软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2016,27(5):1301−1308 [doi: 10.13328/j.cnki.jos.004912]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

可验证的基于词典的可搜索加密方案*

王尚平, 刘利军, 张亚玲

(陕西省网络计算与安全技术重点实验室(西安理工大学),陕西 西安 710048)

通讯作者: 王尚平, E-mail: spwang@mail.xut.edu.cn

摘 要: 针对云存储中数据检索和安全问题,提出了一个可验证的基于词典的可搜索加密方案.该方案能够验证搜索结果的完备性.在适应性不可区分安全模型下证明了该方案的安全性.与现有方案相比,该方案具有陷门大小固定、适应性安全、更新无需重新计算、可验证等优势.

关键词: 可搜索加密;词典;完备性;索引矩阵;校验和

中图法分类号: TP309

中文引用格式: 王尚平,刘利军,张亚玲.可验证的基于词典的可搜索加密方案.软件学报,2016,27(5):1301−1308. http://www.jos.org.cn/1000-9825/4912.htm

英文引用格式: Wang SP, Liu LJ, Zhang YL. Verifiable dictionary-based searchable encryption scheme. Ruan Jian Xue Bao/ Journal of Software, 2016,27(5):1301−1308 (in Chinese). http://www.jos.org.cn/1000-9825/4912.htm

Verifiable Dictionary-Based Searchable Encryption Scheme

WANG Shang-Ping, LIU Li-Jun, ZHANG Ya-Ling

(Shaanxi Key Laboratory for Network Computing and Security Technology (Xi'an University of Technology), Xi'an 710048, China)

Abstract: A verifiable dictionary-based searchable encryption scheme is proposed for verifying the completeness of search results. The security of the proposed scheme is analyzed under the security model of adaptive indistinguishability. Compared with the existing schemes, the proposed scheme has advantages in the following aspects: the size of trapdoor is constant, the updating doesn't require recalculation, and especially the search result is verifiable.

Key words: searchable encryption; dictionary; completeness; index matrix; checksum

为了节省费用和管理方便,越来越多的用户将其数据外包给不可信的第三方存储服务器,如云存储服务和邮件服务.由于外包的数据不在用户的管理之下,因而引起了用户对其数据机密性的担忧.数据加密是解决这一问题的好方法,但是经典的数据加密原语会严重地限制存储服务器处理用户访问请求的能力,如用户希望访问包含某一具体关键词的密文数据.因此,研究可搜索加密,即寻找一种新的密码学原语和协议,能够确保用户数据的机密性和密文数据的可检索性,已经成为近几年的研究热点.

可搜索加密由 Song 等人首次提出[1],目前已经得到了广泛的研究.Goh 给出了对称可搜索加密方案的索引的安全性定义,并利用 Bloom 过滤器提出一个方案,但该方案需要线性搜索时间[2].Boneh 等人首次利用公钥密码系统提出了一个非对称可搜索加密方案[3].Liu 等人提出的非对称可搜索加密方案中允许存储服务器参加解密的过程,从而降低了用户在解密方面的通信与计算负担[4].Golle 等人首次提出了基于连接关键词的可搜索加密方案[5],使得用户一次可以搜索多个关键词,但是方案中需要指定关键词的位置.无需指定关键词位置就可以进行连接关键词搜索的可搜索加密方案[6]由 Kerschbaum 提出,在搜索时仅仅需要一个常量大小的搜索陷门.

* 基金项目: 国家自然科学基金(61572019, 61173192)
Foundation item: National Natural Science Foundation of China (61572019, 61173192)
收稿时间: 2013-04-27; 采用时间: 2015-08-19

Cao[7,8]等人提出了可以对搜索结果进行相关度排序的可搜索加密方案.Curtmola 等人利用广播加密的方法解决了可搜索加密中的多用户问题,但方案中的加密文档是“只读”的,且被撤销的用户会影响未被撤销的用户[9].Yang 等人给出了多用户的可搜索加密的安全性定义,但是他们的方案并没有解决加密密钥共享的问题[10],该问题在文献[11]中得到解决.

