·网络空间安全·　　文章编号：1000-3428(2022)07-0159-09　　文献标志码：A　　中图分类号：TP309.2

基于国密SM9的可搜索加密方案

张　超[1,2]，彭长根[1,2,3]，丁红发[4]，许德权[1,2]

（1.贵州大学 计算机科学与技术学院 公共大数据国家重点实验室，贵阳 550025；2.贵州大学 密码学与数据安全研究所，贵阳 550025；3.贵州大学 贵州省大数据产业应用发展研究院，贵阳 550025；4.贵州财经大学 信息学院，贵阳 550025）

摘　要：为满足密文数据安全级别的要求，现有基于身份的可搜索加密方案多次使用安全参数较大的对称双线性对运算，导致计算效率降低，且其密钥形式难以与国家商用密码算法SM9相结合。针对该问题，设计一种基于SM9密码算法的可搜索加密方案。在离散椭圆曲线的两个子群中分别生成用户的公私钥对，使方案的密钥形式与SM9密码算法保持一致，解决经SM9密码算法加密后数据的检索问题，同时结合SM9密码算法，基于非对称双线性特性在确保方案安全性的同时提高检索效率。根据双线性对的性质分析该方案的正确性和安全性，并验证其满足在随机谕言模型下的适应性密文不可区分性和陷门不可区分性。仿真结果表明，与EdIBEKS、PEAKS、dIBAEKS方案相比，该方案在索引生成算法、陷门生成算法和检索匹配算法上的计算效率分别平均提高了77%、16.67%、28%以上。

关键词：可搜索加密；双线性对；密文数据；SM9密码算法；安全性证明

开放科学（资源服务）标志码（OSID）：

中文引用格式：张超，彭长根，丁红发，等.基于国密SM9的可搜索加密方案[J].计算机工程，2022，48(7)：159-167.

英文引用格式：ZHANG C，PENG C G，DING H F，et al. Searchable encryption scheme based on China state cryptography standard SM9[J].Computer Engineering，2022，48(7)：159-167.

Searchable Encryption Scheme Based on China State Cryptography Standard SM9

ZHANG Chao[1,2]，PENG Changgen[1,2,3]，DING Hongfa[4]，XU Dequan[1,2]

（1.State Key Laboratory of Public Big Data，College of Computer Science and Technology，Guizhou University，Guiyang 550025，China；2.Institute of Cryptography and Data Security，Guizhou University，Guiyang 550025，China；3.Institute of Guizhou Big Data Industries Application and Development，Guizhou University，Guiyang 550025，China；4.College of Information，Guizhou University of Finance and Economics，Guiyang 550025，China）

【Abstract】To meet the requirements of the security level of ciphertext data，existing identity-based Searchable Encryption（SE）schemes use symmetric bilinear pairing with a considerable number of security parameters for many times，which results in a reduced computational efficiency and difficulty in combining their key form with the national commercial cryptographic algorithm，SM9.To solve this problem，a SE scheme based on the SM9 cryptographic algorithm is designed.Public-private key pairing of users are generated in the two subgroups of the discrete elliptic curve to ensure consistency between the scheme key form and the SM9 cryptographic algorithm，and solve the problem of data retrieval after encryption by the SM9 cryptographic algorithm.At the same time，combined with the SM9 cryptographic algorithm，based on the asymmetric bilinear feature，the security of the scheme is ensured，and its retrieval efficiency is improved.According to the properties of the bilinear pairing，the correctness and security of the scheme are verified，and the adaptive ciphertext indistinguishability and trapdoor indistinguishability under the random oracle model are satisfied.Simulation results show that，compared with EdIBEKS，PEAKS，and dIBAEKS，the computational efficiency of the scheme in the index generation algorithm，trapdoor generation algorithm，and search matching algorithm is improved by 77%，16.67%，and 28%，respectively.

【Key words】Searchable Encryption（SE）；bilinear pairing；ciphertext data；SM9 cryptographic algorithm；security proof

DOI：10. 19678/j. issn. 1000-3428. 0062771

基金项目：国家自然科学基金（U1836205）；贵州省科技计划基金（黔科合平台人才[2020]5017）；贵州省教育厅自然科学项目（黔教合KY字[2021]140）。

作者简介：张　超（1990—），男，硕士研究生，主研方向为密码学、密文检索；彭长根，教授、博士；丁红发（通信作者），副教授、博士；许德权，博士。

收稿日期：2021-09-26　　**修回日期：**2021-12-25　　**E-mail：**28568854@qq.com

计算机工程 www.ecice06.com

0 概述

云存储技术的快速发展使得人们存储、管理大量数据更加便捷,但是数据面临泄露风险。为保护云服务器中的数据隐私,云存储技术通常对原始数据进行加密,把加密后的密文数据上传至云服务器,但是会给使用数据的用户带来不便。例如,当用户需要检索数据时,需要将全部密文数据下载到本地,在本地解密之后再执行检索操作。为更好地解决云数据的安全问题,可搜索加密(Searchable Encryption,SE)技术应运而生,并得到研究人员的广泛关注。SE技术允许云服务器在密文场景下进行安全检索,并将满足检索条件的密文数据返回给用户,最后由用户在本地将检索结果解密,从而获得自己需要的明文数据。

本文结合SM9算法构造基于身份的可搜索加密方案。根据非对称双线性对的特性,同时结合SM9加密算法的密钥形式,设计相应的索引生成算法、陷门生成算法和服务器检索匹配算法。基于可证明安全理论,验证该方案在随机谕言模型下满足密文不可区分性和陷门不可区分性。

1 相关工作

文献[1]首次提出对称可搜索加密(Symmetric Searchable Encryption,SSE)的概念,解决在密文环境下的搜索难题,实现了在对称密码体制下密文的搜索。文献[2]基于布隆过滤器(Bloom Filter,BF)[3]设计满足安全索引的SSE方案,将检索的时间复杂度改进为与文档的数目呈线性关系。SSE的计算效率较高,适用于大部分第三方存储。文献[4]提出公钥可搜索加密(Public-key Encryption with Keyword Search,PEKS)方案,并基于文献[5]利用双线性对设计一个具体构造,适用于复杂的加密环境。PEKS不用在发送者与接收者之间建立安全信道,因其实用性较强,而得到快速发展,其中基于身份加密的可搜索加密减少了建立和维护公钥基础设施(Public Key Infrastructure,PKI)的开销,因此成为该领域的研究热点。

