doi:10.11959/j.issn.1000-436x.2017225

多数据拥有者认证的密文检索方案

伍祈应，马建峰，苗银宾，张俊伟，沈丽敏
（西安电子科技大学网络与信息安全学院，陕西 西安 710071）

摘 要：针对共享型多数据拥有者场景，即一个文件被多个数据拥有者拥有，已有可搜索加密方案不能同时支持密文检索和细粒度访问控制。为此，基于线性秘密共享和可搜索加密技术提出一种高效的多数据拥有者认证的密文检索方案，数据用户只有得到多个数据拥有者的授权才能解密返回的结果。严格的安全分析表明该方案在双线性 Diffie-Hellman 假设下能保证安全和隐私，且基于实际数据的实验结果表明本方案在实际应用中是高效、可行的。

关键词：共享型多数据拥有者场景；可搜索加密；细粒度访问控制；线性秘密共享
中图分类号：TN918.4 **文献标识码**：A

Multi-owner accredited keyword search over encrypted data

WU Qi-ying, MA Jian-feng, MIAO Yin-bin, ZHANG Jun-wei, SHEN Li-min
(School of Cyber Engineering, Xidian University, Xi'an 710071, China)

Abstract: A sharing multi-owner setting where data was owned by a fixed number of data owners, the existing searchable encryption schemes could not support ciphertext retrieval and fine-grained access control at the same time. For this end, an efficient cryptographic primitive called as multi-owner accredited keyword search over encrypted data scheme was designed, through combining linear secret-sharing technique with searchable encryption schemes, only the data users authorized bymulti-owner by could decrypt the returned results. The formal security analysis shows that the scheme can protect security and privacy under the bilinear Diffie-Hellman assumption. As a further contribution, an empirical study over real-world dataset was conelucted to show the effectiveness and practicability of the scheme.

Key words: sharing multi-owner setting, searchable encryption, fine-grained access control, linear secret-sharing

1 引言

随着云计算的发展[1,2]，云计算厂商(如微软云、谷歌云、阿里云等)由于其灵活的收费方式和便利的存储计算服务，吸引着越来越多的企业和个人将本地数据迁移到云服务器上。由于存储在云服务器上的数据脱离了数据拥有者的物理控制，因此数据存在严重的安全隐患[3,4]。为保护数据的隐私性，敏感数据在被外包给云端前通常需要进行加密。由于密文数据不再具备一些明文数据的特性，导致传统的明文检索技术无法在密文上进行数据检索。为获取感兴趣的文档，将密文全部下载到本地再进行解密的方式浪费了大量的带宽和计算资源。为此，Boneh 等[5]利用身份加密技术提出了公钥可搜索加密方案，实现了单关键字密文检索，并在邮件系统中得到广泛的应用。为提供丰富的检索方式，之后支持连接关键词[6,7]，模糊关键词[8,9]以及搜索结果排序[10,11]的可搜索加密方案相继被提出。

然而，大部分已有的可搜索加密方案不支持细粒度访问控制，为此，Bethencourt 等[12]提出了基于密文策略的属性加密方案，当密钥的属性集满足密文的访问结构时数据用户即可解密密文数据，从而

收稿日期：2017-05-31；**修回日期**：2017-10-30
基金项目：国家高技术研究发展计划（“863”计划）基金资助项目（No.2015AA016007）；国家自然科学基金资助项目（No.61702404, No.61472310, No. U1405255）；“111 计划”基金资助项目（No.B16037）
Foundation Items: The National High Technology Research and Development Program (863 Program) (No. 2015AA016007), The National Natural Science Foundation of China (No.61702404, No.61472310, No. U1405255), China 111 Project (No.B16037)

实现了细粒度访问控制。接着，从另一方面，Goyal 等[13]提出了基于密钥策略属性加密方案，当密文的属性集满足密钥的访问结构时数据用户即可解密密文数据，由此实现了细粒度访问控制。但是上述方案均不支持密文检索，为此，Zheng 等[14]结合属性加密和可搜索加密技术提出了同时支持密文检索和细粒度访问控制的方案。然而，该方案只适用于单个数据拥有者共享数据的场景，从单数据拥有者场景扩展到多数据拥有者场景会带来许多问题，一方面多数据拥有者利用不同的密钥加密文档，带来了复杂的密钥管理难题；另一方面数据用户需要和多数据拥有者直接交互来获取授权，不仅要求多数据拥有者实时在线，而且增大了隐私泄露的风险。如何设计方案适应于多个数据拥有者场景成为亟待解决的难题，为此，Yin 等[15]设计方案让多数据拥有者利用随机密钥为不同的文件建立索引，数据用户可以选取另外的随机密钥生成陷门而不影响密文匹配，尽管该方案在多数据拥有者场景下减少了密钥管理的复杂度，但是不支持细粒度访问控制。为此，Sun 等[16]利用密文策略的属性加密和代理重加密技术提出了支持细粒度访问控制的密文检索方案。该方案由云服务器对数据用户进行认证，避免了数据用户与多数据拥有者直接交互。

然而，上述方案不能适用于共享型多数据拥有者场景，即一个文件由多个数据拥有者共同控制。如图 1 所示，例如，公司的重要文件由董事会的成员共同控制，普通职员只有获得董事会成员的授权才能获取文件。为此，Layouni 等[17] 在共享型多数据拥有者场景下提出了支持多数据拥有者认证的加密方案，只有经多个数据拥有者授权，数据用户才能解密密文数据。进一步地，Shan 等[18]利用属性加密技术提出了支持多数据拥有者认证的属性加密方案，不仅在共享型多数据拥有者场景下增强了认证的灵活性和效率，而且实现了细粒度访问控制。但是，上述方案均不具备密文检索功能。

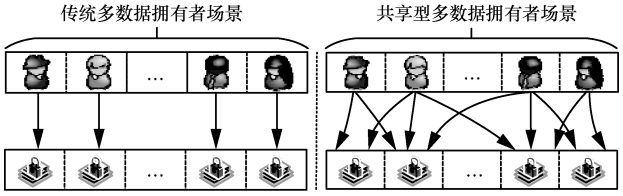


图 1 场景对比

针对上述问题，考虑在共享型多数据拥有者场景下无法同时支持细粒度的访问控制与密文检索。本文方案结合线性秘密共享和可搜索加密技术，使只有获得多个数据拥有者认证的数据用户才能解密返回的结果。表 1 给出了本文方案与其他方案的功能比较。

1) 共享型多数据拥有者场景：在共享型多数据拥有者场景下，只有获得多个数据拥有者认证的数据用户才能解密返回的结果。

2) 安全性：在双线性 Diffie-Hellman 假设下该方案是保护安全和隐私的，并且未经授权的数据用户无法访问密文数据。

3) 效率性：基于实际数据集的性能分析表明文中方案在实际应用场景中是可行的、高效的。

表 1 功能比较

方案	SMOS	TMOS	FGAC
文献[14]方案	×	×	√
文献[16]方案	×	√	√
文献[17]方案	√	√	×
本文方案	√	√	√

如表 1 所示，SMOS (sharing multi-ower setting)表示共享型多数据拥有者场景，TMOS (traditional multi-ower setting)表示传统多数据拥有者场景，FGAC (fine-grained access control)表示细粒度访问控制，“√”表示支持，“×”表示不支持。

