# 云环境下可搜索加密技术安全机制及应用陷阱*

杨 宁[1], 金 逸[2], 刘 丹[1], 彭凝多[3]
(1. 南京南瑞集团公司 北京中电普华信息技术有限公司, 北京 100192; 2. 国网江苏省电力公司, 南京 2100242; 3. 电子科技大学 计算机科学与工程学院, 成都 611731)

**摘 要**: 为了能在云计算环境下安全地使用可搜索加密技术,针对近几年来可搜索加密技术的研究成果,总结了云计算环境下对称/非对称可搜索加密技术的主要算法模型与安全模型,分析了各类安全方案的安全缺陷,并分析了基本可搜索加密方案与加入了密文安全传输、匿名性、第三方代理、隐私保护协议的四类增强方案在基于选择关键字攻击、字典攻击、统计分析攻击、选择明文攻击四类攻击模型下的安全性。最后,总结了可搜索加密技术的应用陷阱,并提出了当前可搜索加密技术面临的安全性挑战。

**关键词**: 可搜索加密; 算法模型; 安全模型; 安全隐患; 应用陷阱
**中图分类号**: TP309 **文献标志码**: A **文章编号**: 1001-3695(2015)08-2254-00
doi:10.3969/j.issn.1001-3695.2015.08.003

# Security and application traps of searchable encryption technique in cloud environment

Yang Ning[1], Jin Yi[2], Liu Dan[1], Peng Ningduo[3]
(1. *Beijing China Power Information Technology Co. Ltd, NARI Group Corporation, Beijing* 100192, *China*; 2. *State Grid Jiangsu Electric Power Company, Nanjing* 210024, *China*; 3. *School of Computer Science, University of Electronic Science & Technology of China, Chengdu* 611731, *China*)

**Abstract**: In order to safely use searchable encryption technique in cloud computing environment, this paper studied the searchable encryption technique which was researched in recent years. It summarized the main scheme model and security model for symmetric/asymmetric searchable encryption schemes, analyzed the security vulnerabilities of each scheme. Under chosen keywords attack, dictionary attack, statistical attack, chosen plaintext attack models, this paper summarized the securities of the basic searchable encryption scheme and the improved schemes with secure ciphertext transmission, anonymity, third party, or private information retrieval protocol. Finally, it summarized the application traps of the searchable encryption technique and proposed the current security challenges.

**Key words**: searchable encryption; application model; security model; security vulnerability; application trap

## 0 引言

云存储为用户提供了具有弹性扩展、高可靠、易使用等特性的数据存储服务,受到越来越广泛的应用[1,2]。然而,数据外包存在着天然的安全隐患。敏感数据(如用户的隐私数据、商业数据等)如果存放在云中,云服务提供商将能够直接读取与使用该数据,从而侵犯用户的隐私与破坏数据的安全[3~5]。对大多数人而言,假定云服务提供商"绝对可信"是不可能的,因此需要从技术层面为用户提供隐私与数据保护,并且要求能在安全性保证的前提下,尽可能保障用户对数据的正常操作(如基于关键字检索文件)。当前,用户隐私保护与数据安全成为了制约数据外包模式是否能被广泛应用的瓶颈。

保护用户数据安全的基本思路是对数据加密。然而,传统的数据加密方案,如 DES(数据加密标准)[6]、AES(高级数据加密标准)[7]等,无法满足数据外包模式的应用需求。其问题在于,云服务提供商不应当拥有加密/解密密钥,因此,在不解密的条件下,用户无法对采用传统加密方案加密的数据进行有效的操作,例如按关键字检索需要的文件。

可搜索加密(searchable encryption)技术[8~10]是当前公认的解决加密邮件转发[9]、云存储文件安全检索[11]、加密日志分析[12]等实际应用问题的重要加密手段。可搜索加密技术的目标是在不影响数据检索功能的条件下,保护用户外包数据的安全与查询隐私。该技术的基本应用场景为:用户加密自己的数据并上传到远程服务器,并在需要检索文件时提交加密的关键字给服务器,随后服务器在不知道用户数据与查询关键字的情况下,用加密的关键字检索加密的数据,并返回给用户符合要求的加密文件。

与传统的封闭加密系统不同,经过可搜索加密算法加密后的数据结构对外提供了公共检索接口,因此导致了许多安全问题。不恰当地使用可搜索加密算法,会导致应用系统存在严重的安全隐患。因此,充分认识可搜索加密技术的安全机制与安全隐患对于正确应用该类技术至关重要。为了探明安全隐患

**收稿日期**: 2014-10-20; **修回日期**: 2014-12-03 **基金项目**: 四川省科技支撑计划资助项目(2013JQ0005,2014JY0010)

**作者简介**:杨宁(1982-),男,江西南昌人,博士,主要研究方向为企业信息化、JavaEE 技术、云计算(freedom.wind@163.com);金逸(1977-),男,江苏南京人,硕士,主要研究方向为电力系统自动化及科技信息化管理;刘丹(1981-),女,湖北荆门人,硕士,主要研究方向为 Java EE 技术、云计算;彭凝多(1985-),男,四川宜宾人,博士研究生,主要研究方向为云安全.

产生的原因,本文从可搜索加密技术的算法模型与安全模型入手,详细分析了安全问题存在的来源,随后分析了功能扩展带来的新的安全隐患,并归纳总结了当前多种安全性弥补方案与存在的问题,最后指出了实际应用陷阱与安全性挑战。

## 1　云环境下对称可搜索加密技术安全机制

在对称可搜索加密方案(symmetric searchable encryption, SSE)的云存储应用场景中,用户既是数据的加密者,也是数据的检索者。由于所有密钥由用户保存,通常利用一些高效的数据结构(如全局文件索引)来加速检索。

### 1.1　算法模型

对称可搜索加密方案最早由文献[13]提出。其中,文件内容被加密成一系列单词的匹配模式;用户提交的关键字经过加密转换后,由服务器进行匹配模式扫描,当发现某文件的某段加密内容出现模式匹配时,则判定该文件满足用户的查询要求。显然,该方案要求对所有加密文件进行线性扫描,系统开销较大。为了加快检索,随后许多基于索引的对称可搜索加密方案被相继提出。