文献[6]提出带关键字搜索的基于身份加密(Identity-Based Encryption with Keyword Search,IBEKS)方案,并给出一般化的转换方法,将任意的选择明文攻击模式下不可区分安全且匿名的IBE方案转换为选择明文攻击模式下不可区分安全的PEKS方案。文献[7]提出一种基于身份的关键字搜索加密方案,不需要安全通道,但要求服务器是可信的。文献[8]提出带关键字搜索的阈值公钥加密(Threshold Public-key Encryption with Keyword Search,TPEKS)和带分布式私钥生成器的基于身份匿名加密。文献[9]提出模糊关键词检索,可以防范内部敌手和外部敌手的关键词猜测攻击。文献[10]提出指定服务器的带关键字搜索的基于身份加密方案,该方案不需要额外的安全信道。文献[11]对可搜索加密进行探讨和总结。文献[12]提出一种具有隐藏结构的可搜索公钥密文(Searchable Public-key Ciphertexts with Hidden Structure,SPCHS),提高了关键字搜索效率。文献[13]对加密数据的关键字搜索技术进行研究,并对他们的效率和安全性进行分析。文献[14]提出基于公钥并且带身份验证功能的PEAKS方案。文献[15]结合格理论提出一种基于身份的可搜索加密方案,可以抵抗量子计算机的攻击。文献[16]设计基于身份的关键词可搜索加密的广播加密方案,以防范内部敌手的攻击。文献[17]提出基于身份的指定服务器用于加密邮件的可搜索认证加密方案。文献[18]提出支持代理重加密的基于身份可搜索加密方案,支持搜索权限的高效共享。文献[19]提出基于身份的动态可搜索加密方案,可以删除指定身份的文件。文献[20]提出在电子邮件系统中指定服务器的关键字搜索加密方案。文献[21]提出一种基于生物识别身份的多关键字搜索(Biometric Identity-Based Multi-Keyword Search,BIBMKS)机制。文献[22]提出身份认证权威辅助的基于身份可搜索加密,在保持效率和安全性的同时,减少存储空间的开销。

上述方案为达到一定的安全级别,多是基于安全参数较大的对称双线性对,且在方案中多次使用双线性对运算,导致方案的计算效率降低。因此,通过优化方案设计,提高可搜索加密方案的检索效率和安全性成为研究热点。

从国家信息安全方面考虑,发展我国自己的加密算法势在必行。2016年,基于身份的SM9密码算法为国家商用密码行业标准(GM/T 0044—2016),并于2018年11月被纳入国际标准。SM9密码算法的密钥长度为256 bit,采用运算速度快、安全性能高的R-ate双线性对[23],可以有效减少Miller算法循环次数,提高计算效率。近年来,SM9密码算法作为基于身份的密码算法,受到越来越多的关注。

SM9密码算法使用安全参数较小的非对称双线性对,用户的公私钥分别在两个循环群中产生,而现有的可搜索加密方案大多通过安全参数较大的对称双线性对来实现,计算效率相对较低[24]。由于用户密钥的形式不同,因此现有的可搜索加密方案并不能很好地与SM9密码算法相结合。

SM9密码算法在邮件系统、数据安全传输协议、物联网安全平台等场景中都有重要的应用,但由于缺少适应性强的可搜索加密方案,用户难以对这些加密数据进行检索,导致这些系统中的加密数据可用性较低。此外,通过为每个用户提供一对额外的密钥,以实现可搜索加密,但繁琐的操作流程会给用户的使用带来不便。为解决该问题,结合以上对SM9密码算法的分析,本文设计基于SM9密码算法的可搜索加密(SM9-based Searchable Encryption,SM9SE)方案。方案中用户的公私钥对分别在不同的子群中生成,其形式与SM9密码算法的密钥对一致,因此用户可以用同一对密钥完成数据的加解密和密文检索操作,并且使用系统参数相对较小的非

计算机工程 www.ecice06.com

对称双线性对运算，从而提高计算效率。

2　基础知识

2.1　双线性对

假设n为正整数，$(G_1,+)$和$(G_2,+)$是两个n阶加法循环群，其零元分别记作O_1和O_2。又设(G_T,g)为n阶乘法循环群，其单位元记为1_T。假设在群G_1，G_2，G_T上计算离散对数是困难的。

定义1　如果一个二元函数$\hat{e}:(G_1\cdot G_2)\to G_T$，满足双线性和非退化性。

双线性是对$\forall P_1、P_2\in G_1$，$Q\in G_2$，都有$\hat{e}(P_1+P_2,Q)=\hat{e}(P_1,Q)\cdot\hat{e}(P_2,Q)$成立，对$\forall P\in G_1$，$Q_1、Q_2\in G_2$，都有$\hat{e}(P,Q_1+Q_2)=\hat{e}(P,Q_1)\cdot\hat{e}(P,Q_2)$成立。

非退化性是对$\forall P\in G_1$，$P\neq O_1$，$\exists Q\in G_2$，使得$\hat{e}(P,Q)\neq 1_T$，$\forall Q\in G_2$，$Q\neq O_2$，$\exists P\in G_1$，使得$\hat{e}(P,Q)\neq 1_T$，则称二元函数$\hat{e}$为双线性对。

2.2　基于双线性对的困难性问题及假设

双线性DH（DBDH）问题判定：在一个双线性映射$\hat{e}:(G_1\cdot G_2)\to G_T$中，若$P_1\in G_1$，$P_2\in G_2$，$Z\in G_T$，对于给定的元组$Y=(P_1,[a]P_1,[b]P_1,P_2,[b]P_2,[c]P_2,Z)$，以及$\beta\in\{0,1\}$，若$Z=\hat{e}(P_1,P_2)^{abc}$，则令$\beta=0$，否则令$\beta=1$，判定$\beta$的值，其中$a$、$b$、$c$为未知的整数。

DBDH假设：对于任何一个概率多项式时间敌手A，解决DBDH问题的概率都是可忽略的，即$\Pr[0\leftarrow A(Y)\,|\,\beta=0]-\Pr[0\leftarrow A(Y)\,|\,\beta=1]$是可忽略的。

3　方案构造与安全模型

3.1　方案构造

本文方案使用的椭圆曲线及曲线参数与SM9加密算法中所使用的参数一致。

本文SM9SE方案由系统参数初始化算法、密钥生成算法、索引生成算法、陷门生成算法和检索匹配算法构成。

1）系统参数初始化算法SysSetup(λ)，输入安全参数λ以生成一个基于循环群上的双线性对映射$\hat{e}:(G_1\cdot G_2)\to G_T$，其中群$G_1$、$G_2$和$G_T$的阶为大素数$q$。PKG随机选择整数$s\in[1,q-1]$作为系统主密钥保密存储，然后计算系统公钥$P_{\text{pub}}=sP_1$，$P_1$为循环群$G_1$的生成元，$P_2$为循环群$G_2$的生成元。此外，PKG选择$H_1:\{0,1\}^*\to Z_q^*$，$H_2:\{0,1\}^*\to Z_q^*$，$H_3:G_T\to\{0,1\}^*$三个哈希函数。系统主密钥对为$(s,P_{\text{pub}})$，系统公共参数为$\text{Params}=(G_1,G_2,G_T,P_1,P_2,\hat{e},P_{\text{pub}},q,H_1,H_2,H_3)$。

