一类新的多关键字检索的公钥加密方案 *

黄大威[1]，杨晓元[1,2]，陈海滨[1]
（1. 武警工程学院 电子技术系 网络与信息安全武警部队重点实验室，西安 710086；2. 西安电子科技大学 计算机网络与信息安全教育部重点实验室，西安 710071）

摘 要：针对带关键字检索的公钥加密体制中多关键字间的关系，分析了 Joonsang Baek 方案在安全性和可用性方面的缺陷，结合拉格朗日插值多项式，提出一种多关键字检索的公钥加密方案。该方案实现了从大量加密数据中选出部分数据进行优先处理，且方案只生成一个陷门信息，效率得到了提升。

关键词：带关键字检索的公钥加密方案；多关键字；拉格朗日插值多项式；陷门
中图分类号：TP309 **文献标志码**：A **文章编号**：1001-3695(2010)07-2629-02
doi:10.3969/j.issn.1001-3695.2010.07.064

New public key encryption with multiple keyword search

HUANG Da-wei[1], YANG Xiao-yuan[1,2], CHEN Hai-bin[1]
(1. *Key Laboratory of Network & Information Security of APF, Engineering College of APF, Xi'an* 710086, *China*; 2. *Key Laboratory of Computer Network & Information Security of Ministry of Education, Xidian University, Xi'an* 710071, *China*)

Abstract: In view of the relationship between multiple keywords in public key encryption with keyword search, this paper analyzed the defects of Joonsang Baek's scheme in security and usability. And proposed a sort of public key encryption with multiple keywords search, making use of the Lagrange polynomial interpolation. This scheme not only gave some datas choosed from the mass encrypted datas priority in disposing, but also produced only one tarpdoor, improving the efficiency.

Key words: PEKS(public key encryption with keyward search); multiple keyword; Lagrange polynomial interpolation; trapdoor

0 引言

2004 年欧密会上，Boneh 等人[1]首次提出带关键字检索的公钥加密方案(PEKS)，提供了一种在公开数据库中检索加密数据的新方法。简单地说，PEKS 允许用户 Alice 向服务器提供陷门信息，从而在服务器中检索包含相应关键字的加密数据，除了 Alice 外其他人不能够获得任何有关该加密数据及关键字的信息。文章基于双线性对首次提出一个单关键字的 PEKS 方案。Gu Chun-xiang 等人[2]提出一个高效的 PEKS 方案，该方案在 PEKS 加密过程中没有双线性对运算，效率得到了一定的提升，并在随机预言机模型下证明了安全性，但方案同样只是针对单关键字。Baek 等人[3]对文献[1]的方案进行了分析，指出方案存在更新关键字、建立安全通道和多关键字处理等问题，提出多关键字间存在两种关系，即现实检索中的“与”“或”关系，并且提出一类多关键字的 PEKS 方案，但该方案只是对每个关键字执行一次文献[1]的操作，这种方案对于多关键字间的这两类关系不太适用。

在现实情况中，服务器中可能储存着来自多个发送方的大量加密数据，这时接收方就得选出部分相对比较紧急的数据进行优先处理，但并不知道发送方具体使用了哪些关键字。本文针对这种情况，仅以关键字的个数为检索限度，结合拉格朗日插值多项式，设计出一种多关键字检索方案，仅考虑多关键字间“与”关系，方案只需要生成一个陷门信息，在匹配阶段只需要进行一次双线性运算，效率得到提升，且方案可以通过生成不同的陷门信息来改变检索的范围。