2 系统模型和安全定义

2.1 系统模型

本文方案的系统模型如图 2 所示，具体包括 4 个实体：可信机构(TA, trust authority)、数据拥有者(DO, data owner)、数据用户(DU, data users)和云服务器(CS, cloud server)。

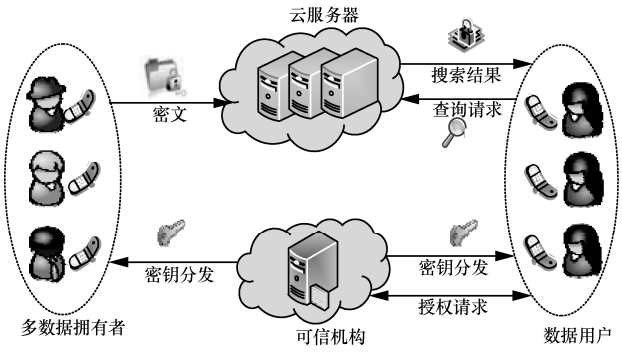


图 2 系统模型

1) 可信机构。TA 是完全可信的，其主要负责系统初始化生成公共参数和主密钥，负责为 DU 和 DO 生成和分发密钥，验证 DU 是否授权。

2) 数据拥有者。DO 根据传统加密方案利用对称密钥集将文档集加密成密文集；利用线性秘密共享技术将对称密钥集加密生成密文密钥集；根据文档提取的关键字集建立索引，并将密文集、密文密钥集和索引上传给 CS。

3) 数据用户。DU 根据查询关键字生成陷门，并将陷门发送给 CS，只有经 DO 授权的 DU 才能正确解密文档。

4) 云服务器。CS 是诚实但好奇的，其诚实地执行既定协议，但又会好奇地获取敏感信息。CS 根据陷门匹配索引得到返回密文集，并发送给 DU。

2.2 安全定义

定义 1 在双线性 Diffie-Hellman 假设下，本文方案是保护隐私安全的，具体要求如下。

1) 数据隐私安全。方案是保护文档集、索引集和对称密钥集隐私安全的，且云服务器通过统计分析无法获得密文中的相关明文信息。

2) 数据拥有者隐私安全。只有经多数据拥有者认证的数据用户才能访问密文数据，且非法的数据拥有者或合谋的数据拥有者无法伪造合法的授权。

3) 数据用户隐私安全。非法的数据用户无法生成有效的陷门去访问敏感数据，且云服务器无法获得陷门中的明文关键字信息。

定义 2 考虑关键字语义安全性，按如下定义选择关键字攻击游戏。

初始化 挑战者 C 选择安全参数 k，并且执行初始化算法生成主密钥 msk 和公共参数 pm，然后把公共参数 pm 发送给敌手 A，主密钥 msk 由挑战者 C 持有。

阶段 1 敌手 A 向挑战者 C 查询关键字 $w_a,\cdots,w_t$ 的密文索引。

挑战 敌手 A 选择 2 个挑战关键字 (w_0,w_1) 并发送给挑战者 C，并要求 w_0 和 w_1 在阶段 1 处未被查询。挑战者 C 随机选择 $b\in\{0,1\}$，并将密文索引返回给敌手 A。

阶段 2 敌手 A 重复阶段 1 适应性地查询关键字 w_i 的密文索引，其中，$w_i\neq w_0,w_1$。

猜测 敌手 A 输出猜测比特 $b'\in\{0,1\}$。如果 $b'=b$，则敌手 A 在游戏中获胜。

定义敌手 A 攻破安全游戏的优势为 $Adv_{\mathrm{A}}(1^k)=\left|\Pr[b\neq b']-\frac{1}{2}\right|$。

3 预备知识

本节给出与本文方案相关的基础知识及相关定义。

3.1 双线性映射

假设 G_1,G_T 是阶为素数 p 的循环群，g 是群 G_1 的生成元，双线性映射 $e:G_1\times G_1\to G_T$ 满足以下性质。

1) 双线性：对任意的 $x,y\in G_1$，$a,b\in Z_p$，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 $x,y\in G_1$，存在有效的算法计算 $e(x,y)$。

3.2 访问结构

令 $P=\{P_1,P_2,\cdots,P_n\}$ 表示参与者集合，若存在访问结构 $\mathrm{P}\subseteq 2^P$，如果对于任意的集合 $B,C\subseteq P$，若 $B\in\mathrm{P}$ 并且 $B\subseteq C$，则有 $C\in\mathrm{P}$，则称 P 是单调的访问结构。访问结构 P 是集合 2^P 的一个非空子集。访问结构 P 中的集合称为授权集合，不在访问结构 P 中的集合称为非授权集合。

3.3 线性秘密共享

定义在实体集 P 上的秘密共享方案是 Z_p 域上线性的，满足以下条件。

1) 各方共享的秘密组成 Z_p 域上的矩阵。

2) 定义 P 是由 (M,ρ) 表示的访问结构，其中，$\boldsymbol{M}$ 是一个 $l\times n$ 的线性矩阵，$\boldsymbol{M}_i$ 表示矩阵 $\boldsymbol{M}$ 的第 i 行向量，$\boldsymbol{M}_i$ 定义为一个实体 $\rho(i)(1\leqslant i\leqslant l)$。$\rho$ 是一个从 $\{1,2,\cdots,l\}$ 到 P 的映射，选取一个随机向量 $\boldsymbol{v}=\{s,r_2,\cdots,r_n\}$，其中，$s\in Z_P$ 表示共享的秘密值，$r_2,\cdots,r_n$ 表示随机值，则 $\boldsymbol{M}\cdot\boldsymbol{v}^{\mathrm{T}}$ 表示利用秘密共享方案得到的关于秘密值 s 的 l 个共享子秘密，其中，共享子秘密 $\lambda_i=(\boldsymbol{M}\cdot\boldsymbol{v}^{\mathrm{T}})_i$ 是属于实体 $\rho(i)$ 的。

假定 $P'\in\mathrm{P}$ 是一个授权集合，定义 $I\subset\{1,2,\cdots,l\}$ 且 $I=\{i\mid\rho(i)\in P'\}$，则一定存在常数集合 $\{\omega_i\in Z_p\}_{i\in I}$，使任意合法的共享秘密 $\{\lambda_i\}_{i\in I}$，有 $\sum_{i\in I}\omega_i\lambda_i=\sum_{i\in I}\omega_i(\boldsymbol{M}_i\cdot\boldsymbol{v}^{\mathrm{T}})=s$。其中，常数集合 $\{\varphi_i\}$ 可以在多项式时间内计算出来。而对于未授权的集合，这些常量值不存在。