2）密钥生成算法UserKeyGen(ID,s)，对于用户A的唯一身份标识ID_A，以及系统主私钥s，PKG在有限域F_N上计算$t_1=H_1(\text{ID}_\text{A})+s$。若$t_1=0$则重新计算系统主密钥并更新已有用户的私钥，否则计算$t_2=s\times t_1^{-1}$。用户私钥$d_\text{A}=[t_2]P_2$，用户公钥的计算方法为$Q_\text{A}=[H_1(\text{ID}_\text{A})]P_1+P_{\text{pub}}=[t_1]P_1$。

3）索引生成算法BuildIndex(ID_R,w)，对于接收者R的唯一身份标识ID_R以及关键字w。该算法随机选择$r\in[1,q-1]$，计算并生成关键字对应的密文索引$I=(I_1,I_2,I_3)$，并将其发送给服务器。其中$I_1=[r]Q_\text{R}$，$I_2=[r]P_2$，$I_3=H_3(\hat{e}(H_2(w)\cdot I_2,P_{\text{pub}}))$。

4）陷门生成算法Trapdoor(w',d_R)，文件接收者R运行该算法以生成检索陷门，输入要搜索的关键字w'和其私钥d_R。该算法选择随机数$t\in[1,q-1]$，然后计算陷门信息$T=(T_1,T_2)$，并将其发送给服务器，其中$T_1=[t]P_{\text{pub}}$，$T_2=[H_2(w')-t]\cdot d_\text{R}$，并且$H_2(w')-t\neq 0$，否则重新选取$t$。

5）检索匹配算法Test(I,T)，由服务器执行，给定安全密文索引$I=(I_1,I_2,I_3)$和陷门$T=(T_1,T_2)$，服务器执行该算法来判断接收到的陷门信息与其存储的关键词密文索引信息是否能正确匹配。如果匹配，则返回密文文件，若$I_3=H_3(\hat{e}(T_2,I_1)\cdot\hat{e}(T_1,I_2))$，则匹配成功。

3.2　安全模型

本节通过以下两个游戏来刻画本文方案在面对内部敌手的离线关键词猜测攻击（Keyword Guessing Attack，KGA）时的语义安全。由于内部敌手所获取的信息大于等于任意外部敌手所能获取的信息，因此不再考虑外部敌手。这两个游戏均由敌手A和挑战者B交互完成。游戏1是密文不可区分性游戏，游戏2是陷门不可区分性游戏。

3.2.1　密文不可区分性游戏

在密文不可区分性游戏中，假设半诚实服务器就是敌手A，它会诚实执行用户预设的操作，但是会尝试获取用户数据的信息。密文不可区分性旨在防止敌手分辨出某个给定的密文是从两个选定的关键词（由敌手自己选定）中的哪一个加密而来，即密文不可区分性可以保证服务器在没有获得用户授权的情况下无法私自对密文进行检索操作。

1）初始化。挑战者B生成系统的公共参数Params和主私钥s，以及各个用户的私钥，然后将系统的公共参数Params发送给敌手A。

2）询问阶段1。敌手A可以适应性地向如下的随机谕言机进行多项式时间次的询问，以获得用户的私钥、密文索引和陷门密文信息。其中，“适应性”是指敌手A每次的询问都会根据当前已知的信息来决定应该询问的信息。挑战者B运行随机谕言机，并按如下规则返回询问结果：（1）Hash谕言机O_H，对于给定的输入值，给敌手A返回其哈希值；（2）密钥谕言机O_E，对于指定的用户，给敌手A返回该用户的密钥；（3）索引谕言机O_C，给定关键词及接收方ID，给敌手A返回其对应的密文索引；（4）陷门谕言机O_T，给定一个关键词，计算并给敌手A返回其对应的陷门。

3）挑战。敌手A向挑战者B声明一个发送者的身份标识ID^*，并将其作为挑战，然后选择两个长度相等的挑战关键字(w_0,w_1)。挑战者B随机选择$\beta\in\{0,1\}$，根据选择结果，运行索引生成算法BuildIndex(ID_R,w_β)，生成密文索引I_β，并将其发送给敌手A。

计算机工程 www.ecice.com

4)询问阶段2。敌手A继续向各个随机谕言机进行询问,该阶段的询问与询问阶段1相同。

5)猜测。敌手A输出一个字节$\beta'\in\{0,1\}$作为对关键字w_β的猜测,假如$\beta=\beta'$,则敌手A赢得游戏。前提是敌手没有询问过ID*对应的私钥,也没有询问过与身份ID*对应的关键字w_0和w_1的陷门密文信息及索引密文信息。

敌手A在该游戏中的优势定义为:

$$\mathrm{Adv}_{\mathrm{A}}^{C}(\lambda)=\left|\Pr[\beta=\beta']-\frac{1}{2}\right|$$

若敌手A赢得游戏的优势可以忽略,即$\mathrm{Adv}_{\mathrm{A}}^{C}(\lambda)\leqslant\mathrm{negl}(\lambda)$,其中$\mathrm{negl}(\lambda)$为可忽略函数,则该方案满足密文不可区分性。

3.2.2 陷门不可区分性游戏

与游戏1相似,游戏2仍然假设半诚实服务器是敌手A。陷门不可区分性旨在防止敌手A从一个给定的陷门信息中获取其对应关键词的信息。陷门不可区分性保证即使是服务器本身,也无法生成针对某接收者的有效密文。

1)初始化。与游戏1的初始化相同。

2)询问阶段1。与游戏1的询问阶段1相同。

3)挑战。敌手A向挑战者B声明一个发送者的身份标识ID*作为挑战,并选择两个长度相等的挑战关键字(w_0,w_1),挑战者B随机选择$\beta\in(0,1)$,并根据选择结果,运行陷门生成算法$\mathrm{Trapdoor}(w_\beta,d_{\mathrm{R}})$,生成陷门$T_\beta$,并将其发送给敌手A。

4)询问阶段2。与游戏1中的询问阶段2相同。

5)猜测。敌手A输出一个字节$\beta'\in(0,1)$,作为对关键字w_β的猜测,假如$\beta=\beta'$,则敌手A赢得游戏。限制条件与游戏1相同。

敌手A在该游戏中的优势定义为:

$$\mathrm{Adv}_{\mathrm{A}}^{T}(\lambda)=\left|\Pr[\beta=\beta']-\frac{1}{2}\right|$$

若敌手A赢得游戏的优势可以忽略,即$\mathrm{Adv}_{\mathrm{A}}^{T}(\lambda)\leqslant\mathrm{negl}(\lambda)$,则该方案满足陷门不可区分性。

4 方案分析与证明

4.1 正确性分析

本文根据双线性对的性质来验证本文方案的正确性。

定义2 假设在本文方案中所有密文索引和陷门都是按照正确的步骤生成,且在传输过程中未遭到恶意攻击,数据没有被篡改。方案正确性是指当用户向服务器发送包含关键词w'的检索陷门时,服务器运行检索匹配算法,以正确匹配到包含该关键词的密文索引,从而返回包含该关键词的密文文件。

