DOI:10.13878/j.cnki.jnuist.2017.05.002

马米米[1]　何德彪[2]　陈建华[1]　刘芹[2]

大数据环境下支持多关键字的可搜索公钥加密方案

摘要

云计算为大数据提供了强大的数据处理平台,而数据的安全和隐私问题也引起了人们的高度关注.本文提出了一个新的大数据环境下支持多关键字的无需安全信道的无证书可搜索公钥加密方案,并且证明了在随机预言机模型下,本文方案能够抵抗关键字猜测攻击.同时,效率分析表明,与 Peng 等方案相比,本文方案降低了计算代价和通信代价.

关键词

大数据;无证书公钥加密;隐私性;可证明安全

中图分类号 TP309

文献标志码 A

收稿日期 2017-05-31

资助项目 国家自然科学基金(61572370,61572379,61501333,U1536204)

作者简介

马米米,女,博士生,研究方向为数论与密码.mamimi421@126.com

何德彪(通信作者),男,博士,教授,主要从事应用密码学、安全协议等领域的教学与科研工作.hedebiao@163.com

1　武汉大学　数学与统计学院,武汉,430072

2　武汉大学　计算机学院,武汉,430072

0　引言

随着互联网、物联网和云计算的兴起,以及移动设备、智能终端、社交网络、电子商务等的快速普及,极大地扩展了互联网的应用范围,必然导致数据量的急剧增长[1].例如,教育、医疗卫生、金融、电力、采矿等各行各业时时刻刻都在产生大数据.信息技术的迅速发展,使得网络信息流量从 GB 增加到 TB、PB 甚至 ZB.云计算为大数据的存储、运算和管理提供了技术平台,使大数据的处理更加便捷高效.越来越多的个体和企业为了节约本地存储空间并方便地进行数据管理,他们更倾向于将数据存储在云服务器上.但是,云服务器可能会对数据进行篡改、删除或损坏.数据的安全和隐私问题受到了人们的高度关注.

为了保护数据的安全性和隐私性,数据拥有者可以将数据加密后再上传到云服务器.然而,在密文上进行高效的搜索将是一个难题.用户可以先将所有数据下载到本地,再进行解密,显然,这将消耗大量的网络带宽和本地存储空间,效率极低.为了解决这个问题,可搜索加密技术应运而生.

可搜索加密技术分为对称加密和非对称加密.Song 等[2]首先提出了可搜索对称加密.但是,可搜索对称加密只适用于单用户模型,应用场景受到限制.为了解决这一问题,Boneh 等[3]提出了第一个可搜索公钥加密技术,它不仅能够有效地进行密文搜索,还能够实现用户之间的数据共享.随后,很多可搜索公钥加密方案被设计出来[4-6].然而,这些方案大部分都是基于公钥基础设施(PKI)或者基于身份的公钥密码系统.

在 PKI 系统中,认证中心(CA)负责分发和管理用户的证书.但是,证书管理是一个棘手的问题.为了解决这一问题,Shamir[7]首次提出基于身份的公钥密码系统(IDPKC).在 IDPKC 中,用户的身份信息都可以作为用户的公钥,用户的私钥由一个可信的第三方即密钥生成中心(KGC)产生.由于 KGC 可能是恶意的,这导致了密钥托管问题.Al-Riyami 等[8]设计了无证书公钥密码系统.在无证书公钥密码系统中,私钥由 KGC 和用户共同产生,这不仅避免了证书管理,也解决了密钥托管问题.至今,将无证书公钥密码系统与可搜索加密技术相结合的方案很少.2014 年,Peng 等[9]提出了第一个无证书可搜索公钥加密方案.然而,Wu 等[10]指出文献[9]不能抵抗恶意 KGC 和离线关键字猜测攻击.

本文的主要贡献总结如下:

1) 提出了一个新的大数据环境下支持多关键字的无需安全信道的无证书可搜索公钥加密方案;

2) 给出了方案的安全性分析.在随机预言模型下,证明了本文方案能够抵抗关键字猜测攻击;

3) 在计算代价和通信代价两方面对方案进行了效率分析.效率分析显示本文方案比较高效.

本文结构安排如下:第 1 章对相关工作进行了描述;第 2 章给出一些预备知识;第 3 章给出本文方案的具体构造;第 4 章和第 5 章分别进行了安全性分析和效率分析;第 6 章是结束语.

1 相关工作

大数据时代,保护数据的隐私性和机密性是一个重要的问题.可搜索加密技术不仅可以在密文上进行搜索,而且能够保护数据的安全性和隐私性.Song 等[2]设计了第一个可搜索对称加密技术:主要是基于扫描思想,先将明文分为"单词",再对其进行加密;关键字搜索阶段,服务器需要扫描整个密文"单词"进行匹配.随后,Golle 等[11]提出了支持连接关键字的可搜索加密机制;Chang 等[12]提出了具有隐私保护的支持关键字搜索的远程数据加密方案,即用户将文件加密后通过电脑上传到云服务器,在移动设备上也可以进行密文文件搜索.然而,这些方案都是基于对称的可搜索加密,不适用于多用户环境.

为了解决上述问题,Boneh 等[3]首次提出了可搜索公钥加密方案:数据拥有者建立文件的关键字索引,使用接收者的公钥加密文件和关键字索引并上传到云服务器;用户用自己的私钥生成待检索关键字的陷门并发送给云服务器,云服务器将加密的关键字索引与陷门进行匹配,最后,将搜索结果返回给用户.随后,Baek 等[4]提出了一个无需安全信道的可搜索公钥加密方案;Xu 等[6]构造了第一个支持模糊关键字搜索的加密方案;Wang 等[13]设计了一个支持多关键字的无需安全信道的可搜索公钥加密方案;Chen 等[14]提出了一个支持关键字搜索的双服务器公钥加密方案;Jiang 等[15]提出了一个具有支持关键字授权的可搜索公钥加密方案.之后,也有很多的可搜索公钥方案被提出[16-17].然而上述方案都存在证书管理或密钥托管问题,而 Peng 等[9]将无证书公钥密码系统与可搜索加密技术相结合,首次构造了一个无证书可搜索公钥加密方案.

2 预备知识

2.1 双线性对

令 G_1 是阶为 q 的循环加法群,G_2 是阶为 q 的循环乘法群.我们称映射

$\hat{e}:G_1 \times G_1 \rightarrow G_2$

为双线性对,如果它满足下面的性质:

1) 双线性性:对于任意的 $P,Q \in G_1,a,b \in Z_q^*$,有 $\hat{e}(aP,bQ)=\hat{e}(bP,aQ)=\hat{e}(P,Q)^{ab}$.

