一个高效的基于连接关键词的可搜索加密方案

王尚平 刘利军* 张亚玲
(西安理工大学密码理论与网络安全研究室 西安 710054)

摘 要：在存储服务中，可搜索加密方案使得用户能够有选择地访问其密文数据，同时还能确保用户搜索数据的机密性。基于连接关键词(即多个关键词的布尔组合)的可搜索加密方案因其更高的搜索精度在安全存储服务中有着重要的应用价值。目前已有的基于连接关键词的可搜索加密方案存在诸如连接关键词的陷门太大、搜索效率不高及不支持多用户等问题。该文采用授权用户和存储服务器先后对关键词加密的方式提出了一个高效的基于连接关键词的可搜索加密方案，使得授权用户能够利用连接关键词的陷门搜索加密文档。在确定性 Diffie-Hellman 问题假设下，证明了方案的安全性。通过与现有方案相比较，提出的方案在通信和计算代价，即搜索陷门大小、关键词加密和搜索的速度等方面的综合效率得到提高。此外，提出的方案支持多用户，即能够动态地增加和撤销用户，使得用户能够直接在存储服务器上进行数据共享。

关键词：密码学；可搜索加密；连接关键词；存储服务；确定性 Diffie-Hellman 问题
中图分类号：TP309 **文献标识码**：A **文章编号**：1009-5896(2013)09-2266-06
DOI: 10.3724/SP.J.1146.2012.01036

An Efficient Conjunctive Keyword Searchable Encryption Scheme

Wang Shang-ping Liu Li-jun Zhang Ya-ling
(*Lab of Cryptography and Network Security*, *Xi'an University of Technology*, *Xi'an* 710054, *China*)

Abstract: In storage service, searchable encryption scheme allows users to access their cipher data selectively, and meanwhile ensures the confidentiality of search data. Since possessing higher search accuracy, conjunctive keyword (namely Boolean combination of multiple keywords) searchable encryption scheme enjoys greater significance in secure storage service application. However, there are some flaws in existing searchable encryption schemes, such as the size of the trapdoor of conjunctive keyword is too large, the search efficiency is slow and there is no support for multiple users search, etc. In this paper, an efficient conjunctive keyword searchable encryption scheme is proposed based on the method that the keywords are encrypted by authorized users and storage server successively，in which authorized users are allowed to search encrypted documents with the trapdoor generated by conjunctive keyword. The scheme is provable secure in the decisional Diffie-Hellman assumption. Compared with the existing schemes, the overall efficiency of the proposed scheme in computation and communication cost, including the size of trapdoor, the speed of keyword encryption and searching, is improved. Moreover, the proposed scheme also supports multiple users, that is, users can be added or revoked dynamically, by this way, and users can share data directly in storage server.

Key words: Cryptography; Searchable encryption; Conjunctive keyword; Storage service; Decisional Diffie-Hellman problem

1 引言

可搜索加密由 Song 等人[1]提出，它能够很好地保护用户外包数据的机密性，同时使得加密数据的高效搜索成为可能，具有很好的扩展性。Goh[2]给出了对称可搜索加密的安全性的形式化定义，并利用 Bloom 过滤器构造了一个方案，但该方案需要线性搜索时间，且搜索结果中存在不符合搜索条件的加密数据。Boneh 等人[3]利用基于身份的加密提出了第 1 个公钥可搜索加密方案。公钥可搜索加密方案允许多个用户利用公钥进行加密，仅有拥有相应私钥的用户才能搜索加密的数据。为了更接近现实世界的搜索情况，Liu 等人[4]利用词典和关键词间的编辑距离(edit distance)提出了一个基于模糊关键词的可搜索加密方案，在关键词拼写错误或格式不一致

2012-08-10 收到，2013-05-11 改回
国家自然科学基金(61173192, 60873268)和陕西省教育厅 2012 年度科学研究计划(12JK0740, 12JK0857)资助课题
*通信作者：刘利军 llj460133921@163.com

的情况下也能搜索出正确的密文。

鉴于一次搜索多个关键词是常见的搜索模式，Golle 等人[5]首次提出了基于连接关键词的可搜索加密方案，其中方案 1(记为 GSW-1)的缺点是关键词的陷门大小与加密文档的数量成线性关系，方案 2(记为 GSW-2)利用双线性映射实现了常量大小的关键词陷门，但是判断一个文档需要计算两次双线性对。Ballard 等人[6]也构造了两个基于连接关键词的可搜索加密方案，其缺点分别与文献[5]中的相同。Byun 等人[7]和 Ryu 等人[8]分别利用双线性对构造了一个基于连接关键词的可搜索加密方案(分别记为 BLL 和 RT)，方案的特点都是关键词陷门大小固定，但是判断每个文档都需要计算两次双线性对。Kerschbaum[9]首次提出了在非结构化文本上的基于连接关键词的可搜索加密方案，即在搜索加密文档时无需指定关键词的位置(记为 FK)。此外，Cao 等人[10]和 Chuah 等人[11]提出了其它具有特殊功能的基于多关键词的可搜索加密方案，扩展了多关键词可搜索加密的应用。