假设用户ID′的公私钥对为$Q'=[t_1']P_1$,$d'=[t_2']P_2$,用户ID′为关键词w'生成的陷门$T'(T_1'=[t']P_{\mathrm{pub}},T_2'=[H_2(w')-t']\cdot d')$。当服务器运行检索匹配算法并逐个对比密文索引,由双线性对的性质可知:

$$H_3(\hat{e}(T_2',I_1)\cdot\hat{e}(T_1',I_2))=$$
$$H_3(\hat{e}([H_2(w')-t']\cdot d',[r]Q')\cdot\hat{e}([t']P_{\mathrm{pub}},[r]P_2))=$$
$$H_3(\hat{e}(H_2(w')\cdot[t_2']P_2,[r][t_1']P_1)\cdot\hat{e}(P_1,P_2)^{-t'sr}\cdot$$
$$\hat{e}(P_1,P_2)^{t'sr})=H_3(\hat{e}(H_2(w')\cdot P_2,P_1)^{rt_1't_2'})=$$
$$H_3(\hat{e}(H_2(w')\cdot P_2,P_1)^{rs})=$$
$$H_3(\hat{e}(H_2(w')\cdot[r]P_2,[s]P_1))=H_3(\hat{e}(H_2(w')\cdot I_2,P_{\mathrm{pub}}))$$

如果$w=w'$,则有:

$$H_3(\hat{e}(T_2',I_1)\cdot\hat{e}(T_1',I_2))=H_3(\hat{e}(H_2(w)\cdot I_2,P_{\mathrm{pub}}))=I_3$$

则服务器可以正确匹配到密文索引,正确性成立。

4.2 安全性证明

本文SM9SE方案在随机谕言模型下满足适应性密文不可区分性以及适应性陷门不可区分性。具体证明过程如下:

定理1 若敌手A能以优势$\mathrm{Adv}_{\mathrm{A}}^{C}(\lambda)$赢得密文不可区分性游戏(最多进行$q_{H_1}$次$H_1$询问),则存在一个挑战者B能利用敌手A以$[1/(2\cdot q_{H_1})]\cdot\mathrm{Adv}_{\mathrm{A}}^{C}(\lambda)$的优势解决DBDH问题。

证明 若挑战者B解决一个DBDH实例元组$Y=(G_1,G_2,G_T,\hat{e},q,P_1,[x]P_1,[y]P_1,P_2,[y]P_2,[z]P_2,Z)$,则挑战者B按如下过程与敌手A进行密文不可区分性游戏。

1)初始化。挑战者B选择随机数$x\in Z_q^*$,将系统参数设为$\mathrm{Params}=(G_1,G_2,G_T,P_1,P_2,\hat{e},q,P_{\mathrm{pub}}=[x]P_1)$。

2)询问阶段1。敌手A适应性地询问由挑战者B模拟的随机谕言机,假设敌手A不会对同一个随机谕言机进行相同询问,并且在向H_1询问ID值之前不会对ID做任何计算。随机谕言机主要有以下6个:

(1)Hash谕言机O_{H_1}:挑战者B建立并维护一张初始为空的表$L_{H_1}=<\cdot,\cdot>$,设敌手A最多进行q_{H_1}次询问,挑战者B随机选择$i\in\{1,2,\cdots,q_{H_1}\}$,并猜测敌手A的第$i$次询问正是挑战的ID*,如果敌手A询问ID的哈希值,挑战者B按以下情况将值返回,若这是第i次询问,即$\mathrm{ID}=\mathrm{ID}^*$,随机返回值$h\in Z_q^*$,并将$<\mathrm{ID},h,[y]P_1,[z]P_2,\perp>$添加到表$L_{H_1}$中,否则随机返回$h_1\in Z_q^*$,并且在表中添加$<\mathrm{ID},h_1,[h_1+x]P_1=[t_1]P_1,[x/(h_1+x)]P_2=[t_2]P_2,h_1>$。

(2)Hash谕言机O_{H_2}:对于关键词w,若$\mathrm{ID}=\mathrm{ID}^*$,则返回$x$,否则随机选择$h_2\in Z_q^*$作为$H_2(w)$的值输出。

(3)Hash谕言机O_{H_3}:对于群G_T中的元素,随机选择任意长度的随机位串作为$H_3(\cdot)$的值输出。

(4)密钥谕言机O_E:对于ID,若$\mathrm{ID}=\mathrm{ID}^*$,中断模拟,并随机输出字节$\beta'$,否则查找表$L_{H_1}$,并返回$d_{\mathrm{ID}}=[t_2]P_2,Q_{\mathrm{ID}}=[t_1]P_1$。

(5)索引谕言机O_C:对于$(\mathrm{ID}_{\mathrm{R}},w)$,随机选择$r\in Z_q^*$,并按以下情况计算索引$I=(I_1,I_2,I_3)$,若$\mathrm{ID}=\mathrm{ID}^*$,则$I_1=[r]P_1$,$I_2=[r]P_2$,$I_3=H_3(Z')$;否则查找表$L_{H_1}$并计算$I_1=[r][t_1]P_1$,$I_2=[r]P_2$,以及$I_3=H_3(\hat{e}(H_2(w)\cdot I_2,[x]P_1))$。

www.ecicc.com

（6）陷门谕言机 O_T：随机选择 $t \in Z_q^*$，并按以下情况计算陷门 $T=(T_1,T_2)$，若 $\mathrm{ID}=\mathrm{ID}^*$，则 $T_1=[t][y]P_1$，$T_2=[H_2(w)-t][y]P_2$；否则查找表 L_{H_1} 并计算 $T_1=[t]P_{\mathrm{pub}}$，$T_2=[H_2(w)-t][t_2]P_2$。

3）挑战。敌手A提交两个挑战关键词 w_0、w_1 以及要挑战的 ID^*，挑战者B随机选择字节 $\hat{\beta}\in\{0,1\}$ 和元素 $r\in Z_q^*$，返回密文索引 $I^*=(I_1^*,I_2^*,I_3^*)$，其中，$I_1^*=[r]P_1$，$I_2^*=[r]P_2$，$I_3^*=H_3(\hat{e}(H_2(w)\cdot I_2,[x]P_1))$。

4）询问阶段2。与询问阶段1相同。

5）猜测。敌手A输出猜测的字节 $\hat{\beta}'$，如果 $\hat{\beta}'=\hat{\beta}$，挑战者B输出 $\beta'=0$，否则挑战者B输出 $\beta'=1$。

若挑战者B对挑战身份的猜测不正确，则模拟中断，将此事件表示为F。若挑战者B模拟中断，则挑战者B随机输出字节 $\beta'=\beta$ 的概率为1/2。由于挑战者B的猜测是随机的，因此F不发生的概率为 $1/q_{H_1}$，即 $\Pr[\bar{\mathrm{F}}]=1/q_{H_1}$。