文献[14]最早提出基于索引的可搜索加密方案。该方案将文件的所有关键字加密后放到布隆过滤器(Bloom filter)中,查询的关键字被加密为布隆过滤器的输入值后上传到服务器,服务器测试布隆过滤器,如果匹配则表示该文件满足查询条件并返回相应文件。除此之外,该文献首次提出了基本的语义安全模型,即抗选择关键字攻击语义安全性(semantic security against chosen keyword attack,CKA)。

文献[8]提出了两个重要的对称可搜索加密方案。方案均以加密的关键字为索引,构造匹配该关键字的所有文件的标志符(identifier,ID)链表。服务器收到用户提交的加密关键字后,直接得到加密的文件 ID 链表并解密,返回对应的文件。对于安全性而言,第一个方案为非自适应的(静态历史记录),第二个方案为自适应的(动态历史记录)。该方案在文献[11]的实际应用系统中得到了实现,并在文献[15]中得到了抽象建模,形成一个算法体系。

为了清晰阐述对称可搜索加密方案的应用场景与应用接口,以下给出对称可搜索加密方案基本算法模型。

**定义 1**　对称可搜索加密方案算法模型。一个通用的对称可搜索加密方案(SSE)为五个多项式时间算法 Gen、Enc、Token、Search、Dec 的集合。

a) $K \leftarrow \mathrm{Gen}(1^k)$,为一个概率算法,用于生成密钥。输入安全参数 $k$,输出密钥 $K$。算法由用户执行,输出的密钥由用户保存。

b) $(I,C) \leftarrow \mathrm{Enc}_K(D)$,为一个概率算法,用于加密文件数据。输入密钥 $K$ 与具有 $n$ 个文档的用户文档集 $D=(d_1,\cdots,d_n)$,输出可搜索结构 $I$(如安全索引)与加密的文档集 $C=(c_1,\cdots,c_n)$。可搜索结构使得用户可以查询关键字,并由服务器返回匹配的文档。算法由用户执行,$(I,C)$ 发送给服务器存储。

c) $t \leftarrow \mathrm{Token}_K(w)$,为一个确定性算法,用于生成安全的查询令牌。输入密钥 $K$ 与查询的关键字 $w$,输出一个查询令牌 $t$(token,部分文献称为后门(trapdoor)或能力(capacity))。算法由用户执行。

d) $X \leftarrow \mathrm{Search}(I,t)$,为一个确定性算法,用于检索加密的文件。输入可搜索结构 $I$ 与查询的令牌 $t$,输出匹配的 $m$ 个加密文件的标志符 $X=(x_1,\cdots,x_m)$。算法由服务器执行,并将结果(即匹配的加密文档)返回给用户。

e) $d_i \leftarrow \mathrm{Dec}_K(c_i)$,为一个确定性算法,用于解密加密的文件内容。输入密钥 $K$ 与加密文档 $c_i$,输出恢复的明文 $d_i$。算法由用户执行。

### 1.2　安全模型

为了更清晰地表述对称可搜索加密方案的安全模型,首先介绍一些额外的概念。称某一时间段内这些信息的实体为一个历史信息;称服务器接收到令牌并返回匹配文档的过程为访问模式;称用户查询的关键字历史为查询模式。令 $\Delta$ 表示一个包含所有单词的字典,令 $D \subseteq 2^{\Delta}$ 表示基于该字典的一个文档集,以上概念的严格定义如下。

**定义 2**　历史信息 $H$。对于 $D$,一个具有 $q$ 次查询的历史定义为一个二元组 $H=(D,W)$,该元组包括文档集 $D$ 与 $q$ 个查询的关键字 $W=(w_1,\cdots,w_q)$。

**定义 3**　访问模式 $\alpha(H)$。基于 $q$ 次查询的历史 $H=(D,W)$ 的访问模式定义为一个多元组 $\alpha(H)=(D(w_1),\cdots,D(w_q))$,其中 $D(w)$ 表示文档集 $D$ 中匹配查询关键字 $w$ 的所有文档。

**定义 4**　查询模式 $\sigma(H)$。基于 $q$ 次查询的历史 $H=(D,W)$ 的查询模式为一个 $q \times q$ 对称二元矩阵 $\sigma(H)$。对于 $1 \leqslant i,j \leqslant q$,第 $i$ 行第 $j$ 列的值为 1 当且仅当 $w_i=w_j$;其他情况下值为 0。

**定义 5**　轨迹 $\tau(H)$。基于 $q$ 次查询的历史 $H=(D,W)$ 的轨迹定义为一个多元组 $\tau(H)=(|d_1|,\cdots,|d_n|,\alpha(H),\sigma(H))$。其中包括 $n$ 个文档的大小信息、访问模式与查询模式。

**定义 6**　非单历史信息。称一个历史 $H$ 为非单的,当且仅当至少存在一个历史 $H' \neq H$ 满足 $\tau(H')=\tau(H)$,且该历史信息可以在给定 $\tau(H)$ 的条件下,在多项式时间内找出。

基本的安全模型是非自适应安全模型,该模型指静态产生的所有历史。换言之,当历史没有构造完成时,不允许敌手观察文件集的索引构造与任何令牌构造的信息。在一般应用中,该模型能满足用户的基本安全需求。基于非自适应抗选择关键字攻击(non-adaptive indistinguishability against chosen keyword attack,IND-CKA)的安全模型定义如下。

**定义 7**　非自适应不可分辨性。令 SSE = (Gen,Enc,Token,Search,Dec) 为基于字典 $\Delta$ 的对称可搜索加密方案,令 $k \in \mathbf{N}$ 为安全参数,$A=(A_1,A_2)$ 为多项式时间敌手(即敌手被限制在多项式时间内完成计算),$st_A$ 表示算法的状态信息,$x \leftarrow_U X$ 表示从集合 $X$ 中均匀随机抽取元素 $x$。考虑以下概率实验 $\mathrm{Ind}_{\mathrm{SSE},A}(k)$。

$\mathrm{Ind}_{\mathrm{SSE},A}(k)$
　$K \leftarrow \mathrm{Gen}(1^k)$
　敌手计算 $(st_A,H_0,H_1) \leftarrow A_1(1^k)$
　$b \leftarrow_U \{0,1\}$
　将 $H_b$ 解析为 $(D_b,W_b)$
　$(I_b,C_b) \leftarrow \mathrm{Enc}_K(D_b)$
　对于 $1 \leqslant i \leqslant a$,计算 $t_{b,i} \leftarrow \mathrm{Token}_K(w_{b,i})$
　令 $T_b=(t_{b,1},\cdots,t_{b,a})$
　敌手计算 $b' \leftarrow A_2(st_A,I_b,C_b,T_b)$
　如果 $b'=b$,则输出 1,否则输出 0

