云计算中基于属性的可搜索加密电子病历系统

李晓蓉 宋子夜 任婧怡 徐 磊 许春根
(南京理工大学理学院 南京 210094)

摘 要 电子病历是医学报告在云计算技术迅速发展下的一个重要产物,它的出现方便了医院和患者对病历的管理。然而,患者的相关隐私数据存储在云上,就必然面临着隐私泄露、非法访问等隐患。为保护存储在云端的电子病历数据的私密性,提出了一个基于属性的可搜索加密方案,并给出了它在电子病历系统中的重要应用。与传统的可搜索加密方案相比,本方案降低了多用户环境下密钥管理的难度,且方案中的陷门可以在非安全信道上传输。此外,该方案可隐藏访问结构,具有细粒度访问控制,可根据数据拥有者的请求增加或撤销用户的访问权限。安全性分析表明,该方案保护了关键词的隐私性,可抵抗关键词猜测攻击,能有效防止隐私数据的泄露。关键词陷门匹配算法仅需一次双线性对运算,大大提高了搜索效率。

关键词 电子病历,可搜索加密,细粒度访问控制,关键词猜测攻击
中图法分类号 TP309 **文献标识码** A

Attribute-based Searchable Encryption of Electronic Medical Records in Cloud Computing

LI Xiao-rong SONG Zi-ye REN Jing-yi XU Lei XU Chun-gen
(School of Science, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract Electronic medical record is an important product of the medical report under the rapid development of cloud computing technology, whose presence is convenient for the hospitals and patients to manage the medical reports. However, it would be inevitably faced with potential hazards such as the privacy disclosure, illegal access, etc. if the privacy data about patients are stored in the cloud. To protect the privacy of the electronic medical records stored in the cloud, an attribute-based searchable encryption scheme was proposed, and its important application in the electronic medical record system was given. Compared with the traditional searchable encryption scheme, this scheme does not need a secure channel for the trapdoor transmitting, and the difficulties in key management for multiple users could be reduced. Moreover, the scheme can hide the access structure and has fine-grained access control to add and revoke access rights of users according to the request of data owners. Security analysis shows that the scheme can not only protect the privacy of keywords, but also effectively resist the attack of keyword guessing and prevent the private data from leaking out. The keyword-based trapdoor matching algorithm requires only one computation of bilinear pairing, greatly improves the searching efficiency.

Keywords Electronic medical record, Searchable encryption, Fine-grained access control, Attack of keyword guessing

1 引言

1.1 背景介绍

随着电子医疗系统和云计算的迅速发展,许多用户开始将数据迁移到云端服务器,以避免繁琐的本地数据管理并获得更加便捷的服务。对于存储在云中的数据,用户并不希望云服务中心的员工能够随意查询用户自己的信息,此时数据的安全性就变得尤为重要。为保证数据安全和用户隐私,一般是将数据以密文的形式存储在云端服务器中,但是用户将会遇到如何在密文上进行查找的难题。可搜索加密是近年发展起来的一种支持用户在密文上进行关键字查找的密码学原语,它能够为用户节省大量的网络和计算开销,并充分利用云端服务器庞大的计算资源进行密文上的关键字查找。

1.2 相关工作

可搜索加密机制最早由Song等人[1]提出,它采用对称可搜索加密技术实现了加密数据的高效搜索,但是该模型只允许持有私钥的用户对数据进行加密和搜索,在数据频繁交流的今天,这种模式显然不能满足人们的需求。可搜索加密过程如图1所示。针对公钥系统加密数据的安全、高效搜索问题,Boneh等人[2]于2004年开创性地提出带关键词的公钥可搜索加密方案,利用基于关键词的公钥可搜索加密技术和双线性对实现了对使用公钥加密的密文数据的搜索功能,并首次证明了其在随机预言模型下的安全性。2005年,针对基于身份密码体制的局限性,Goyal等人[3]首次提出了基于属性的加密体制,采用访问结构来控制系统中用户的解密和签名,实现了细粒度、灵活的访问策略和高效的属性撤销操作。

本文受江苏省自然科学基金(BK20141405),南京理工大学本科生科研训练'百千万'计划国家级项目立项资助。
李晓蓉(1995-),女,主要研究方向为信息与计算科学;**宋子夜**(1996-),女,主要研究方向为信息与计算科学;**任婧怡**(1997-),女,主要研究方向为信息与计算科学;**徐 磊**(1990-),男,博士,主要研究方向为公钥密码学;**许春根**(1969-),男,博士,教授,CCF会员,主要研究方向为信息安全与密码技术应用,E-mail:xuchung@njust.edu.cn(通信作者)。

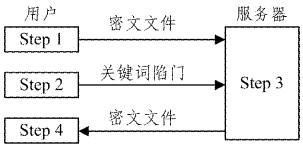

图 1　可搜索加密过程

2009 年，针对病人与其数据的隐私性和安全性问题，文献[4]提出了一个由病人控制加密、保护个人隐私的可搜索电子病历系统，采用可搜索加密技术实现了对病人信息的安全存储及高效搜索，但是该系统需要高昂的密钥管理成本，检索操作需多次交互且效率不高。2013 年，针对访问控制的灵活性问题，Kulvaibhavh 等人[5]采用密文策略并基于属性加密技术构造了基于 CP-ABE(密文策略属性基加密系统)的可搜索加密方案，实现了为授权用户提供细粒度访问控制的功能，但该方案在没有加密的情况下直接将密钥上传给云服务器，存在密钥泄露的问题。