假设挑战者B模拟不中断，若 $Z=\hat{e}(P_1,P_2)^{xyz}$，则挑战者B的模拟和敌手A在真实攻击中的视图相同，此时敌手A赢得游戏的概率为 $\mathrm{Adv}_{\mathrm{A}}^{C}(\lambda)+1/2$。若 Z 是 G_T 中的随机元素，由于挑战者B的模拟是完备的，因此挑战密文会将字节 $\hat{\beta}$ 完全隐藏，则敌手A赢得游戏的概率最多为1/2。因此，挑战者B在解决DBDH问题中的优势可表示为：

$$\mathrm{Adv}_{\mathrm{B}}^{\mathrm{DBDH}}(\lambda)=$$

$$\left|\Pr[\beta'=\beta\,|\,\mathrm{F}]\cdot\Pr[\mathrm{F}]+\Pr[\beta'=\beta\,|\,\bar{\mathrm{F}}]\cdot\Pr[\bar{\mathrm{F}}]-1/2\right|=$$

$$\left|\frac{1}{2}\cdot(1-\Pr[\bar{\mathrm{F}}])+(\Pr[\beta'=0\,|\,\bar{\mathrm{F}}\wedge\beta=0]\cdot\Pr[\beta=0]+\right.$$

$$\left.\Pr[\beta'=1\,|\,\bar{\mathrm{F}}\wedge\beta=1]\cdot\Pr[\beta=1])\cdot\Pr[\bar{\mathrm{F}}]-\frac{1}{2}\right|\geqslant$$

$$\left|\frac{1}{2}\cdot(1-\Pr[\bar{\mathrm{F}}])+\Pr[\bar{\mathrm{F}}]\cdot((\mathrm{Adv}_{\mathrm{A}}^{C}(\lambda)+\frac{1}{2})\cdot\frac{1}{2}+\frac{1}{4})-\frac{1}{2}\right|=$$

$$\frac{1}{2}\cdot\Pr[\bar{\mathrm{F}}]\cdot\mathrm{Adv}_{\mathrm{A}}^{C}(\lambda)=\frac{1}{2q_{H_1}}\cdot\mathrm{Adv}_{\mathrm{A}}^{C}(\lambda)$$

若 $\mathrm{Adv}_{\mathrm{A}}^{C}(\lambda)$ 不可忽略，则 $\mathrm{Adv}_{\mathrm{B}}^{\mathrm{DBDH}}(\lambda)$ 也不可忽略，即挑战者B能以不可忽略的优势解决DBDH问题，与DBDH假设矛盾。因此，敌手A无法以不可忽略的优势赢得密文不可区分性游戏。

定理2　若敌手A能以优势 $\mathrm{Adv}_{\mathrm{A}}^{T}(\lambda)$ 赢得陷门不可区分性游戏（最多进行 q_{H_1} 次 H_1 询问），则存在一个挑战者B能利用敌手A以 $[1/(2\times q_{H_1})]\times\mathrm{Adv}_{\mathrm{A}}^{T}(\lambda)$ 的优势解决DBDH问题。

陷门不可区分性的证明与密文不可区分性的证明相似。陷门不可区分性只是在挑战阶段挑战者B返回陷门 $T^*=(T_1^*,T_2^*)$。首先随机选择 $t\in Z_q^*$，然后计算 $T_1^*=[t][y]P_1$，$T_2^*=[H_2(w)-t][y]P_2$。

定理3　若一个敌手 A_1 能以不可忽略的优势 $\mathrm{Adv}_{\mathrm{A}_1}^{\mathrm{KGA}}(\lambda)$ 对SM9SE方案进行离线关键词猜测攻击，则存在另一个敌手 A_2 能以 $\frac{1}{2}+\frac{1}{2}\times\mathrm{Adv}_{\mathrm{A}_1}^{\mathrm{KGA}}(\lambda)$ 的优势赢得密文不可区分性游戏。

证明　当敌手 A_2 和挑战者B进行密文不可区分性游戏时，在挑战阶段，敌手 A_2 可将挑战者B返回的挑战密文发送给敌手 A_1，若敌手 A_1 猜测的结果为敌手 A_2 挑选的关键词之一，则敌手 A_2 向挑战者B输出对应的结果作为猜测，否则随机输出猜测结果。可知敌手 A_2 的优势为：

$$\mathrm{Adv}_{\mathrm{A}_2}^{C}=\frac{1}{2}+\frac{1}{2}\cdot\mathrm{Adv}_{\mathrm{A}_1}^{\mathrm{KGA}}(\lambda)$$

该证明与定理1矛盾。因此，如果DBDH假设成立，则不存在敌手能以不可忽略的优势对SM9SE方案发起离线关键词猜测攻击。

5　仿真实验与分析

5.1　效率分析

本文将SM9SE方案与其他现有的基于身份可搜索加密方案进行对比分析。本文在同一标准下将EdIBEKS方案[10]、PEAKS方案[14]以及dIBAEKS方案[17]进行效率分析。基于文献[25]中的定义，表1定义了若干不同符号及其换算方法。

表1　符号定义与换算

Table 1　Symbol definition and conversion

符号	定义
T_m	模乘运算所需时间
T_b	双线性配对所需时间，$T_b\approx 86T_m$
T_{be}	双线性配对的幂运算所需时间，$T_{be}\approx 43.5T_m$
T_{em}	曲线标量点乘法运算所需时间，$T_{em}\approx 29T_m$

EdIBEKS方案、PEAKS方案以及dIBAEKS方案所有的运算都建立在对称双线性对上，使用的曲线为嵌入度2，以1 024 bit的大素数 p 为阶的超奇异椭圆曲线 $E(F_p)$：$y^2=x^3+x$。本文方案的运算建立在非对称双线性对上，使用的曲线为嵌入度12，阶为256 bit的大素数 p 的BN曲线 E：$y^2=x^3+5$。以上选择参数能够确保所有方案都可以实现与2 048 bit的RSA密钥相当的安全级别。表2表示不同方案的安全性假设、安全级别及其参数大小，以及本文方案与其他基于身份可搜索加密方案中用户执行一次索引生成算法、陷门生成算法和服务器执行一次检索匹配算法所需要执行的操作，其中忽略了方案中都存在的哈希函数运算。当计算椭圆曲线上的双线性对时，曲线的嵌入次数越多，计算越复杂，因此安全性也更高。相比其他方案采用嵌入度为2的超奇异椭圆曲线，本文方案采用嵌入度为12的BN曲线，因此本文方案的安全性优于其他方案。由于BN曲线的双线性对具有友好性[26]，在BN曲线上计算双线性对更加高效。从表2可以看出，本文SM9SE方案在保证计算效率的前提下，采用更小的安全参数实现与其他方案同等的安全级别，并且在索引生成算法上的效率优于其他三个方案，在陷门生成算法上的效率优于PEAKS方案和dIBAEKS方案，与EdIBEKS方案持平，都需要一次双线性对运算。由于本文方

www.ecice06.com 计算机工程

案采用参数更小的BN曲线,因此计算效率会更高,检索匹配算法的效率优于EdIBEKS方案和dIBAEKS方案,与PEAKS方案持平,都需要两次双线性对运算和一次椭圆曲线标量点乘法运算。由于本文方案采用参数更小的BN曲线,因此计算效率会更高。因此,本文方案的总体性能优于其他方案。