相对于一次只能搜索一个关键词，基于连接关键词的可搜索加密方案能够产生更精确的搜索结果，如在邮件外包服务中，相对于搜索所有来自于“Bob”的邮件，用户可能仅仅想要那些被标记为“urgent”的来自于“Bob”的邮件，因而存储服务器需要对关键词“Bob”和“urgent”的连接进行搜索。但是这些已经存在的基于连接关键词的可搜索加密方案或多或少的存在如下几个问题：(1)连接关键词的陷门大小与加密文档的数量成线性关系；(2)存储服务器搜索的效率过低；(3)方案不适合多用户环境，即无法增加和撤销用户。

考虑到外包数据可能由多个用户操作和访问，Curtmola 等人[12]利用广播加密解决了可搜索加密中的多用户问题。但是，方案中的加密数据是“只读”的，并且由于加密密钥共享从而导致“不完全的”用户撤销问题。Bao 等人[13]利用双线性映射提出了一个多用户环境下的可搜索加密方案，解决了文献[12]中的“只读”问题。Yang 等人[14]又在文献[13]的基础上解决了“不完全的”用户撤销问题。Dong 等人[15]利用 El Gamal 代理加密技术提出了一个具有更新数据能力的多用户环境下的可搜索加密方案，不仅解决了“不完全的”用户撤销问题，而且相对于文献[13,14]，方案具有更高的计算效率，但需要更多的存储服务器的存储空间。

针对连接关键词可搜索加密方案中的问题，结合现有的多用户环境下的可搜索加密方案，本文采用授权用户和存储服务器先后对关键词加密的方式在对称密钥环境下提出了一个高效的基于连接关键词的可搜索加密方案，使得用户能够利用连接关键词的陷门搜索加密文档。在确定性 Diffie-Hellman 问题假设下，证明了方案的安全性。通过对比分析，本文提出的方案具有以下优势：(1)连接关键词的陷门大小固定；(2)存储服务器的搜索效率高；(3)方案适用于多用户环境。

2　预备知识

2.1　系统模型

系统参与者包括 $\{D,\mathrm{UM},\mathrm{Serv},U\}$，其中 D 为用户要外包存储的数据集合；UM 是授权用户的管理机构，负责管理用户，如用户的增加与撤销；Serv 是外包存储服务器，负责存储与搜索服务；U 是授权用户的身份集合，其中用户的身份唯一，如用户的邮箱地址等。

假设用户有 n 个文档 $D=(D_1,\cdots,D_n)$ 需要以加密的形式存储在不完全可信的 Serv 上。为了简化方案的描述，假设每个文档都有 m 个关键词字段，如邮件可以定义 4 个关键词字段：“From”，“To”，“Subject”，“Date”。另外，假设：

(1)每个文档中都不包含两个相同的关键词，这可以通过在关键词前加上关键词所属字段来满足。如关键词“From: Bob”属于“From”字段，不会与属于“To”字段的关键词“To: Bob”相混淆。

(2)若某个关键词字段没有内容，则将该关键词字段的内容设为空，如在邮件中，对于那些“Subject”关键词字段没有内容的邮件，可以定义关键词为“Subject: NULL”。

(3) UM 是完全可信的，并且所有与 UM 的会话都是安全的。

(4) U 中的用户不会与 Serv 发起合谋攻击。

记文档 $D_i,1\le i\le n$ 的关键词列表为 $W_i=(w_{i,1},\cdots,w_{i,m})$，其中 $w_{i,j}$ 为 D_i 的第 j 个关键词字段的关键词。I_i 表示 W_i 加密后生成的 D_i 的索引。对 D_i 的加密采用标准的对称加密算法，如 AES，记密钥 k 下的加解密算法分别为 $\mathrm{Enc}_k(\cdot)$ 和 $\mathrm{Dec}_k(\cdot)$。$\mathrm{negl}(\cdot)$ 表示可忽略的函数，即对任意的多项式 $p(\cdot)$，存在 N，使得对任意的整数 $n>N$，$\mathrm{negl}(n)<(1/p(n))$ 成立。

基于连接关键词的可搜索加密方案的系统模型如下：

一个基于连接关键词的可搜索加密方案由以下几个多项式时间算法组成：

(1) $\mathrm{Init}(1^k)$：该算法由 UM 执行以初始化系统，输入安全参数 k，输出系统参数 params，主密钥 msk 和语义安全的对称加密算法 $\mathrm{Enc}(\cdot)$ 的加密密钥

ek，两个随机种子 s' 和 s''。

(2) Enroll($\mathrm{msk}, u_{\mathrm{ID}}$)：该算法由 UM 执行以添加用户，输入 msk 和用户身份 $u_{\mathrm{ID}} \in U$，输出用户 u_{ID} 的密钥和辅助密钥 $(\mathrm{sk}_{u_{\mathrm{ID}}}, \mathrm{ComK}_{u_{\mathrm{ID}}})$，将 $(\mathrm{sk}_{u_{\mathrm{ID}}}, \mathrm{ek}, s', s'')$ 安全地发送给用户 u_{ID}，$(u_{\mathrm{ID}}, \mathrm{Com}K_{u_{\mathrm{ID}}})$ 安全地发送给 Serv。