其中,满足非单历史限制条件 $\tau(H_0)=\tau(H_1)$。称对称可搜索加密方案具有非自适应抗选择关键字攻击不可分辨性,当且仅当对于任意一个多项式时间敌手 $A=(A_1,A_2)$,满足概率

$Pr[\mathrm{Ind}_{\mathrm{SSE},A}(k)=1]\leqslant 1/2+\varepsilon$。其中，$\varepsilon$ 为一个可忽略小量，其概率基于概率函数 Gen 与 Enc。

从安全模型中可以看出，任何一个满足非自适应不可分辨性的可搜索加密方案都是静态安全的，即用户加密数据并查询数据，而敌手在某一时刻截获用户在服务器上的所有数据与历史记录，此时用户的数据是基本安全的（这里的基本安全是指在满足安全模型的限制条件下的安全，事实上有可能该限制条件并不能有效地满足）。可见，该安全模型很好地模拟了实际场景。

### 1.3 安全隐患

#### 1.3.1 易受选择明文攻击

取消限制条件 $\tau(H_0)=\tau(H_1)$ 即为抗自适应的选择明文攻击（chosen plaintext attack，CPA）安全模型。增加该限制条件的主要来源是 $t\leftarrow\mathrm{Token}_K(w)$ 为一个确定性算法。该算法模型表明，如果敌手两次查询相同的内容，则得到的令牌相同。显然，传统的抗自适应 CPA 的加密方案要求加密是概率性的，加密同样的内容应该会得到不同的看似随机的值。

事实上，在实际应用中用户每次查询的内容都不同的概率很小，因此敌手（包括窃听者、存储服务提供商的内部攻击者等）可以通过记录令牌，从而判断该次查询与历史查询之间的关系（即出现 $\tau(H_0)\neq\tau(H_1)$ 的情况）。同时，由于令牌与索引 $I$ 中的条目是对应的，因此索引 $I$ 中对应的条目也将受到同样的影响，即索引也易受选择明文攻击。与传统的不抗 CPA 的密码系统类似，此时若敌手具有明文或密钥的某些先验信息，则可推断明文信息乃至获取密钥。

根据以上分析，令牌与索引的安全性可以归纳为介于抗选择明文攻击与唯密文攻击（ciphertext only attack，COA）之间，即安全模型的安全性之间具有以下关系：

$$\mathrm{COA}<\mathrm{CKA}<\mathrm{CKA2}<\mathrm{CPA}$$

#### 1.3.2 易受统计分析攻击

用户查询的历史记录易受统计分析攻击（statistical attack，SA）。用户的查询统计信息包括访问模式 $\sigma(H)$ 与查询模式 $\sigma(H)$，而这些信息为敌手（包括窃听者、存储服务提供商的内部攻击者等）提供了精确的查询词频、每个查询单词对应的文件、多个文件是否存在共同关键字、用户查询的关键字顺序等。因此，与易受 CPA 攻击类似，在拥有某些先验信息的情况下，敌手可从统计信息中推断明文信息。例如，在得知用户兴趣模式的情况下，根据查询的词频推断出查询的信息与文件的内容主题信息。

导致该问题的来源为两个方面：a）与无法抵抗选择明文攻击相同，令牌的确定性与索引中匹配的确定性导致了查询模式得以积累统计（即 $w_i=w_j$ 的情况下，两者的令牌、索引中的匹配条目与返回的文件均相同）；b）即使令牌是概率性的（即敌手无法分辨两个由同一关键字产生的令牌 $t_1,t_2\leftarrow\mathrm{Token}_K(w)$，其中 $w$ 为相同的关键字），访问模式 $\alpha(H)=(D(w_1),\cdots,D(w_q))$ 也为敌手提供了每个关键字与文档的对应情况、文档之间关键字的包含情况等信息。

## 2 云环境下非对称可搜索加密技术安全机制

### 2.1 算法模型

在基于公钥机制的非对称可搜索加密方案（asymmetric searchable encryption，ASE）的云存储应用场景中，用户生成公钥与私钥，并将公钥公布给数据发送者（用于加密数据）与存储服务提供商（用于检索加密的数据）。数据发送者使用用户的公钥对数据进行加密，并发送加密数据到服务器；用户使用自己的私钥生成基于关键字的查询令牌与解密数据。

非对称可搜索加密方案最早由文献[9]提出。该方案利用双线性运算中指数的可互换性，通过将公钥与私钥加密的数据进行指数合并，使得只有通过公钥与私钥加密的相同关键字才能得到匹配。在该方案中，可搜索的加密结构称为 PEKS（public encryption with keyword search），其扩展定义与安全模型在文献[16]中给出；文献[17]增加了私有信息检索协议，给出了最强的安全模型。以下给出非对称可搜索加密方案基本算法模型。

**定义 8** 非对称可搜索加密方案算法模型。一个通用的非对称可搜索加密方案（ASE）为四个多项式时间算法 Gen、PEKS、Token、Test 的集合：

a）$K\leftarrow\mathrm{Gen}(1^k)$，为一个概率算法，用于生成密钥。输入安全参数 $k$，输出公钥/私钥对 $K=(K_{\mathrm{pub}},K_{\mathrm{priv}})$。算法由用户执行，私钥 $K_{\mathrm{priv}}$ 由用户保密。

b）$s\leftarrow\mathrm{PEKS}(K_{\mathrm{pub}},w)$，为一个概率算法，用于生成安全的可搜索结构。输入公钥 $K_{\mathrm{pub}}$ 与一个关键字 $w$，输出可搜索结构 $s$。算法由数据发送者执行，生成的可搜索结构 $s$ 链接到加密的消息（文档）后提交给服务器。

c）$t\leftarrow\mathrm{Token}(K_{\mathrm{priv}},w)$，为一个确定性算法，用于生成安全的查询令牌。输入私钥 $K_{\mathrm{priv}}$ 与一个关键字 $w$，输出查询令牌 $t$。算法由用户执行。

d）$b\leftarrow\mathrm{Test}(K_{\mathrm{pub}},s,t)$，为一个确定性算法，用于检测查询令牌与可搜索结构的匹配情况。输入公钥 $K_{\mathrm{pub}}$、可搜索结构 $s\leftarrow\mathrm{PEKS}(K_{\mathrm{pub}},w')$、查询令牌 $t\leftarrow\mathrm{Token}(K_{\mathrm{priv}},w)$，如果 $w=w'$ 则输出 $b=1$，否则输出 $b=0$。算法由服务器执行，服务器返回满足 $b=1$ 条件的对应文档。