2) 非退化性:存在 $P,Q \in G_1$,使得

$\hat{e}(P,Q) \neq 1 \in G_2$.

3) 可计算性:对于任意的 $P,Q \in G_1$,存在有效算法能够计算 $\hat{e}(P,Q)$.

2.2 数学困难问题

定义 1 计算性 Diffie-Hellman 问题(CDH 问题):给定 $P,aP,bP \in G_1$,其中 $a,b \in Z_q^*$ 是未知的随机数,计算 $abP \in G_1$.

定义 2 p- 双线性 Diffie-Hellman 求逆问题(p-BDHI):设 G_1,G_2 为 q 阶循环群,P 是 G_1 的生成元,$\hat{e}:G_1 \times G_1 \rightarrow G_2$ 为双线性映射.给定 $(P,xP,x^2P,\cdots,x^pP)$,计算 $\hat{e}(P,P)^{\frac{1}{x}}$.

2.3 系统模型

本小节,我们定义无需安全信道的无证书可搜索公钥加密方案的系统模型,如图 1 所示.在这个系统模型中,存在 4 个参与者:云服务器、数据拥有者、接收者、密钥生成中心(KGC).

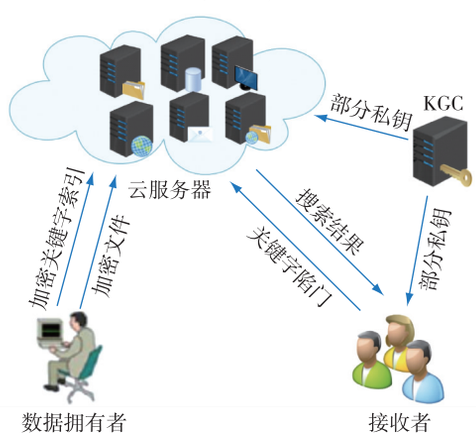

图 1 系统模型

Fig. 1 System model

1) KGC:主要负责产生系统参数、服务器和接收者的部分私钥.

2) 数据拥有者:主要负责建立关键字索引,并将

文件和关键字索引加密,将生成的文件密文和索引密文上传到云服务器.

3)接收者:当进行关键字搜索时,需要生成关键字陷门,然后将陷门发送给云服务器.

4)云服务器:主要负责数据的处理,比如数据存储、计算以及搜索.

3 方案的具体构造

本文提出的支持多关键字的无需安全信道的无证书可搜索公钥加密方案中,共有 8 个多项式时间算法.它们分别是:系统建立、产生部分私钥、产生秘密值、产生私钥、产生公钥、关键字加密、陷门生成和测试.具体描述如下:

1)系统建立:给定安全参数 k,KGC 选择 2 个阶都为 $q > 2^k$ 的群 G_1 和 G_2,一个双线性对 $\hat{e}:G_1 \times G_1 \rightarrow G_2$.$P$ 是 G_1 的生成元.KGC 选择一个随机数 $s \in Z_q^*$ 作为系统主密钥,计算系统主公钥 $P_{\text{pub}} = sP$,并且选择 2 个 Hash 函数 $h_1:\{0,1\}^* \rightarrow Z_q^*$,$h_2:\{0,1\}^* \times \{0,1\}^* \times G_1 \times G_1 \rightarrow Z_q^*$.KGC 公布系统参数 $params = \{k,G_1,G_2,\hat{e},q,P,P_{\text{pub}},h_1,h_2\}$,并将 s 秘密保存.

2)产生部分私钥:输入公共参数 $params$、系统主密钥 s、服务器身份标识 ID_S 和接收者身份标识 ID_R.KGC 执行下面的操作产生服务器的部分私钥 D_{ID_S} 和接收者的部分私钥 D_{ID_R}:

① 计算 $D_{ID_S} = \dfrac{1}{s + h_1(ID_S)}P$;

② 计算 $D_{ID_R} = \dfrac{1}{s + h_1(ID_R)}P$.

KGC 发送 D_{ID_S} 给服务器,发送 D_{ID_R} 给接收者.为了简单起见,我们定义 $T_{ID_R} = P_{\text{pub}} + h_1(ID_R)P$,$T_{ID_S} = P_{\text{pub}} + h_1(ID_S)P$.

当接收者收到部分私钥后,通过验证等式 $e(D_{ID_R},T_{ID_R}) = e(P,P)$ 是否成立来判断部分私钥 D_{ID_R} 是否合法.如果等式成立,部分私钥 D_{ID_R} 合法,否则就不合法.

3)产生秘密值:输入公共参数 $params$、服务器身份标识 ID_S 和接收者身份标识 ID_R.

① 服务器选择一个随机数 $x_{ID_S} \in Z_q^*$,并且将 x_{ID_S} 设置为自己的秘密值;

② 接收者选择一个随机数 $x_{ID_R} \in Z_q^*$,并且将 x_{ID_R} 设置为自己的秘密值.

4)产生私钥:给定 $params,D_{ID_S},x_{ID_S},D_{ID_R}$ 和 x_{ID_R},输出 $SK_{ID_S} = (D_{ID_S},x_{ID_S})$ 作为服务器的私钥,$SK_{ID_R} = (D_{ID_R},x_{ID_R})$ 作为接收者的私钥.

5)产生公钥:输入公共参数 $params$、服务器身份标识 ID_S、接收者身份标识 ID_R、服务器的秘密值 x_{ID_S} 和接收者的秘密值 x_{ID_R},KGC 运行以下步骤:

① 计算

$T_{ID_S} = P_{\text{pub}} + h_1(ID_S)P$, $T_{ID_R} = P_{\text{pub}} + h_1(ID_R)P$;

② 计算 $PK_{ID_S} = x_{ID_S}T_{ID_S}$, $PK_{ID_R} = x_{ID_R}T_{ID_R}$.

KGC 输出 PK_{ID_S} 作为服务器的公钥,输出 PK_{ID_R} 作为接收者的公钥.