(3) U-Enc($\mathrm{sk}_{u_{\mathrm{ID}}}, \mathrm{ek}, s', D_i, W_i$)：用户 u_{ID} 执行的加密算法，输入用户密钥 $\mathrm{sk}_{u_{\mathrm{ID}}}$、加密密钥 ek、随机种子 s'、文档 D_i 及其关键词列表 W_i，，输出密文 $C_i^* = (\mathrm{Enc}_{\mathrm{ek}}(D_i), I_i^*)$，将 (u_{ID}, C_i^*) 发送给 Serv。

(4) S-Enc(u_{ID}, C_i^*)：Serv 执行以对 C_i^* 中的 I_i^* 进行重加密，输入用户身份 u_{ID} 和接收到的 C_i^*，Serv 根据 u_{ID} 查找 $(u_{\mathrm{ID}}, \mathrm{ComK}_{u_{\mathrm{ID}}})$，若无，则返回 $\perp$；否则重加密 I_i^* 得到索引 I_i，最后将 $C_i = (\mathrm{Enc}_{\mathrm{ek}}(D_i), I_i)$ 存储在 Serv 上。

(5) Trapdoor($\mathrm{sk}_{u_{\mathrm{ID}}}, s', s'', l_1, \cdots, l_d, w_1', \cdots, w_d'$)：用户 u_{ID} 执行以生成连接关键词的陷门，输入 $\mathrm{sk}_{u_{\mathrm{ID}}}$，$s'$，$s''$ 和要检索的关键词位置 $1 \le l_1, \cdots, l_d \le m$ 及相应的关键词 $w_1', \cdots, w_d'$，输出陷门 T。

(6) Search(T, C_i)：Serv 执行用于搜索加密文档，输入陷门 T 及密文 $C_i, 1 \le i \le n$，输出搜索到的密文集合 Ω。最后将 Ω 发送给用户 u_{ID}。

(7) Dec(ek, Ω)：用户 u_{ID} 执行以解密密文，输入对称密钥 ek 及接收到的 Ω，解密得明文。

(8) RevokeUser(u_{ID})：UM 执行以撤销用户，输入用户身份 u_{ID}，UM 向 Serv 发送撤销用户 u_{ID} 的命令，Serv 删除 $(u_{\mathrm{ID}}, \mathrm{ComK}_{u_{\mathrm{ID}}})$。

2.2 安全性定义

直观上，系统的安全性定义应该被概要为：给定访问索引的权限和关键词陷门，存储服务器无法知道对应关键词列表的任何信息，在连接关键词的背景下，这意味着存储服务器无法利用已知的连接关键词 $(w_1, \cdots, w_d)$ 的陷门构造出一个新的连接关键词 $(w_1', \cdots, w_\delta')$ 的陷门，即使 $\{w_1', \cdots, w_\delta'\} \subset \{w_1, \cdots, w_d\}$[6]。此外，即使存储服务器能够欺骗用户为其选择的任何关键词生成陷门，该安全性定义也必须成立，这种安全性定义被称为在适应性选择关键词攻击下的不可区分性(IND-CKA)[2]。但是，本文通过游戏只给出了轻量级的 IND-CKA 的定义，如下：

安全性游戏 IND-CKA(INDistinguishability against adaptive Chosen-Keyword Attack)对于任何概率多项式时间攻击者 A (存储服务器)和挑战者 C (授权用户 u_{ID})：

(1) Setup：运行系统初始化函数 Init(1^k)，对 $\forall u_{\mathrm{ID}} \in U$，执行用户添加函数 Enroll($\cdot, u_{\mathrm{ID}}$)，将 $\mathrm{sk}_{u_{\mathrm{ID}}}$ 发送给挑战者 C，$\mathrm{ComK}_{u_{\mathrm{ID}}}$ 发送给攻击者 A。

(2) Query_1：A 适应性选择多项式多个关键词列表 W_i，并向 C 请求 W_i 的加密索引 I_i。

(3) Challenge：A 选择两个未询问过加密索引的关键词 $\widetilde{W}_0, \widetilde{W}_1$ 并发送给 C。C 随机选择 $b \in_R \{0,1\}$，返回 W_b 的用户加密索引 $\tilde{I}_b$ 给攻击者 A。

(4) Query_2：A 再次向 C 适应性选择询问关键词列表的加密索引，除了不能询问关于关键词列表 W_0, W_1 的加密索引，整个询问的次数为安全参数 k 的多项式。