3.4 双线性 Diffie-Hellman（BDH, bilinear Diffie-Hellman）假设

假设 G_1,G_T 是阶为素数 p 的循环群，g 是群 G_1

的生成元，k 是安全参数，给定四元组 (g,g^a,g^b,g^c)，其中，$a,b,c\in Z_p$，计算 $e(g,g)^{abc}\in G_T$。给定五元组 (g,g^a,g^b,g^c,D)，其中，$a,b,c\in Z_p$，$D\in G_T$，判断是否 $D=e(g,g)^{abc}$。多项式时间敌手 A 以 ε 的优势解决双线性 Diffie-Hellman 问题，如果不等式 $|\Pr[A(g,g^a,g^b,g^c,e(g,g)^{abc})=0]-\Pr[A(g,g^a,g^b,g^c,D)=0]|\geqslant\varepsilon(k)$ 成立。如果不存在多项式时间敌手 A 以不可忽视的优势 ε 解决双线性 Diffie-Hellman 问题，则称双线性 Diffie-Hellman 假设成立。

4 方案描述

4.1 方案定义

1) $\text{Setup}(1^k)\rightarrow(msk,pm)$。给定安全参数 k，完全可信的 TA 输出双线性映射参数 (G,G_T,e,p,g)，并计算生成主密钥 msk 和公共参数 pm，其中，主密钥 msk 被 TA 私有。

2) $\text{KeyGen}(pm,msk)\rightarrow(sk_{y_j},pk_{y_j},sk_{ID})$。给定数据拥有者集合 $Y=\{y_j\}$，TA 首先为每个数据拥有者 y_j 生成授权私钥 sk_{y_j} 和授权公钥 pk_{y_j}，然后，为身份为 ID 的数据用户生成用户私钥 sk_{ID}。

3) $\text{Enc}(pm,K,W,F,\text{P},pk_{y_j})\rightarrow(C,I)$。给定文件集 $F=\{f_i\}$，数据拥有者集合根据传统对称加密方案，利用对称密钥集 $K=\{k_i\}$ 将文件集 $F=\{f_i\}$ 加密为密文集，利用基于线性秘密共享技术将对称密钥集 $K=\{k_i\}$ 加密为密文密钥集 $C=\{C_i\}$；给定关键字集 $W=\{w_t\}$，数据拥有者集合生成索引集 $I=\{I_t\}$，并将密文集，密文密钥集 $C=\{C_i\}$ 和索引集 $I=\{I_t\}$ 发送给 CS。

4) $\text{Trap}(pm,sk_{ID},w')\rightarrow(T_{w'})$。身份是 ID 的数据用户根据查询关键字 w' 生成陷门 $T_{w'}$，并将陷门 $T_{w'}$ 发送给 CS。

5) $\text{Search}(pm,T_{w'},I,C)\rightarrow(C')$。云服务器匹配索引集 I 和陷门 $T_{w'}$ 得到返回密文集，并将返回密文集和对应的返回密文密钥集 C' 发送给身份是 ID 的数据用户。

6) $\text{Dec}(sk_{ID},C',pm,ID)\rightarrow(\{k_i\})$。TA 验证身份是 ID 的数据用户是否在授权用户列表中，若不在，则输出 0；否则，TA 将生成解密授权集 $\{A_{y_j}\}$，并发送给身份是 ID 的数据用户。身份是 ID 的数据用户利用线性秘密共享技术恢复对称密钥 $\{k_i\}$，从而解密得到明文文件。

4.2 方案构造

在介绍本文方案的具体定义之前，给出方案用到的符号定义，如表 2 所示。

表 2 符号定义

符号	定义
$F=\{f_i\},1\leqslant i\leqslant m$	文件集
$W=\{w_t\},1\leqslant t\leqslant\tau$	关键字集
$Y=\{y_j\},1\leqslant j\leqslant l$	数据拥有者集合
$K=\{k_i\},1\leqslant i\leqslant m$	对称密钥集
$C=\{C_i\},1\leqslant i\leqslant m$	密文密钥集
$I=\{I_t\},1\leqslant t\leqslant\tau$	索引集
$\{A_{y_j}\},1\leqslant j\leqslant l$	解密授权集
I_t	关键字 w_t 的索引
$T_{w'}$	陷门
C'	返回密文密钥集

$\text{Setup}(1^k)\rightarrow(msk,pm)$：初始化阶段，给定安全参数 k，完全可信的 TA 首先输出双线性映射参数 (G,G_T,e,p,g)，其中，G 和 G_T 是阶为素数 p 的乘法循环群，g 是群 G 的生成元，双线性映射 $e:G\times G\rightarrow G_T$，接着定义散列函数 $H:\{0,1\}^*\rightarrow G$，选取随机数 $a,b\in Z_p$，计算生成主密钥 msk 和公共参数 pm，其中，主密钥 msk 被 TA 私有。

$$msk=(a,b) \tag{1}$$

$$pm=(G,G_T,e,p,g,g^b,e(g,g)^a,H) \tag{2}$$

$\text{KeyGen}(pm,msk)\rightarrow(sk_{y_j},pk_{y_j},sk_{ID})$。私钥生成，给定数据拥有者集合 $Y=\{y_j\}$，TA 首先为每个数据拥有者 y_j 选取随机数 $u_j\in Z_p$ 并计算 g^{u_j}，记授权私钥 $sk_{y_j}=u_j$ 和授权公钥 $pk_{y_j}=g^{u_j}$；然后，对身份为 ID 的数据用户选取随机数 $u\in Z_p$ 并计算生成用户私钥 $sk_{ID}=\{sk_{ID,1},sk_{ID,2},sk_{ID,3}\}$。其中，$sk_{ID,1}=u$，$sk_{ID,2}=g^ag^{bu}$，$sk_{ID,3}=g^aH(ID)^b$。

$\text{Enc}(pm,K,W,F,\text{P},pk_{y_j})\rightarrow(C,I)$。加密阶段，给定文件集 $F=\{f_i\}$，数据拥有者集合根据传统对称加密方案，利用对称密钥集 $K=\{k_i\}$ 将文件集 $F=\{f_i\}$ 加密为密文集；接着，根据线性秘密共享技术，定义访问结构 P。其中，P 由 $(\boldsymbol{M},\rho)$ 表示，$\boldsymbol{M}$ 是一个 $l\times n$ 的线性矩阵，$\boldsymbol{M}_j$ 表示矩阵 $\boldsymbol{M}$ 的第

j 行向量，ρ 是一个单映射函数，将矩阵的每一行映射成一个数据拥有者，选取一个随机向量 $\boldsymbol{v}=\{s,r_2,\cdots,r_n\}$，其中，$s\in Z_P$ 表示共享的秘密值，计算共享子秘密 $\lambda_j=\boldsymbol{M}_j\cdot\boldsymbol{v}^{\mathrm{T}}$，将对称密钥 k_i 加密为密文密钥 $C_i=\{C_{i,j},C_{i,\alpha},C_{i,\beta}\}$，其中，$C_{i,j}=g^{b\lambda_j}g^{-u_js}$，$1\leqslant j\leqslant l$，$C_{i,\alpha}=k_ie(g,g)^{as}$，$C_{i,\beta}=g^s$。