表2 不同方案的计算效率与安全性对比

Table 2 Computational efficiency and security comparison among different schemes

方案	索引生成算法	陷门生成算法	检索匹配算法	安全级别/bit	安全性假设	参数q/bit	使用曲线
EdIBEKS方案	T_b+3T_{be}	$2T_{be}$	$3T_b$	112	BDH假设	1 024	超奇异椭圆曲线
PEAKS方案	$3T_{be}+T_{em}$	T_b+T_{be}	$2T_b+T_{em}$	112	DBDH假设	1 024	超奇异椭圆曲线
dIBAEKS方案	$2T_b+3T_{be}$	$T_b+2T_{be}+T_{em}$	$2T_b+2T_{be}$	112	DBDH假设	1 024	超奇异椭圆曲线
SM9SE方案	T_b+2T_{em}	$2T_{be}$	$2T_b+T_{em}$	112	DBDH假设	256	BN曲线

5.2 仿真结果分析

本文对EdIBEKS、PEAKS、dIBAEKS和SM9SE方案进行仿真实验,在2.40 GHz的4核64位Intel® Core™i5-10200H处理器、8 GB内存(RAM)、Windows 10操作系统的实验环境进行实验,以myeclipse 10作为实验平台、Java作为实验编程语言,对各方案的索引生成算法、陷门生成算法和匹配检索算法进行模拟运行。本文首先使用不同长度的关键词对各个方案进行模拟运行,以分析关键词长度对各方案效率的影响,然后分别抽取Enron邮件数据集中的部分数据和Kaggle邮件数据集中的部分数据作为测试输入,使用不同数量的关键词进行对比实验。关键词长度对各方案的算法运行时间的影响如图1所示。

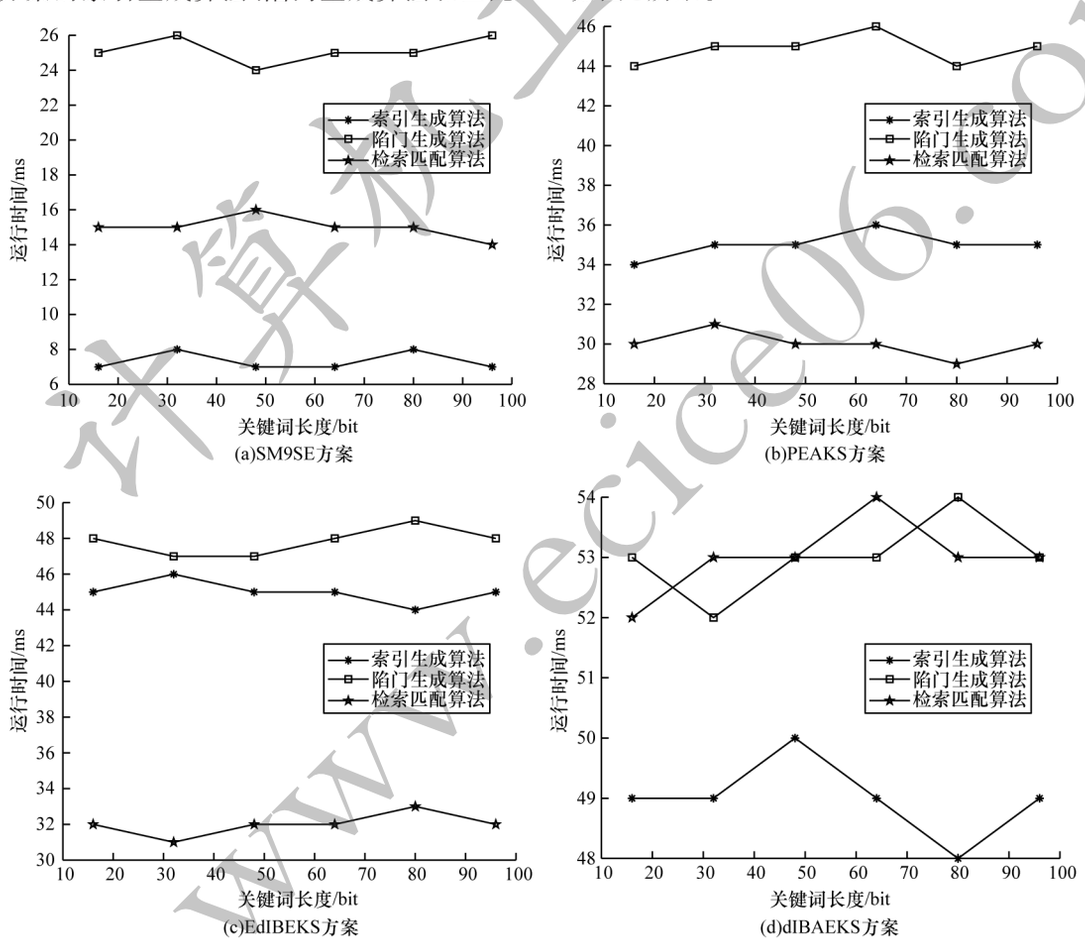

图1 关键词长度对各方案中算法运行时间的影响

Fig.1 Influence of keyword length on the running time of algorithms in each schemes

从图1可以看出,关键词长度对各算法运行效率的影响较小(关键词长度影响哈希函数运行时间,但哈希运算速度极快),略受各算法所选择的随机数影响,但总体运行时间较为平稳。

图2表示各个方案中索引生成算法的运行时间,其中横坐标是关键词的个数,纵坐标为建立索引所需要的时间。从图2可以看出,SM9SE方案在索引生成算法上的运行效率优于其他三种方案。SM9SE方案生成一个密文索引的平均时间在8 ms左右,而其他方案的密文索引时间则在35~68 ms之间。因此,本文方案在索引生成算法上的效率相比其他三种方案提高了77%以上。

计算机工程 www.ecice06.com

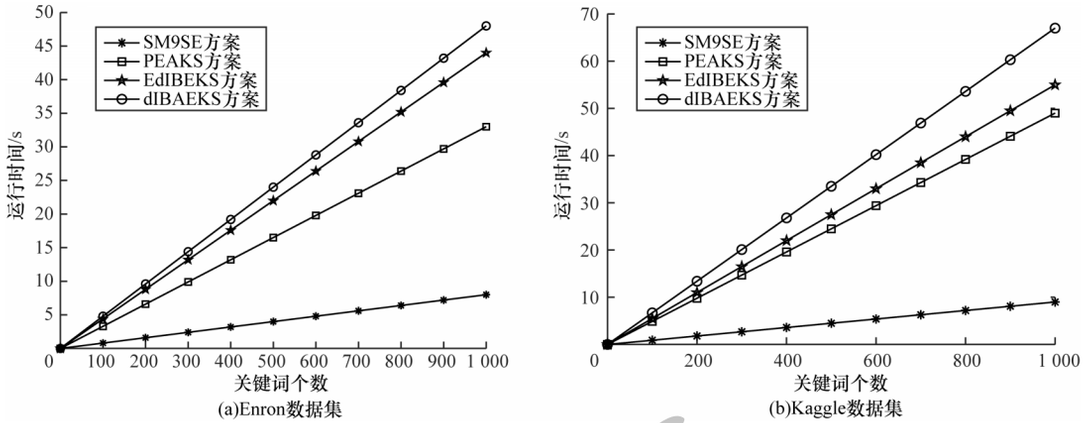

图2　索引生成算法的运行时间对比

Fig.2　Comparison of running time of index generation algorithms

图3表示在各个方案中陷门生成算法的运行时间对比,其中横坐标是关键词的个数,纵坐标为生成检索陷门所需要的时间。从图3可以看出,SM9SE方案在陷门生成算法上的运行效率也优于其他三种方案。其他三种方案生成一个陷门的运行时间平均为30~53 ms之间,而SM9SE方案生成一个陷门的平均时间在25 ms左右,效率提高了16.67%以上。