(5) Response：A 输出关于加密索引 $\tilde{I}_b$ 中 b 的判断 $b_A \in \{0,1\}$。若 $b_A = b$，则攻击者 A 成功。攻击者 A 的优势定义为 $\mathrm{Adv}_A(1^k) = \left|\Pr[b_A = b] - 1/2\right|$。

在上述的安全性游戏中没有考虑到对文档 $D_i (1 \le i \le n)$ 的加密的安全性，因为 D_i 的加密采用的是标准的语义安全的对称加密算法(如 AES)，这种加密算法能确保 D_i 的加密的安全性。此外，在安全性游戏中没有考虑到陷门的安全性，即攻击者无法通过陷门获得任何有关关键词的信息及无法通过已有的陷门构造出新的有效的陷门，这一点我们在安全性分析过程中给出了详细分析。

定义 1 如果对于任意的多项式时间攻击者 A，$\mathrm{Adv}_A(1^k)$ 是安全参数 k 的一个可忽略函数，则称基于连接关键词的可搜索加密方案是 IND-CKA 安全的。

2.3 方案的困难性假设

本文方案的安全性证明基于确定性 Diffie-Hellman 问题(DDHP)，其详细定义如下：

定义 2 确定性 Diffie-Hellman 问题(DDHP) 令 G 是一个阶为素数 q 的群，g 是 G 的生成元。给定两个三元组 (g^a, g^b, g^{ab}) 和 (g^a, g^b, g^c)，其中 $a, b, c \in_R Z_q^*$，DDHP 是判断 $c \stackrel{?}{=} ab$。

定义 3 (DDH 假设) 群 G 中 DDHP 是困难的，如果对于任意的概率多项式时间攻击者 A，使得

$$\left|\Pr[A(G, q, g, g^a, g^b, g^{ab}) = 1] - \Pr[A(G, q, g, g^a, g^b, g^c) = 1]\right| < \mathrm{negl}(k)$$

3 一个高效的基于连接关键词的可搜索加密方案

本节将详细描述构造的方案及方案的安全性证明和效率分析。

3.1 一个高效的基于连接关键词的可搜索加密方案

假设用户将文档集合 $D = (D_1, \cdots, D_n)$ 外包到存储服务器 Serv，文档 D_i 的关键词列表为 $W_i = (w_{i,1}, \cdots, w_{i,m})$，$1 \le i \le n$，其中 $w_{i,j}$ 为 D_i 的第 j 个关键词

字段的关键词，$1 \le j \le m$，构造的方案包括 8 个多项式时间算法，详细描述如下：

(1) $\mathrm{Init}(1^k)$：该算法由用户管理机构 UM 执行以初始化系统，输入安全参数 k，输出阶为素数 q 的循环群 G，g 为 G 的生成元，并且 G 中的 DDHP 是困难的。随机选择 $x \in_R Z_q^*$ 作为 UM 的主密钥，记为 $k_{\mathrm{UM}} = x$，计算 $h = g^x$；UM 选择两个伪随机函数 $f' : \{0,1\}^k \times \{0,1\}^* \to Z_q^*$ 和 $f'' : \{0,1\}^k \times Z_q^* \to Z_q^*$ 及其随机种子分别为 $s', s'' \in_R \{0,1\}^k$，并为语义安全的对称加密算法 $\mathrm{Enc}(\cdot)$ 选择加密密钥 ek，发布 $\mathrm{params} = (G, g, q, f', f'', h, \mathrm{Enc})$ 作为系统参数。

(2) $\mathrm{Enroll}(k_{\mathrm{UM}}, u_{\mathrm{ID}})$：该算法由用户管理机构 UM 执行以添加用户，输入 UM 的主密钥 k_{UM} 和用户身份 $u_{\mathrm{ID}} \in U$ (用户身份是唯一的，如用户的电子邮件地址)，输出 u_{ID} 的密钥和辅助密钥 $(\mathrm{sk}_{u_{\mathrm{ID}}}, \mathrm{ComK}_{u_{\mathrm{ID}}}) = (x_{u_{\mathrm{ID}}} \in_R Z_q^*, k_{\mathrm{UM}}/x_{u_{\mathrm{ID}}}) = (x_{u_{\mathrm{ID}}}, x/x_{u_{\mathrm{ID}}})$。将 $(\mathrm{sk}_{u_{\mathrm{ID}}}, \mathrm{ek}, s', s'')$ 安全地发送给用户 u_{ID}，$(u_{\mathrm{ID}}, \mathrm{ComK}_{u_{\mathrm{ID}}})$ 安全地发送给 Serv，Serv 在其用户列表 U-ComK 中加入 $(u_{\mathrm{ID}}, \mathrm{ComK}_{u_{\mathrm{ID}}})$。