给定关键字集 $W=\{w_t\}$，如图 3 所示，数据拥有者集合提取文档的关键字，并对包含关键字 w_t 的文件生成索引 I_t，选取随机数 $\theta\in Z_p$ 并计算生成索引 $I_t=\{I_{t,1},I_{t,2}\}$，其中，$I_{t,1}=e(g,g)^{a\theta}e(g,H(w_t))^{\theta}$，$I_{t,2}=g^{\theta}$，最后，数据拥有者集合将密文集，密文密钥集 $C=\{C_i\}$ 和索引集 $I=\{I_t\}$ 发送给 CS。

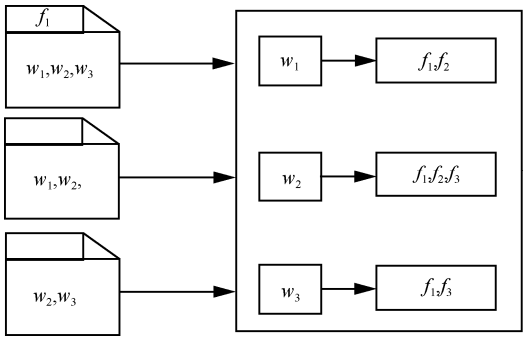


图 3　索引结构

$\mathrm{Trap}(pm,sk_{ID},w')\rightarrow(T_{w'})$。陷门生成，身份是 ID 的数据用户根据查询关键字 w' 生成陷门 $T_{w'}=\{T_{w',1},T_{w',2}\}$，其中，$T_{w',1}=H(w')g^ag^{bu}$，$T_{w',2}=g^{bu}$，并将陷门 $T_{w'}$ 发送给 CS。

$\mathrm{Search}(pm,T_{w'},I,C)\rightarrow(C')$。密文搜索，云服务器根据式(3)匹配索引集 I 和陷门 $T_{w'}$ 得到返回密文集，并将返回密文集和对应的返回密文密钥集 C' 发送给身份是 ID 的数据用户。

$$I_{i,1}\cdot e(I_{i,2},T_{w',2})=e(I_{i,2},T_{w',1}) \tag{3}$$

$\mathrm{Dec}(sk_{ID},C',pm,ID)\rightarrow(\{k_i\})$。密文解密，TA 验证身份是 ID 的数据用户是否在授权用户列表中，若不在，则输出 0；否则 TA 将生成解密授权集 $\{A_{y_j}=H(ID)^{u_j}\}$，并发送给身份是 ID 的数据用户。假定 $P'\in$P 是一个授权的数据拥有者集合，$\eta=\{j\mid\rho(j)\in P'\}\subset\{1,2,\cdots,l\}$，根据线性秘密共享技术，则一定存在常数集合 $\{\varphi_j\in Z_p\}_{j\in\eta}$，使 $\sum_{j\in\eta}\varphi_j\lambda_j=s$。身份是 ID 的数据用户根据解密授权集 $\{A_{y_j}\}_{y_j\in P'}$ 和用户私钥 sk_{ID} 按照式(4)计算得到 $e(g,g)^{as}$，接着，计算 $k_i=\dfrac{C_{i,\alpha}}{e(g,g)^{as}}$ 得到对称密钥，从而解密得到明文文件。

$$\frac{e(C_{i,\beta},sk_{ID,3})}{\prod_{j\in\eta}(e(C_{i,j},H(ID))e(A_{y_j},C_{i,\beta}))^{\varphi_j}}=e(g,g)^{as} \tag{4}$$

4.3　正确性分析

为了验证式(3)的正确性，当数据用户提交的关键字满足 $w'=w_t$，有

$$\begin{aligned}e(I_{i,2},T_{w',1})&=e(g^{\theta},H(w')g^ag^{bu})\\&=e(g,H(w'))^{\theta}e(g,g)^{(bu+a)\theta}\\e(I_{i,2},T_{w',2})&=e(g^{\theta},g^{bu})=e(g,g)^{bu\theta}\\I_{i,1}&=e(g,g)^{a\theta}e(g,H(w_t))^{\theta}\\e(I_{i,2},T_{w',1})&=I_{i,1}\cdot e(I_{i,2},T_{w',2})\end{aligned}$$

由此可以验证式(3)的正确性。

为了验证式(4)的正确性，当数据用户是被授权访问文件，数据用户可以得到 $e(g,g)^{as}$

$$\begin{aligned}&e(C_{i,j},H(ID)e(A_{y_j},C_{i,\beta})\\&=e(g^{b\lambda_j}g^{-u_js},H(ID))\cdot e(g,H(ID))^{su_j}\\&=e(g,H(ID))^{b\lambda_j-u_js}\cdot e(g,H(ID))^{su_j}\\&=e(g,H(ID))^{b\lambda_j}\\&\frac{e(C_{i,\beta},sk_{ID,3})}{\prod_{j\in\eta}(e(C_{i,j},H(ID))e(A_{y_j},C_{i,\beta}))^{\varphi_j}}\\&=\frac{e(g^s,g^aH(ID)^b)}{e(g,H(ID))^{\sum_{j\in\eta}b\lambda_j\varphi_j}}=\frac{e(g,g)^{as}e(g,H(ID))^{bs}}{e(g,H(ID))^{b\sum_{j\in\eta}\lambda_j\varphi_j}}\\&=\frac{e(g,g)^{as}e(g,H(ID))^{bs}}{e(g,H(ID))^{bs}}=e(g,g)^{as}\end{aligned}$$

由此可以验证式(4)的正确性。

5　安全性

基于 BDH 假设，本文方案能保证 3 种类型的安全，分别介绍如下。

1) 数据隐私安全。本文方案利用传统加密算法对文档集 $F=\{f_i\}$ 进行加密，文档集以密文形式存储在云服务器上，从而保证了文档集的隐私安全。本文方案利用散列函数，加密生成密文索引 $I_t=\{I_{t,1}=e(g,g)^{a\theta}e(g,H(w_i))^{\theta},I_{t,2}=g^{\theta}\}$，不会泄露任

何明文信息给云服务器，从而保证了索引集的隐私安全，此外，基于 BDH 假设，本文方案是抵抗选择关键字攻击的。本文方案利用线性秘密共享技术将对称密钥集 $K=\{k_i\}$ 加密为密文密钥集 $C=\{C_i\}$，假定身份为 ID 的数据用户获得部分用户私钥 $sk_{ID,3}=g^aH'(ID)^b$ 和解密授权集 $\{A_{y_j}=H'(ID)^{u_j}\}$。当解密授权集不满足访问结构，令 $H'(ID)^{u_j}=H'(x_j)^{u^*}$, $g^aH'(ID)^b=g^ag^{bu^*}$，这样的 u^*,x_j 可以被证明是存在的。由此，密文密钥集 $C_{i,j}=g^{b\lambda_j}g^{-u_js}$ 可以被写成 $C_{i,j}=g^{b\lambda_j}H'(x_j)^{-s}$。根据CP-ABE方案[12]，在 BDH 假设成立的前提下，本文方案是保护密钥集隐私安全的。特别地，本文给出密文索引的安全性证明，如定理 1 所示。