上述所有方案都是假设存储服务器是“半诚实但好奇的(semi-honest-but-curious server)”,即,这样的存储服务器会:(1) 存储外包数据并且不会篡改它;(2) 公正地执行每一个搜索操作并且会返回符合搜索条件的密文数据;(3) 尝试获得有关用户密文数据的明文信息.但是,那些为了营利的外包存储服务提供商很可能会通过降低计算代价和节省下载宽带来降低成本,如存储服务器只执行一部分的搜索操作和返回一部分搜索结果.为了解决该问题,Chai 等人提出了一个可验证的可搜索加密方案[12],能够验证搜索结果的完备性.Wang 等人又提出了可验证的基于模糊关键词的可搜索加密方案[13],不仅能够验证结果的正确性和完备性,同时还能搜索模糊的关键词.Chang 等人提出了一个保持私有性的可搜索加密方案[14],该方案不使用公钥密码,支持文件的增量服务检索.Kamara 等人提出了动态可搜索加密方案[15],并证明了该方案是 CPA2-security.在该方案中,应用两种加密算法分别对明文和索引进行加密,同时生成搜索令牌,用加密后的索引构造一个 KRB 树,当需要添加或者删除文件时,用生成的更新令牌来更新加密后的索引和密文.Cash 等人提出了在超大规模的数据库中构造动态对称可搜索方案[16],该方案的基本理论构造支持单关键字搜索和提供渐进最优化的服务器索引大小、完全并行搜索和最少的泄露.该方案给出了一个用动态可搜索加密机制作为基础来支持最近的 SSE 进展的实现,包括复杂的搜索询问和丰富的操作设置.王尚平等人提出了一个高效的基于连接关键词的可搜索加密方案[17],使得授权用户能够利用连接关键词的陷门搜索加密文档.提出的方案在搜索陷门大小、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.Lü 等人提出了一个安全非对称的可搜索加密方案[18],该方案支持连接、分离和否定搜索操作,并证明了该方案在标准模型下是安全的.李经纬等人给出了可搜索加密技术研究综述[19],总结和展望了待解决的关键性问题和未来的研究方向.

本文提出了一个可验证的基于词典的可搜索加密方案 VDSES.这里的可验证主要是指搜索结果的正确性和搜索完备性,其中,搜索正确性是指只有符合搜索条件的加密文档才被返回,搜索完备性是指所有符合搜索条件的加密文档都被返回.搜索正确性是目前所有方案的都必须满足的性质,而搜索完备性本文方案是通过增加关键词的检验和来完成.根据安全性定义,证明了方案的安全性达到了适应性不可区分.与已提出的方案相比,本文方案具有陷门大小固定、适应性安全、更新无需重新计算、可验证等优势.此外,本文方案主要使用的是伪随机函数,故有很高的计算效率.

1 预备知识

1.1 系统模型

系统由$\{D,Serv,\Delta,u\}$组成,其中,D 为用户 u 要外包存储的文档集合;$Serv$ 是存储服务器,负责存储与搜索服务;Δ为关键词词典,包括所有可能的有意义的关键词,D 为其上的文档集合,即 $D\subseteq 2^{\Delta}$.

假设用户 u 有 n 个文档 $D=(D_1,\ldots,D_n)$要外包到可能会发生恶意行为的存储服务器 $Serv$ 上,记文档 D_i $(1\leqslant i\leqslant n)$的关键词列表为 $W_i=(w_{i,1},\ldots,w_{i,m},\ldots)\subset\Delta$,其中,$w_{i,j}(1\leqslant j\leqslant|W_i|)$为 D_i 的第 j 个关键词.令 $SKE=(Gen,Enc,Dec)$表示一个对称加密方案(如 AES),D_i在密钥 ek 下的加解密算法分别为 $SKE.Enc_{ek}(D_i)$和 $SKE.Dec_{ek}(Enc_{ek}(D_i))$,$|S|$表示集合 S 的元素个数,$D(w)\subset D$ 表示含有关键 w 的所有文档,$a\|b$ 表示两个字符串 a 和 b 的级联,$b\in_R B$ 表示从集合 B 中随机均匀的选取元素 b.$negl(\cdot)$表示可忽略的函数,即,对任意的多项式 $p(\cdot)$,存在 N_0,使得对任意的整数 $n>N_0$, $negl(n)<1/p(n)$ 成立.

为了能够让用户 u 验证搜索结果的完备性,u 选择两个秘密的数:一个大素数 p 和一个随机整数 $1<x<p$,u 为每个文档 $D_i(1\leqslant i\leqslant n)$随机均匀地选择一个唯一标识符 $id_i\in_R Z_p^*$,对于给定的关键词 $w_j\in\Delta(1\leqslant j\leqslant|\Delta|)$,$u$ 存储一个

w_j的检验和 $c_j=\prod_{id_i\in IDS(w_j)}(id_i+x) \bmod p$,其中,$IDS(w_j)$表示为包含 w_j的文档的标识符集合,该检验和使得文档的增加(乘以(id_i+x))和删除(乘以$(id_i+x)^{-1}$)都很容易.

为使存储服务器 *Serv* 能够搜索密文数据,对每个关键词 $w_j\in\Delta(1\leqslant j\leqslant|\Delta|)$都建立一个 n 维数组 A_j,记 A_j 中位置 i 的值为 $A_j[i]$,$A_j[i]$的形式为$\langle v_1,v_2\rangle$,其中,v_2 是随机均匀选取的 k-bit 大小的字符串.对于文档 D_i,若 D_i 包含关键词 w_j,则 $A_j[i]$中的 v_1 由伪随机函数生成,否则随机均匀选取 v_1 的值.将所有的 A_j 根据伪随机置换函数组成一个 $|\Delta|\times n$ 的矩阵,记为索引矩阵 M,如图 1 所示.