6)关键字加密:设 $W = (w_1,w_2,\cdots,w_n)$ 是一个关键字向量集合.输入公共参数 $params$、服务器身份标识 ID_S、接收者身份标识 ID_R、服务器公钥 PK_{ID_S} 和接收者公钥 PK_{ID_R}.数据拥有者执行以下步骤加密每个关键字 $w_i \in W(i = 1,2,\cdots,n)$;

① 选择一个随机数 $r_i \in Z_q^*$;

② 计算

$T_{ID_S} = P_{\text{pub}} + h_1(ID_S)P$, $T_{ID_R} = P_{\text{pub}} + h_1(ID_R)P$;

③ 计算

$$U_i = r_i(PK_{ID_R} + h_2(w_i \| ID_R \| P_{\text{pub}} \| PK_{ID_R})\,T_{ID_R}),$$

$$V_i = r_i(PK_{ID_S} + h_2(w_i \| ID_S \| P_{\text{pub}} \| PK_{ID_S})\,T_{ID_S}).$$

W 对应的密文为 $C = (C_1,C_2,\cdots,C_n)$,其中 $C_i = (U_i,V_i)$.

7)陷门生成:输入 $params$、接收者身份标识 ID_R、接收者公私钥对 (PK_{ID_R},SK_{ID_R}) 和关键字 w',计算陷门

$$T_{w'} = \frac{1}{x_{ID_R} + h_2(w' \| ID_R \| P_{\text{pub}} \| PK_{ID_R})}D_{ID_R}.$$

8)测试:输入公共参数 $params$、服务器的私钥 SK_{ID_S}、陷门 $T_{w'}$ 和关键字密文 C.验证下面等式是否成立:

$$\hat{e}(T_{w'},U_i) = \hat{e}\left(V_i,\frac{1}{x_{ID_S} + h_2(w_i \| ID_S \| P_{\text{pub}} \| PK_{ID_S})}D_{ID_S}\right),$$

如果等式成立,返回“1”;否则,返回“0”.

等式的正确性验证:

证明 假设 $w' = w_i$,下面证明方案的正确性.

$$\hat{e}(T_{w'},U_i) =$$

$$\hat{e}\left(\frac{1}{x_{ID_R} + h_2(w' \| ID_R \| P_{\text{pub}} \| PK_{ID_R})}D_{ID_R},\right.$$

$$\left.r_i(PK_{ID_R} + h_2(w_i \| ID_R \| P_{\text{pub}} \| PK_{ID_R})\,T_{ID_R})\right) =$$

$$\hat{e}\left(\frac{1}{(x_{ID_R}+h_2(w'\|ID_R\|P_{pub}\|PK_{ID_R}))(s+h_1(ID_R))}P,\right.$$
$$\left.r_i(x_{ID_R}T_{ID_R}+h_2(w_i\|ID_R\|P_{pub}\|PK_{ID_R})T_{ID_R})\right)=$$
$$\hat{e}\left(\frac{1}{(x_{ID_R}+h_2(w'\|ID_R\|P_{pub}\|PK_{ID_R}))(s+h_1(ID_R))}P,\right.$$
$$\left.r_i(x_{ID_R}+h_2(w_i\|ID_R\|P_{pub}\|PK_{ID_R}))T_{ID_R}\right)=$$
$$\hat{e}\left(\frac{1}{s+h_1(ID_R)}P,r_iT_{ID_R}\right)=$$
$$\hat{e}\left(\frac{1}{s+h_1(ID_R)}P,r_i(P_{pub}+h_1(ID_R)P)\right)=$$
$$\hat{e}\left(\frac{1}{s+h_1(ID_R)}P,r_i(s+h_1(ID_R))P\right)=\hat{e}(P,P)^{r_i},$$
$$\hat{e}\left(V_i,\frac{1}{x_{ID_S}+h_2(w_i\|ID_S\|P_{pub}\|PK_{ID_S})}D_{ID_S}\right)=$$
$$\hat{e}\left(r_i(PK_{ID_S}+h_2(w_i\|ID_S\|P_{pub}\|PK_{ID_S})T_{ID_S}),\right.$$
$$\left.\frac{1}{x_{ID_S}+h_2(w_i\|ID_S\|P_{pub}\|PK_{ID_S})}D_{ID_S}\right)=$$
$$\hat{e}\left(r_i(x_{ID_S}+h_2(w_i\|ID_S\|P_{pub}\|PK_{ID_S}))T_{ID_S},\right.$$
$$\left.\frac{1}{x_{ID_S}+h_2(w_i\|ID_S\|P_{pub}\|PK_{ID_S})}D_{ID_S}\right)=$$
$$\hat{e}(r_iT_{ID_S},D_{ID_S})=$$
$$\hat{e}\left(P_{pub}+h_1(ID_S)P,\frac{1}{s+h_1(ID_S)}P\right)^{r_i}=$$
$$\hat{e}\left((s+h_1(ID_S))P,\frac{1}{s+h_1(ID_S)}P\right)^{r_i}=\hat{e}(P,P)^{r_i},$$

由上面2个等式可知:

$$\hat{e}(T_{w'},U_i)=\hat{e}\left(V_i,\frac{1}{x_{ID_S}+h_2(w_i\|ID_S\|P_{pub}\|PK_{ID_S})}D_{ID_S}\right).$$

4 安全性分析

4.1 安全模型

在无证书密码系统中[1],存在2种类型的敌手A_1,A_2.

第Ⅰ类敌手A_1模拟的是恶意用户.A_1不知道系统主密钥及用户的部分私钥,但是可以替换用户的公钥.

第Ⅱ类敌手A_2模拟的是恶意KGC.A_2掌握系统主密钥及用户的部分私钥,但是不能替换用户的公钥.

敌手的能力在安全模型中被确定为敌手能否执行相关的查询操作.下面我们首先对敌手有可能会执行的一些查询操作给出相应的描述.

1)Hash询问:敌手A可以询问所有的Hash预言机,并得到相应的Hash值.

2)部分私钥询问:通过对用户ID的部分私钥询问,敌手A能够获得用户ID的部分私钥D_{ID}.

3)公钥询问:敌手A通过执行该操作能够获得用户ID的公钥PK_{ID}.

4)私钥询问:敌手A通过执行该操作能够获得用户ID的私钥SK_{ID}.

5)公钥替换询问:通过执行该查询操作,敌手A能够使用自己选择的新的公钥PK_{ID}'替换用户ID的公钥PK_{ID}.

6)陷门询问:通过执行该查询操作,敌手A能够获得关键字w的陷门.