定理 1 基于 BDH 假设，本文方案是抵抗选择关键字攻击游戏的。

证明 若敌手 A 以不可忽略的优势赢得选择关键字攻击游戏，则挑战者 C 利用敌手 A 能以不可忽略的优势解决 BDH 难题。具体过程如下。

初始化 给定 BDH 参数为 $u_1=g^{\alpha}$，$u_2=g^{\beta}$，$u_3=g^{\gamma}\in G$，其中 g 是群 G 的生成元，目的是计算 $e(g,g)^{\alpha\beta\gamma}\in G_T$。挑战者 C 选择安全参数 k 以及散列函数 $H:\{0,1\}^*\to G$，执行初始化算法得到公共参数 $pm=(G,G_T,e,p,g,H)$，并发送给敌手 A。

阶段 1 敌手 A 询问随机预言机 H，挑战者 C 维护一个 (w_i,h_i,a_i,c_i) 的 H 列表，列表初始化为空。敌手 A 向 H 发送 $w_i\in\{0,1\}^*$，挑战者 C 做如下处理：如果 w_i 已经在 H 列表中，挑战者 C 将 $H(w_i)=h_i\in G$ 发送给敌手 A。否则，挑战者 C 选择随机数 $c_i\in\{0,1\}$，$a_i\in Z_p$，如果 $c_i=0$，挑战者 C 计算 $h_i=u_2^{a_i}=g^{\beta a_i}\in G$。如果 $c_i=1$，挑战者 C 计算 $h_i=g^{a_i}\in G$。挑战者 C 将此四元组 (w_i,h_i,a_i,c_i) 添加到 H 列表中，同时将 $H(w_i)=h_i$ 作为回复发送给敌手 A。

挑战 敌手 A 选择 2 个挑战关键字 (w_0,w_1) 发送给挑战者 C，挑战者 C 从 H 列表中 (w_i,h_i,a_i,c_i) 可知 $H(w_0)=h_0$，$H(w_1)=h_1$。如果 $c_0=1$ 并且 $c_1=1$，挑战者 C 回复失败并终止此过程。否则，挑战者 C 选择 $b\in\{0,1\}$，使 $c_b=0$，并回复 w_b 的挑战密文 I_b^*。挑战者 C 选取随机数 $\gamma\in Z_p$，挑战者 C 执行加密算法得到密文索引 $I_b^*=\{I_{b,1}^*,I_{b,2}^*\}$，$I_{b,1}^*=e(g^{a_b},g^{\gamma})e(g,H(w_b))^{\gamma}=e(g,g)^{a_b\gamma}e(g,u_2^{a_b})^{\gamma}=e(g,g)^{a_b\gamma}e(g,g)^{a_b\beta\gamma}=e(g,g)^{a_b(1+\beta)\gamma}\in G_T$，$I_{b,2}^*=u_3=g^{\gamma}$，根据定义可知 I_b^* 是合法的密文索引。

阶段 2 敌手 A 重复阶段 1 适应性地查询关键字 w_i 的密文索引，其中，$w_i\neq w_0,w_1$。

猜测 敌手 A 输出索引 I_b^* 中对 b 猜测 $b'\in\{0,1\}$，如果 $b'=b$，则敌手 A 在游戏中获胜，意味着敌手 A 能以可以忽略的优势输出 $e(g,g)^{a_b(1+\beta)\gamma}\in G_T$，否则敌手 A 在游戏中失败。因此，敌手 A 在游戏中没有任何优势获胜。

因此，可以得出，在 BDH 假设成立的前提下，敌手 A 只能以可以忽略的优势赢得选择关键字攻击游戏。

2) 数据拥有者隐私安全。在密文解密阶段，只有获得多数据拥有者认证的数据用户才能恢复秘密值 $e(g,g)^{as}$，从而才能正确解密密文文档。且非法的数据拥有者或合谋的数据拥有者无法伪造合法的授权，解密授权集 $\{A_{y_j}=H(ID)^{u_j}\}$ 类似于签名技术[19]，由于敌手无法从解密授权集 $\{A_{y_j}=H(ID)^{u_j}\}$ 推断出授权私钥集 $\{u_j\}$，因此，敌手无法伪造有效的解密授权集。此外，数据用户与可信机构（TA）直接交互，避免了直接与数据拥有者交互而带来的信息泄露的风险。因此，在 BDH 假设成立的前提下，本文方案是保护数据拥有者隐私安全的。证明过程类似于定理 1，省略详细证明，具体可以参考文献[19]。

3) 数据用户隐私安全。在密文搜索阶段，非法的数据用户由于身份标志 ID 和用户私钥 sk_{ID} 无法伪造有效的陷门去访问敏感数据。此外，数据用户利用散列函数 H 和用户私钥 sk_{ID} 加密查询关键字 w' 生成陷门 $T_{w'}=\{H(w')g^ag^{bu},g^{bu}\}$，云服务器收到陷门 $T_{W'}$ 后无法从中获取查询关键字的明文信息。因此，在 BDH 假设成立的前提下，本文方案是保护数据用户隐私安全的。证明过程类似于定理 1，省略详细证明，具体可以参考文献[5]。

6 性能分析

分别从理论性能和实际性能对比分析文献[16]方案和本文方案的优劣性。

6.1 计算开销分析

表 3 是文献[16]方案和本文方案之间的理论计

算开销比较，主要考虑几种比较耗时的密码运算，包括群 G 中的指数运算 E 、群 G_T 中的指数运算 E_T 、散列运算 H 以及双线性对运算 P 。

表 3　　计算开销分析

算法阶段	文献[16]方案	本文方案
KeyGen	$(2\|S\|+1)E+E_T$	$(l+2)E+H$
Enc	$(\|S\|+1)E+E_T$	$(2l+2)E+3E_T+H+P$
Trap	$(2\|S\|+1)E$	$E+\|R\|H$
Search	$(\|S\|+1)P+E_T$	$(\|R\|+1)P$
Dec	—	$lE_T+lE+3P+2H$

其中，l 表示数据拥有者个数；$|S|$ 表示系统属性个数；$|R|$ 表示查询关键字个数；—表示没有该项操作。

在 KeyGen 阶段，本文方案的密钥生成时间随着数据拥有者个数 l 的增加而增加，文献[16]方案随着系统属性个数 $|S|$ 的增加而增加。在 Enc 阶段，本文方案的加密时间随着数据拥有者个数 l 的增加而增加，文献[16]方案的加密时间随着系统属性个数 $|S|$ 的增加而增加。在 Trap 阶段，本文方案的陷门生成时间随着查询关键字个数 $|R|$ 的增加而增加，文献[16]方案随着数据用户属性个数 $|S|$ 的增加而增加。在 Search 阶段，本文方案的密文搜索时间随着查询关键字个数 $|R|$ 的增加而增加，文献[16]方案随着系统属性个数 $|S|$ 的增加而增加。此外，在 Dec 阶段，假设 η 集合中最多包含 l 个数据拥有者，本文方案的密文解密时间随着数据拥有者个数 l 的增加而增加。