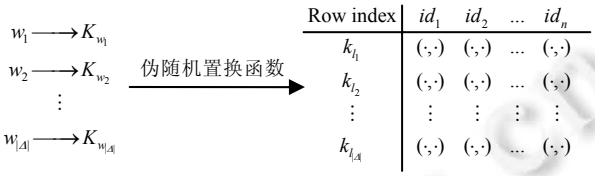


Fig.1 Storage structure of index matrix

图 1 索引矩阵的存储结构

可验证的基于词典的可搜索加密方案:位于词典Δ上的可验证的基于词典的可搜索加密方案 *VDSES*=(*Init*,*Enc*,*Trapdoor*,*Search*,*Verify*,*Dec*)由 6 个多项式时间算法构成,具体如下:

- *Init*(1^k):是一个概率密钥生成算法,由用户 u 执行以初始化系统.输入安全参数 k,输出系统密钥 K 和系统参数 *params*.
- *Enc*(D,K):是一个概率算法,由用户 u 执行以生成加密文档集合、生成索引矩阵及关键词检验和集合.输入文档集合 D 和系统密钥 K,输出密文集合 $C=(C_1,\ldots,C_n)$、索引矩阵 M 和关键词检验和集合 $CS=(c_1,\ldots,c_{|\Delta|})$.
- *Trapdoor*(w,K):是一个确定性算法,由用户 u 执行以获得要搜索的关键词的陷门.输入关键词 w 和系统密钥 K,输出关键词 w 的陷门 T_w.
- *Search*(T_w,M):是一个确定性算法,由存储服务器 *Serv* 执行以搜索包含关键词 w 的文档标识符.输入陷门 T_w 和索引矩阵 M,输出包含关键词 w 的文档标识符集合 *IDS*(w).
- *Verfy*(*IDS*(w),*CS*,K):是一个确定性算法,由用户 u 执行以验证搜索结果的完备性.输入文档标识符集合 *IDS*(w)、检验和集合 *CS* 和系统密钥 K,输出验证结果“1”或“0”.
- *Dec*(C_i,K):是一个确定性算法,由用户 u 执行以解密密文.输入密文 C_i 和系统密钥 K,输出明文 D_i.

方案的正确性:一个可验证的基于词典的可搜索加密方案是正确的,如果对于$\forall k\in\mathbf{N}$,$\forall K\leftarrow Init(1^k)$,$\forall D\subseteq 2^{\Delta}$,$\forall(M,C,CS)\leftarrow Enc(D,K)$及$\forall w\in\Delta$:

$$(Search(Trapdoor(w,K),M)=IDS(w))\wedge Verfy(IDS(w),K)=1\wedge(Dec_K(C_i)=D_i)(1\leqslant i\leqslant n).$$

1.2 安全性定义

提出的方案采用文献[9]中的安全性定义.文献[9]中定义了搜索历史、访问模式、搜索模式和迹(trace)的概念,并将迹作为愿意泄露的有关搜索历史的信息,详细定义如下:

定义 1(搜索历史,query history)[9]. 一个在文档集合 D 上的 q 次搜索历史为 $H=(D,\omega)$,其中,ω为含有 q 个关键词的向量,即,$\omega=(w_1,\ldots,w_q)$.

定义 2(访问模式,access pattern)[9]. 由 q 次搜索历史 $H=(D,\omega)$产生的访问模式是$\alpha(H)=(D(w_1),\ldots,D(w_q))$.

定义 3(搜索模式,search pattern)[9]. 由 q 次搜索历史 $H=(D,\omega)$产生的搜索模式是一个对称的二元矩阵 $\sigma(H)$,对于$\forall 1\leqslant i,j\leqslant q$,若 $w_i=w_j$,则$\sigma(H)$中的第 i 行第 j 列的元素值为 1,否则为 0.

定义 4(迹,trace)[9]. 由 q 次搜索历史 $H=(D,\omega)$产生的迹$\tau(H)$包括 D 中的文档长度、由 H 产生的访问模式和搜索模式,即,$\tau(H)=(|D_1|,\ldots,|D_n|,\alpha(H),\sigma(H))$.

此外,文献[9]又假设所有在词典Δ上的搜索历史H都是非奇异的,即定义5.

定义5(非奇异搜索历史,non-singular query history)[9]. 对于搜索历史H,若① 至少存在一个搜索历史$H'\neq H$,使得$\tau(H)=\tau(H')$;② 并且这样的搜索历史在给定$\tau(H)$时可以在多项式时间内找到.