定义3 一个无证书可搜索公钥加密方案在适应性选择关键字攻击下是语义安全的,如果敌手A_1,A_2在下面2个游戏中获胜的概率$Adv_{A_i}(k)(i\in\{1,2\})$是可以忽略的,其中,

$$Adv_{A_i}(k)=|\Pr[\beta'=\beta]-1/2|.$$

游戏1 这种游戏是在挑战者C和第Ⅰ类敌手A_1之间来进行的.游戏的交互过程如下:

1)系统初始化:C输入安全参数k,运行系统建立算法产生系统主密钥s和公共参数$params$,然后发送$params$给A_1,并秘密保存s.

2)询问操作:A_1能够适应性地选择进行下列一系列询问操作:Hash询问、部分私钥询问、私钥询问、公钥询问、公钥替换询问、陷门询问.当收到这些询问时,C返回相应的值给A_1.

3)挑战阶段:A_1输出2个不同的挑战关键字w_0,w_1,C随机选择$\beta\in\{0,1\}$,然后运行陷门生成算法产生T_{w_β},并将该值返回给A_1.

4)猜测:最后,A_1输出$\beta'\in\{0,1\}$作为它的猜测.如果$\beta'=\beta$,则A_1获胜.

游戏2 这种游戏是在挑战者C和第Ⅱ类敌手A_2之间来进行的.他们的交互过程如下:

1)系统初始化:C输入安全参数k,运行系统建立算法产生系统主密钥s和公共参数$params$,然后发送s和$params$给A_2.

2)询问操作:A_2能够适应性地选择进行下列询

问操作:Hash 询问、部分私钥询问、私钥询问、公钥询问、陷门询问.当收到这些询问时,C 返回相应的值给 A_2.

3) 挑战阶段:A_2 输出2个不同的挑战关键字 w_0, w_1,C 随机选择 $\beta \in \{0,1\}$,然后运行陷门生成算法产生 T_{w_β},并将该值发送给 A_2.

4) 猜测:最后,A_2 输出 $\beta' \in \{0,1\}$ 作为它的猜测.如果 $\beta' = \beta$,则 A_2 获胜.

4.2 方案的安全性证明

在这个部分,将给出上述提出的方案在随机预言机模型下的安全性证明.

定理 1 在随机预言机模型下,如果 p-BDHI 及 CDH 问题是困难的,那么上述提出的方案在适应性选择关键字攻击下是语义安全的.

上述定理可以由下面 2 个引理推导出.

引理 1 假设敌手 A_1 以 ε 的优势攻破了上述方案,则存在一个算法 C 以

$$\varepsilon' \geqslant \left(\frac{\varepsilon}{q_{h_1}}\right)\left(1-\frac{1}{q_{h_1}}\right)^{q_{Ext}+q_E+q_T}$$

的优势解决 p-BDHI 问题,其中,$q_{h_1}, q_{Ext}, q_E, q_T$ 分别表示 A_1 访问 h_1 预言机、部分私钥询问预言机、私钥询问预言机和陷门询问预言机的次数.

证明 算法 C 输入一个 p-BDHI 问题的随机实例:给定 $(P, xP, x^2P, \cdots, x^pP)$,$C$ 的目标是通过调用 A_1 为子程序,最终计算出 $\hat{e}(P,P)^{\frac{1}{x}}$.

首先,C 运行系统建立算法.C 随机选择 $l_0, l_1, \cdots, l_{p-1} \in Z_q^*$,令

$$f(z) = \prod_{i=1}^{p-1}(z+l_i) = \sum_{i=0}^{p-1} c_i z^i,$$

由于 $l_i \neq 0 (1 \leqslant i \leqslant p-1)$,所以 $c_0 \neq 0$.计算

$$P_1 = f(x)P = \sum_{i=0}^{p-1} c_i x^i P,$$

$$\tilde{P}_1 = xP_1 = xf(x)P = \sum_{i=0}^{p-1} c_i x^{i+1} P.$$

如果 $P_1 = 0$,则存在某个 l_i,使得 $x = -l_i$,从而可以直接求解 p-BDHI 问题.所以,不妨假设 $P_1 \neq 0$.C 定义多项式

$$f_i(z) = \frac{f(z)}{z+l_i} = \sum_{j=0}^{p-2} d_{i,j} z^j, \quad 1 \leqslant i \leqslant p-1,$$

计算

$$\frac{1}{x+l_i}P_1 = \frac{f(x)}{x+l_i}P = f_i(x)P = \sum_{j=0}^{p-2} d_{i,j} x^j P,$$

设集合

$$S = \left\{ l_i + l_0, \frac{1}{x+l_i}P_1 \,\middle|\, i = 1,2,\cdots,p-1 \right\},$$

设

$$P_{\text{pub}} = (x - l_0)P_1 = \tilde{P}_1 - l_0 P_1.$$

C 随机选择一个身份 ID_I 作为挑战身份,并发送系统参数 $params = \{G_1, G_2, \hat{e}, q, P, P_{\text{pub}}, h_1, h_2\}$ 给敌手 A_1.

1) h_1- 询问:C 维护一个含数组 $\langle ID_i, h_{1i}, D_{ID_i}\rangle$ 的列表 h_1^{list},该表初始化为空.当 A_1 对 ID_i 进行询问时,C 执行如下操作:

① 如果 $\langle ID_i, h_{1i}, D_{ID_i}\rangle$ 已经包含在 h_1^{list} 中,C 将 h_{1i} 发送给 A_1;

② 否则,如果 $ID_i = ID_I$,C 发送 l_0 给 A_1,并添加 $\langle ID_i, l_0, \perp\rangle$ 到 h_1^{list};

③ 否则($ID_i \neq ID_I$),从集合 S 中挑选一个元素 $\left(l_i + l_0, \left(\frac{1}{(x+l_i)}\right)P_1\right)$,将 $l_i + l_0$ 发送给 A_1,并添加新的数组 $\langle ID_i, l_i + l_0, \left(\frac{1}{(x+l_i)}\right)P_1\rangle$ 到 h_1^{list}.

2) h_2- 询问:C 维护一个初始化为空的列表 h_2^{list},表中含数组 $\langle w_i, ID_i, P_{\text{pub}}, PK_{ID_i}, h_{2i}\rangle$.当 C 接收到 A_1 对 $\langle w_i, ID_i, P_{\text{pub}}, PK_{ID_i}\rangle$ 的询问时,C 执行如下:

① 如果 $\langle w_i, ID_i, P_{\text{pub}}, PK_{ID_i}, h_{2i}\rangle$ 已经包含在列表 h_2^{list} 中,C 将 h_{2i} 返回给 A_1;

② 否则,C 选择一个随机数 $h_{2i} \in Z_q^*$,添加新的数组 $\langle w_i, ID_i, P_{\text{pub}}, PK_{ID_i}, h_{2i}\rangle$ 到 h_2^{list},并将 h_{2i} 发送给 A_1.