王少辉等人[6]于 2014 年针对在密文上进行搜索的权限问题，采用基于身份的可搜索加密技术设计了一个高效的 dIBEKS(对指定测试者的基于身份可搜索加密)新方案，但该方案不能实现一对多的共享功能。文献[7]针对一对多的共享功能的需求，采用基于属性的可搜索加密技术，提出了两种在公钥可搜索加密和属性加密下基于属性加密的关键词搜索和数据访问控制方案，但是其无法抵御关键词猜测攻击。林鹏等人[8]针对关键词猜测攻击，提出了一种指派搜索服务器的关键词搜索属性加密技术，该技术实现了关键词搜索访问权限控制，可抵抗离线或在线关键词猜测攻击，但其方案的复杂度较高。

李双等人[12]针对一对多的共享通信问题，利用文献[13]中提出的访问结构定义和基于属性的加密技术，构造了一个基于属性的可搜索加密方案，但是该方案要求在安全信道下传输陷门，若要达到在安全信道内传输陷门的目的，则会大大增加通信服务的成本。刘全明等人[10]于 2016 年在李双方案的基础上进行改进，基于将私钥和关键词再加密的技术，构造了基于属性的关键字搜索加密，其可防止关键词明文泄露，并且关键词在传输过程不再需要安全信道，但在交互过程中加入了一个多项式算法，增加了算法复杂度。陈燕俐等人[11]于 2016 年针对目前用户在检索密文时直接将自身密钥提交给服务器而存在密钥泄露这一问题，采用向服务器发送门限值使密钥盲化的技术，构造了一个可支持属性撤销的基于 CP-ABE 的可搜索加密方案，保证了用户密钥的保密性和安全性。文献[16]针对应用场景问题，将 CP-ABE 应用于电子病历系统，实现了方案的实用性。

2017 年 Qiu[9]针对关键词攻击的问题，通过隐藏访问策略的技术提出了一种可抵抗关键字猜测攻击的隐藏策略属性可检索加密方案。然而，该方案需要数据拥有者与数据用户共享密钥，且在加密关键词的过程中没有更换密钥，降低了方案的安全性。

1.3　本文的贡献

针对目前电子病历多用户系统存在的问题，本文提出了一种基于属性的可搜索加密系统(MOD-ATT-PEKS)，其特点如下：

(1)计算复杂度低。加密算法、解密算法、关键词陷门匹配算法仅需要 1～2 次双线性对运算，降低了数据拥有者和查询者的计算要求。

(2)细粒度访问控制。本方案采用访问控制树的结构，不同文件有不同的访问策略，加密算法与访问控制相关联，只有满足相应属性的用户才拥有检索权限。

(3)密钥安全性高。在本方案中，每位用户拥有不同的属性，这些属性与用户的私钥相关联，即不同的用户拥有独立的私钥，在多用户场景下避免了共享密钥的情形，降低了密钥泄露带来的风险。

(4)安全性高。本方案基于 q-ABDHE 问题的困难性假设，可抗关键词猜测攻击，保证了用户私钥安全以及关键词陷门安全。

(5)关键词隐蔽性。云服务器中的关键词以密文形式存储，用户以关键词陷门的形式向云服务器提交关键词信息，在可搜索功能实现的交互过程中，服务器无法获知关键词明文信息。

本方案与其他方案的功能对比如表 1 所列。

表 1　方案功能的对比

方案	不需要安全信道	密钥生成中心	细粒度访问控制	访问权限撤销	抵抗关键词猜测	是否可以实现共享
文献[10]	是	有	是	无	否	是
文献[15]	是	有	是	无	是	是
文献[12]	否	有	是	无	否	是
本方案	是	有	是	是	是	是

2　理论基础

2.1　双线性对定义

定义 1　设 G 和 G_T 是两个阶数为 p 的乘法循环群。g 是 G 的生成元。称 $e:G\times G\rightarrow G_T$ 是一个双线性映射，若映射 $e:G\times G\rightarrow G_T$ 满足如下性质：

(1)双线性：对于任意的 $u,v\in G$ 和 $a,b\in Z_p$，有 $e(u^a,u^b)=e(u,v)^{ab}$。

(2)非退化性：$e(g,g)\neq 1$。

(3)可计算性：存在有效的多项式时间算法，对任意的 $u,v\in G$ 计算 $e(u,v)$ 的值。

2.2　困难性假设

本方案的安全性是基于判定性增广的双线性 Diffie-Hellman 困难性假设(q-ABDHE)，其具体定义如下。

定义 2(q-ABDHE 问题)[14]　任给一个 $2q+2$ 维元组 $(g',g'^{(\alpha^{q+2})},g,g^{\alpha},g^{(\alpha^2)},\cdots,g^{(\alpha^q)},g^{(\alpha^{q+2})},\cdots,g^{(\alpha^{2q})})\in G^{2q+2}$，计算 $e(g,g')^{(\alpha^{q+1})}\in G_T$ 的值。

为将 q-ABDHE 问题用于 MOD-ATT-PEKS，用另一种观点来看 q-ABDHE 问题。

定义 3[14]　在 q-ABDHE 问题中，省略元素 $g^{(\alpha^{q+2})},\cdots,g^{(\alpha^{2q})}$ 得到缩短性 q-ABDHE[14] 问题，求解缩短性 q-ABDHE 问题与求解 q-ABDHE 问题的复杂度相同。

假设算法 $\mathcal{A}$ 具有优势 ε 求解缩短性 q-ABDHE 问题，即

$$\Pr[\mathcal{A}(g',g'_{q+2},g,g_1,\cdots,g_q)=e(g_{q+1},g')]\geqslant\varepsilon$$

其中，$g_i=g^{(\alpha^i)}$，$g'_i=g'^{(\alpha^i)}$，g,g' 是 G 中随机生成的元素，α 是 $\mathbb{Z}_p$ 中随机生成的元素。