令$VDSES=(Init,Enc,Trapdoor,Search,Verify,Dec)$为上述可验证的基于词典的可搜索加密方案,类似文献[9]中的安全性定义,给出本文方案的安全性定义.因为算法$Enc(D,K)$生成的检验和集合CS由用户自己保存,故在定义方案的安全性时,$Enc(D,K)$的输出中将省略CS.

定义6(适应性不可区分性安全)[9]. 令$k\in\mathbf{N}$为安全参数,$A=(A_0,\ldots,A_{q+1})$为攻击者,其中,$q\in\mathbf{N}$,进行如下的概率实验$Ind^*_{A,VDSE}(k)$:

$Ind^*_{VDSES,A}(k)$
- $K\leftarrow Init(1^k)$
- $b\in_R\{0,1\}$
- $(st_A,D_0,D_1)\leftarrow A_0(1^k)$
- $(M_b,C_b)\leftarrow Enc(D_b,K)$
- $(st_A,w_{0,1},w_{1,1})\leftarrow A_1(st_A,M_b)$
- $T_{b,1}\leftarrow Trapdoor(w_{b,1},K)$
- for $2\leqslant i\leqslant q$,
  - $(st_A,w_{0,i},w_{1,i})\leftarrow A_i(st_A,M_b,C_b,T_{b,1},...,T_{b,i-1})$
  - $T_{b,i}\leftarrow Trapdoor(w_{b,i},K)$
- let $T_b=(T_{b,1},...,T_{b,q})$
- $b'\leftarrow A_{q+1}(st_A,M_b,C_b,T_b)$
- if $b'=b$, output 1
- otherwise output 0

其中,st_A是一个字符串表示A的状态,并且要满足$\tau(D_0,w_{0,1},\ldots,w_{0,q})=\tau(D_1,w_{1,1},\ldots,w_{1,q})$的限制.称VDSES在适应性不可区分性语义下是安全的,如果对于任意多项式大小的攻击者$A=(A_0,\ldots,A_{q+1}),q=poly(k)$:

$$\Pr[Ind^*_{VKSES,A}(k)=1]\leqslant 1/2+negl(k).$$

定义7(适应性语义安全)[9]. 令$k\in\mathbf{N}$为安全参数,$A=(A_0,\ldots,A_q)$为攻击者,其中,$q\in\mathbf{N}$,$S=(S_0,\ldots,S_q)$为模拟器,进行如下的概率实验$Real^*_{VDSES,A}(k)$和$Sim^*_{VDSES,A,S}(k)$:

$Real^*_{VDSES,A}(k)$
- $K\leftarrow Init(1^k)$
- $(st_A,D)\leftarrow A_0(1^k)$
- $(M,C)\leftarrow Enc(D,K)$
- $(st_A,w_1)\leftarrow A_1(st_A,M,C)$
- $T_1\leftarrow Trapdoor(w_1,K)$
- for $2\leqslant i\leqslant q$,
  - $(st_A,w_i)\leftarrow A_i(st_A,M,C,T_1,...,T_{i-1})$
  - $T_i\leftarrow Trapdoor(w_i,K)$
- let $T=(T_1,...,T_q)$
- output $V=(M,C,T)$ and st_A

$Sim^*_{VDSES,A,S}(k)$
- $(st_A,D)\leftarrow A_0(1^k)$
- $(st_S,M,C)\leftarrow S_0(\tau(D))$
- $(st_A,w_1)\leftarrow A_1(st_A,M,C)$
- $(st_S,T_1)\leftarrow S_1(st_S,\tau(D,w_1))$
- for $2\leqslant i\leqslant q$,
  - $(st_A,w_i)\leftarrow A_i(st_A,M,C,T_1,...,T_{i-1})$
  - $(st_S,T_i)\leftarrow S_i(st_S,\tau(D,w_1,...,w_i))$
- let $T=(T_1,...,T_q)$
- output $V=(M,C,T)$ and st_A

称VDSES是适应性语义安全的,如果对于任意的多项式大小的攻击者$A=(A_0,\ldots,A_{q+1}),q=poly(k)$,都存在一个非均匀的模拟器$S=(S_0,\ldots,S_q)$,使得对任意的多项式大小的区分器$\bar{D}$:

$$|\Pr[\bar{D}(st_A,V)=1:(st_A,V)\leftarrow Real^*_{VKSES,A}(k)]-\Pr[\bar{D}(st_A,V)=1:(st_A,V)\leftarrow Sim^*_{VKSES,A,S}(k)]|\leqslant negl(k).$$

文献[9]中指出,适应性语义安全性意味着适应性不可区分性,故有:

定理 1. VDSES 方案的适应性语义安全性意味着 VDSES 方案的适应性不可区分性.

2 可验证的基于关键词的可搜索加密方案

本节将详细地描述所构造的方案,并给出安全性证明和效率分析.