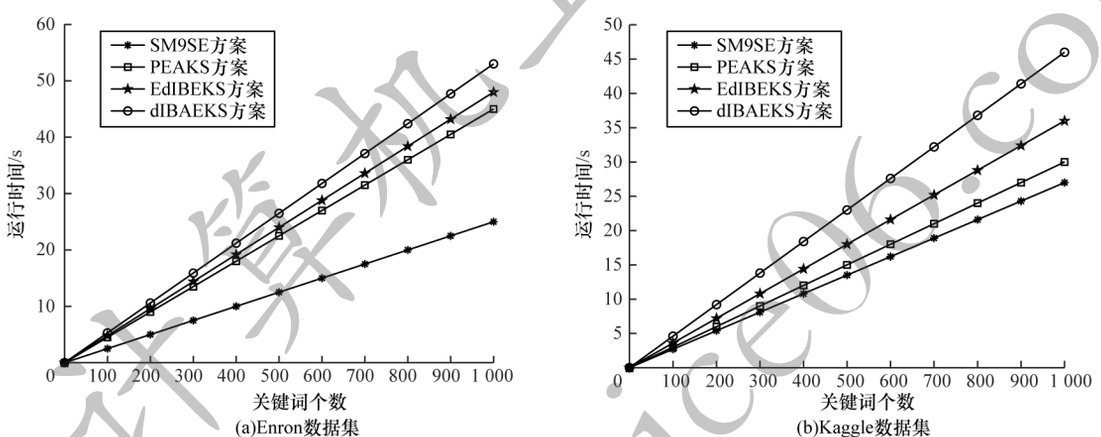

图3　陷门生成算法的运行时间对比

Fig.3　Comparison of running time of trapdoor generation algorithms

图4表示在各个方案中检索匹配算法所运行的时间,其中横坐标是密文索引的个数,纵坐标为所有密文索引的运行时间。从图4可以看出:SM9SE方案单次检索匹配算法的平均运行时间为18 ms左右;dIBAEKS方案的单次检索匹配算法的平均运行时间为53 ms左右;PEAKS方案和EdIBEKS方案的单次检索匹配算法的平均运行时间在25~45 ms之间。因此,SM9SE方案的检索匹配算法运行效率相对其他三种方案提高了28%以上。

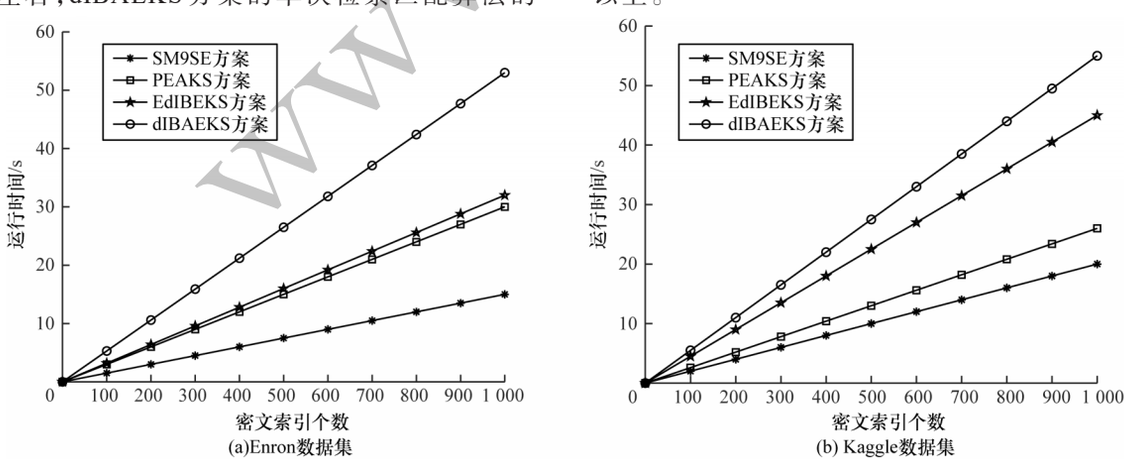

图4　检索匹配算法的运行时间对比

Fig.4　Comparison of running time of retrieval matching algorithms

计算机工程 www.ecice06.com

图5表示在不同方案中各个算法的平均运行时间。从图5可以看出，本文方案执行一次索引生成算法、陷门生成算法和检索匹配算法的全过程所需要的平均时间总和为52 ms，其他三种方案则在105~168 ms之间。相比其他三种方案，本文方案的总体效率提高了50.4%以上。

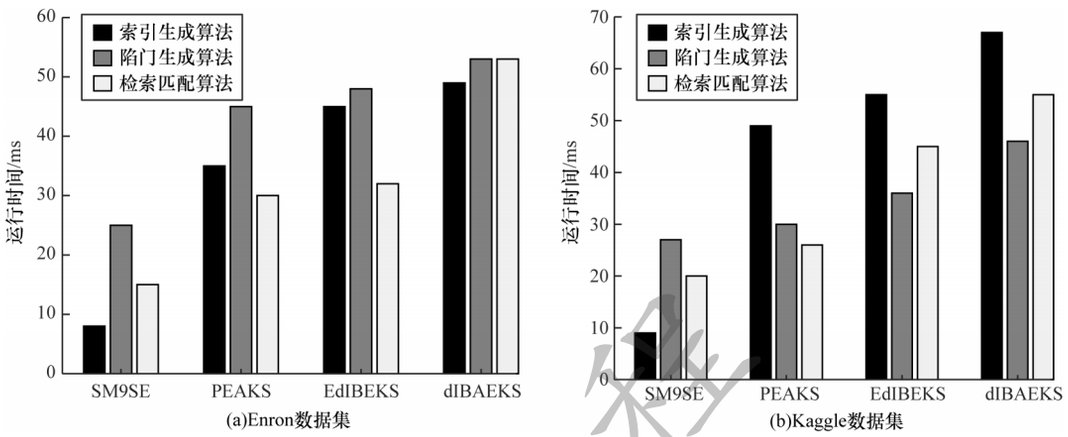

图5 各个算法的平均运行时间对比

Fig.5 Comparison of average running time of each algorithms

6 结束语

针对现有基于身份的可搜索加密方案计算效率低、与SM9密码算法难以结合等问题，本文提出一种基于SM9密码算法的可搜索加密方案。根据非对称双线性对的性质，结合SM9加密算法的密钥形式，设计索引生成算法、陷门生成算法和服务器检索匹配算法，并验证方案的正确性和安全性。该方案在随机谕言模型下满足适应性密文不可区分性和适应性陷门不可区分性。仿真结果表明，相比EdIBEKS、PEAKS、dIBAEKS方案，本文方案具有较高的计算效率，能够实现可搜索加密与国家商用密码算法SM9的结合，并扩展了SM9密码算法的应用。下一步将在本文方案中增加指定身份检索、身份验证等更灵活的功能，使其适用于较复杂的应用场景。此外，还将在安全索引中嵌入数据发送者的私钥，并在陷门中嵌入发送者的公钥，通过设计相应的匹配算法实现精确检索目标文件的功能。