### 2.2 安全模型

对于非对称加密算法而言，安全模型主要针对存储在服务器的 PEKS，其定义如下。

**定义 9** PEKS 语义安全性。令 $k\in\mathrm{N}$ 为安全参数，$A$ 为多项式时间敌手。考虑以下概率实验：

a）挑战者运行 $\mathrm{Gen}(1^k)$ 得到密钥 $K=(K_{\mathrm{pub}},K_{\mathrm{priv}})$，并将 $K_{\mathrm{pub}}$ 发送给敌手。

b）敌手可以自适应地向挑战者查询关键字。对于查询的任意关键字 $w$，敌手从挑战者处得到令牌 $t$。

c）在某个时候，敌手发送给挑战者两个查询的关键字 $w_0$、$w_1$。这里唯一的限制为该关键字在之前并没有被查询过。挑战者选择一个随机数 $b=\{0,1\}$，并返回给敌手 $s=\mathrm{PEKS}(K_{\mathrm{pub}},w_b)$，称 $s$ 为挑战的可搜索结构。

d）敌手可以继续查询除 $w_0$、$w_1$ 外的任意关键字。

e）最终，敌手输出判断 $b'=\{0,1\}$。

称非对称可搜索加密方案为语义安全的，当且仅当对于任意多项式时间敌手，满足概率 $Pr[b=b']\leqslant 1/2+\varepsilon$。其中，$\varepsilon$ 为一个可忽略小量，其概率基于概率函数 Gen 与 PEKS。

从安全模型中可以看出，存储在服务器上的数据是动态安全的，即允许敌手截获服务器上的所有数据（PEKS 与加密的数据），并允许敌手持续监控用户的所有查询。该场景很好地

模拟了实际的应用情况。

### 2.3　安全隐患

#### 2.3.1　易受字典攻击

字典攻击(dictionary attack,DA)[18]是指敌手利用一个常见关键字组成的字典,对目标进行离线测试攻击。由于敌手可以获得公钥,所以给定一个令牌 $t$,敌手可以使用用户的公钥产生指定的可搜索结构 $s \leftarrow \text{PEKS}(K_{\text{pub}}, w)$,并调用函数 $b \leftarrow \text{Test}(K_{\text{pub}}, s, t)$ 来判断猜测的关键字是否与用户的查询令牌匹配,如果匹配即找出了用户查询的关键字。例如,如果用户采用英语,则对于敌手而言,根据常用的 3 万个左右的英文单词做成字典并执行以上步骤,可以通过离线的字典攻击在极短的时间内获得用户查询的内容。

导致该问题的来源为两个方面:a)最初的安全模型中仅考虑了 PEKS 的安全性,没有考虑令牌的安全性,因此尽管存储在服务器上的 PEKS 和数据文件是静态安全的,然而用户一旦开始查询,则会导致查询令牌产生动态安全隐患;b)敌手通常能够获取用户的查询令牌(包括窃听者、存储服务提供商的内部攻击者等),因此在敌手能同时获取用户公钥的情况下,攻击者能够有效地执行加密—测试匹配算法来进行对猜测的判断。

#### 2.3.2　易受部分统计分析攻击

较对称可搜索加密算法而言,该类攻击的难度更大,因为令牌与 PEKS 都是概率加密的;即使是同一个关键字,加密后的每个结果都不一样。然而,以下两类统计信息可以被有效记录并被敌手分析。

a)部分查询模式 $\sigma(H)$。当两个不同的令牌 $t_i$、$t_j$ 对同一个 PEKS 满足条件 $1 \leftarrow \text{Test}(K_{\text{pub}}, s, t)$ 时,$t_i$、$t_j$ 对应的关键字是一致的,即查询模式 $\sigma(H)$ 满足 $w_i = w_j$。因此,对查询模式的统计分析攻击也是有效的,并能得到部分关键字的词频、顺序等信息。

b)访问模式 $\alpha(H)$。该问题与对称可搜索加密算法相同,即反映出多个匹配文件之间包含共同关键字的情况。由于记录访问模式信息的隐含条件为 $1 \leftarrow \text{Test}(K_{\text{pub}}, s, t)$,结合查询模式可以精确地确定多个匹配文件之间包含的共同关键字数量等信息。

## 3　扩展功能的安全性

### 3.1　多关键字查询与短语查询

文献[19~21]提出的可搜索加密算法支持连续的多关键字查询功能。安全的多关键字查询并非单关键字查询结果的简单组合(如两次单关键字查询结果取交集)。对于连续的多关键字检索而言,不仅需要保证文件与查询内容的安全性,同时要保证敌手无法从单次查询的统计信息中获得有用的信息,包括每个关键字的查询结果统计信息、查询的关键字数量、关键字之间的关系(如属于同一短语)等信息[22]。

文献[23]基于可信服务、文献[24]基于密文的二叉查找算法提出了短语查询功能,支持用户一次性查询一个短语,而不需要多次查询。该功能与一般的多关键字查询功能有些差别。例如,查询短语“cloud computing”要求不仅每一个关键字“cloud”与“computing”需要出现在每个返回的文档中(即等价于多关键字查询),其顺序也需要满足,即“cloud”后面要求跟随单词“computing”。

事实上,由于多关键字的查询结果增加了查询的精确性,减少了访问模式 $\alpha(H)$ 能记录的信息数量,安全性较单关键字查询更强。然而,易受的各类攻击与单关键字查询相同。

### 3.2　结果排序查询

基于保序加密(order preserving encryption)[25]算法的基本思想,文献[20,26~27]提出的算法支持对返回的文件按关键字匹配度(如基于关键字在文件中的词频)进行排序。在通常情况下,排序查询功能使得所有匹配的文档会根据一个标准进行排序,最终仅返回给用户最相关的 $k$ 个文档。