(3) $U\text{-Enc}(\mathrm{sk}_{u_{\mathrm{ID}}}, \mathrm{ek}, s', D_i, W_i)$：用户 u_{ID} 执行的加密算法，输入用户密钥 $\mathrm{sk}_{u_{\mathrm{ID}}}$、加密密钥 ek、随机种子 s'、文档 D_i 及其关键词列表 $W_i = (w_{i,1}, \cdots, w_{i,m})$，$1 \le i \le n$，随机选择 $r_i \in_R Z_q$，计算 g^{r_i} 和 h^{r_i}，对 $\forall w_{i,j} \in W_i$，计算 $\sigma_{i,j} = f'(s', w_{i,j})$，$w_{i,j}^* = (g^{\mathrm{sk}_{u_{\mathrm{ID}}}})^{r_i \sigma_{i,j}}$，$1 \le j \le m$，令 $I_i^* = (g^{r_i}, h^{r_i}, w_{i,1}^*, \cdots, w_{i,m}^*)$，记 $C_i^* = (\mathrm{Enc}_{\mathrm{ek}}(D_i), I_i^*)$，将 (u_{ID}, C_i^*) 发送给 Serv。

(4) $S\text{-Enc}(u_{\mathrm{ID}}, C_i^*)$：Serv 执行对 C_i^* 中的 I_i^* 的重加密，输入用户身份 u_{ID} 和接收到的 C_i^*，Serv 根据 u_{ID} 在 U-ComK 中查找 $(u_{\mathrm{ID}}, \mathrm{ComK}_{u_{\mathrm{ID}}})$，若无，则返回 $\perp$；否则重新计算 C_i^* 中的 I_i^* 得索引 $I_i = (g^{r_i}, h^{r_i}, (g^{\mathrm{sk}_{u_{\mathrm{ID}}}})^{r_i \sigma_{i,1} \cdot \mathrm{ComK}_{u_{\mathrm{ID}}}}, \cdots, (g^{\mathrm{sk}_{u_{\mathrm{ID}}}})^{r_i \sigma_{i,m} \cdot \mathrm{ComK}_{u_{\mathrm{ID}}}}) = (g^{r_i}, h^{r_i}, h^{r_i \sigma_{i,1}}, \cdots, h^{r_i \sigma_{i,m}})$，将 $C_i = (\mathrm{Enc}_{\mathrm{ek}}(D_i), I_i)$ 存储在 Serv 上。

(5) $\mathrm{Trapdoor}(\mathrm{sk}_{u_{\mathrm{ID}}}, s', s'', l_1, \cdots, l_d, w_1', \cdots, w_d')$：用户 u_{ID} 执行以生成连接关键词的陷门，输入 $\mathrm{sk}_{u_{\mathrm{ID}}}$，$s'$，$s''$ 和要检索的关键词位置 $1 \le l_1, \cdots, l_d \le m$ 及对应的关键词 $w_1', \cdots, w_d'$，随机选择 $t_1, t_2 \in_R Z_q^*$，计算：$T_1 = \left(t_1 + f''(s'', t_2) \sum_{j=1}^d f'(s', w_j')\right) \mathrm{sk}_{u_{\mathrm{ID}}} = \left(t_1 + f''(s'', t_2) \cdot \sum_{j=1}^d f'(s', w_j')\right) x_{u_{\mathrm{ID}}}$，$T_2 = t_1$，$T_3 = f''(s'', t_2)$ 将陷门 $T = (u_{\mathrm{ID}}, T_1, T_2, T_3, l_1, \cdots, l_d)$ 发送给 Serv。

(6) $\mathrm{Search}(T, C_i)$：Serv 执行用于搜索加密文档，输入陷门 $T = (u_{\mathrm{ID}}, T_1, T_2, T_3, l_1, \cdots, l_d)$ 及密文 $C_i = (\mathrm{Enc}_{\mathrm{ek}}(D_i), I_i)$，Serv 首先根据 u_{ID} 在 U-ComK 中查找 $(u_{\mathrm{ID}}, \mathrm{ComK}_{u_{\mathrm{ID}}})$，若无，则返回 $\perp$；否则 Serv 初始化空集 Ω，计算 $v = T_1 \cdot \mathrm{ComK}_{u_{\mathrm{ID}}} = \left(t_1 + f''(s'', t_2) \cdot \sum_{j=1}^d f'(s', w_j')\right) x$，对 $C_i, 1 \le i \le n$，判断如下等式是否成立

$$(g^{r_i})^v / (h^{r_i})^{T_2} = h^{r_i \left(\sum_{j=1}^d f'(s', w_j')\right) f''(s'', t_2)} \stackrel{?}{=} \left(\prod_{j=1}^d h^{r_i \sigma_{i,l_j}}\right)^{T_3}$$