1 预备知识

1.1 双线性映射

设 G_1 和 G_2 分别是阶为大素数 q 的循环加法群和乘法群，P 是 G_1 的生成元。随机数 $a,b\in Z_q$，假设离散对数问题在 G_1 和 G_2 中都是难解的。双线性对是这两个循环群之间的一个映射 $e: G_1\times G_1\rightarrow G_2$，满足如下性质：

a)双线性。$e(aP,bQ)=e(P,Q)^{ab}$，其中 $a,b\in Z_q$。

b)非退化性。存在 $P,Q\in G_1$，使 $e(P,Q)\neq 1$。

c)可计算性。存在有效的算法计算 $e(P,Q)$，其中 $P,Q\in G_1$。

1.2 复杂性假设

1)离散对数问题(DLP) 任取 $P,Q\in G$，找整数 $n\in Z_q^*$，使满足 $Q=nP$。

2)计算 Diffie-Hellman 问题(CDHP) 给定 P,aP,bP，计算 abP。

3)双线性 Diffie-Hellman 问题 假设 P 是加法群 G_1 的生成元，给定 G_1 中的元素 aP,bP,cP 作为输入，计算 $e(P,P)^{abc}\in G_2$。假如所有多项式时间算法拥有可以忽略的概率去解决双

收稿日期：2009-12-09；**修回日期**：2010-01-25 **基金项目**：国家自然科学基金资助项目(60842006)；武警部队科研基金课题(wjk2009014)

作者简介：黄大威(1987-)，男，硕士研究生，主要研究方向为密码学、网络与信息安全(hdw2004@yeah.net)；杨晓元(1959-)，男，教授，主要研究方向为密码学与信息安全、数字水印；陈海滨(1986-)，男，硕士研究生，主要研究方向为密码学与信息安全.

线性 Diffie-Hellman 问题,笔者认为双线性 Diffie-Hellman 问题是难以解决的。

1.3 拉格朗日插值多项式

设 q 为素数的幂,α 是有限域 $GF(q)$ 中的本原元,设要共享的秘密为 $D\in GF(q)$,随机选取 $GF(q)$ 上的 $k-1$ 次多项式 $f(x)$ 使得

$$f(x)=D+a_1x+a_2x^2+\cdots+a_{k-1}x^{k-1}$$

其中 $a_i\in GF(q)$,$i=1,2,\cdots,k-1$,且 $a_{k-1}\neq 0$。对 n 个互不相同的 α_i,$i=1,2,\cdots,n$,计算 $d_i=f(\alpha_i)$,则集合 $\{d_i\}_{i=1}^n$ 即构成一个 (k,n) 门限方案。假设 k 个参与者提供了 k 个份额 d_i,$i=1,2,\cdots,k$,则根据拉格朗日插值公式 $f(x)=\sum d_i\prod_{j\neq i}\frac{x-a_j}{a_i-a_j}$,进而得到 $f(x)$ 的常数项,即秘密 $D=f(0)$。

2 MPEKS 的定义及安全性要求

2.1 定义

一个非交互性的多关键字检索公钥加密方案包含如下多项式算法:

a) KeyGen(s)。输入安全参数 s,产生公私钥对 A_{pub},A_{priv}。

b) MPEKS(A_{pub},W)。给定 A 的公钥和关键字集合 W,产生可检索的秘密值 y。

c) Trapdoor(A_{priv},W)。给定 A 的私钥和关键字 W,产生一个陷门 T_W。

d) Test(A_{pub},y,T_W)。输入陷门 T_W、公钥 A_{pub} 和关键字检索秘密值 y,输出 1 表示匹配成功,输出 0 表示匹配失败。

2.2 安全性要求

1) 计算一致性 保证匹配算法的结论总是正确的,具体是对所有复杂性参数 k 有

$$P[(pk,sk)\leftarrow KG(k);w\leftarrow\{0,1\};$$
$$\text{Test}(pk,Td(sk,w),\text{PEKS}(pk,w))=1]=1$$

2) 安全性 保证关键字陷门信息不会泄露关键字本身。考虑以下对抗实验:

$\text{Exp}_A^{CPA}(k)$:
$(pk,sk)\leftarrow KG(k)$;
$(w_0,w_1,St)\leftarrow A_1^{Td}(pk)$,$w_0\neq w_1$;
$b\leftarrow\{0,1\}$;$y^*\leftarrow\text{PEKS}(pk,w_b)$;$d\leftarrow A_2^{Td}(y^*,St)$;
output$(d\oplus b)$;

