新一个密文数据大约耗费 4.25 ms，更新的计算开销与访问策略和噪声因子数量皆无关，因此数据更新效率高。

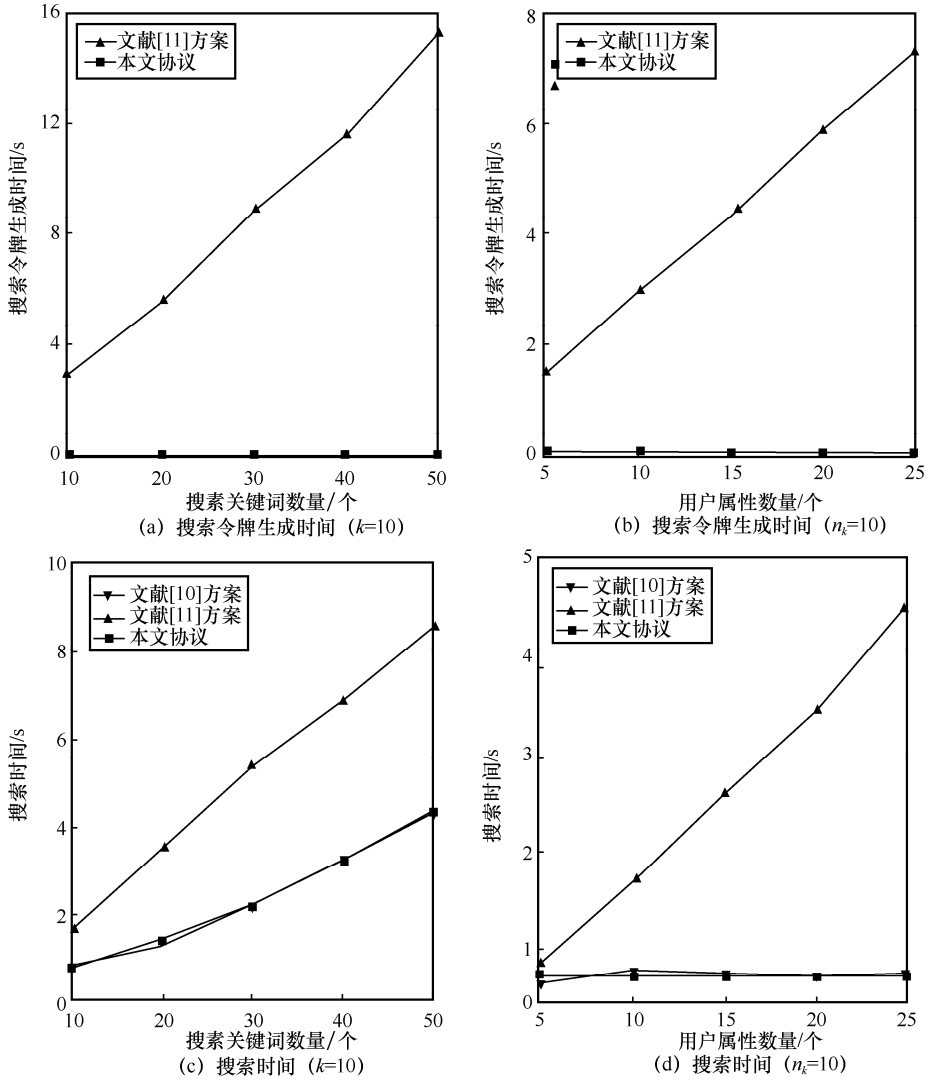

图 5　搜索令牌的生成时间和搜索时间

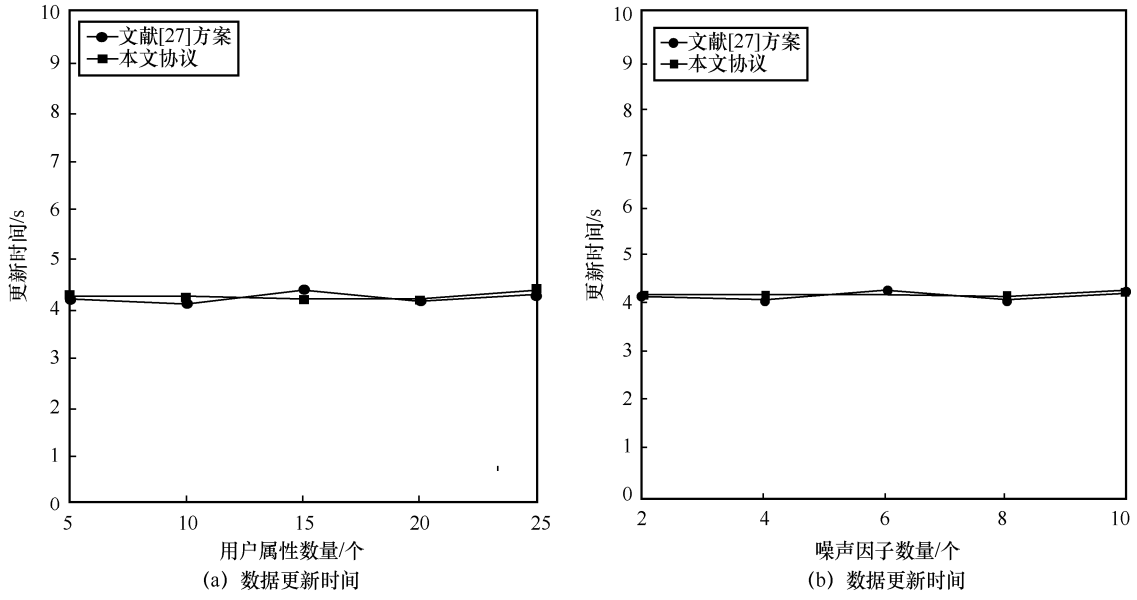

图 6　数据更新时间

7　结束语

本文采用密文策略属性基加密算法和智能合约技术，构造了一套安全且高效的噪声化数据分享控制协议。本文协议实现了对不同类型的数据用户可以访问到的数据精度的控制。首先，为了抵抗恶意服务器，本文协议利用智能合约机制实现数据搜索；其次，为了减轻客户端的计算开销，属性加解密过程被外包给云服务器；其次，本文协议支持密文的快速更新，更新的时间开销与访问策略复杂度和噪声因子数量无关；最后，安全性分析证明本文协议具有密文不可区分性和适应性选择关键词语义的安全性，性能分析与实验评估证明了本文协议在客户端的计算开销、搜索效率和数据更新开销等方面具有优良的性能。

参考文献：

[1] 代闯闯, 栾海晶, 杨雪莹, 等. 区块链技术研究综述[J]. 计算机科学, 2021, 48(S2): 500-508.
DAI C C, LUAN H J, YANG X Y, et al. Overview of blockchain technology[J]. Computer Science, 2021, 48(S2): 500-508.

[2] LIU H, YANG B, XIONG X R, et al. A financial management platform based on the integration of blockchain and supply chain[J]. Sensors, 2023, 23(3): 1497.

[3] 王晨旭, 程加成, 桑新欣, 等. 区块链数据隐私保护: 研究现状与展望[J]. 计算机研究与发展, 2021, 58(10): 2099-2119.
WANG C X, CHENG J C, SANG X X, et al. Data privacy-preserving for blockchain: state of the art and trends[J]. Journal of Computer Research and Development, 2021, 58(10): 2099-2119.

[4] 牛淑芬, 刘文科, 陈俐霞, 等. 基于联盟链的可搜索加密电子病历数据共享方案[J]. 通信学报, 2020, 41(8): 204-214.