若成立，则 $\Omega = \Omega \cup \{C_i\}$。最后将搜索结果 Ω 发送给用户 u_{ID}。

(7) $\mathrm{Dec}(\mathrm{ek}, \Omega)$：用户 u_{ID} 执行以解密密文，输入对称密钥 ek 及接收到的 Ω，对 $\forall C_i \in \Omega$，计算 $D_i = \mathrm{Dec}_{\mathrm{ek}}(\mathrm{Enc}_{\mathrm{ek}}(D_i))$。

(8) $\mathrm{RevokeUser}(u_{\mathrm{ID}})$：UM 执行以撤销用户，输入用户身份 u_{ID}，UM 向 Serv 发送撤销用户 u_{ID} 的命令，Serv 执行操作 $U\text{-ComK} = U\text{-ComK} \setminus \{u_{\mathrm{ID}}\}$。

3.2 方案的安全性分析

定理 1　上述方案满足完备性。

证明　若所有数据都是按照方案中描述生成的，并且如果 $f'(s', w_j') = \sigma_{i,l_j}$，其中 $1 \le i \le m$，$1 \le j \le d$，则

$$\begin{aligned}(g^{r_i})^v / (h^{r_i})^{T_2} &= (g^{r_i})^{\left(t_1 + f''(s'', t_2) \sum_{j=1}^d f'(s', w_j')\right) x} / (h^{r_i})^{t_1} \\ &= h^{r_i \left(\sum_{j=1}^d f'(s', w_j')\right) f''(s'', t_2)} \\ &= h^{r_i \left(\sum_{j=1}^d \sigma_{i,l_j}\right) f''(s'', t_2)} = \left(\prod_{j=1}^d h^{r_i \sigma_{i,l_j}}\right)^{T_3}\end{aligned}$$

证毕

定理2　如果群 G 中的DDH假设成立，则上述方案是IND-CKA安全的。

证明　若攻击者 A (存储服务器)以不可忽略的优势 ε 赢得游戏IND-CKA(2.2节定义)，则我们构造一个算法 A' 利用攻击者 A 能以不可忽略的优势 ε 解决DDHP。给定 $(g_1 = g^a, g_2 = g^b, g_3 = g^c)$ 为DDHP挑战，A' 选择两个伪随机函数 f', f''，并令 $h = g_1 = g^a$，将 (G, g, q, f', f'', h) 作为公共参数 params。对 $\forall u_{\mathrm{ID}} \in U$，$A'$ 随机选择 $\mathrm{sk}_{u_{\mathrm{ID}}} = x_{u_{\mathrm{ID}}} \in_R Z_q$。

Query_1：当 A 向 A' 询问关键词列表 $W_i = (w_{i,1}, \cdots, w_{i,m})$，$i = 1, \cdots, n$ 的加密索引时，A' 按如下方式对 W_i 进行加密。A' 随机选择 $r_i \in_R Z_q$，计算 g^{r_i} 和 h^{r_i}，对 $W_i = (w_{i,1}, \cdots, w_{i,m})$ 中每个 $w_{i,j}$，计算 $\sigma_{i,j} = f'(s', w_{i,j})$，输出正确的加密索引 $I_i = (g^{r_i}, h^{r_i}, h^{r_i \sigma_{i,1}}, \cdots, h^{r_i \sigma_{i,m}})$。

Challenge：A 选择两个未询问过的关键词 $\widetilde{W}_0, \widetilde{W}_1$ 发送给 A'，其中 $\widetilde{W}_i = (\widetilde{w}_{i,1}, \cdots, \widetilde{w}_{i,m})$，$i \in \{0,1\}$。

A' 随机选择 $\xi \in_R \{0,1\}$，对 $\widetilde{W}_\xi = (\widetilde{w}_{\xi,1}, \cdots, \widetilde{w}_{\xi,m})$ 中每个 $\widetilde{w}_{\xi,j}$，计算 $\tilde{\sigma}_{\xi,j} = f'(s', \widetilde{w}_{\xi,j})$，随机选择 $r_i' \in_R Z_q$，输出 $\widetilde{W}_\xi = (\widetilde{w}_{\xi,1}, \cdots, \widetilde{w}_{\xi,m})$ 的加密索引 $\tilde{I}_\xi =$

$(g_2^{r_i'}, g_3^{r_i'}, g_3^{r_r'\tilde{\sigma}_{\xi,1}}, \cdots, g_3^{r_r'\tilde{\sigma}_{\xi,m}})$。

Query_2：A 可以再次适应性向 A' 询问关键词的加密索引，除了不能询问关于关键词 $\widetilde{W}_0, \widetilde{W}_1$ 的加密索引，整个询问的次数为安全参数的多项式。

Response：A 输出关于加密索引 $\tilde{I}_\xi = (g_2^{r_i'}, g_3^{r_i'}, g_3^{r_r'\tilde{\sigma}_{\xi,1}}, \cdots, g_3^{r_r'\tilde{\sigma}_{\xi,m}})$ 中 ξ 的判断 $\xi_A \in \{0,1\}$。若 $\xi_A = \xi$，则攻击者 A 成功。若攻击者 A 成功，则 A' 回答 DDHP 挑战 $(g_1 = g^a, g_2 = g^b, g_3 = g^c)$ 中 $c = ab$；若攻击者 A 失败，则 A' 回答 $c \neq ab$。