算法 $\mathcal{B}$ 输出 $b\in\{0,1\}$，且具有优势 ε 求解缩短性 q-ABDHE 问题，即

$$|\Pr[\mathcal{B}(g',g'_{q+2},g,g_1,\cdots,g_q,e(g_{q+1},g'))=0]-\Pr[\mathcal{B}(g',g'_{q+2},g,g_1,\cdots,g_q,Z)]=0|\geqslant\varepsilon$$

其中，g,g'是G中随机生成的元素，α是$\mathbb{Z}_p$中随机生成的元素，Z是G_T中随机生成的元素。记左端的概率为P_{ABDHE}，右端的概率为R_{ABDHE}。

定义 4[14] 称定义在G上的缩短性(t,ε,q)-ABDHE问题是难解的，若不存在多项式时间算法可以在t时间内以至少优势ε求解在G上的缩短性q-ABDHE问题。

2.3 访问结构与访问树

定义 5(访问结构[13]) 访问结构的工作原理类似于秘密共享，定义授权访问子集可以重构秘密，非授权访问子集不可以重构秘密。

设$P=\{P_1,P_2,\cdots,P_n\}$是一个实体集，$A\subseteq 2^P$是2^P上的一个非空子集，$\forall B,C$，若当$B\in A$且$B\subseteq C$时，有$C\in A$，则称$A\subseteq 2^P$是单调的。一个访问结构A(通常指单调的)是$P=\{P_1,P_2,\cdots,P_n\}$的一个非空子集，即$A\subseteq 2^P\backslash\{\Phi\}$。包含于$A$中的集合被称为授权集合，否则称为非授权集合。

定义 6(访问树 Γ)[12] 树Γ表示一个访问结构，每个非叶子结点都由子结点个数和陷门值表示，记num_x为子结点的个数，k_x为陷门值且满足$0<k_x<num_x$。当$k_x=1$时，门限表示“或”门；当$k_x=num_x$时，门限表示“与”门。每个叶子结点都表示一个属性，并且其陷门值$k_x=1$。

为叙述方便，定义如下函数：

(1)$parent(x)$表示结点x的父节点。

(2)$att(x)$表示与x关联的属性值，其中x是叶子结点。

(3)对树Γ中每个结点的子结点定义了顺序，按1到num的顺序编号，函数$index(x)$返回与x相关联的编号。

定义 7(满足访问树[12]) 设访问树Γ的根结点为r，记Γ_x表示根结点为x的子树，则Γ也可表示成Γ_r。若属性集合ω满足访问树结构Γ_x，则记为$\Gamma_x(\omega)=1$。可以采用递归的方式计算$\Gamma_x(\omega)$，若x是非叶子结点，计算x的每个子结点x'的$\Gamma_{x'}(\omega)$，当且仅当至少有k_x个子节点返回1时，$\Gamma_x(\omega)=1$；若x是叶子结点，当且仅当$att(x)\in\omega$时，$\Gamma_x(\omega)=1$。

例如，在图2所示的访问控制树结构中，主治医生为电子病历系统的数据拥有者，主治医生上传患者的病历之后，满足如下属性的三类人可以对其进行访问：本医院的医生、患者本人以及本科室内的护士长。若用户拥有的属性不满足上述三类中的任何一种，则其无权访问这份病历。

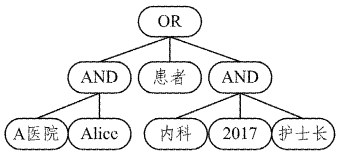

图2 访问结构

2.4 基于属性的可搜索加密算法的定义

定义 8[12] 基于属性的可搜索加密方案主要由5个概率多项式算法组成，即全局参数生成算法$Setup$、密钥生成算法$KeyGen$、关键词加密算法$PEKS$、关键词陷门生成算法$Trapdoor$、判断算法$Test$。

$Setup(1^\lambda):(Pub,Msk)\leftarrow Setup(1^\lambda)$

由系统参数λ生成公共参数Pub和主密钥Msk。

$KeyGen(Msk,A):Priv_A\leftarrow KeyGen(Msk,A)$

密钥生成中心根据系统主密钥Msk和访问结构A生成用户的私钥$Priv_A$。

$PEKS(Pub,w,\gamma):C\leftarrow PEKS(Pub,w,\gamma)$

由系统公共参数Pub、关键词w和属性γ生成密文C，其中发送方通过Pub和γ加密关键词w，从而生成可搜索的关键词密文C，且只有属性γ满足访问结构A的用户才有权对C进行搜索。

$Trapdoor(Priv_A,w):T_w\leftarrow Trapdoor(Priv_A,w)$

由用户私钥$Priv_A$和关键词w生成关键词陷门T_w，其中接收方利用个人私钥$Priv_A$计算查询关键词$w\in\{0,1\}^*$的陷门值T_w并发送给网关服务器。

$Test(Pub,T_w,C):b\leftarrow Test(Pub,T_w,C)$

判断密文C是否是陷门T_w对应的关键词密文，若已知公共参数Pub、关键词w'的密文$C=PEKS(Pub,w',\gamma)$和关键词w的陷门值$T_2=Trapdoor(Priv_A,w)$，如果$w'=w$，则输出$b=1$，否则$b=0$。