该功能引入的新安全隐患为泄露了结果文件对关键字的匹配度信息。即两个文件 $a$、$b$ 对关键字的匹配度满足 $a > b$,则对应的加密文件(返回的结果)的匹配度也满足 $a > b$,反之亦然。该问题存在的来源为保序加密算法的天然隐患,即可以通过密文的大小比较来推断明文之间的大小关系。这部分信息无疑将提高统计分析攻击的有效性。

### 3.3　模糊查询

基于模糊集的概念与多关键字检索算法的基本思想,文献[28~30]提出的算法支持对模糊数据集进行检索。该功能允许用户仅提交关键字的一个片断(该关键字可能不是一个有效的单词),服务器返回与该片断最接近的关键字所匹配的相关文档。同样基于模糊集的概念,文献[31,32]提出的算法支持服务器返回与检索关键字内容相近的文件。虽然所采用的具体方法有所不同,但其基本的数据结构和算法思路与模糊查询方案是一致的。

该功能的引入增加了文件之间关联信息的透露。根据统计信息,敌手能够基于模糊集中的距离大小(该距离信息可能为公共已知的)来确定密文之间的关联性大小,从而能够对密文进行有效的分类并能更精确地推测密文的内容。显然,该功能的引入减少了可搜索加密算法的安全性。

## 4　安全增强方案与存在的问题

安全增强方案的主要目的是为了弥补抗选择关键字攻击模型的安全强度的不足,即力图消除基本可搜索加密方案的安全隐患。综合而言,弥补的手段包括增加安全传输通道、增加匿名性、增加第三方代理和增加隐私保护协议。表 1 中列出了在选择关键字攻击(CKA)、字典攻击(DA)、统计分析攻击(SA)、选择明文攻击(CPA)四类攻击模型下,各类增强方案抵抗外部攻击(外)与存储服务提供商的内部攻击(内)的情况。

### 4.1　增加密文安全传输

增加密文安全传输包括两类方案:a)最基本的方式为加密用户与服务器间的传输数据[9],以对抗来自窃听者对可搜索加密结构与令牌的获取,如采用基于安全套接层(SSL)的传输;b)将服务器的公钥嵌入可搜索加密结构与令牌中,以在无安全传输通道的情况下,对抗来自窃听者的字典攻击等攻击手段[33~37]。第二类解决方案是当前研究的热点,其算法模型定义如下。

**定义 10**　基于密文安全传输的非对称可搜索加密方案算法模型。一个通用的基于密文安全传输的非对称可搜索加密方案为五个多项式时间算法 $\text{Gen}_{\text{sever}}$、$\text{Gen}_{\text{user}}$、PEKS、Token、Test 的集合。

表 1 基本方案与增强方案安全性一览

| 方案类型 | CKA 安全性 | | DA 安全性 | | SA 安全性 | | CPA 安全性 | |
|---|---|---|---|---|---|---|---|---|
| | 可搜索结构 | 令牌 | 可搜索结构 | 令牌 | 可搜索结构 | 令牌 | 可搜索结构 | 令牌 |
| 基本对称方案[8,11,13~15] | 外/内 | 外/内 | 外/内 | 外/内 | 无 | 无 | 外/内满足条件:无查询历史 | 外/内满足条件:无查询历史 |
| 基本非对称方案[9,16,17] | 外/内 | 外/内 | 外/内 | 无 | 无 | 无 | 外/内满足条件:仅能获取可搜索结构 | 外/内满足条件:仅能获取查询令牌 |
| 增加密文安全传输[9,33~37] | 外/内 | 外/内 | 外/内 | 外 | 外 | 外 | 外/内满足条件:无查询历史 | 外/内满足条件:无查询历史 |
| 增加匿名性[38~42] | 外/内 | 外/内 | 外/内 | 无 | 外/内满足条件:输入关键字具有不可预测性 | 外/内满足条件:输入关键字具有不可预测性 | 外/内满足条件:无查询历史 | 外/内满足条件:输入数据具有随机性 |
| 增加第三方代理[43~45] | 外/内 | 外/内 | 外/内 | 外/内 | 外/内 | 外/内 | 外/内 | 外/内 |
| 增加隐私保护协议[46~49] | 外/内 | 外/内 | 外/内 | 无 | 外/内 | 外/内 | 外/内满足条件:无查询历史 | 外/内满足条件:无查询历史 |

a) $SK \leftarrow \mathrm{Gen}_{\mathrm{sever}}(1^k)$,为一个概率算法,用于生成服务器密钥。输入安全参数 $k$,服务器(server)输出公钥/私钥对 $SK=(SK_{\mathrm{pub}},SK_{\mathrm{priv}})$。私钥 $SK_{\mathrm{priv}}$ 由服务器保密。

b) $UK \leftarrow \mathrm{Gen}_{\mathrm{user}}(1^k)$,为一个概率算法,用于生成用户密钥。输入安全参数 $k$,用户(user)输出公钥/私钥对 $UK=(UK_{\mathrm{pub}},UK_{\mathrm{priv}})$。私钥 $UK_{\mathrm{priv}}$ 由用户保密。

c) $s \leftarrow \mathrm{PEKS}(SK_{\mathrm{pub}},UK_{\mathrm{pub}},w)$,为一个概率算法,用于生成安全的可搜索结构。输入服务器公钥 $SK_{\mathrm{pub}}$、用户公钥 $UK_{\mathrm{pub}}$ 与一个关键字 $w$,输出可搜索结构 $s$。算法由数据发送者执行,生成的可搜索结构 $s$ 链接到加密的消息(文档)后提交给服务器。

d) $t \leftarrow \mathrm{Token}(SK_{\mathrm{pub}},UK_{\mathrm{priv}},w)$,为一个确定性算法,用于生成安全的查询令牌。输入服务器公钥 $SK_{\mathrm{pub}}$、用户私钥 $UK_{\mathrm{priv}}$ 与一个关键字 $w$,输出查询令牌 $t$。算法由用户执行。

e) $b \leftarrow \mathrm{Test}(SK_{\mathrm{priv}},UK_{\mathrm{pub}},s,t)$,为一个确定性算法,用于检测查询令牌与可搜索结构的匹配情况。输入服务器私钥 $SK_{\mathrm{priv}}$、用户公钥 $K_{\mathrm{pub}}$、可搜索结构 $s \leftarrow \mathrm{PEKS}(SK_{\mathrm{pub}},UK_{\mathrm{pub}},w')$、查询令牌 $t \leftarrow \mathrm{Token}(SK_{\mathrm{pub}},UK_{\mathrm{priv}},w)$,如果 $w=w'$ 则输出 $b=1$,否则输出 $b=0$。算法由服务器执行,服务器返回满足 $b=1$ 条件的文档。