6.2　存储开销分析

表 4 是 ABKS-UR 方案和本文方案之间的理论存储开销比较，其中，$|G_T|$ 表示群 G_T 中元素的长度大小，$|G|$ 表示群 G 中元素的长度大小以及 $|Z_P|$ 表示群 Z_P 中元素的长度大小。

表 4　　存储开销分析

算法阶段	文献[16]方案	本文方案
KeyGen	$(2\|S\|+1)\|G\|+\|Z_P\|$	$(l+2)\|G\|+(l+1)\|Z_P\|$
Enc	$(2\|S\|+1)\|G\|+\|G_T\|$	$(l+2)\|G\|+(l+1)\|Z_P\|+2\|G_T\|$
Trap	$(2\|S\|+1)\|G\|+\|Z_P\|$	$(\|R\|+1)\|G\|$
Search	$(\|S\|+3)\|G_T\|$	$(\|R\|+1)\|G_T\|$
Dec	—	$(2l+2)\|G_T\|+\|G\|$

其中，l 表示数据拥有者个数；$|S|$ 表示系统属性个数；$|R|$ 表示查询关键字个数；—表示没有该项操作。

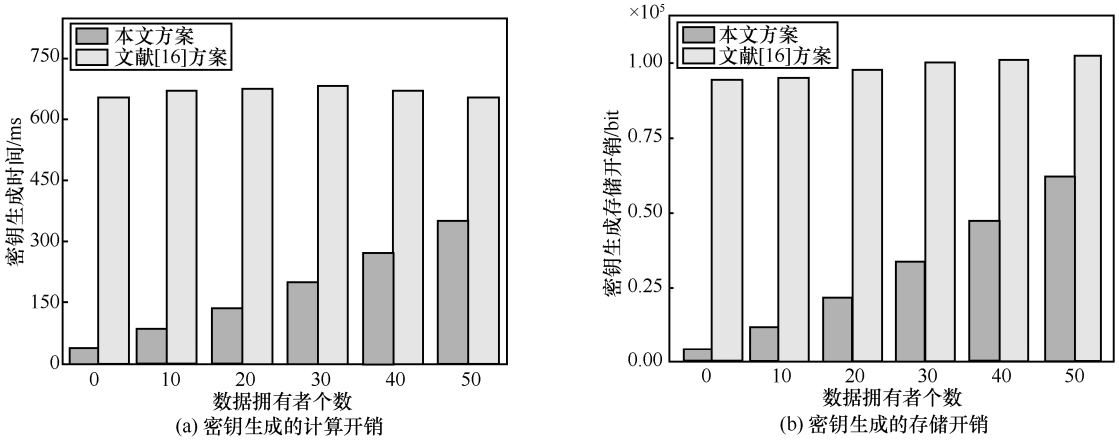
在 KeyGen 阶段，本文方案的密钥生成存储开销随着数据拥有者个数 l 的增加而增加，文献[16]方案随着系统属性个数 $|S|$ 的增加而增加。在 Enc 阶段，本文方案的加密存储开销随着数据拥有者个数 l 的增加而增加，文献[16]方案随着系统属性个数 $|S|$ 的增加而增加。在 Trap 阶段，本文方案的陷门生成存储开销随着查询关键字个数 $|R|$ 的增加而增加，文献[16]方案随着系统属性个数 $|S|$ 的增加而增加。在 Search 阶段，本文方案的密文搜索存储开销随着查询关键字个数 $|R|$ 的增加而增加，文献[16]方案随着系统属性个数 $|S|$ 的增加而增加。此外，在 Dec 阶段，假设 η 集合中最多包含 l 个数据拥有者，本文方案的密文解密存储开销随着数据拥有者个数 l 的增加而增加。

6.3　实验性能分析

为测试本文方案的实际性能，基于实际数据集，做了一系列的仿真实验。实验平台是 CPU 为酷睿 i5，2.3 GHz，内存为 4.0 GB，操作系统为 Ubuntu 15.04 的笔记本电脑。在密码函数库（PBC, pairing-based cryptography）中，A 类椭圆曲线可表示为 $E(F_q): y^2=x^3+x$，阶为 p 的群 G 和群 G_T 是群 $E(F_q)$ 的子群，其中，参数 p 为 160 bit，参数 q 为 512 bit，即 $|G|=|G_T|=1\,024$ bit，$|Z_P|=160$ bit。

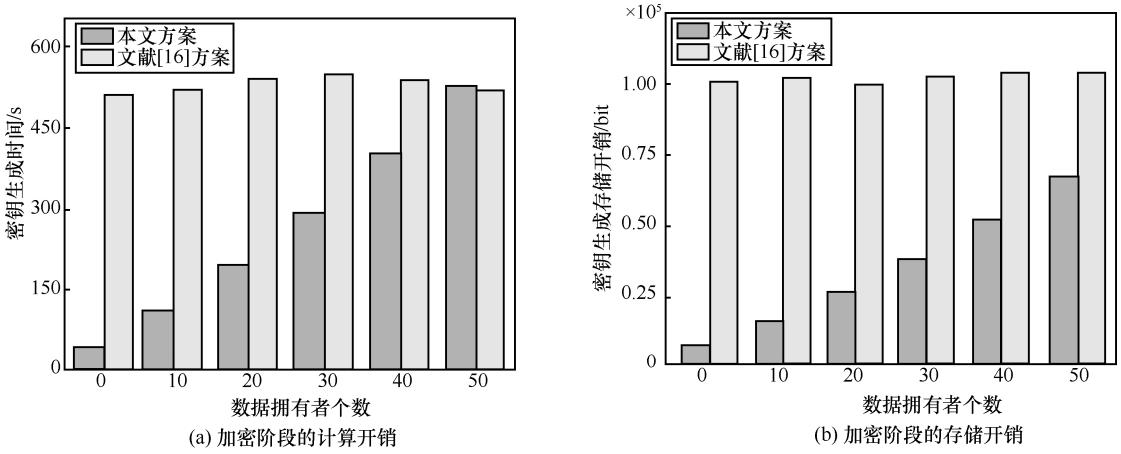
本文方案和文献[16]方案在 KeyGen 阶段的计算开销和存储开销的对比如图 4 所示，数据拥有者个数 $l\in[1,50]$，系统属性个数固定 $|S|=50$。图 4(a)是密钥生成时间开销对比，本文方案的是 $(2|S|+1)E+E_T$。随着数据拥有者个数 l 的增加，本文方案的密钥生成时间会增加，但是总体上本文方案的密钥生成时间少于文献[16]方案。特别地，数据拥有者个数取 $l=50$，本文方案的密钥生成时间为文献[16]方案的 53%。图 4(b)是密钥生成存储开销对比，本文方案的密钥生成存储开销是 $(l+2)|G|+(l+1)|Z_P|$，文献[16]方案的是 $(2|S|+1)|G|+|Z_P|$。随着数据拥有者个数 l 的增加，本文方案的密钥生成存储开销会增加，但是总体上本文方案的密钥生成存储开销少于文献[16]方案。特别地，数据拥有者个数取 $l=50$，本文方案的密钥生成存储开销为文献[16]方案的 60%。