为了通过一致性，设对于所有的$\lambda\in N$、属性γ和关键词$w\in\{0,1\}^*$，有：

$$\Pr[Test(Pub,Trapdoor(Priv_A,w),PEKS(Pub,w,\gamma))=1]=1$$

该函数取遍所有公共参数Pub、主密钥Msk和概率算法。

2.5 攻击游戏

本节定义在选择属性集选择关键词模型下的基于属性的可搜索加密方案的安全性，需要确保算法$PEKS(Pub,w,\gamma)$在不提供陷门T_w的情况下不会泄露任何关于关键词w的信息。基于属性的可搜索加密方案的安全目标是在属性集合模型下实现选择关键词密文攻击的不可区分性，即假设一个活跃的攻击者可以得到他选择的任何关键词w的陷门T_w，即使在这样的攻击下，攻击者也无法区分他未询问过的关键词w_0和关键词w_1的陷门。由此，在攻击者选择他想挑战的属性集γ并将此告知挑战者的前提下，定义在选择属性集选择关键词模型下挑战者和攻击者之间的攻击游戏[2]如下：

Setup：利用基于属性的可搜索加密方案的$Setup$算法产生公共参数Pub并将其发给攻击者。

Phase1：攻击者选定访问结构A_j进行陷门查询，其中$\forall j,\gamma\notin A_j$，攻击者可随机选择关键词$w\in\{0,1\}^*$并得到其陷门$T_w$。

Challenge：攻击者向挑战者提交两个他想挑战的关键词w_0和w_1，唯一的限制是攻击者不可以提前得知陷门T_{w_0}和T_{w_1}，挑战者挑出一个随机数$b\in\{0,1\}$并将其加密后的密文$C=PEKS(Pub,w,\gamma)$发送给攻击者。

Phase2：攻击者重复$Phase1$的操作，继续询问任意关键词w的陷门，其中$w\neq w_0,w_1$。

Guess：最后，攻击者$\mathcal{A}$输出$b'\in\{0,1\}$，当$b=b'$时，攻击者获胜。

具体来说，攻击者如果能正确猜出他得到的$PEKS$是关键词w_0或w_1，则获胜。定义$\mathcal{A}$在攻击游戏中的优势为：

$$Adv_{MOD\text{-}ATT\text{-}PEKS}=\left|\Pr[b'=b]-\frac{1}{2}\right|$$

定义 9 称基于属性的可搜索加密方案在选择属性集选择关键词模型下是安全的，如果对任意多项式攻击时间的攻击者，破解方案的优势$Adv_{MOD\text{-}ATT\text{-}PEKS}$是可忽略的。

3　电子病历系统下的 MOD-ATT-PEKS 算法的构造

本节主要描述基于 Gentry 等人[14]的基于身份的加密方案而提出的 MOD-ATT-PEKS 方案（Modified Attribute Based PEKS)在电子病历系统中的应用。电子病历系统包含 4 个实体：密钥分发中心、云存储中心、数据拥有者和数据访问者。其中，密钥分发中心是一个可信的第三方服务器，负责生成公开参数、生成和分发密钥、增加或撤销用户访问权限等；云存储中心主要负责数据的存储和检索；数据拥有者拥有增加或召回其文件访问者的权限；数据访问者拥有一些指定的属性集合以及相应的密钥，满足属性要求时可以访问数据。系统中算法的运行过程如下：

定义一个以 g 为生成元的 p 阶乘法群 $G,e:G\times G\to G_T$ 为其上的双线性映射。

$Setup(n)$：输入属性集包含元素的个数 n，随机选取 $y,\alpha\in\mathbb{Z}_p,g',t_1,\cdots,t_{n+1}\in G$，令 $g_0=g^y,g_1=g^\alpha$，输出公共参数和主密钥：

$$Pub=(g_0,g_1,g',t_1\cdots,t_{n+1}),Msk=(y,\alpha)$$

定义拉格朗日系数[9] $\Delta_{i,s}(X)=\prod_{j\in s,j\neq i}\frac{x-j}{i-j}$，其中 $i\in\mathbb{Z}_p$，S 是由 $\mathbb{Z}_p$ 中元素构成的一个集合。令 $i,j\in\{1,\cdots,n+1\}$，定义函数 T：

$$T(X)=g'^{X^n}\prod_{i=1}^{n+1}t_i^{\Delta i,N(X)}$$

密钥分发中心执行 $Setup(n)$，生成公共参数 Pub，为用户生成和分发主密钥 Msk。

$KeyGen(Pub,Msk,\Gamma)$：首先将用户的所有属性对应到 $\mathbb{Z}_p^*$ 中的唯一元素，并按下述方法构造访问树 Γ[11]：从根结点开始，为访问树中每个非叶子结点 x 选取一个多项式 q_x，其阶数 d_x 应满足 $d_x=k_x-1$，k_x 为该结点的阈值。对于根结点 r，令 $q_r(0)=y$，而 q_r 在其他 d_r 个点随机选取；在非根结点 x 处，$parent(x)$ 表示访问树 Γ 中 x 的父节点，$q_x(0)=q_{parent}(index(x))$，其他 d_x 个点的值随机定义。若满足 $\Gamma(\gamma)=1$，则用户可解密由属性集 γ 加密的消息。

算法从 $\mathbb{Z}_p$ 中为用户 x 随机选取 r_x，为每个叶子结点计算 D_x 和 R_x：$D_x=g'^{q_x(0)}\cdot T(i)^{r_x},i\leftarrow att(x),R_x=g^{r_x}$，并返回 $(\alpha,\{Priv_x\}_{x\in\Gamma})$ 给用户，其中 $Priv_x=(D_x,R_x)$。

用户注册账号时，密钥分发中心执行 $KeyGen(Pub,Msk,\Gamma)$ 算法，由公共参数 Pub、主密钥 Msk 及用户的属性 γ 生成含有用户属性的密钥。