阶段 1 A_1 将进行下列一系列询问.

1) 部分私钥询问:当 C 接收到 A_1 对 ID_i 的询问时,C 首先从 h_1^{list} 中查找数组 $\langle ID_i, h_{1i}, D_{ID_i}\rangle$,然后执行下列操作:

① 如果 $ID_i = ID_I$,C 终止模拟(该事件用 E_1 表示);

② 否则,C 发送 D_{ID_i} 给 A_1.

2) 公钥询问:C 维护一个初始化为空的列表 PK^{list},该表含数组 $\langle ID_i, h_{1i}, PK_{ID_i}, x_i\rangle$.当 A_1 对身份 ID_i 进行公钥询问时,C 操作如下:

① 如果 $\langle ID_i, h_{1i}, PK_{ID_i}, x_i\rangle$ 已经在 PK^{list} 中,C 发送 PK_{ID_i} 给 A_1;

② 否则,C 在列表 h_1^{list} 中查找数组 $\langle ID_i, h_{1i}, D_{ID_i}\rangle$,选择一个随机数 $x_i \in Z_q^*$,计算 $PK_{ID_i} = x_i(P_{\text{pub}} + h_{1i}P_1)$,添加 $\langle ID_i, h_{1i}, PK_{ID_i}, x_i\rangle$ 到 PK^{list},并输

出 PK_{ID_i}.

3) 公钥替换询问:当 C 接收到 A_1 对 (ID_i, PK_{ID_i}') 的公钥替换询问时,C 从列表 PK^{list} 中恢复数组 $\langle ID_i, h_{1i}, PK_{ID_i}, x_i\rangle$,设置 $PK_{ID_i} = PK_{ID_i}', x_i = \perp$.

4) 私钥询问:假设 ID_i 的公钥没有被替换,当 A_1 对身份 ID_i 进行询问时,C 执行如下:

① 如果 $ID_i = ID_l$,C 终止(该事件用 E_2 表示);

② 否则,C 检查列表 h_1^{list} 和 PK^{list} 中是否已含有数组 $\langle ID_i, h_{1i}, D_{ID_i}\rangle$ 和 $\langle ID_i, h_{1i}, PK_{ID_i}, x_i\rangle$. 如果有,$C$ 令 $SK_{ID_i} = (D_{ID_i}, x_i)$,并将 SK_{ID_i} 发送给 A_1;否则,C 对 ID_i 分别执行 h_1- 询问和公钥询问来获得数组 $\langle ID_i, h_{1i}, D_{ID_i}\rangle$ 和 $\langle ID_i, h_{1i}, PK_{ID_i}, x_i\rangle$,输出 $SK_{ID_i} = (D_{ID_i}, x_i)$.

5) 陷门询问:当 A_1 对 (ID_i, w_i) 进行陷门询问时,C 操作如下:

① 如果 $ID_i = ID_l$,C 终止(该事件用 E_3 表示);

② 否则($ID_i \neq ID_l$),C 从列表 $h_1^{\text{list}}, h_2^{\text{list}}$ 和 PK^{list} 中分别调出数组 $\langle ID_i, h_{1i}, D_{ID_i}\rangle$,$\langle w_i, ID_i, P_{\text{pub}}, PK_{ID_i}, h_{2i}\rangle$ 和 $\langle ID_i, h_{1i}, PK_{ID_i}, x_i\rangle$,计算 $T_{w_i} = \dfrac{1}{x_i + h_{2i}} D_{ID_i}$,并将 T_{w_i} 发送给 A_1.

挑战阶段 A_1 输出对身份 ID^* 的2个不同的挑战关键字 w_0, w_1.C 执行如下:

1) 如果 $ID^* \neq ID_l$,C 终止(该事件用 E_4 表示).

2) 否则,C 随机挑选 $\beta \in \{0,1\}$.C 随机选择 $r \in Z_q^*$,$V^* \in G_1$,计算 $U^* = rP_1$,设挑战密文 $C^* = (U^*, V^*)$,并将 C^* 发给 A_1.

设

$$Q^* = \hat{e}\left(V^*, \frac{1}{x_{ID_S} + h_2(w_i \| ID_S \| P_{\text{pub}} \| PK_{ID_S})} D_{ID_S}\right),$$

注意到 $C^* = (U^*, V^*)$ 是一个有效密文,根据定义,下面等式成立:

$$Q^* = \hat{e}(T_{w_\beta}, U^*) = \hat{e}\left(\left(\frac{1}{((x_i + h_{2i})x)}\right)P_1, rP_1\right) = \hat{e}(P_1, P_1)^{\frac{r}{((x_i+h_{2i})x)}}.$$

阶段 2 A_1 可以继续对关键字 w_i 进行陷门询问,C 按照阶段1中描述的方式回答 A_1 的询问.但是,A_1 不能询问 w_0 和 w_1 的陷门(该事件用 E_5 表示).

猜测 最后,A_1 输出 $\beta' \in \{0,1\}$ 作为它的猜测.如果 $\beta' = \beta$,那么 A_1 在游戏中获胜.

设

$$\widetilde{Q} = \hat{e}\left(\frac{f(x)P - c_0 P}{x}, P_1 + c_0 P\right) = \hat{e}\left(\frac{f(x) - c_0}{x}P, P_1 + c_0 P\right),$$

$$\widetilde{Q}^* = (\hat{e}(P_1, P_1)^{\frac{r}{((x_i+h_{2i})x)}})^{\frac{(x_i+h_{2i})}{r}} = \hat{e}(P_1, P_1)^{\frac{1}{x}} = \hat{e}(P, P)^{\frac{f(x)^2}{x}},$$

则有:

$$\frac{\widetilde{Q}^*}{\widetilde{Q}} = \frac{\hat{e}(P_1, P_1)^{\frac{1}{x}}}{\hat{e}\left(\left(\frac{(f(x) - c_0)}{x}\right)P, P_1 + c_0 P\right)} = \frac{\hat{e}(P, P)^{\frac{f(x)^2}{x}}}{\hat{e}(P, P)^{\frac{(f(x)^2 - c_0^2)}{x}}} = \hat{e}(P, P)^{\frac{c_0^2}{x}},$$

所以,C 可以输出 $\hat{e}(P, P)^{\frac{1}{x}} = \left(\dfrac{\widetilde{Q}^*}{\widetilde{Q}}\right)^{\frac{1}{c_0^2}}$.