本文方案和文献[16]方案在 Enc 阶段的计算开销和存储开销的对比如图 5 所示，数据拥有者个数 $l\in[1,50]$，系统属性个数固定 $|S|=50$，明文文件数量固定是 10 000。图 5(a)是加密阶段时间开销对比，



(a) 密钥生成的计算开销　　(b) 密钥生成的存储开销

图 4　密钥生成



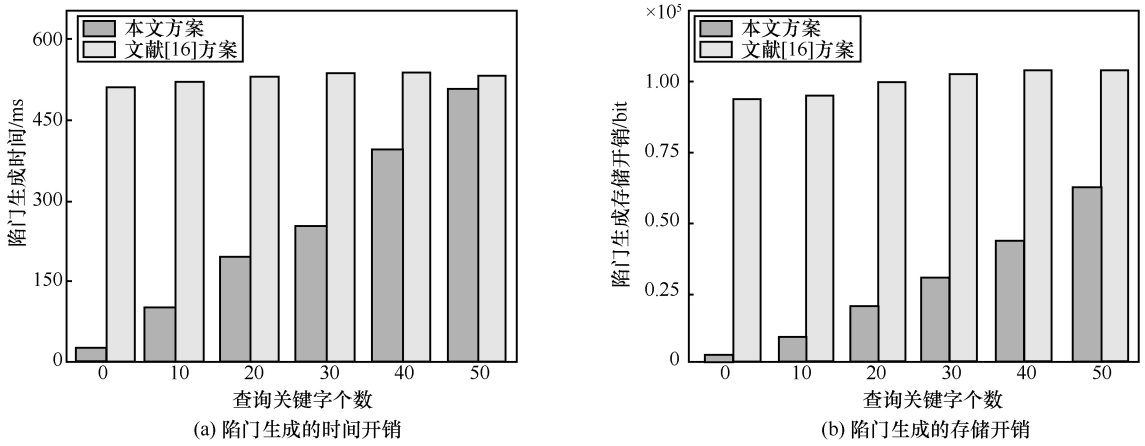
(a) 加密阶段的计算开销　　(b) 加密阶段的存储开销

图 5　加密阶段

本文方案的加密时间是 $(2l+2)E+3E_T+H+P$，文献[16]方案的加密时间是 $(|S|+1)E+E_T$。随着数据拥有者个数 l 的增加，本文方案的加密时间会增加，但是总体上本文方案的加密时间少于文献[16]方案。特别地，数据拥有者个数取 $l=50$，本文方案的加密时间与文献[16]方案近似相等。图 5(b)是加密阶段存储开销对比，本文方案的加密阶段存储开销是 $(l+2)|G|+(l+1)|Z_P|+2|G_T|$，文献[16]方案的是 $(2|S|+1)|G|+|G_T|$。随着数据拥有者个数 l 的增加，本文方案的加密阶段存储开销会增加，但是总体上本文方案的加密阶段存储开销少于文献[16]方案。特别地，数据拥有者个数取 $l=50$，本文方案的密文生成存储开销为文献[16]方案的 64%。

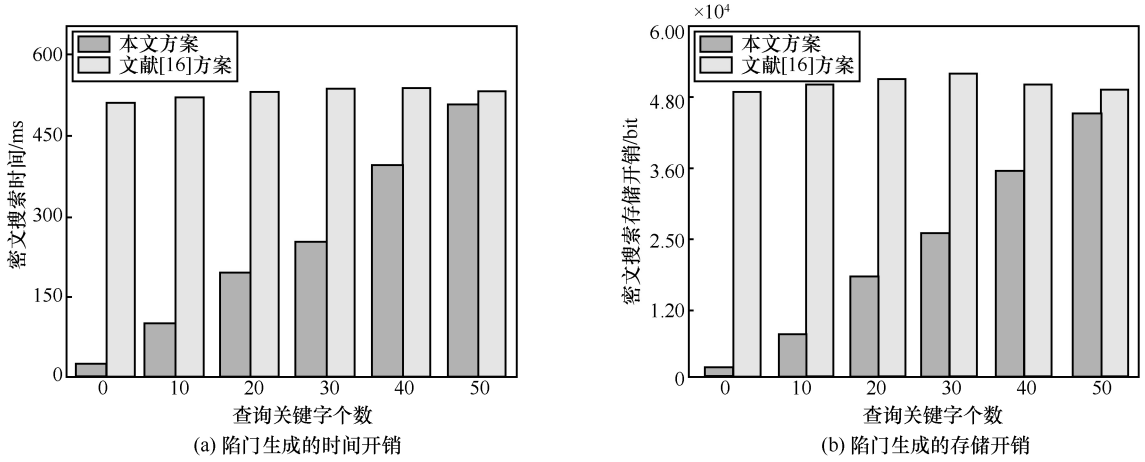
本文方案和文献[16]方案在 Trap 阶段的计算开销和存储开销的对比如图 6 所示，查询关键字个数 $|R|\in[1,50]$，系统属性个数固定 $|S|=50$。图 6(a)是陷门生成时间开销对比，本文方案的陷门生成时间是 $E+|R|H$，文献[16]方案的是 $(2|S|+1)E$。随着查询关键字个数 $|R|$ 的增加，本文方案的陷门生成时间会增加，但是总体上本文方案的陷门生成时间少于文献[16]方案。特别地，查询关键字个数取 $|R|=50$，本文方案的陷门生成时间与文献[16]方案近似相等。图 6(b)是陷门生成存储开销对比，本文方案的陷门生成存储开销是 $(|R|+1)|G|$，文献[16]方案的是 $(2|S|+1)|G|+|Z_P|$。随着查询关键字个数 $|R|$ 的增加，本文方案的陷门生成存储开销会增加，但是总体上本文方案的陷门生成存储开销少于文献[16]方案。特别地，查询关键字个数取 $|R|=50$，本文方案的陷门生成存储开销是文献[16]方案的 52%。

本文方案和文献[16]方案在 Search 阶段的计算开销和存储开销的对比如图 7 所示，查询关键字个数 $|R|\in[1,50]$，系统属性个数固定 $|S|=50$。图 7(a)是密文搜索时间开销对比，本文方案的密文搜索时间是 $(|R|+1)P$，文献[16]方案的是 $(|S|+1)P+E_T$。随着查询关键字个数 $|R|$ 的增加，本文方案的密文搜索时间会增加，但是总体上本文方案的密文搜索