如果一个攻击者在该对抗实验中攻击成功的概率可以忽略不计,则方案满足安全性。

3 Joonsang Baek 方案的缺陷

a) KenGen。选择阶为大素数 q 的循环加法群 G_1 和乘法群 G_2,构造双线性映射 e: $G_1\times G_1\rightarrow G_2$,选择随机数 $y\in Z_q^*$,计算公钥 $Y=yP$,随机数 xy 作为私钥。另外定义哈希函数 H_1: $\{0,1\}^*\rightarrow G_1$,H_2: $G_2\rightarrow\{0,1\}^k$。

b) MPEKS。选择随机数 $r\in Z_q^*$,对于 n 个关键字的集合,计算 $S=(U,V_1,\cdots,V_n)$,$U=rP$,$V_i=H_2(e(H_1(W_i),Y)^r)$,输出 PEKS $=S$。

c) Trapdoor。对于每个关键字计算陷门信息 $T_W=yH_1(W)$。

d) Test。分解 PEKS 为 $[U,V_i]$,对每个关键字运行匹配算法,验证是否满足 $H_2(e(T_W,U))=V_i$。

分析可知该方案只是对每个关键字运行一次 Boneh 方案,只是为多个关键字生成一个公用的 U,减少了群中的乘法运算,效率得到一定提升。方案假设加密值和陷门信息是一一对应的,但由于方案的非交互性,接收方应当不知道发送方具体使用了哪几个关键字方案,方案对于多关键字间的两类关系不太适用。同时根据文献[4]中的分析方法,用于“与”关系时,该方案还会无意泄露部分信息。如服务器中有三个 PEKS 值:$[U_1,V_1,V_2]$,$[U_2,V_1,V_3]$,$[U_3,V_2,V_3]$,V_1、V_2、V_3 分别对应关键字 W_1、W_2、W_3,那么服务器从第三个 PEKS 值中陷门信息同样可以得到前两个 PEKS 值中分别包含 W_2 和 W_3,这应当是一个 PEKS 方案应当避免的。

4 一种多关键字检索的公钥加密方案

首先选择阶为大素数 q 的循环加法群 G_1 和乘法群 G_2 构造双线性映射 e: $G_1\times G_1\rightarrow G_2$,$P$ 是群 G_1 的生成元,随机选择 $Q\in G_1$。定义哈希函数 H_1: $\{0,1\}^*\rightarrow Z_q^*$,拉格朗日系数 $\Delta_{i,w}(x)=\prod_{\substack{j\in H(W),\\ j\neq i}}\frac{x-j}{i-j}$,消息的接收方和发送方事先共享一个多关键字的集合 $W=\{W_1,W_2,\cdots,W_n\}$。

KeyGen 的步骤如下:

a) 消息接收方随机选择 $a\in Z_q^*$,计算 $A_{pub}=aQ$,$A_{pri}=a$。

b) 随机选择 $r\in Z_q^*$,生成一个 $d-1$ 阶多项式 $f(x)=r+a_1x+a_2x^2+\cdots+a_{d-1}x^{d-1}$。其中 $f(0)=r$。

c) 对于关键字集合 $W=\{W_1,W_2,\cdots,W_n\}$ 中的每个关键字 W_i 计算 $y_i=f(H_1(W_i))$,$P_i=y_iP$,并公开 P_i 作为关键字的公钥。

PEKS: 随机选择 $l\in Z_q^*$,发送方任意选择多个关键字,对于每个关键字计算 $t_i=e(P_i,lA_{pub})^{\Delta_{i,w}(0)}$;输出 PEKS $=[lQ,\{t_i\}]$。