参考文献

[1] SONG D X, WAGNER D, PERRIG A. Practical techniques for searches on encrypted data[C]//Proceedings of IEEE Symposium on Security and Privacy. Washington D. C., USA: IEEE Press, 2000: 1-10.

[2] GOH E J. Secure indexes[EB/OL]. [2021-08-24]. http://xenon. stanford. edu/~eujin/papers/secureindex/2004mar-secureindex. pdf.

[3] BLOOM B H. Space/time trade-offs in hash coding with allowable errors[J]. Communications of the ACM, 1970, 13(7): 422-426.

[4] BONEH D, DI CRESCENZO G, OSTROVSKY R, et al. Public key encryption with keyword search[EB/OL][2021-08-24]. https://link. springer. com/content/pdf/10. 1007%2F978-3-540-24676-3_30. pdf.

[5] BONEH D, FRANKLIN M. Identity-based encryption from the Weil pairing [C]//Proceedings of International Conference on Advances in Cryptology. Berlin, Germany: Springer, 2001: 213-229.

[6] ABDALLA M, BELLARE M, CATALANO D, et al. Searchable encryption revisited: consistency properties, relation to anonymous IBE, and extensions[J]. Journal of Cryptology, 2008, 21(3): 350-391.

[7] TIAN X X, WANG Y. ID-based encryption with keyword search scheme from bilinear pairings[C]//Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing. Washington D. C., USA: IEEE Press, 2008: 1-4.

[8] SIAD A. Anonymous identity-based encryption with distributed private-key generator and searchable encryption[C]//Proceedings of the 5th International Conference on New Technologies, Mobility and Security. Washington D. C., USA: IEEE Press, 2012: 1-8.

[9] XU P, JIN H, WU Q H, 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] WU T Y, TSAI T T, TSENG Y M. Efficient searchable ID-based encryption with a designated server[J]. Annals of Telecommunications-Annales Des Telecommunications, 2014, 69(7/8): 391-402.

[11] 李经纬，贾春福，刘哲理，等. 可搜索加密技术研究综述[J]. 软件学报，2015，26(1)：109-128.
LI J W, JIA C F, LIU Z L, et al. Survey on the searchable encryption[J]. Journal of Software, 2015, 26(1): 109-128. (in Chinese)

[12] XU P, WU Q H, WANG W, et al. Generating searchable public-key ciphertexts with hidden structures for fast keyword search [J]. IEEE Transactions on Information Forensics and Security, 2015, 10(9): 1993-2006.

[13] RAMYA C P, VIPIN K M. Comparison of different techniques for keyword searching over encrypted data[C]//Proceedings of Online International Conference on Green Engineering and Technologies. Washington D. C., USA: IEEE Press, 2016: 1-4.

计算机工程 www.ecice06.com

[14] HUANG Q，LI H B. An efficient public-key searchable encryption scheme secure against inside keyword guessing attacks[J]. Information Sciences，2017，403/404：1-14.

[15] ZHANG X J，XU C X，MU L M，et al. Identity-based encryption with keyword search from lattice assumption[J]. China Communications，2018，15(4)：164-178.

[16] JIANG P，GUO F C，MU Y. Efficient identity-based broadcast encryption with keyword search against insider attacks for database systems[J]. Theoretical Computer Science，2019，767(45)：51-72.

[17] LI H B，HUANG Q，SHEN J，et al. Designated-server identity-based authenticated encryption with keyword search for encrypted emails[J]. Information Sciences，2019，481(19)：330-343.

[18] 朱敏惠，陈燕俐，胡媛媛. 支持代理重加密的基于身份可搜索加密方案[J]. 计算机工程，2019，45(1)：129-135，140.
ZHU M H，CHEN Y L，HU Y Y. Identity-based searchable encryption scheme supporting proxy re-encryption[J]. Computer Engineering，2019，45(1)：129-135，140.（in Chinese）

[19] 倪绿林，许春根. 基于身份的动态可搜索加密方案[J]. 计算机工程，2019，45(1)：136-140.
NI L L，XU C G. Dynamic searchable encryption scheme based on identity[J]. Computer Engineering，2019，45(1)：136-140.（in Chinese）

[20] 牛淑芬，杨平平，谢亚亚，等. 电子邮件系统中指定服务器的关键字搜索加密方案[J]. 计算机工程，2020，46(10)：137-142，150.
NIU S F，YANG P P，XIE Y Y，et al. Keyword search encryption scheme for designated server in email system[J]. Computer Engineering，2020，46(10)：137-142，150.（in Chinese）

[21] ZHANG X J，HUANG C，GU D W，et al. BIB-MKS：post-quantum secure biometric identity-based multi-keyword search over encrypted data in cloud storage systems[J]. IEEE Transactions on Services Computing，2021，99(14)：1-10.

[22] LIU Z Y，TSENG Y F，TSO R，et al. Identity-certifying authority-aided identity-based searchable encryption framework in cloud systems[J]. IEEE Systems Journal，2021，99(15)：1-12.

[23] LEE E，LEE H S，PARK C M. Efficient and generalized pairing computation on abelianvarieties [J]. IEEE Transactions on Information Theory，2009，55(4)：1793-1803.

[24] CHATTERJEE S，MENEZES A. On cryptographic protocols employing asymmetric pairings—the role of Ψ revisited[J]. Discrete Applied Mathematics，2011，159(13)：1311-1322.

[25] 彭长根，张小玉，丁红发，等. 基于Cocks身份密码体制的高效签密方案[J]. 通信学报，2020，41(12)：128-138.
PENG C G，ZHANG X Y，DING H F，et al. Efficient signcryption scheme based on Cocks' identity cryptosystem[J]. Journal on Communications，2020，41(12)：128-138.（in Chinese）

[26] BARRETO P S L M，NAEHRIG M. Pairing-friendly elliptic curves of prime order [C]//Proceedings of International Conference on Selected Areas. Berlin，Germany：Springer，2006：319-331.

编辑　薛晋栋

计算机工程 www.ecw.com