从该类方案的算法模型中可以看出,由于可搜索结构与令牌与服务器的公钥相关,只有拥有私钥的服务器才能正确运算相关算法来判断密文与令牌的匹配性,从而可以有效抵抗字典攻击与统计分析攻击。显然,该类方案主要针对外部攻击者。由于服务器本身拥有部分私钥(服务器私钥),所以该类方案并不能防止来自存储服务提供商的内部攻击。目前可能的解决方案为采用可信计算的方式来增加对服务提供商的可信性,防止内部攻击。

### 4.2 匿名通信与匿名加密

字典攻击与统计分析攻击实现的前提是能够明确目标对象,即能确定可搜索加密结构或令牌的真实来源,并能准确统计每一次查询对应的数据与用户信息。基于匿名加密算法的思想[38],文献[39~42]提出了匿名可搜索加密方案,有效地避免了对密文的定位与对查询历史的统计。

该类方案的基本思想是采用基于匿名身份的加密(identity based encryption,IBE)系统,将身份 ID(在可搜索加密算法中为关键字)进行随机化(匿名化)。具体而言,给定一个匿名 IBE 系统 $\Pi=(\mathrm{Setup},\mathrm{KeyGen},\mathrm{Enc},\mathrm{Dec})$,与传统算法中产生一个与身份 ID 关联的密钥不同,方案中将密钥生成算法 KeyGen 改为:使用一个随机选择的种子 s,输入一个强随机数产生函数 ext()生成随机身份 $\mathrm{ID}_s=\mathrm{ext}(\mathrm{ID},s)$,并根据该随机身份输出密钥 $K_{\mathrm{ID}_s}$ 作为用户的秘密。显然,只要每次使用的身份 $\mathrm{ID}_s$ 具有随机性,则敌手即使获取密钥 $K_{\mathrm{ID}_s}$ 与种子 $s$ 都不会泄露用户的真实 ID。将该系统应用到非对称可搜索加密算法中,用户查询的关键字(即 ID)将具有匿名性,即服务器在观察用户查询的时候,无法得知用户查询的关键字是什么。显然,该类方案能有效抵抗来自外部与存储服务器提供商的内部统计攻击。

然而,具有高安全性的匿名可搜索加密算法要求关键字具有不可预测性[42](等价于在匿名 IBE 系统中,敌手无法猜测用户的真实 ID)。而在实际应用中,用户选择的关键字往往都是简单常见的内容,因此在特定的查询场景中,对敌手而言令牌依然具有可预测性(尤其是敌手获得部分先验信息时)。一种可能的解决方案为增加可信代理服务器来进行关键字的转换与模糊化。

### 4.3 基于第三方代理的概率令牌

文献[43~45]提出了基于第三方代理存储文件的对称可搜索加密方案。该类解决方案一般基于三方主体:用户、数据查询服务提供商(用于文件查询)、数据存储服务提供商(如存储索引与文件服务器)。典型的方案如文献[45]提出的令牌随机协议。在该查询协议中,用户提交加密 $N$ 次的令牌,然后索引服务器和文件服务器(数据存储服务提供商)更新令牌和文件 ID(再次加密)。因此,对于数据查询服务提供商而言,同样的关键字每次对应的令牌都不一样(加密系统本质上基于上次令牌的向前加密安全方案),因此无法对用户的查询历史进行有效的定位与统计分析,从而达到保护用户隐私的目的。

该类方案的问题较明显,即需要增加第三方参与交互,这无疑大大增加了系统的复杂性。通常而言,在云存储应用中,用户往往希望只与云服务提供商交互。因此,云服务提供商应该既是唯一被授权存储用户数据的地方,同时也是唯一被授权执行文件查询的地方。

### 4.4 基于私有信息检索协议的概率查询

私有信息检索(private information retrieval,PIR)[46~49]技术是一种隐秘获取目标数据的协议。通过该协议,用户可以从远程服务器的数据库中获取数据,且在过程中服务器不知道用户获取的是哪项数据,即用户可以从一个 $n$ bit 的数据库中隐秘地获取第 $i$ bit 的数据。

统计分析攻击的数据来源是用户的查询历史,主要为访问模式与查询模式。因此,如果采用私有信息检索协议来执行查询,则可以有效抵抗统计分析攻击。在基于该协议的可搜索加密算法中[17],协议一般同时用于两个方面:查询索引(隐藏查询模式)和文献获取(隐藏访问模式)。在协议的整个执行过程中,服务器无法得知用户查询的信息 $i$,而只有用户才能正确解读返回的结果信息并获得第 $i$ 条记录信息。

此类安全增强方案会极大增加系统的开销,应用性较弱,因此在文献[17]中作者也将之作为一种理论性的最强安全模型来讨论。例如在文件获取部分,一个具有 $N$ bit 的文件集每次获取可能会导致传输 $\sqrt{N}$ bit 的数据。具体的网络流量消耗与采用的私有信息检索协议相关。一般而言,随着目标大小的增加,网络流量为亚线性(sub-linear)增长。

## 5　应用陷阱与安全性挑战

综上所述,可搜索加密技术在实际应用中,安全性较传统密码系统弱,因此以下应用思路是存在安全隐患的。

陷阱 1)只要密钥安全,则数据安全

该思路通常来自对传统密码系统的认识,即认为只要通过加密数据,保护好密钥,存放在云端的数据即能保障其安全性。然而由前面的讨论可知,由于用户对存储的加密数据进行查询,而查询的模式与获取文件的访问模式是确定性的,所以易受外部攻击者或内部攻击者的统计分析攻击。此外,部分方案还易受字典攻击(如非对称可搜索加密算法)、对比攻击(增加结果排序的可搜索加密算法)等。而这些攻击者在具有某些先验信息的条件下,能有效推测出用户的查询内容乃至文件内容。解决该类问题的安全性挑战在于如何构造具有抗选择明文攻击(CPA)的可搜索加密结构与查询令牌。