$PEKS(Pub,w,\gamma)$：输入关键词 w，从 $\mathbb{Z}_p$ 中随机选取元素 s，计算 $t=g^{(\alpha+w)s}$，并据此获得用属性集 γ 标记的关键词密文 $E=(\gamma,t,E',E'',E''',\{E_i\}_{i\in\gamma})$：

$$E=(\gamma,t,e(gg_0,g')^s,g^s,e(g,g)^s,\{E_i=T(i)^s\}_{i\in\gamma})$$

用户基于公共参数 Pub、所拥有的指定属性集合 γ 和关键词 w，根据算法 $PEKS(Pub,w,\gamma)$ 得到关键词密文 E。

$Trapdoor(Priv_x,w)$：用户 x 输入密钥 $Priv_x$ 和关键词 w，算法随机选取 $\mathbb{Z}_p$ 上的元素 r，生成关键词陷门 $T_w=(U,r,D_x,R_x)$：

$$T_w=((g'g^{-r})^{\frac{1}{\alpha+w}},r,D_x,R_x)$$

并将此陷门 T_w 返回给用户 x。

接着，用户根据 $Trapdoor(Priv_x,w)$ 算法生成关键词陷门 T_w（即搜索令牌），并将此令牌发送给云存储中心。

$Test(T_w,E,x)$：输入 T_w,E,x，输出判断值 $b,b\in\{0,1\}$。作如下计算：

关于关键词陷门、密文、以及结点 T_w,E,x，令 $i=att(x)$。

若 x 为叶子结点，定义如下函数：

$$F(E,T_w,x)=\begin{cases}\frac{e(D_x,E'')}{e(R_x,E_i)}=e(g,g')^{s\cdot q_x(0)},i\in\gamma\\ \perp,\text{其他}\end{cases}$$

若 x 为非叶子结点时，作如下递归运算：

对于结点 x 的所有子结点 z，记函数 $F(E,T_w,x)$ 的输出值为 N_z。令所有子结点 z 中满足 $N_z\neq\perp$ 且大小为 k_x 的结点的集合为 S_x，若结点 $z\notin S_x$，输出 $N_z=\perp$；否则令 $i=index(x),S_x'=\{index(z):z\in S_x\}$，且：

$$\begin{aligned}N_x&=\prod_{z\in S_x}N_z^{\Delta i,S_x'(0)}\\&=\prod_{z\in S_x}(e(g,g')^{s\cdot q_z(0)})^{\Delta i,S_x'(0)}\\&=\prod_{z\in S_x}(e(g,g')^{s\cdot q_{parent(index(z))}})^{\Delta i,S_x'(0)}\\&=\prod_{z\in S_x}(e(g,g')^{s\cdot q_x(0)})^{\Delta i,S_x'(0)}\\&=e(g,g')^{s\cdot q_x(0)}\end{aligned}$$

下面进行基于属性集的密文搜索运算。

如果 x 结点的属性 $\Gamma(\gamma)=1$，则计算：

$$F(E,T_w,x)=e(g,g')^{ys}=e(g_0,g)^s$$

若有 $B:=\frac{E'}{F(E,T_w,x)}=\frac{e(U,t)}{(E''')^{-r}}$ 成立，则输出 $b=1$，否则 $b=0$。

云存储中心接受用户的查询请求后，根据陷门 T_w 和查询关键词的密文 E 以及用户 x 的属性，由 $Test(T_w,E,x)$ 算法判断用户是否满足加密文件的访问树。

上述电子病历系统中的交互过程如图 3 所示。

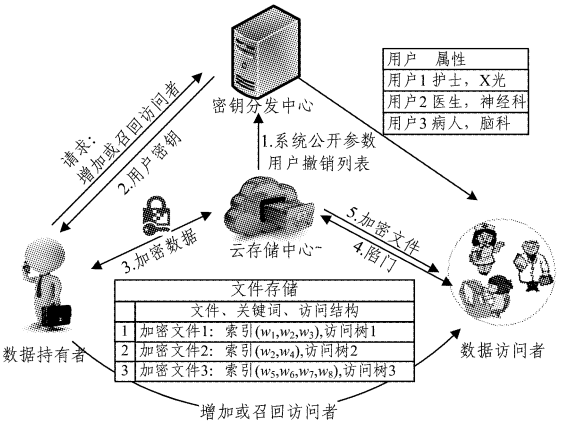

图 3　电子病历系统模型

3.1　MOD-ATT-PEKS 计算一致性

对于结点 x，当 $i\in att(x)\in\gamma$ 时，由算法中的定义得：

$$B=\frac{E'}{F(E,T_w,x)}=\frac{e(g,g_2)^s e(g_0,g_2)^s}{e(g_0,g_2)^s}=e(g,g_2)^s$$

另一方面，由 T_w 和 E''' 可计算：

$$\frac{e(U,t)}{E'''^{(-r)}}=\frac{e((g_2g^{-r})^{\frac{1}{\alpha+\omega}},g^{(\alpha+\omega)s})}{E'''^{(-r)}}=\frac{e(g_2g^{-r},g)^s}{E'''^{(-r)}}$$

$$=\frac{e(g_2,g)^s e(g^{-r},g)^s}{e(g,g)^{-rs}}=e(g_2,g)^s=B$$

因此，当 $\frac{e(U,t)}{E'''^{(-r)}}=B$ 时，说明关键词陷门 Γ_ω 与密文对应同一个关键词，且用户的属性满足 $i\in att(x)\in\gamma$。

3.2　MOD-ATT-PEKS 算法的复杂度

MOD-ATT-PEKS 方案的复杂度分析结果如表 2 所列。

设 B 表示双线性对运算，E 代表指数运算，H 代表 Hash 函数运算。记属性集 γ 的元素个数为 n，PEKS/Crypt 表示加密关键词密文的算法，Test/Decrypt 表示解密关键词陷门值的算法，Trapdoor 表示计算关键词陷门值的算法。