分析 C 在上述游戏中获胜的优势:

1) 对 A_1 的 h_1, h_2 询问的回答是有效的,且与现实世界是不可区分的,因为每个回答在 Z_q^* 中都是均匀独立分布的;

2) 如果 $E_i (1 \leqslant i \leqslant 5)$ 都不发生,则 C 没有终止,C 就可以解决 p-BDHI 问题.

显然,

$$\Pr[\neg E_1 \wedge \neg E_2 \wedge \neg E_3 \wedge \neg E_4] = \left(1 - \frac{1}{q_{h_1}}\right)^{q_{Ext}+q_E+q_T}\left(\frac{1}{q_{h_1}}\right).$$

下面证明 $\Pr[\neg E_5] \geqslant 2\varepsilon$.一方面,

$$\Pr[\beta' = \beta] = \Pr[\beta' = \beta | E_5]\Pr[E_5] + \Pr[\beta' = \beta | \neg E_5]\Pr[\neg E_5] \leqslant \Pr[\beta' = \beta | E_5]\Pr[E_5] + \Pr[\neg E_5] = \frac{1}{2} + \frac{1}{2}\Pr[\neg E_5],$$

另一方面,

$$\Pr[\beta' = \beta] \geqslant \Pr[\beta' = \beta | E_5]\Pr[E_5] = \frac{1}{2} - \frac{1}{2}\Pr[\neg E_5].$$

所以,

$$\Pr[\neg E_5] \geqslant 2\left|\Pr[\beta' = \beta] - \frac{1}{2}\right| \geqslant 2\varepsilon.$$

算法 C 解决 p-BDHI 问题的优势:

$$\varepsilon' \geqslant \frac{1}{2}\Pr[\neg E_5]\Pr[\neg E_1 \wedge \neg E_2 \wedge \neg E_3 \wedge \neg E_4] \geqslant \frac{1}{2} \cdot 2\varepsilon \cdot \left(1 - \frac{1}{q_{h_1}}\right)^{q_{Ext}+q_E+q_T} \cdot \left(\frac{1}{q_{h_1}}\right) =$$

$$\left(\frac{\varepsilon}{q_{h_1}}\right)\left(1-\frac{1}{q_{h_1}}\right)^{q_{Ext}+q_E+q_T}.$$

引理2 假设敌手 A_2 以 ε 的优势攻破了上述方案,则存在一个算法 C 以

$$\varepsilon' \geqslant \left(\frac{\varepsilon}{q_{h_1}}\right)\left(1-\frac{1}{q_{h_1}}\right)^{q_E+q_T}$$

的优势解决 CDH 问题.其中,q_{h_1},q_E,q_T 分别表示 A_2 访问 h_1 预言机、私钥询问预言机和陷门询问预言机的次数.

证明 假定给算法 C 一个 CDH 问题的实例:给定 $P,aP,bP \in G_1$,其中 $a,b \in Z_q^*$ 是随机选取的未知数.C 的目标是通过与 A_2 交互来计算出 $abP \in G_1$.

C 运行系统建立算法,选择一个随机数 $s \in Z_q^*$ 作为系统主密钥,计算 $P_{pub}=sP$.C 随机选择一个身份标识 ID_I 作为挑战身份,发送系统参数 $params=\{k,G_1,G_2,\hat{e},q,P,P_{pub},h_1,h_2\}$ 和 s 给敌手 A_2.

1)h_1- 询问:C 维护一个初始化为空的列表 h_1^{list},表中含数组$\langle ID_i,h_{1i}\rangle$.当 A_2 对 ID_i 进行询问时,C 按照下列情况响应 A_2 的询问请求:

① 如果列表 h_1^{list}中已经包含有数组$\langle ID_i,h_{1i}\rangle$,$C$ 将 h_{1i} 返回给 A_2;

② 否则,C 选择一个随机数 $h_{1i} \in Z_q^*$,添加新的数组$\langle ID_i,h_{1i}\rangle$ 到 h_1^{list},并将 h_{1i} 返回给 A_2.

2)h_2- 询问:C 维护一个初始化为空的列表 h_2^{list},表中含数组$\langle w_i,ID_i,P_{pub},PK_{ID_i},h_{2i}\rangle$. 当 A_2 对$\langle w_i,ID_i,P_{pub},PK_{ID_i}\rangle$ 进行询问时,C 操作如下:

① 如果$\langle w_i,ID_i,P_{pub},PK_{ID_i},h_{2i}\rangle$ 已经包含在列表 h_2^{list}中,C 将 h_{2i} 返回给 A_2;

② 否则,C 选择一个随机数 $h_{2i} \in Z_q^*$,添加新的数组$\langle w_i,ID_i,P_{pub},PK_{ID_i},h_{2i}\rangle$ 到 h_2^{list},并将 h_{2i} 返回给 A_2.

阶段1 A_2 将进行下列一系列询问,C 按照下面情况分别给出相应的响应:

1) 公钥询问:C 维护一个初始化为空的列表 PK^{list},表中含数组形式为$\langle ID_i,h_{1i},PK_{ID_i},x_i\rangle$.当 A_2 对身份 ID_i 进行公钥询问时,C 操作如下:

① 如果$\langle ID_i,h_{1i},PK_{ID_i},x_i\rangle$ 已经在 PK^{list} 中,输出 PK_{ID_i};

② 否则,如果 $ID_i=ID_I$,C 在列表 h_1^{list}中查找数组$\langle ID_I,h_{1I}\rangle$,计算 $PK_{ID_I}=(s+h_{1I})aP$,添加$\langle ID_I,h_{1I},PK_{ID_I},\perp\rangle$ 到 PK^{list},并输出 PK_{ID_I};

③ 否则($ID_i \neq ID_I$),C 在列表 h_1^{list}中查找数组$\langle ID_i,h_{1i}\rangle$,选择一个随机数 $x_i \in Z_q^*$,计算 $PK_{ID_i}=x_i(P_{pub}+h_{1i}P)$, 添加$\langle ID_i,h_{1i},PK_{ID_i},x_i\rangle$ 到列表 PK^{list},并将 PK_{ID_i} 发送给 A_2.