陷阱 2)在服务器绝对可信的条件下,只要保障了数据传输安全,则数据是安全的

该思路来自当前云存储服务的典型应用模式(未采用可搜索加密算法),如各类云盘。这类应用的基本模式为云服务提供商与用户之间的数据传输通过安全通道(如 SSL),数据上传到云服务提供商后被加密并存储。用户需要查询文件时,服务器可以解密数据并提供检索服务。

可见,这类应用要求用户相信云服务提供商不存在内部攻击的现象,而加密的数据由服务器采用传统的加密算法进行保护。因此,即使数据被盗(典型的例子为当前较频繁的云存储数据泄露事件),用户数据也是安全的。然而,如果用户采用了可搜索加密算法,当数据被盗时,数据则不可避免地易受字典攻击(针对历史令牌)与统计分析攻击(针对日志)。事实上,一个良好的密码系统产生的密文(如 AES)的传输,不需要保护传输安全,也不需要担心密文的泄露。因此解决该问题的方案依然是增强可搜索加密算法的安全性。同时,从可搜索加密技术的发展目的而言,主要目标是针对云服务提供商,因此一般并不假定云服务提供商是绝对可信的。

陷阱 3)只要数据最初是安全的,则一直保持安全

该思路来源于对传统加密系统的认识。而对于可搜索加密算法,加密后的文件也会参与后期查询的运算,因此是动态的。未来任何一个环节的安全性缺陷都会给最初安全的数据带来危险,如易受针对查询历史的统计分析攻击。

因此,认识到可搜索加密技术的安全性问题是持续的至关重要,对于设计安全的可搜索加密算法需要从整个数据的生命周期与应用周期全面考虑。

陷阱 4)假定服务器是半可信的

绝大部分的可搜索加密算法都假定服务器是半可信的(honest-but-curious),换言之,即使是来源于服务器的内部攻击,也只是被动地观察用户的数据与查询过程,而不会主动进行数据更改等。

事实上,该思路违背传统密码系统的安全需求:敌手不仅会被动地观察数据,而且会主动地窜改、重放、消除数据。文献[50,51]讨论了在可搜索加密算法中有效鉴别服务器对数据窜改的方法,然而在可搜索加密技术领域,该类解决方案处于发展初期。该问题的安全性挑战在于,如何在正确执行用户请求但充满恶意攻击(honest-but-malicious)的数据外包环境下,保障用户数据的安全。

## 6　结束语

本文详细讨论了云环境下对称/非对称可搜索加密技术各类代表性方案的算法模型与安全模型,分析了各类方案在面对选择关键字攻击、字典攻击、统计分析攻击、选择明文攻击四种安全模型下的安全性,最后归纳总结了对可搜索加密技术的错误认识与在应用中不恰当使用该技术导致的安全隐患。综合而言,可搜索加密技术仍然面临巨大的安全挑战,需要有效解决安全隐患的方案,才能真正在实际中广泛应用。

**参考文献:**


[1] Fox A, Griffith R, Joseph A, *et al*. Above the clouds: a Berkeley view of cloud computing, UCB/EECS-2009-28[R]. Berkeley: Electrical Engineering and Computer Sciences Department, University of California, 2009.

[2] Chakraborty R, Ramireddy S, Raghu T S, *et al*. The information assurance practices of cloud computing vendors[J]. IT Professional, 2010, 12(4): 29-37.

[3] Takabi H, Joshi B D, Ahn G J. Security and privacy challenges in cloud computing environments[J]. IEEE Security & Privacy, 2010, 8(6): 24-31.

[4] Subashini S, Kavitha V. A survey on security issues in service delivery models of cloud computing[J]. Journal of Network and Computer Applications, 2011, 34(1): 1-11.

[5] Popovic K, Hocenski Z. Cloud computing security issues and challenges[C]//Proc of the 33rd International Convention Conference. 2010: 344-349.

[6] Coppersmith D. The data encryption standard (DES) and its strength against attacks[J]. IBM Journal of Research and Development, 1994, 38(3): 243-250.

[7] Daemen J, Rijmen V. The design of Rijndael: AES, the advanced encryption standard[M]. New York: Springer, 2002.

[8] Curtmola R, Garay J, Kamara S, *et al*. Searchable symmetric encryption: improved definitions and efficient constructions[C]//Proc of the 13th ACM Conference on Computer and Communications Security. New Yokr: ACM Press, 2006: 79-88.

[9] Boneh D, Di Crescenzo G, Ostrovsky R, *et al*. Public key encryption with keyword search[C]//Advances in Cryptology. Berlin: Springer, 2004: 506-522.

[10] 沈志荣,薛巍,舒继武. 可搜索加密机制研究与进展[J]. 软件学报, 2014, 25(4): 880-895.

[11] Kamara S, Papamanthou C, Roeder T. CS2: a searchable cryptographic cloud storage system, MSR-TR-2011-58[R]. [S.l.]: Microsoft Research, 2011.

[12] Waters B R, Balfanz D, Durfee G, *et al*. Building an encrypted and searchable audit log[C]//Proc of the 11th Annual Network and Distributed System Security Symposium. 2004: 5-6.

[13] Song D X, Wagner D, Perrig A. Practical techniques for searches on encrypted data[C]//Proc of IEEE Symposium on Security and Privacy. 2000: 44-55.

[14] Goh E J. Secure indexes, 2003/216[R]//Cryptology ePrint Archive. 2003.

[15] Chase M, Kamara S. Structured encryption and controlled disclosure [C]//Advances in Cryptology. Berlin: Springer, 2010: 577-594.

[16] Abdalla M, Bellare M, Catalano D, *et al*. Searchable encryption revisited: consistency properties, relation to anonymous IBE, and extensions[C]//Advances in Cryptology (CRYPTO). Berlin: Springer, 2005: 205-222.

[17] Boneh D, Kushilevitz E, Ostrovsky R, *et al*. Public key encryption that allows PIR queries[C]//Advances in Cryptology. Berlin: Springer, 2007: 50-67.

[18] Byun J W, Hyun A P, Dong H L. Off-line keyword guessing attacks on recent keyword search schemes over encrypted data[M]//Secure Data Management. Berlin: Springer, 2006: 75-83.