2.1 可验证的基于关键词的可搜索加密方案

- *Init*(1^k):该算法由用户 u 执行以初始化系统,输入安全参数 k,随机选择大素数 p 及 $1<x<p$;令 $F:\{0,1\}^k\times\{0,1\}^*\to\{0,1\}^k$, $G:\{0,1\}^k\times\{0,1\}^*\to\{0,1\}^{k+\log_2 p}$ 为伪随机函数, $Q:\{0,1\}^k\times\{0,1\}^*\to\{0,1\}^{\log_2|\Delta|}$ 为伪随机置换,随机均匀地选择 3 个 k-bit 长的字符串 K_1,K_2 和 *flag*,其中,K_1 和 K_2 分别作为伪随机函数 F 和 Q 的随机种子;为语义安全的对称加密算法 SKE 生成加密密钥 $ek\leftarrow SKE.Gen(1^k)$,发布 $params=(F,G,Q,SKE,flag)$作为系统参数,系统密钥为 $K=(K_1,K_2,ek,x,p)$.
- *Enc*(D,K):该算法由用户 u 执行以加密数据集合、生成索引矩阵及关键词的检验和集合,输入文档集合 D 和系统密钥 K,用户 u 按如下步骤计算:
  ① 为每个文档 $D_i\in D(1\leqslant i\leqslant n)$随机均匀地选择一个唯一的标识符 $id_i\in_R Z_p^*$,加密文档 D_i 为
  $$C_i=SKE.Enc_{ek}(D_i);$$
  ② 为每个关键词 $w_j\in\Delta(1\leqslant j\leqslant|\Delta|)$生成一个 n 维数组 A_j;
  ③ 每个数组 $A_j(1\leqslant j\leqslant|\Delta|)$按如下过程执行:对每个文档 $D_i\in D(1\leqslant i\leqslant n)$,随机均匀地选择一个 k-bit 的字符串 $r_{j,i}$,若 $w_j\in W_i$,这里,$W_i=(w_{i,1},\ldots,w_{i,m},\ldots)\subset\Delta$为文档 $D_i(1\leqslant i\leqslant n)$的关键词列表,计算 $K_{w_j}=F_{K_1}(w_j)$ 和检验和 $c_j:=c_j\times(x+id_i) \bmod p$,这里,$c_j$ 的初始值为 1,将 $(\langle flag\|id_i\rangle\oplus G(K_{w_j},r_{j,i}),r_{j,i})$ 存储在 $A_j[i]$,其中,*flag* 为系统参数,是一个固定的 k-bit 长的字符串;否则,随机均匀地选择字符串 $v_1\subset\{0,1\}^{k+\log_2 p}$,将$(v_1,r_{j,i})$存储在 $A_j[i]$;
  ④ 将所有数组 A_j 组成一个$|\Delta|\times n$ 的索引矩阵 M,其中,A_j 位于 M 的 $Q_{K_2}(w_j)$ 行;
  ⑤ 将索引矩阵 M 和密文集合 $C=(C_1,\ldots,C_n)$发送给存储服务器 *Serv* 存储,检验和集合 $CS=(c_1,\ldots,c_{|\Delta|})$由用户 u 保存.
- *Trapdoor*(w,K):该算法由用户 u 执行以获得关键词的陷门,输入要搜索的关键词 $w\in\Delta$和系统密钥 K,计算陷门 $T_w=(Q_{K_2}(w),F_{K_1}(w))$, 将 T_w 发送给云存储服务器 *CSS*.
- *Search*(T_w,M):该算法由存储服务器 *Serv* 执行以搜索包含关键词 w 的文档标识符,输入陷门 T_w 和索引矩阵 M,*CSS* 首先定位到 M 的第 $Q_{K_2}(w)$ 行,记该行为数组 A_w,若无,则返回⊥;否则,初始化一个空集 *IDS*(w),对 A_w 中的每个元素的值$(v_{i,1},v_{i,2})(1\leqslant i\leqslant n)$,计算 $\upsilon=G(F_{K_1}(w),v_{i,2})\oplus v_{i,1}$, 并判断下式是否成立:
  $$first\_k\_bit(\upsilon)=flag,$$
  其中,*first_k_bit*(·)为取字符串前 k-bit 的函数.若成立,则 $IDS(w)=IDS(w)\cup\{get\_id(\upsilon)\}$,其中 *get_id*(·)为取字符串中的文档标识符函数,即获得υ的后 $\log_2 p$bit;若不成立,检验下一个数组$(v_{i+1,1},v_{i+1,2})(1\leqslant i+1\leqslant n)$,直到最后将 *IDS*($w$)发送给用户 u.
- *Verfy*(*IDS*(w),*CS*):由用户 u 执行以验证搜索结果的完备性,输入 *IDS*(w)和用户自己保存的 *CS*,用户 u 首先从 *CS* 中获得关键词 w 的检验和,记为 c_w,并判断下式是否成立:
  $$c_w=\prod_{id_i\in IDS(w)}(id_i+x) \bmod p,$$
  若成立,则根据 $id_i\in IDS(w)$向 *CSS* 获得相应的密文 C_i;否则,返回⊥.
- *Dec*(C_i,K):由用户 u 执行,用于解密密文数据,输入密文 C_i 和系统密钥 K,用户 u 执行解密算法得明文:
  $$D_i=SKE.Dec_{ek}(C_i).$$

2.2 方案的安全性分析

定理 2. 提出的方案是正确的.