事实上，若 $c = ab$，则 $\widetilde{W}_\xi = (\widetilde{w}_{\xi,1}, \cdots, \widetilde{w}_{\xi,m})$ 的加密索引

$$\begin{aligned}\tilde{I}_\xi &= (g_2^{r_i'}, g_3^{r_i'}, g_3^{r_r'\tilde{\sigma}_{\xi,1}}, \cdots, g_3^{r_r'\tilde{\sigma}_{\xi,m}}) \\ &= (g^{br_i'}, h^{br_i'}, h^{br_i'\tilde{\sigma}_{\xi,1}}, \cdots, h^{br_i'\tilde{\sigma}_{\xi,m}})\end{aligned}$$

是一个正确的加密索引，注意这里 $h = g_1 = g^a$，$g_2 = g^b, g_3 = g^c = h^b$。反之，若 $c \neq ab$，则 $\tilde{I}_\xi = (g_2^{r_i'}, g_3^{r_i'}, g_3^{r_r'\tilde{\sigma}_{\xi,1}}, \cdots, g_3^{r_r'\tilde{\sigma}_{\xi,m}})$ 不是一个正确的密文索引。因此，假如攻击者 A 以不可忽略的优势 ε 赢得成功，则 $\tilde{I}_\xi = (g_2^{r_i'}, g_3^{r_i'}, g_3^{r_r'\tilde{\sigma}_{\xi,1}}, \cdots, g_3^{r_r'\tilde{\sigma}_{\xi,m}})$ 一定是一个正确的加密索引，所以必然 $c = ab$。若 $c \neq ab$，攻击者 A 在上述的游戏中没有任何的优势。

综上所述，假如攻击者 A 以不可忽略的优势 ε 赢得成功，则 A' 可以以不可忽略的优势解决DDHP挑战。因此，假如群 G 中的DDH假设成立，本方案是满足IND-CKA安全的。 证毕

此外在陷门的安全性方面，我们注意到，关键词是经过伪随机函数 f' 在密钥 s' 下处理的，由于攻击者不知道密钥 s'，从而不会泄露任何有关关键词的明文信息；由于攻击者不知道用户密钥 $\mathrm{sk}_{u_{\mathrm{ID}}}$，因此攻击者无法推断出两个陷门是否针对同一个连接关键词(搜索模式[10])；又由于攻击者不知道 s''，并且对任意两个随机数 $t_1, t_2 \in_R Z_q^*$，$f''(s'', t_1 + t_2) \neq f''(s'', t_1) + f''(s'', t_2)$，故攻击者无法从已知的有效陷门中构造出新的连接关键词的有效陷门。因此，方案中的陷门是安全的。

我们的方案是适合多用户环境的，这里的多用户是指可以增加和撤销用户，用户增加可以通过算法 $\mathrm{Enroll}(k_{\mathrm{UM}}, u_{\mathrm{ID}})$ 实现，用户的撤销可以通过算法 $\mathrm{RevokeUser}(u_{\mathrm{ID}})$ 实现。当然这里需要 UM 和 Serv 之间有安全信息通道，如果需要，可以通过其他身份认证协议实现。

3.3 方案的效率分析

本节将构造的方案与已提出的基于连接关键词的可搜索加密方案在计算和通信代价方面做了详细比较，比较结果如表 1 所示。记“exp”为指数运算，“pr”为双线性运算，n 为外包存储的文档数量，m 为每个文档的关键词列表中的关键词的数量。对于每个方案，给出了每次搜索的陷门大小、每个文档的关键词列表的加密代价、每次搜索的陷门计算代价、搜索每个文档的计算代价及是否支持多用户。

通过比较可以发现，本文的方案的陷门生成效率高，搜索效率与 GSW-1 相近，但比他它 4 个方案的搜索效率都要高；FK 中的陷门最小，GSW-2, BLL, RT 和本文方案的陷门大小间仅相差 $|q|$，要远远低于 GSW-1 的陷门大小(因为 GSW-1 中的陷门大小与加密文档的数量成线性关系)；在本文的方案中，客户端在关键词加密的过程中实际上只需要 $(m+2)$ 次 exp 计算，其余的计算都由存储服务器执行，而其他方案中的关键词加密的计算则都是在客户端完成的，故本文的方案的关键词加密效率实际上是与 GSW-1 相近，而比其他方案的效率要高。此外，FK 中的连接关键词的陷门中虽不用指定关键词的位置，但每次搜索都是针对每个文档中的所有关键词的，而不是有选择的进行搜索。因此，本文的方案在通信和计算代价方面拥有较好的性能。此外，在本文的方案中，用户管理机构 UM 可以添加和撤销用户，只有授权的用户能够搜索和解密加密文档，因而本文的方案是适用于多用户的，使得用户能够直接在存储服务器上进行数据共享。