[19] Boneh D, Waters B. Conjunctive, subset, and range queries on encrypted data[M]//Theory of Cryptography. Berlin: Springer, 2007: 535-554.

[20] Cao Ning, Wang Cong, Li Ming, *et al*. Privacy-preserving multi-keyword ranked search over encrypted cloud data[J]. IEEE Trans on Parallel and Distributed Systems, 2011, 25(1): 222-233.

[21] Örencik C, Savas E. Efficient and secure ranked multi-keyword search on encrypted cloud data[C]//Proc of Joint EDBT/ICDT Workshops. New York: ACM Press, 2012: 186-195.

[22] Kerschbaum F. Secure conjunctive keyword searches for unstructured text[C]//Proc of the 5th International Conference on Network and System Security. 2011: 285-289.

[23] Zittrower S, Zou C C. Encrypted phrase searching in the cloud[C]//Proc of IEEE Conference and Exhibition on Global Telecommunications. 2012: 764-770.

[24] Tang Yinqi, Gu Dawu, Ding Ning, *et al*. Phrase search over encrypted data with symmetric encryption scheme[C]//Proc of the 32nd International Conference on Distributed Computing Systems[S. l.]: IEEE, 2012: 471-480.

[25] Agrawal R, Kiernan J, Srikant R, *et al*. Order preserving encryption for numeric data[C]//Proc of ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2004: 563-574.

[26] Wang Cong, Cao Ning, Li Jin, *et al*. Secure ranked keyword search over encrypted cloud data[C]//Proc of the 30th International Conference on Distributed Computing Systems. 2010: 253-262.

[27] Swaminathan A, Mao Y, Su G M, *et al*. Confidentiality-preserving rank-ordered search[C]//Proc of ACM Workshop on Storage Security and Survivability. New York: ACM Press, 2007: 7-12.

[28] Bösch C, Brinkman R R, Hartel P, *et al*. Conjunctive wildcard search over encrypted data[M]//Secure Data Management. Berlin: Springer, 2011: 114-127.

[29] Bringer J, Chabanne H. Embedding edit distance to allow private keyword search in cloud computing[M]//Secure and Trust Computing, Data Management and Applications. Berlin: Springer, 2011: 105-113.

[30] Li Jin, Wang Qian, Wang Cong, *et al*. Fuzzy keyword search over encrypted data in cloud computing[C]//Proc of INFOCOM. 2010: 1-5.

[31] Wang Cong, Ren Kui, Yu Shucheng, *et al*. Achieving usable and privacy-assured similarity search over outsourced cloud data[C]//Proc of INFOCOM. 2012: 451-459.

[32] Kuzu M, Islam M S, Kantarcioglu M. Efficient similarity search over encrypted data[C]//Proc of the 28th International Conference on Data Engineering. 2012: 1156-1167.

[33] Baek J, Reihaneh S N, Willy S. Public key encryption with keyword search revisited[C]//Computational Science and Its Applications. Berlin: Springer, 2008: 1249-1259.

[34] Hyun S R, Jong H P, Susilo W, *et al*. Trapdoor security in a searchable public-key encryption scheme with a designated tester[J]. Journal of Systems and Software, 2010, 83(5): 763-771.

[35] Hu Chengyu, Liu Pengtao. A secure searchable public key encryption scheme with a designated tester against keyword guessing attacks and its extension[C]//Advances in Computer Science, Environment, Ecoinformatics, and Education. Berlin: Springer, 2011: 131-136.

[36] Hyun R S, Jong H P, Lee D H. Generic construction of designated tester public-key encryption with keyword search[J]. Information Sciences, 2010, 205(1): 93-109.

[37] Yan W C, Phan R C, Heng S H, *et al*. Keyword guessing attacks on secure searchable public key encryption schemes with a designated tester[J]. International Journal of Computer Mathematics, 2013, 90(12): 2581-2587.

[38] BOYEN X, WATERS B. Anonymous hierarchical identity-based encryption (without random oracles)[C]//Advances in Cryptology. Berlin: Springer, 2006: 290-307.

[39] Gentry C, Peikert C, Vaikuntanathan V. Trapdoors for hard lattices and new cryptographic constructions[C]//Proc of the 40th Annual ACM Symposium on Theory of Computing. New York: ACM Press, 2008: 197-206.

[40] Agrawal S, Boneh D, Boyen X. Efficient lattice (H) IBE in the standard model[C]//Advances in Cryptology. Berlin: Springer, 2010: 553-572.

[41] Kurosawa K, Phong L T. Maximum leakage resilient IBE and IPE, 2011/628[R]//CryptologyePrint Archive. 2011.

[42] Boneh D, Ananth R, Gil S. Function-private identity-based encryption: hiding the function in functional encryption[C]//Advances in Cryptology. Berlin: Springer, 2013: 461-478.

[43] Bao Feng, Deng R H, Ding Xuhua, *et al*. Private query on encrypted data in multi-user settings[M]//Information Security Practice and Experience. Berlin: Springer, 2008: 71-85.

[44] DongG Changyu, Giovanni R, Naranker D. Shared and searchable encrypted data for untrusted servers[J]. Journal of Computer Security, 2011, 19(3): 367-397.

[45] Choi D, Hyun S K, Lee Y. Address permutation for privacy-preserving searchable symmetric encryption[J]. ETRI Journal, 2012, 34(1): 66-75.

[46] Chor B, Kushilevitz E, Goldreich O, *et al*. Private information retrieval[J]. Journal of the ACM, 1998, 45(6): 965-981.

[47] Ostrovsky R, Shoup V. Private information storage[C]//Proc of the 29th Annual ACM Symposium on Theory of Computing. New York: ACM Press, 1997: 294-303.

[48] Ostrovsky R, Skeith W E. A survey of single-database private information retrieval: techniques and applications[M]//Public Key Cryptography. Berlin: Springer, 2007: 393-411.

[49] Yekhanin S. Private information retrieval[J]. Communications of the ACM, 2010, 53(4): 68-73.

[50] Kurosawa K, Ohtaki Y. UC-secure searchable symmetric encryption[M]//Financial Cryptography and Data Security. Berlin: Springer, 2012: 285-298.

[51] Chai Qi, Gong Guang. Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers[C]//Proc of IEEE International Conference on Communications. 2012: 917-922.