2) 私钥询问:当 C 接收到 A_2 对身份 ID_i 进行私钥询问时,C 执行如下:

① 如果 $ID_i=ID_I$,C 终止(该事件用 E_1 表示);

② 否则,C 在列表 h_1^{list}和 PK^{list} 中分别查找数组$\langle ID_i,h_{1i}\rangle$ 和$\langle ID_i,h_{1i},PK_{ID_i},x_i\rangle$, 然后,$C$ 令 $SK_{ID_i}=\left(\frac{1}{s+h_{1i}}P,x_i\right)$,并将 SK_{ID_i} 返回给 A_2.

3) 陷门询问:当 A_2 对(ID_i,w_i) 进行陷门询问时,C 执行下列操作:

① 如果 $ID_i=ID_I$,C 终止(该事件用 E_2 表示);

② 否则($ID_i \neq ID_I$),C 在列表 PK^{list} 中查找数组$\langle ID_i,h_{1i},PK_{ID_i},x_i\rangle$, 再从列表 h_2^{list} 中查找$\langle w_i,ID_i,P_{pub},PK_{ID_i},h_{2i}\rangle$,计算

$$T_{w_i}=\frac{1}{(x_i+h_{2i})(s+h_{1i})}P,$$

并将 T_{w_i} 发送给 A_2.

挑战阶段 A_2 输出 ID^* 和2个不同的挑战关键字 w_0,w_1.C 按照下列条件执行:

1) 如果 $ID^* \neq ID_I$,C 终止(该事件用 E_3 表示).

2) 否则,C 随机选择 $\beta \in \{0,1\}$.C 随机选择 $r \in Z_q^*$,$U^* \in G_1$,计算 $V^*=rbP$,设挑战密文 $C^*=(U^*,V^*)$,并将 C^* 发送给 A_2.

注意到定义 $C^*=(U^*,V^*)$ 是一个有效密文使得

$$\begin{aligned}U^* &= rb(PK_{ID_I}+h_2(w_\beta \| ID_I \| P_{pub} \| PK_{ID_I})\,T_{ID_I}) = \\ &rb((s+h_{1i})aP+h_{2i}(s+h_{1i})P) = \\ &r(s+h_{1i})abP+rh_{2i}(s+h_{1i})aP.\end{aligned}$$

阶段2 A_2 可以继续对关键字 w_i 进行陷门询问,C 按照阶段1中描述的陷门询问方式回答 A_2 的询问请求.但是,A_2 不能询问 w_0 和 w_1 的陷门(该事件用 E_4 表示).

猜测 最后,A_2 输出 $\beta' \in \{0,1\}$ 作为它的猜测.如果 $\beta'=\beta$,那么 A_2 在游戏中获胜.C 可以成功计算出 abP:

$$(U^*-rh_{2i}(s+h_{1i})aP)(r(s+h_{1i}))^{-1}=abP.$$

分析 C 在这个游戏中获胜的优势:

1) 由于对 A_2 的 h_1,h_2 询问的每个应答在 Z_q^* 中是均匀独立分布的,所以对 h_1,h_2 询问的应答与现实世界是不可区分的,并且是有效的.

2) 对 A_2 的私钥询问和陷门询问的应答是有效

的,除非 E_1 或 E_2 发生.如果 $E_i(1 \leqslant i \leqslant 4)$ 都不发生,则 C 没有终止.

显然,

$$\Pr[\neg E_1 \wedge \neg E_2 \wedge \neg E_3] = \left(1 - \frac{1}{q_{h_1}}\right)^{q_E+q_T}\left(\frac{1}{q_{h_1}}\right).$$

下面证明 $\Pr[\neg E_4] \geqslant 2\varepsilon$.一方面,

$$\Pr[\beta' = \beta] =$$

$$\Pr[\beta' = \beta \mid E_4]\Pr[E_4] + \Pr[\beta' = \beta \mid \neg E_4]\Pr[\neg E_4] \leqslant$$

$$\Pr[\beta' = \beta \mid E_4]\Pr[E_4] + \Pr[\neg E_4] =$$

$$\frac{1}{2}\Pr[E_4] + \Pr[\neg E_4] =$$

$$\frac{1}{2} + \frac{1}{2}\Pr[\neg E_4],$$

另一方面,

$$\Pr[\beta' = \beta] \geqslant \Pr[\beta' = \beta \mid E_4]\Pr[E_4] =$$

$$\frac{1}{2} - \frac{1}{2}\Pr[\neg E_4].$$

所以

$$\Pr[\neg E_4] \geqslant 2\left|\Pr[\beta' = \beta] - \frac{1}{2}\right| \geqslant 2\varepsilon.$$

算法 C 解决 CDH 问题的优势:

$$\varepsilon' \geqslant \frac{1}{2}\Pr[\neg E_4]\Pr[\neg E_1 \wedge \neg E_2 \wedge \neg E_3] \geqslant$$

$$\frac{1}{2} \cdot 2\varepsilon \cdot \left(1 - \frac{1}{q_{h_1}}\right)^{q_E+q_T} \cdot \left(\frac{1}{q_{h_1}}\right) =$$

$$\left(\frac{\varepsilon}{q_{h_1}}\right)\left(1 - \frac{1}{q_{h_1}}\right)^{q_E+q_T}.$$

5 效率分析

下面我们将从计算代价和通信代价 2 个方面分析效率,并将本文方案与 Peng 等[9]提出的无证书可搜索公钥方案进行比较.

5.1 计算代价

首先,我们给出一些定义:

1) T_{sm}:标量乘操作的运行时间.

2) T_{bp}:双线性对操作的运行时间.

3) T_H:Hash 函数映射到点的运行时间.

4) T_h:一次普通 Hash 函数操作的运行时间.

5) T_{pa}:点加运算操作的运行时间.

本文使用的基本操作运行时间如表 1 所示.实验环境为戴尔个人笔记本电脑(I5-4460S 2.90 GHz 处理器,4 GB 内存和 Window 8 操作系统),依赖 MIRACL 库[18].