4 结论

本文提出了一个新的基于连接关键词的可搜索加密方案，使得授权用户能够利用连接关键词的陷门去搜索加密的文档，并且在 DDH 假设下证明了

表 1 方案的效率比较

方案	|trapdoor|	encryption	trapdoor	search	multi-users
GSW-1[5]	$(n+1)\|q\| + \log m$	$(m+1)$ exp	n exp	1 exp	No
GSW-2[5]	$3\|q\| + \log m$	$(2m+1)$ exp	3 exp	3pr	No
BLL[7]	$2\|q\| + \log m$	2 exp+ $m$pr	2 exp	2pr	No
RT[8]	$2\|q\| + \log m$	$(m+1)$ exp+ 1pr	2 exp	2pr	No
FK[9]	$3\|q\|$	$3m$ exp	$(m+2)$ exp+ 1pr	2pr	No
本文方案	$3\|q\| + \log m$	$(2m+2)$ exp	0	3 exp	Yes

方案是 IND-CKA 安全的并分析了方案中陷门的安全性。通过对比分析证明了本文方案在通信和计算代价方面具有较好的性能。然而，在方案中我们假设用户和存储服务器不会发起合谋攻击，对于这种攻击的一般解决办法是增加一个可信的管理机构，对于每个用户都将系统主密钥分成 3 份分别发给用户、存储服务器和新增的管理机构，详细的介绍请阅读文献[15]。

参 考 文 献

[1] Song D, Wagner D, and Perrig A. Practical techniques for searches on encrypted data[C]. Proceedings of IEEE Symposium on Security and Privacy, Berkeley, 2000: 44–55.

[2] Goh E. J. Secure indexes[OL]. http://eprint.iacr.org/2003/216/, 2003.

[3] Boneh D, Crescenzo G, Ostrovsky R, *et al*.. Public key encryption with keyword search[C]. Proceedings of International Conference on the Theory and Applications of Cryptographic Techniques, Interlaken, 2004: 506–522.

[4] Liu Chang, Zhu Lie-huang, Li Long, *et al*.. Fuzzy keyword search on encrypted cloud storage data with small index[C]. Proceedings of International Conference on Cloud Computing and Intelligence Systems, Beijing, 2011: 269–273.

[5] Golle P, Staddon J, and Waters B. Secure conjunctive keyword search over encrypted data[C]. Proceedings of the 2nd International Conference on Applied Cryptography and Network Security, Huangshan, 2004: 31–45.

[6] Ballard L, Kamara S, and Monrose F. Achieving efficient conjunctive keyword searches over encrypted data[C]. Proceedings of the 7th International Conference on Information and Communications Security, Beijing, 2005: 414–426.

[7] Byun J W, Lee D H, and Lim J. Efficient conjunctive keyword search on encrypted data storage system[C]. Proceedings of EuroPKI, Turin, 2006: 184–196.

[8] Ryu E K and Takagi T. Efficient conjunctive keyword-searchable encryption[C]. Proceedings of 21st International Conference on Advanced Information Networking and Application Workshops, Niagara, 2007: 409–414.

[9] Kerschbaum F. Secure conjunctive keyword searches for unstructured text[C]. Proceedings of 5th International Conference on Network and System Security, Milan, 2011: 285–289.

[10] Cao Ning, Wang Cong, Li Ming, *et al*.. Privacy-preserving multi-keyword ranked search over encrypted cloud data[C]. Proceedings of IEEE INFOCOM, Shanghai, 2011: 829–837.

[11] Chuah M and Hu W. Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data[C]. Proceedings of the 31st International Conference on Distributed Computing Systems Workshops, Minneapolis, 2011: 273–281.

[12] Curtmola R, Garay J, Kamara S, *et al*.. Searchable symmetric encryption: improved definitions and efficient constructions[J]. *Journal of Computer Security*, 2011, 19(5): 895–934.

[13] Bao Feng, Deng R H, Ding Xu-hua, *et al*.. Private query on encrypted data in multi-user setting[C]. Proceedings of the 4th International Conference on Information Security and Experience, Sydney, 2008: 71–85.

[14] Yang Yan-jing, Lu Hai-bing, and Weng Jian. Multi-user private keyword search for cloud computing[C]. Proceedings of the 3rd International Conference on Cloud Computing Technology and Science, Athens, 2011: 264–271.

[15] Dong Chang-yu, Russello G, and Dulay N. Shared and searchable encrypted data for untrusted servers[J]. *Journal of Computer Security*, 2011, 19(21): 367–397.

王尚平：　男，1963年生，博士，教授，博士生导师，主要研究方向为密码理论与网络安全.

刘利军：　男，1987年生，硕士生，研究方向为密码理论与网络安全.

张亚玲：　女，1966年生，博士，教授，硕士生导师，主要研究方向为网络信息安全.