表 2 MOD-ATT-PEKS 方案运算量统计表

Algorithm	Operator	文献[10]	文献[12]
PEKS/Crypt	B	3	2
	E	n+4	n+3
	H	1	2
Test/Decrypt	B	0	1
	E	1	0
	H	1	1
Trapdoor	B	1	0
	E	2	1
	H	0	1
Algorithm	Operator	文献[15]	MOD-ATT-PEKS
PEKS/Crypt	B	2	2
	E	2n+2	n+4
	H	3	0
Test/Decrypt	B	2	1
	E	1	1
	H	1	0
Trapdoor	B	0	0
	E	1	1
	H	1	0

由表 2 可知，MOD-ATT-PEKS 方案在实现基本功能的前提下，减少了计算量，改变了文献[10-12]中使用 Hash 函数加密的方式，避免了冲突的发生，减少了数据存储量。相较于其他基于属性的可搜索加密方案，MOD-ATT-PEKS 方案实现的功能更为全面，更能适应电子病历系统中多用户场景的需要。本方案能保护明文信息不被泄露，可在密文上进行关键词查找且可抵抗关键词猜测攻击，实现了加密数据的数据共享，基本满足了当前实际应用的需要。

4 安全性证明

定理 1 假设缩短性 (t,ε,q)-ABDHE 困难问题在 (G,G_T,e) 上是难解的，则 MOD-ATT-PEKS 是 (t',ε',q') 选择属性集在选择关键词攻击下是不可区分性安全，其中 $t'=t-O(t_{exp}\cdot m^2)$，$m=\max\{q,n\}$，$\varepsilon'=\varepsilon+\frac{2}{p}$，$t_{exp}$ 表示在 G 中取幂的次数。

证明：假设攻击者 $\mathcal{A}$ 在选择属性集选择关键词模型下以 ε' 的优势攻击 MOD-ATT-PEKS，我们构造一个算法 $\mathcal{B}$ 可以解决缩短性 q-ABDHE 问题。设 $G=\langle g\rangle$，已知 $g',g''=g'^{a^{q+2}}$，$g_i=g^{a^i}$，目标是输出 $e(g_{q+1},g')\in G_T$，真正的挑战者随机选取 $b',c'\in\{0,1\}$，输入 $(g',g'',g,g_1,g_2,\cdots,g_q,Z)$。当 $\mathcal{B}$ 输出 0 时，$Z=e(g_{q+1},g')$；当 $\mathcal{B}$ 输出 1 时，随机选取 $\mathbb{Z}_p\in G_T$。算法 $\mathcal{B}$ 模拟挑战者与攻击者 $\mathcal{A}$ 之间的交互过程如下：

Init：攻击者 $\mathcal{A}$ 选择属性集 γ（包含 $\mathbb{Z}_p$ 中 n 个元素的集合），算法 $\mathcal{B}$ 令 $g_0=g^y$，$g_1=g^\alpha$，$t_i=g'^{u(i)}g^{f(i)}$，其中 $i=1,2,\cdots,n+1$，$f(x)$ 是随机选取的 n 次多项式。$u(x)$ 定义如下：

$$\begin{cases}u(x)=-x^n, & \forall x\in\gamma\\ u(x)\neq -x^n, & \forall x\notin\gamma\end{cases}$$

而 t_i 是随机独立选取的，因此有：

$$T(i)=g'^{i^n+u(i)}g^{f(i)}$$

Setup：算法 $\mathcal{B}$ 随机生成 q 次多项式 $f(x)\in\mathbb{Z}_p[x]$，令 $g'=g^{f(\alpha)}$，且 g' 可由 $g,g_1,g_2,\cdots,g_q$ 计算得出，并将元组 $(g_0,g_1,g',t_1,\cdots,t_{n+1})$ 发送给敌手 $\mathcal{A}$。

Phase1：攻击者 $\mathcal{A}$ 询问一些属性集 γ 不满足的访问结构。假设 $\mathcal{A}$ 请求访问结构满足 $\Gamma(\gamma)=0$ 的密钥。为生成密钥，算法 $\mathcal{B}$ 为每个访问树的非叶子结点指定一个阶数为 d_x 的多项式 Q_x。

设 $PolyUnsat(\Gamma_x,\gamma,g^{\lambda_x})$ 是一个确定访问树中节点 x 的多项式算法[12]（属性 γ 不满足以 x 为根节点的访问树，即 $\Gamma_x(\gamma)=0$），该算法访问树 Γ_x（x 为根节点），属性集 γ 和 $g^{\lambda_x}\in G_1$ 为输入，其中 $\lambda_x\in\mathbb{Z}_p$，输出访问树中结点 x 的多项式。首先定义结点 x 的次数为 d_x 的多项式 q_x，使得 $q_x(0)=\lambda_x$；由于 $\Gamma_x(\gamma)=0$，结点 x 存在少于 d_x 个的子节点满足属性，令 $h_x\leqslant d_x$ 为 x 中满足属性集的子结点数。对于 x 的每个满足属性集 γ 的子结点 x'，随机选取 $\lambda_{x'}\in\mathbb{Z}_p$，令 $q_x(index(x'))=\lambda_{x'}$，然后随机选取 q_x 中剩余的 d_x-h_x 个点，直到 q_x 被完全确定。依照此方法递归地定义访问树中不满足属性 γ 的点。在 $g^{q_x(0)}$ 已知的情况下，$g^{q_x(index(x'))}$ 可以通过插值法得到，且 $q_{x'}(0)=q_x(index(x'))$。