NIU S F, LIU W K, CHEN L X, et al. Electronic medical record data sharing scheme based on searchable encryption via consortium blockchain[J]. Journal on Communications, 2020, 41(8): 204-214.

[5] 薛腾飞, 傅群超, 王枞, 等. 基于区块链的医疗数据共享模型研究[J]. 自动化学报, 2017, 43(9): 1555-1562.
XUE T F, FU Q C, WANG C, et al. A medical data sharing model via blockchain[J]. Acta Automatica Sinica, 2017, 43(9): 1555-1562.

[6] CHEN Y W, HU B W, YU H J, et al. A threshold proxy re-encryption scheme for secure IoT data sharing based on blockchain[J]. Electronics, 2021, 10(19): 2359.

[7] SAHAI A, WATERS B. Fuzzy identity-based encryption[C]//Lecture Notes in Computer Science. Berlin: Springer, 2005: 457-473.

[8] GOYAL V, PANDEY O, SAHAI A, et al. Attribute-based encryption for fine-grained access control of encrypted data[C]//Proceedings of the 13th ACM conference on Computer and Communications Security. New York: ACM Press, 2006: 89-98.

[9] 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.

[10] 汪玉江, 曹成堂, 游林. 基于区块链和属性基加密的个人隐私数据保护方案[J]. 密码学报, 2021, 8(1): 14-27.
WANG Y J, CAO C T, YOU L. A novel personal privacy data protection scheme based on blockchain and attribute-based encryption[J]. Journal of Cryptologic Research, 2021, 8(1): 14-27.