证明:给定关键词 $w\in\Delta$的陷门 $T_w=(Q_{K_2}(w),F_{K_1}(w))$,*Serv* 可以定位到索引矩阵 M的 $Q_{K_2}(w)$ 行,记该行为数组 A_w,将 A_w 中的每个元素的值记为$(v_{i,1},v_{i,2})(1\leqslant i\leqslant n)$,针对各个$(v_{i,1},v_{i,2})$,利用 $F_{K_1}(w)$ 计算 $\upsilon=G(F_{K_1}(w),v_{i,2})\oplus v_{i,1}$,若 W_i 中包含关键词 w,则 $first\_k\_bit(\upsilon)=flag$,从而 $IDS(w)=IDS(w)\cup\{get\_id(\upsilon)\}$.用户 u 在获得 $IDS(w)$后,首先在检验和集合 *CS* 中找到关键词 w 的检验和,记为 c_w,若 *Serv* 没有发生恶意行为且 $IDS(w)$在传输中没有被篡改,则可以验证 $c_w=\prod_{id_i\in IDS(w)}(id_i+x)\bmod p$,验证通过后,就可以通过 *IDS* 中的文档标识符获得加密的文档并解密.

因此,提出的方案是正确的. □

定理 3. 如果 F 和 G 是伪随机函数,Q 是伪随机置换,*SKE* 是语义安全的,则 VDSSE 是适应性安全的.

证明:我们描述一个多项式大小的模拟器 $S=(S_0,\ldots,S_q)$,使得对任意多项式大小的攻击者 $A=(A_0,\ldots,A_q)$,$Real^*_{VDSES,A}(k)$ 和 $Sim^*_{VDSES,A,S}(k)$ 的输出是计算不可区分的.

模拟器 $S=(S_0,\ldots,S_q)$按如下方式适应性生成字符串 $V^*=(M^*,C^*,T^*)=(M^*,C_1^*,...,C_n^*,T_1^*,...,T_n^*)$:

- $S_0(1^k,\tau(D))$:S_0 生成一个大小为$|\Delta|\times n$ 的索引矩阵 M^*,在矩阵的每个位置处都插入一个随机值对 $(rv_{i,j}^*,cv_{i,j}^*)$ $(1\leqslant i\leqslant|\Delta|,1\leqslant j\leqslant n)$,$S_0$ 将 M^*包含在 st_S 中,输出(M^*,C^*,st_S),其中,$C_i^*\in_R\{0,1\}^{|D_i|}(1\leqslant i\leqslant n)$.除了可忽略的概率之外,因为 st_A 中不包括 K_1 和 K_2,F 和 G 是伪随机函数,Q 是伪随机置换,故 M^*与真实的索引矩阵是不可区分的.同样,除了可忽略的概率之外,因为 st_A 中不包括 *ek*,语义安全的 *SKE* 可以确保 C_i^* 与真实的密文不可区分.
- $S_1(st_S,\tau(D,w_1))$:在本文的方案中,对每个关键词 $w_j\in\Delta(1\leqslant j\leqslant|\Delta|)$构造一个数组 A_j,针对每个文档 D_i $(1\leqslant i\leqslant n)$,根据其关键词列表 W_i 中是否包含关键词 w_j,$A_j[i]$是 $(\langle flag\,\|\,id_i\rangle\oplus G(K_{w_j},r_{j,i}),r_{j,i})$ 或$(random\ selected,r_{j,i})$.$S_1$ 从 M^*中随机选择一行 $t_{1,1}$,对应的数组为 $A_{t_{1,1}}$,并随机选择 $t_{1,2}\in_R\{0,1\}^k$,对每个文档 D_i,若 $id_i\in\tau(D,w_1)$,则随机选择 k-bit 的字符串 $r^*_{t_{1,1},i}$,记为 $cv_{t_{1,1},i}$,计算 $\langle flag\,\|\,id_i\rangle\oplus G(t_{1,2},r^*_{t_{1,1},j})$,记为 $rv_{t_{1,1},i}$,令 $A_{t_{1,1}}[i]$ 为 $(rv_{t_{1,1},i},cv_{t_{1,1},i})$ 并存储在 $A_{t_{1,1}}[i]$;否则,随机选择 $rv_{t_{1,1},i}$.令 $T_1^*=(t_{1,1},t_{1,2})$,S_1 将 T_1^* 与 w_1 的关系包含在 st_S 中,输出 (T_1^*,st_S).除了可忽略的概率之外,因为 st_A 中不包括 K_1 和 K_2,F 是伪随机函数,Q 是伪随机置换,故 $t_{1,1}$ 与真实的 $Q_{K_2}(w_1)$ 不可区分,$t_{1,2}$ 与真实的 $F_{K_1}(w_1)$ 不可区分.
- $S_i(st_S,\tau(D,w_1,\ldots,w_i))(1\leqslant i\leqslant q)$:$S_i$ 首先检查 w_i 是否已搜索过,这可以通过检查是否存在 $1\leqslant j\leqslant i-1$ 使得 $\sigma[i,j]=1$,若 w_i 没有被搜索过,则 S_i 按照 S_1 生成 w_i 的陷门 T_i^*;否则,将 w_i 已经使用过的陷门记为 T_i^*.最后,S_i 输出 (T_i^*,st_S).很明显,与 S_1 类似,$t_{i,1}$ 与真实的 $Q_{K_2}(w_1)$ 不可区分,$t_{i,2}$ 与真实的 $F_{K_1}(w_1)$ 不可区分.