表 1 基本操作的运行时间

Table 1 Running time of basic operations ms

T_{sm}	T_{bp}	T_H	T_h	T_{pa}
2.165	5.427	5.493	0.007	0.013

在表 2 和图 2 中,将本文方案与 Peng 等[9]方案的计算代价进行了比较.从表 2 和图 2 中可以看出:在密钥生成阶段、关键字加密阶段和陷门生成阶段,本文方案比 Peng 等[9]方案的计算代价分别降低了 53.93%,70.02%和 81.89%.虽然,在测试阶段,本文方案的运行时间比 Peng 等方案的运行时间略高,但是,测试阶段是由计算能力强大的云来完成的,本文方案中用户端的计算代价比较小.另外,本文方案能够抵抗关键字猜测攻击,而 Peng 等方案不能抵抗关键字猜测攻击,本文方案比 Peng 等的方案更安全.

表 2 计算代价比较

Table 2 Computation cost comparison ms

	Peng 等[9]的方案	本文方案
密钥生成	$2T_H+8T_{sm}=28.306$	$4T_h+6T_{sm}+2T_{pa}=13.044$
关键字加密	$3T_H+2T_h+5T_{sm}+3T_{bp}=43.599$	$4T_h+6T_{sm}+4T_{pa}=13.07$
陷门生成	$T_H+T_h+3T_{sm}=11.995$	$T_h+T_{sm}=2.172$
测试	$T_h+T_{sm}+2T_{pa}+T_{bp}=7.625$	$T_h+T_{sm}+2T_{bp}=13.026$

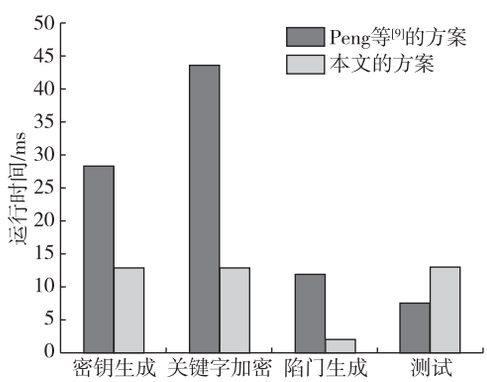

图 2 计算代价比较

Fig. 2 Computation cost comparison

5.2 通信代价

通信代价方面,我们将本文方案与 Peng 等[9]方案进行了比较,结果如表 3 所示.

从表 3 中可以看出本文方案降低了通信代价,效率更高.

表 3　通信代价比较

Table 3　Communication cost comparison

	Peng 等[9]的方案	本文方案
公钥长度	$4\|G_1\|$	$2\|G_1\|$
密文长度	$\|G_1\|+\|Z_q\|$	$2\|G_1\|$
陷门长度	$3\|G_1\|$	$\|G_1\|$

注:$|G_1|$表示群 G_1 中点的长度,$|Z_q|$表示 Z_q 中数的长度.

6　结束语

针对大数据的环境下,本文设计了一个新的支持多关键字的无需安全信道的无证书可搜索公钥加密方案,并在随机预言模型下,对该方案进行了安全性分析.安全性分析结果表明,该方案能够抵抗关键字猜测攻击;同时,效率分析显示,该方案具有较低的计算代价和较低的通信代价.因此,本文的方案更安全更有效.

参考文献

References

[1] Mayer-Schönberger V, Cukier K. Big data: A revolution that will transform how we live, work, and think[M]. Boston, NY: Houghton Mifflin Harcourt, 2013

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

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

[4] Baek J, Safavi-Naini R, Susilo W. Public key encryption with keyword search revisited[C]// International Conference on Computational Science and Its Applications, 2008: 1249-1259

[5] Rhee H S, Park J H, 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

[6] 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

[7] Shamir A. Identity-based cryptosystems and signature schemes[C]// Workshop on the Theory and Application of Cryptographic Techniques, 1984: 47-53

[8] Al-Riyami S S, Paterson K G. Certificateless public key cryptography[C]// International Conference on the Theory and Application of Cryptology and Information Security, 2003: 452-473

[9] Peng Y G, Cui J T, Peng C G, et al. Certificateless public key encryption with keyword search[J]. China Communications, 2014, 11(11): 100-113

[10] Wu T Y, Meng F, Chen C M, et al. On the security of a certificateless searchable public key encryption scheme[C]// International Conference on Genetic and Evolutionary Computing, 2016: 113-119

[11] Golle P, Staddon J, Waters B. Secure conjunctive keyword search over encrypted data[C]// International Conference on Applied Cryptography and Network Security, 2004: 31-45

[12] Chang Y C, Mitzenmacher M. Privacy preserving keyword searches on remote encrypted data[C]// International Conference on Applied Cryptography and Network Security, 2005: 442-455

[13] Wang T, Au M H, Wu W. An efficient secure channel free searchable encryption scheme with multiple keywords[C]// International Conference on Network and System Security, 2016: 251-265

[14] Chen R M, Mu Y, Yang G M, et al. Dual-server public-key encryption with keyword search for secure cloud storage[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(4): 789-798

[15] Jiang P, Mu Y, Guo F C, et al. Public key encryption with authorized keyword search[C]// Australasian Conference on Information Security and Privacy, 2016: 170-186

[16] Chen R M, Mu Y, Yang G M, et al. Server-aided public key encryption with keyword search[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(12): 2833-2842

[17] Li J G, Lin X N, Zhang Y C, et al. KSF-OABE: Outsourced attribute-based encryption with keyword search function for cloud storage[C]// IEEE Transactions on Services Computing, 2016, DOI: 10.1109/TSC.2016.2542813

[18] Shamus Software Ltd. MIRACL library[EB/OL]. (2015-05-01)[2017-05-31]. http://www.shamus.ie/index.php?page=home

南京信息工程大学学报(自然科学版),2017,9(5):462-471
Journal of Nanjing University of Information Science and Technology(Natural Science Edition),2017,9(5):462-471

A public key encryption with multiple keywords scheme for big data

MA Mimi[1] HE Debiao[2] CHEN Jianhua[1] LIU Qin[2]
1 School of Mathematics and Statistics, Wuhan University, Wuhan 430072
2 School of Computer, Wuhan University, Wuhan 430072

Abstract The continuous improvement of information technology will inevitably promote the arrival of the era of big data. Cloud computing provides a powerful data processing platform for big data. However, data security and privacy issues have aroused great concern. In this paper, we propose a new certificateless public key encryption scheme with multiple keywords search and free of secure channel for big data. We also prove that the proposed scheme can resist chosen keyword attack in random oracle model. The performance analysis shows that, the proposed scheme reduces the computation cost as well as the communication cost compared with the scheme proposed by Peng et al. in 2014.

Key words big data; certificateless public key encryption; privacy; provably secure