[11] 牛淑芬, 杨平平, 谢亚亚, 等. 区块链上基于云辅助的密文策略属性基数据共享加密方案[J]. 电子与信息学报, 2021, 43(7): 1864-1871.
NIU S F, YANG P P, XIE Y Y, et al. Cloud-assisted ciphertext policy attribute based encryption data sharing encryption scheme based on blockchain[J]. Journal of Electronics & Information Technology, 2021, 43(7): 1864-1871.

[12] ZHANG Q Y, ZHAO Z Y. Distributed storage scheme for encryption speech data based on blockchain and IPFS[J].The Journal of Supercomputing, 2023, 79(1): 897-923.

[13] WU N, XU L, ZHU L. A blockchain based access control scheme with hidden policy and attribute[J]. Future Generation Computer Systems, 2023, 141: 186-196.

[14] YIN H J, CHEN E, ZHU Y, et al. Attribute-based private data sharing with script-driven programmable ciphertext and decentralized key management in blockchain Internet of things[J]. IEEE Internet of Things Journal, 2022, 9(13): 10625-10639.

[15] ZHANG Y, ZHANG L, WU Q, et al. Blockchain-enabled efficient distributed attribute-based access control framework with privacy-preserving in IoV[J]. Journal of King Saud University - Computer and Information Sciences, 2022, 34(10): 9216-9227.

[16] YU G S, ZHA X, WANG X, et al. Enabling attribute revocation for fine-grained access control in blockchain-IoT systems[J]. IEEE Transactions on Engineering Management, 2020, 67(4): 1213-1230.

[17] LI F Q, LIU K M, ZHANG L P, et al. EHRChain: a blockchain-based EHR system using attribute-based and homomorphic cryptosystem[J]. IEEE Transactions on Services Computing, 2022, 15(5): 2755-2765.

[18] LIU J W, WU M L, SUN R, et al. BMDS: a blockchain-based medical data sharing scheme with attribute-based searchable encryption[C]//Proceedings of IEEE International Conference on Communications. Piscataway: IEEE Press, 2021: 1-6.

[19] LONG H Q, HOU J, LI Q M, et al. Data privacy protection of industrial blockchain[C]//International Conference on Cloud Computing. Berlin: Springer, 2021: 83-99.

[20] HUANG L, LEE H H. A medical data privacy protection scheme based on blockchain and cloud computing[J]. Wireless Communica-

tions and Mobile Computing, 2020, 2020: 1-11.

[21] REGUEIRO C, SECO I, DIEGO S D, et al. Privacy-enhancing distributed protocol for data aggregation based on blockchain and homomorphic encryption[J]. Information Processing & Management, 2021, 58(6): 102745.

[22] GOCHHAYAT S P, BANDARA E, SHETTY S, et al. Yugala: blockchain based encrypted cloud storage for IoT data[C]//Proceedings of IEEE International Conference on Blockchain (Blockchain). Piscataway: IEEE Press, 2020: 483-489.

[23] AGYEKUM K O B O, XIA Q, SIFAH E B, et al. A proxy re-encryption approach to secure data sharing in the Internet of things based on blockchain[J]. IEEE Systems Journal, 2021, 16(1): 1685-1696.

[24] ZHANG L Y, ZHANG Y, WU Q, et al. A secure and efficient decentralized access control scheme based on blockchain for vehicular social networks[J]. IEEE Internet of Things Journal, 2022, 9(18): 17938-17952.

[25] LIU J, ZHAO J, HUANG H H, et al. A novel logistics data privacy protection method based on blockchain[J]. Multimedia Tools and Applications, 2022, 81(17): 23867-23887.

[26] ZHANG P, CHEN Z H, LIU J K, et al. An efficient access control scheme with outsourcing capability and attribute update for fog computing[J]. Future Generation Computer Systems, 2018, 78: 753-762.

[27] XIE Q Q, HOU Y T, CHENG K, et al. Flexibly and securely shape your data disclosed to others[C]//Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security. New York: ACM Press, 2019: 160-167.

[作者简介]

谢晴晴（1990− ），女，安徽宿州人，博士，江苏大学讲师，主要研究方向为区块链、应用密码学。

杨念民（1998− ），男，江西赣州人，江苏大学硕士生，主要研究方向为区块链、应用密码学等。

冯霞（1983− ），女，江苏镇江人，博士，江苏大学副教授，主要研究方向为物联网认证协议、区块链和应用密码学。