(a) 陷门生成的时间开销　　(b) 陷门生成的存储开销

图 6　陷门生成



(a) 陷门生成的时间开销　　(b) 陷门生成的存储开销

图 7　密文搜索

时间少于文献[16]方案。特别地，查询关键字个数取$|R|=30$，本文方案的密文搜索时间是文献[16]方案的 54%。图 7(b)是密文搜索存储开销对比图，本文方案的密文搜索存储开销是$(|R|+1)|G_T|$，文献[16]方案的是$(|S|+3)|G_T|$。随着查询关键字个数$|R|$的增加，本文方案的密文搜索存储开销会增加，但是总体上本文方案的密文搜索存储开销少于文献[16]方案。特别地，查询关键字个数取$|R|=30$，本文方案的密文搜索存储开销是文献[16]方案的 50%。

本文方案在 Dec 阶段的计算开销和存储开销的对比如图 8 所示，假设η集合中最多包含l个数据拥有者，数据拥有者个数$l\in[1,50]$。本文方案在 Dec 阶段的密文解密时间$lE_T+lE+3P+2H$，随着数据拥有者个数l的增加而增加。本文方案在 Dec 阶段的密文解密存储开销是$(2l+2)|G_T|+|G|$，随着数据拥有者个数l的增加而增加。特别地，数据拥有者个数取$l=50$，本文方案的密文解密时间为 265 ms，本文方案的密钥生成存储开销为 101 520 bit。

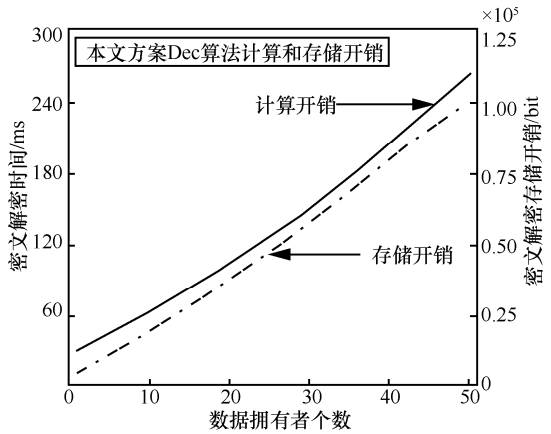


图 8　密文解密

本文方案的计算性能分析和存储开销分析表明，本文方案在实际应用场景中可以高效地实现多数据拥有者认证的密文检索。并且，本文方案在双线性 Diffie-Hellman 假设下是保护数据隐私安全、数据拥有者隐私安全和数据用户隐私安全的。然而，本文方案在降低存储和计算开销的同时不可避

免地引入额外的通信开销，且要求可信机构必须实时在线以验证数据用户合法性。

7 结束语

本文方案在共享型多数据拥有者场景下，提出了支持多数据拥有者认证的密文检索方案。本文方案结合线性秘密共享和可搜索加密技术，只有经多数据拥有者授权的数据用户才能正确解密密文。本文方案是保护数据隐私安全的，未经授权的数据用户无法访问密文数据。并且本文方案在实际应用场景中是高效可行的。未来的工作会改进方案中可信机构实时在线的缺点，设计支持丰富查询功能的密文检索方案。

参考文献：

[1] WEI L, ZHU H, CAO Z, et al. Security and privacy for storage and computation in cloud computing[J]. Information Sciences, 2014, 258: 371-386.

[2] MIAO Y, MA J, LIU Z. Revocable and anonymous searchable encryption in multi-user setting[J]. Concurrency and Computation: Practice and Experience, 2016, 28(4): 1204-1218.

[3] REN K, WANG C, WANG Q. Security challenges for the public cloud[J]. IEEE Internet Computing, 2012, 16(1): 69-73.

[4] LI J, MA R, GUAN H. Tees: an efficient search scheme over encrypted data on mobile cloud[J]. IEEE Transactions on Cloud Computing, 2017, 5(1):126-139.

[5] BONEH D, DI CRESCENZO G, OSTROVSKY R, et al. Public key encryption with keyword search[C]//International Conference on the Theory and Applications of Cryptographic Techniques. 2004: 506-522.

[6] GOLLE P, STADDON J, WATERS B. Secure conjunctive keyword search over encrypted data[C]//Second International Conference on Applied Cryptography and Network Security. 2004:31-45.

[7] YANG Y, MA M. Conjunctive keyword search with designated tester and timing enabled proxy re-encryption function for e-health clouds[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(4): 746-759.

[8] WANG B, YU S, LOU W, et al. Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud[C]//Conference on Computer Communications. 2014: 2112-2120.

[9] XU P, JIN H, WU Q, et al. Public-key encryption with fuzzy keyword search: a provably secure scheme under keyword guessing attack[J]. IEEE Transactions on Computers, 2013, 62(11):2266-2277.

[10] XIA Z, WANG X, SUN X, et al. A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data[J]. IEEE Transactions on Parallel and Distributed Systems, 2016, 27(2): 340-352.

[11] CAO N, WANG C, LI M, et al. Privacy-preserving multi-keyword ranked search over encrypted cloud data[J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(1): 222-233.

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

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

[14] ZHENG Q, XU S, ATENIESE G. VABKS: verifiable attribute-based keyword search over outsourced encrypted data[C]//IEEE Conference on Computer Communications. 2014:522-530.

[15] YIN H, QIN Z, ZHANG J, et al. Secure conjunctive multi-keyword search for multiple data owners in cloud computing [C]//The 22th IEEE International Conference on Parallel and Distributed Systems. 2016: 761-768.

[16] SUN W, YU S, LOU W, et al. Protecting your right: verifiable attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud[J].IEEE Transactions on Parallel and Distributed Systems. 2016, 27(4): 1187-1198.

[17] LAYOUNI M, YOSHIDA M, OKAMURA S. Efficient multi-authorizer accredited symmetrically private information retrieval[C]//The 10th International Conference on Information and Communications Security. 2008: 387-402.

[18] SHAN Y, CAO Z. Extended attribute based encryption for private information retrieval[C]//The 6th IEEE International Conference on Mobile Ad Hoc and Sensor Systems. 2009: 702-707.

[19] DAN B, LYNN B, SHACHAM H. Short signatures from the weil pairing[J]. Journal of Cryptology, 2004, 17(4):297-319.

作者简介：

伍祈应（1994-），女，湖南邵阳人，西安电子科技大学硕士生，主要研究方向为网络与信息安全。

马建峰（1963-），男，陕西西安人，西安电子科技大学教授、博士生导师，主要研究方向为密码学、无线和移动安全等。

苗银宾（1988-），男，河南驻马店人，博士，西安电子科技大学讲师，主要研究方向为应用密码学、无线网络安全等。

张俊伟（1982-），男，陕西西安人，博士，西安电子科技大学副教授，主要研究方向为密码学、网络安全等。

沈丽敏（1978-），女，江苏南京人，西安电子科技大学博士生，主要研究方向为密码学、信息论等。