综上,$Real^*_{VDSES,A}(k)$ 输出的 V 和 $Sim^*_{VDSES,A,S}(k)$ 输出的 V^*是不可区分的.根据定理 1,构造的方案的安全性是适应性不可区分的. □

此外,注意到由于检验和集合是存储在用户端的,恶意的存储服务器 *Serv* 返回的文档标识符集合 $IDS(w)$通过 $Verfy(IDS(w),CS,K)$检验的概率为 $1/p$,因此,方案中的检验和是安全的.

2.3 方案的效率分析

本节将提出的方案与已有的可搜索加密方案[2,9,14](属于 SSE)进行比较,因为提出的方案与相比较的方案使用的主要是伪随机函数和异或运算,故方案都有很高的计算效率.下面就方案的陷门大小、搜索复杂性、用户存储、是否适应性安全、更新是否需要重新计算和是否可验证几个方面进行比较,比较结果见表 1,其中,r 表示 Bloom 过滤器的大小.

通过比较可以发现,我们的方案具有陷门大小固定、适应性安全、更新无需重新计算、可验证等优势.在搜索复杂性方面,除了 SSE-2 的效率最高外,其余的方案都相同;但是 Goh 方案因为使用的是 Bloom 过滤器,故其搜索结果中会存在误报.在用户存储方面,本文方案因为为每个关键词都存储了一个检验和,从而导致需要更

多的用户存储空间.

Table 1 Compared with other searchable encryption schemes

表 1 与其他可搜索加密方案的比较

Scheme	|Trapdoor|	Search complexity	User storage	Adaptive security	Update need recalculation	Is verified
Goh[2]	$o(r)$	$O(\|D\|)$	$o(r)$	No	No	No
SSE-1[9]	$O(1)$	$O(\|D(w)\|)$	$O(1)$	No	Yes	No
SSE-2[9]	$O(\|D\|)$	$O(\|D\|)$	$O(1)$	Yes	Yes	No
Chang[14]	$O(1)$	$O(\|D\|)$	$O(\|\Delta\|)$	No	Yes	No
Ours	$O(1)$	$O(\|D\|)$	$O(\|\Delta\|)$	Yes	No	Yes

3 总 结

本文提出了一个可验证的基于词典的可搜索加密方案,使得用户不仅可以搜索加密文档,同时还能让用户验证搜索结果的完备性,从而可以在一定程度上预防存储服务器的恶意行为.根据安全性定义,证明了方案是适应性不可区分的.此外,通过与已有的方案相比较,本文方案具有陷门大小固定、适应性安全、更新无需重新计算、可验证等优势.

References:

[1] Song DX, Wagner D, Perrig A. Practical techniques for searches on encrypted data. In: Proc. of the IEEE Symp. on Security and Privacy. Berkeley, 2000. 44−55. [doi: 10.1109/SECPRI.2000.848445]

[2] Goh EJ. Secure indexes. 2003. http://eprint.iacr.org/2003/216/

[3] Boneh D, Kushilevitz E, Ostrovsky R, Persiano G. Public key encryption with keyword search. In: Cachin C, Camenisch J, eds. Proc. of the Eurocrypt 2004. LNCS 3027, Berlin, 2004. 506−522. [doi: 10.1007/978-3-540-24676-3_30]

[4] Liu Q, Wang GJ, Wu J. Secure and privacy preserving keyword searching for cloud storage services. Journal of Network and Computer Applications, 2012,35(3):927−933. [doi: 10.1016/j.jnca.2011.03.010]

[5] Golle P, Staddon J, Waters B. Secure conjunctive keyword search over encrypted data. In: Jakobsson M, Yung M, Zhou JY, eds. Proc. of the Applied Cryptography and Network Security. LNCS 3089, 2004. 31−45. [doi: 10.1007/978-3-540-24852-1_3]