模拟器通过运行 $PolyUnsat(\Gamma_x,\gamma,g^{\lambda_x})$ 为访问树的每个结点定义多项式。对于每个结点 x，若 x 满足属性集，则 q_x 能完全确定；若 x 不满足属性集，则至少知道 $g^{q_x(0)}$。模拟器为每个结点 x 定义 $Q_x(\cdot)=q_x(\cdot)$，$y=Q_r(0)=\alpha^{q+1}$，每个叶子结点对应的密钥按照如下方式给出：

令 $i=att(x)$，

If $i\in\gamma$

$$D_x=g_2^{Q_x(0)}T(i)^{r_x}=g_2^{q_x(0)}T(i)^{r_x},R_x=g^{r_x}$$

其中，随机选取 $r_x\in\mathbb{Z}_p$。

If $i\notin\gamma$

$$D_x=g^{\frac{-f(i)}{i^n+u(i)}}(g_2{}^{i^n+u(i)}g^{f(i)})^{r_{x'}}$$

$$R_x=g_3^{\frac{-1}{i^n+u(i)}}g^{r_{x'}}$$

其中，随机选取 $r_{x'}\in\mathbb{Z}_p$。

$i^n+u(i)$ 没有非零点，$\forall i\notin\gamma$，设 $r_x=r_{x'}-\frac{q_x(0)}{i^n+u(i)}$，则

$$\begin{aligned}D_x&=g_3^{\frac{-f(i)}{i^n+u(i)}}(g_2^{i^n+u(i)}g^{f(i)})^{r_{x'}}\\&=g^{\frac{-q_x(0)\cdot f(i)}{i^n+u(i)}}(g_2^{i^n+u(i)}g^{f(i)})^{r_{x'}}\\&=g_2^{q_x(0)}(g_2^{i^n+u(i)}g^{f(i)})^{\frac{-q_x(0)}{i^n+u(i)}}(g_2^{i^n+u(i)}g^{f(i)})^{r_{x'}}\\&=g_2^{q_x(0)}(g_2^{i^n+u(i)}g^{f(i)})^{r_{x'}-\frac{q_x(0)}{i^n+u(i)}}\\&=g_2^{q_x(0)}(T(i))^{r_x}\\&=g_2^{Q_x(0)}T(i)^{r_x}\end{aligned}$$

$$R_x=g_3^{\frac{-1}{i^n+u(i)}}g^{r_{x'}}=g^{r_{x'}-\frac{q_x(0)}{i^n+u(i)}}=g^{r_x}$$

因此，算法 $\mathcal{B}$ 可以构造具有访问结构 Γ 的密钥。除此之外，访问结构 Γ 的密钥的分配方式与原方案相同。

定义 $q-1$ 阶多项式 $L(x)=\frac{f(x)-f(\omega)}{x+\omega}$，攻击者 $\mathcal{A}$ 询问关键词 w 的陷门值 T_w，算法 $\mathcal{B}$ 返回 $(f(\omega),g^{L(\alpha)})$ 作为 (r,U)。

$$g^{L(\alpha)}=g^{\frac{f(\alpha)-f(\omega)}{\alpha+\omega}}=(g'g^{-f(\omega)})^{\frac{1}{\alpha+\omega}}$$

综上，给出关键词 w 的陷门值：

$$T_w=(g^{L(\alpha)},f(\omega),g_2^{Q_x(0)}T(i)^{r_x},g^{r_x})$$

Challenge：攻击者 $\mathcal{A}$ 将提交两个挑战关键词 w_0,w_1。算法 $\mathcal{B}$ 随机选择 $b\in\{0,1\}$，给出 w_b 的密文：

$C=(\gamma, E'=e(g,g_2)^s Z, E''=g^s, E'''=e(g,g)^s, \{E_i=T(i)^s\}_{i\in\gamma})$

其中，Z 的定义同前文。

Phase2：重复 Phase1，攻击者 $\mathcal{A}$ 可以询问关键词 w_0, w_1 以外的陷门。

Guess：攻击者 $\mathcal{A}$ 输出猜测 $b', c'\in\{0,1\}$。若 $b=b'$ 且 $c=c'$，$\mathcal{B}$ 输出 0(表示 $Z=e(g_{q+1},g')$)，否则输出 1。

概率分析[2]：当 $Z=e(g_{q+1},g')$ 时，说明仿真实验可行，攻击者 $\mathcal{A}$ 对 (b,c) 猜测正确的概率为 $\frac{1}{4}+\varepsilon'$。此外，Z 均匀随机。当 $(g',g',g,g_1,\cdots,g_{q+1},Z)$ 是从 R_{ABDHE} 抽取时，$|\Pr[\mathcal{B}(g',g'',g,g_1,\cdots,g_{q+1},Z)=0]-\frac{1}{4}|\leqslant\frac{p}{2}$；而当 $(g',g'',g,g_1,\cdots,g_{q+1},Z)$ 是从 p_{ABDHE} 中抽取时，$|\Pr[\mathcal{B}(g',g'',g,g_1,\cdots,g_{q+1},Z)=0]-\frac{1}{4}|\geqslant\varepsilon$。

因此，对于均匀随机的 g,g',α 和 Z，有：

$$|\Pr[\mathcal{B}(g',g'',g,g_1,\cdots,g_{q+1},e(g_{q+1},g'))=0]-\Pr[\mathcal{B}(g',g'',g,g_1,\cdots,g_{q+1},Z)=0]|\geqslant\varepsilon'-\frac{2}{p}$$