Trapdoor:消息接收方计算陷门 $T_W=arP$,发送 $[T_W,d]$ 给服务器。

Test 的步骤如下:

a) 分解 PEKS 值为 $[A,B]$。其中 $A=lQ$。

b) 服务器匹配 $e(T_W,A)\overset{?}{=}\prod t_i$。如果匹配成功,输出 1;否则输出 0。

5 安全性及效率性分析

5.1 安全性分析

1) 计算一致性 如果服务器收到的 PEKS 值少于 d 个,那么按照秘密共享的知识显然不能匹配成功;如果服务器收到的 PEKS 值多于 d 个,那么任选 d 个运行匹配算法:

$$\prod t_i=\prod e(P_i,lA_{pub})^{\Delta_{i,w}(0)}=$$
$$e(lQ,\Sigma y_i\Delta_{i,w}(0)aP)=e(lQ,arP)=e(A,T_W)$$

另外,假如用户只需要关键字数目满足 $d-1$ 个就进行处理时,用户可以在生成陷门信息的同时自行生成一个 t_i 值,就可以完成一类新的检索。使用类似方法可以调整检索范围,使得方案更具有实用性。

2) 关键字保密性 a) 陷门信息中接收方的私钥 a 与关键字没有任何联系,而随机数 r 的安全性又是基于秘密共享方案的,要想恢复出 r,那么至少得知道 d 个 (下转第 2635 页)

通道,信息流策略已由定义 15 指定,非法流的信息流特性即包括流特性 1 ~4。

4)与 SRM 方法结果对比分析

如果在初始化 SRM 矩阵时不区分元流,原 SRM 方法将得到 67 组发送原语—中间变量—接收原语待人工分析。如果不对信息流组合过程加以监控,原始 SRM 方法将得到 57 组发送原语—中间变量—接收原语待人工分析。由于原 SRM 方法并不能记录完整的信息流路径,它们其实代表了远超此数的疑似非法流需要重建信息流序列;疑似非法流数量随流序列长度激增。初始矩阵实施的逆向传递闭包程序历经四轮传递,最后生成 31 条信息流序列。其中,21 条已在组合过程中验明是伪流;仅剩 10 条作为疑似隐蔽通道有待下一步分析;有 26 条流序列只经过两轮传递就停止延伸。

4　结束语

传统的信息流分析技术无论语义信息流分析还是安全类型分析,要求(明确或隐含)给变量附加安全标签或类型,通过推导、检查流路径中的安全标签来验证信息流安全性。然而,通常很难合理地确定单个变量安全标签,因此这类分析技术容易汇报大量伪流。伪流会耗费巨大的无效分析工作量,本文针对不同隐蔽通道信息流特性提出的检查规则框架与标志优化规则,可以有效地限制隐蔽通道信息流组合扩散,降低伪流的误报率,从而简化并加快整个隐蔽通道的分析过程。本文基于一个 B2 级安全操作系统 SLinux 的设计开发实践,重点讨论了基于动态信息流分析的隐蔽通道分类方法,描述了各类隐蔽通道的信息流表达式和信息流分类特征,分析了不同类型隐蔽通道相互间的关联关系。在此基础上,提出了隐蔽通道信息流通用检查框架,并基于不同类型的隐蔽通道特性设计了相关标志优化规则。本文研究成果与先前的一些分析技术对比取得了明显优势,在安全操作系统的隐蔽信道分析工作中收到了良好效果,特别是能够有效地限制隐蔽信道标志过程的伪非法流问题,从而降低人工分析复杂度和工作量。

参考文献:

[1] TASI C R, GLIGOR V D, CHANDERSCKARAN C S. A formal method for the identification of covert storage channels in source code [J]. IEEE Trans on Software Engineering, 1990, 16(6): 569-580.

[2] 卿斯汉, 沈昌祥. 高等级安全操作系统的设计[J]. 中国科学 E 辑:信息科学,2007,3(2):238-253.

[3] DENNING D E. A lattice model of secure information flow[J]. Communications of the ACM,1976,19(5):236-243.

[4] PORRAS P A, KEMMERER R A. Covert flow tree:a technique for identifying and analyzing covert storage channels[C]//Proc of IEEE Computer Society Symposium on Research in Security and Privacy. 1991:36-51.