[6] Kerschbaum F. Secure conjunctive keyword searches for unstructured text. In: Proc. of the 5th Int'l Conf. on Network and System Security. Milan, 2011. 285−289. [doi: 10.1109/ICNSS.2011.6060016]

[7] Cao N, Wang C, Li M, Lou W. Privacy-Preserving multi-keyword ranked search over encrypted cloud data. In: Proc. of the IEEE INFOCOM. Shanghai, 2011. 829−837. [doi: 10.1109/INFCOM.2011.5935306]

[8] Wang C, Cao N, Ren K, Li M, Ren K, Lou W. Enabling secure and efficient ranked keyword search over outsourced cloud data. IEEE Trans. on Parallel and Distributed Systems, 2012,23(8):1467−1479. [doi: 10.1109/TPDS.2011.282]

[9] Curtmola R, Garay JA, Kamara S, Ostrovsky R. Searchable symmetric encryption: Improved definitions and efficient constructions. Journal of Computer Security, 2011,19(5):895−934. [doi: 10.1145/1180405.1180417]

[10] Yang YJ, Ding XH, Deng RH, Bao F. Multi-User private queries over encrypted database. Journal of Applied Cryptography, 2009, 1(4):309−319. [doi: 10.1504/IJACT.2009.028029]

[11] Yang YJ, Lu HB, Weng J. Multi-User private keyword search for cloud computing. In: Lambrinoudakis C, Rizomiliotis P, Wlodarczyk TW, eds. Proc. of the 3rd Int'l Conf. on Cloud Computing Technology and Science. Athens, 2011. 264−271. [doi: 10.1109/CloudCom.2011.43]

[12] Chai Q, Gong G. Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers. In: Proc. of the IEEE Int'l Conf. on Communications. Ottawa, 2012. 917−922. [doi: 10.1109/ICC.2012.6364125]

[13] Wang JF, Ma H, Li J, Zhu H, Ma SQ, Chen XF. A new efficient verifiable fuzzy keyword search scheme. Journal of Wireless Mobile Networks, Ubiquitious Computing, and Dependable Applications, 2012,3(4):61−71.

[14] Chang YC, Mitzenmacher M. Privacy preserving keyword searches on remote encrypted data. In: Ioannidis J, Keromytis AD, Yung M, eds. Proc. of the 3rd Int'l Conf. on Applied Cryptography and Network Security. LNCS 3531, Berlin: Springer-Verlag, 2005. 442−455. [doi: 10.1007/11496137_30]

[15] Kamara S, Papamanthou C. Parallel and dynamic searchable symmetric encryption. In: Proc. of the Financial Cryptography and Data Security. Berlin, Heidelberg: Springer-Verlag, 2013. 258−274. [doi: 10.1007/978-3-642-39884-1_22]

[16] Cash D, Jaeger J, Jarecki S, Jutla C, Krawczyk H. Dynamic searchable encryption in very-large databases: Data structures and implementation. In: Proc. of the NDSS 2014. San Diego, 2014. [doi: 10.14722/ndss.2014.23264]

[17] Wang SP, Liu LJ, Zhang YL. An efficient conjunctive keyword searchable encryption scheme. Journal of Electronics & Information Technology, 2013,35(9):2266−2271 (in Chinese with English abstract). [doi: 10.3724/SP.J.1146.2012.01036]

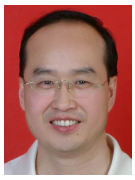
[18] Lü ZQ, Hong C, Zhang M, Feng DG. Expressive and secure searchable encryption in the public key setting (full version). Lecture Notes in Computer Science, 2014,8783:364−376. [doi: 10.1007/978-3-319-13257-0_21]

[19] Li JW, Jia CF, Liu ZL, Li J, Li M. Survey on the searchable encryption. Ruan Jian Xue Bao/Journal of Software, 2015,26(1): 109−128 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4700.htm [doi: 10.13328/j.cnki.jos.004700]

附中文参考文献:

[17] 王尚平,刘利军,张亚玲.一个高效的基于连接关键词的可搜索加密方案.电子与信息学报,2013,35(9):2266−2271. [doi: 10.3724/SP.J.1146.2012.01036]

[19] 李经纬,贾春福,刘哲理,李进,李敏.可搜索加密技术研究综述.软件学报,2015,26(1):109−128. http://www.jos.org.cn/1000-9825/4700.htm [doi: 10.13328/j.cnki.jos.004700]



王尚平(1963－),男,陕西扶风人,博士,教授,博士生导师,主要研究领域为密码理论,网络安全.

张亚玲(1966－),女,博士,教授,主要研究领域为云存储中的信息安全,计算机软件理论.

刘利军(1986－),男,硕士,主要研究领域为云存储中的数据安全.