时间复杂度：仿真实验中，$\mathcal{B}$ 的运算消耗主要来源于在 $\mathcal{A}$ 的陷门生成查询中对 $g^{P(x)}$ 的计算，其中 $P(x)$ 是 $q-1$ 阶多项式或 n 阶多项式，且 $g^{P(x)}$ 运算的时间复杂度为 $O(m)$ 次，其中 $m=\max\{q,n\}$；而 $\mathcal{A}$ 最多需要 m 次询问，因此时间复杂度 $t=t'+O(t_{\exp}\cdot m^2)$。

定理 1 证毕。

结束语　本文针对云计算环境下的电子病历系统提出了一种改进的基于属性的可搜索加密方案，该方案隐藏了访问结构，且在关键词搜索加密的权限控制中不需要安全信道，能抵抗关键词猜测攻击，它既实现了用户对加密后的关键词的检索，也让所有满足属性结构要求的用户能通过私钥解密文档；同时，在 q-ABDHE 假设下证明了 MOD-ATT-PEKS 方案是安全的，并给出了本方案的算法复杂度分析。与原方案相比，本文使用群上双线性运算替代 Hash 函数，在保证其安全性的同时降低了计算复杂度，提高了检索的效率。本文将 MOD-ATT-PEKS 方案与电子病历场景相结合，这具有很强的实用性和现实意义。MOD-ATT-PEKS 方案虽然降低了计算复杂度，但是用户的密钥传输需要安全信道，因此如何在降低计算复杂度的同时保证密钥传输的安全性是今后需要进一步解决的问题。

参 考 文 献

[1] SONG D, WAGNER D, PERRI G. A Practical techniques for searches on encrypted data [C] // Proceedings of IEEE Symposium on Securityand Privacy. Berkeley, 2000: 44-55.

[2] BONEH D, CRESCENZO G D, OSTROVSKY R, et al. Public key encryption with keyword search [C] // Proceedings of the EUROCRYPT'04. Interlaken, Switzerland, 2004: 506-522.

[3] 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, USA, 2006: 89-98.

[4] BENALOH J, CHASE M, HORVITZ E, et al. Patient Controlled Encryption: Ensuring Privacy of Electronic Medical Records [C] // Proceeding of the ACM Workshop Cloud Computing Security (CCSW'09). 2009: 103-114.

[5] KULVAIBHAV K, VIJAYARAGHAVAN V, RAJARATHNAMN. Multi-users Attribute Based Searchable Encryption [C] // Proceeding of International Conference on Mobile Data Management. Milan: IEEE Computer Society, 2013: 200-205.

[6] 王少辉，韩志杰，肖甫，等. 指定测试者的基于身份的可搜索加密方案[J]. 通信学报，2014，35(7)：22-32.

[7] LI J, ZHANG L. Attribute-based keyword search and data access control in cloud [C] // 2014 Tenth International Conference on Computational Intelligence and Security (CIS). IEEE, 2014: 382-386.

[8] 林鹏，江颉，陈铁明. 云环境下关键词搜索加密算法研究[J]. 通信学报，2015，36(Z1)：259-265.

[9] QIU S, LIU J, SHI Y, et al. Hidden policy ciphertext-policy attribute-based encryption with keyword search against keyword guessing attack [J]. Science China Information Sciences, 2017, 60(5): 052105.

[10] 刘全明，赵宝，高富强. 基于属性的可搜索加密方案[J]. 山西大学学报(自然科学版)，2016，39(4)：593-600.

[11] 陈燕俐，杨华山. 可支持属性撤销的基于 CP-ABE 可搜索加密方案[J]. 重庆邮电大学学报(自然科学版)，2016，28(4)：545-554.

[12] 李双，徐茂智. 基于属性的可搜索加密方案[J]. 计算机学报，2014，37(5)：1017-1024.

[13] BEIMEL A, et al. Secure schemes for secret sharing and key distribution [R]. Technion-Israel Institute of Technology, Faculty of Computer Science, 1996.

[14] GENTRY C. Practical identity-based encryption without random oracles [C] // Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer Berlin Heidelberg, 2006: 445-464.

[15] 郭璐璐. 对密文查询和基于查询的可搜索加密方案的研究[D]. 南京：南京理工大学，2014.

[16] 杨旸，林柏钢，马懋德. 具有细粒度访问控制的隐藏关键词可搜索加密方案[J]. 通信学报，2013，34(Z1)：92-100.

[17] HAN F, QIN J, ZHAO H, et al. A general transformation from KP-ABE to searchable encryption [J]. Future Generation Computer Systems, 2014, 30: 107-115.

[18] BONEH D, BOYEN X. Efficient selective-ID secure identity-based encryption without random oracles [C] // International Conference on the Theory and Applications of Cryptographic Techniques. Springer Berlin Heidelberg, 2004: 223-238.

[19] BONEH D, BOYEN X, GOH E J. Hierarchical identity based encryption with constant size ciphertext [C] // Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer Berlin Heidelberg, 2005: 440-456.

[20] BONEH D, GENTRY C, WATERS B. Collusion resistant broadcast encryption with short ciphertexts and private keys [C] // Annual International Cryptology Conference. Springer Berlin Heidelberg, 2005: 258-275.

[21] BONEH D, KATZ J. Improved efficiency for CCA-secure cryptosystems built using identity-based encryption [C] // Cryptographers' Track at the RSA Conference. Springer Berlin Heidelberg, 2005: 87-103.

[22] 马春光，石岚，汪定. 基于访问树的属性基签名算法[J]. 电子科技大学学报，2013，42(3)：410-414.

[23] 李经纬，贾春福，刘哲理. 可搜索加密技术研究综述[J]. 软件学报，2015，26(1)：109-128.