[5] 聊斯汉, 朱继锋. 安胜安全操作系统的隐蔽通道分析[J]. 软件学报,2004,15(9):1385-1392.

[6] 国家质量技术监督局. GB 17859-1999,计算机信息系统安全保护等级划分准则[S]. 1999.

[7] 刘文清,陈　. 隐蔽通道标志与处理[J]. 计算机工程,2006,32(8):1-3.

[8] 刘文清,韩乃平,陈　. 一个安全操作系统 SLinux 隐蔽通道标志与处理[J]. 电子学报,2007,35(1):153-156.

(上接第 2630 页)关键字的私钥信息,而获得任一关键字的私钥就是一个双线性对中的离散对数问题,因此可以知道从陷门信息中服务器不可能获得任何有关关键字的信息。b)服务器要想从单个 PEKS 值获得关键字的信息,就相当于验证 $t_i = e(P_i\Delta_{i,w}(0), cQ)$,而 $cQ = laQ$,也就是说服务器要完成计算 Diffie-Hellman 问题(CDHP)的运算,对于多个 PEKS 值,服务器同样需要完成一个类似运算。另外即使服务器可以获得对抗实验中的陷门询问,也只是知道加密值是由多于 d 个关键字集合生成,并不能具体确定集合中有哪些关键字,因此不能获得任何有关关键字的信息,所以方案满足于关键字保密性。

5.2　效率性分析

仅考虑一些运算量较大的运算,如群里的乘法运算和双线性运算。设 g_1 为加法群 G_1 中的乘法运算,g_2 为乘法群 G_2 中的乘法运算,e 为双线性运算,设本方案中发送方选择的关键字数为 $d(d \leqslant n)$,而 n 为文献[4]中发送方选择的关键字数。由于涉及多关键字的只有文献[4]的方案,仅对本文和文献[4]中的方案作出对比(见表 1)。

表 1　方案对比

方案	KeyGen	PEKS	Trapdoor	Test
文献[4]	g_1	$g_1 + ne$	ng_1	ne
本文	$(n+1)g_1$	$de + g_1$	g_1	$e + dg_2$

6　结束语

带关键字检索的公钥加密方案提供了一种在加密数据中检索的好方法。本文仅考虑多关键字间“与”的关系,结合拉格朗日插值多项式,设计出一种多关键字检索方案。方案实现了对多关键字间满足一定条件的加密数据进行优先处理。设计多关键字的检索方案非常有必要,对于多关键字间更复杂、通用的关系还有待进一步研究。

参考文献:

[1] BONEH D, CRESCENZO G D, OSTRPVSKY R, *et al*. Public key encryption with keyword search[C]//Proc of EUROCRYPT. Interlaken, Switerland: Springer-Verlag, 2004:506-522.

[2] GU Chun-xiang, ZHU Yue-fei, ZHANG Ya-juan. Efficient public key encryption with keyword search schemes from pairings[M]. Berlin:Springer-Verlag, 2008:372-383.

[3] BAEK J, SAFIAVI-NAINI R, SUSILO W. Public key encryption with keyword search revisited[M]. Berlin: Springer-Verlag, 2008: 1249-1259.

[4] GOLLE P, STADDON J, WATERS B. Secure conjunctive search over encrypted data[C]//Proc of the 2nd International Conference on Applied Cryptography and Network Security. London:Springer-Verlag, 2004:31-45.

[5] BELLARE M, BOLDYREVA A, DESAI A, *et al*. Key-privacy in public-key encryption[C]//Proc of the 7th Conference on Theory and Application of Cryptology and Information Security. London:Springer-Verlag, 2001:566-582.

[6] BALLARD L, KAMARA S, MONROSE F. Achieving efficient conjunctive keyword searches over encrypted data[C]//Proc of the 7th International Conference on Information Communications Security. Berlin:Springer-Verlag, 2005:414-426.