一种基于纠错码的叛逆者追踪模型

周 斌,王 箭
(南京航空航天大学 信息科学与技术学院,江苏 南京 210016)

摘 要:当一些共谋者试图构造一个盗版的解码器以获取对数据内容非法访问的时候,叛逆者追踪模型可以用来发现至少其中的一个叛逆者。Chor 和 Fiat 提出了一种将数据划分为若干块并且每一块用一些不同密钥加密的叛逆者追踪模型。文中研究相同的问题,给出了一种基于纠错码并可防止最多 k 个用户共谋的追踪模型。基于纠错码的叛逆者追踪模型是通过纠错码来构造的。它是一种公开的、确定的并且有很好性能的模型,此模型只需将数据内容划分为 $8k\log N$ 块,每个数据块需要 $2k$ 个密钥。

关键词:加密;叛逆者追踪;纠错码
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2007)07-0163-04

An ECC-Based Traitor Tracing Scheme

ZHOU Bin, WANG Jian
(College of Information Science & Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Traitor tracing schemes allow detection of at least one traitor when a group of colluders attempt to construct a pirate decoder and gain illegal access to content. Chor and Fiat proposed some schemes which divide the content into some sessions and each session key is encrypted under some keys. We consider the same scenario as Chor and Fiat and present an ECC-based traitor tracing scheme, that is good against coalitions of at most k corrupt users. ECC-based traitor tracing scheme is constructed with an error correcting code. It is open, deterministic and efficient which requires session length $8k\log N$ and $2k$ keys each session.

Key words: encryption; traitor tracing; ECC

0 引 言

近年来,很多关于如何安全分发电子数据产品的模型被提了出来。一个典型的应用就是付费电视:只有付费用户可以收看相应的节目,而其他用户则不允许。显然,对特定用户可见而对其他用户不可见的任何数据都可以通过加密的方式来保护。广播加密系统(Broadcast encryption systems)[1]可以实现将信息加密并传送给一些特定的接受者。授权用户通过数据提供者给的密钥解开数据。但加密的方法不能解决共谋的问题。一些共谋者可以将明文或者将构造的可以解开加密数据的盗版解码器传送给未授权的用户。这种对数据的未授权的访问被称作盗版(piracy)。叛逆者追踪模型就是用来解决网络传输或者广播中存在的盗版问题。

叛逆者(traitor or traitors)是指一个或者一群帮助未授权用户获得数据的授权用户。这些未授权用户被称为盗版者(pirate users)。叛逆者可以通过两种方式帮助盗版者:传送解密后的数据明文;构造一个可以用来解密的盗版解码器。

目前很多叛逆者追踪模型已经被提了出来。Boneh 和 Shaw 提出了一种模型[2]:该模型通过向每一份电子数据内容的拷贝中插入不同的水印来达到区分的目的。该模型更加适合对数据明文(而非盗版解码器)泄漏的追踪。Chor 和 Fiat 给出了一些防止共谋者传送解密密钥的模型[3]。每一个合法解码器是由数据提供者分配并被一个授权用户所私有的。当一个盗版解码器被获取时,叛逆者追踪模型将被用于追踪参与构造此盗版解码器的授权用户。文献[2]和[3]中的模型都是盖然论的(probabilistic)。也就是说这些模型追踪出的叛逆者有可能出错(出错概率可以通过改变参数按照要求降到任意小)。

Fiat 和 Tassa 提出了动态追踪模型(dynamic tracing schemes)[4]。它可以实时地根据叛逆者上一次的动作改变密钥分配。进一步,文献[5]提出了顺序追踪

收稿日期:2006-09-12
作者简介:周 斌(1982-),男,江苏南京人,硕士研究生,研究方向为信息安全;王 箭,副教授,研究方向为信息安全。

模型(sequential scheme)。在该模型中密钥只被分配一次并在以后的数据传输过程中不改变,而追踪过程是动态进行的。

Dwork,Lotspiech 及 Naor 给出了自强制模型(self - enforcement)[6]。该模型将用户的一些敏感信息(比如信用卡号码)嵌入到他们得到的密钥中,因为用户不愿意把自己的密钥提供给其他人。

各种公钥追踪模型也被提了出来[7,8]。公钥追踪模型采用公钥加密的方式,因此性能比对称加密模型更好。

1 文中的工作

文中探讨了 Chor,Fiat 在文献[3]中研究的问题:防止共谋者传送解密密钥。和文献[3]一样,我们的模型是对称加密(symmetric)和静态的(static)。

对称加密和非对称加密(公钥加密)不同之处在于数据提供者和用户共享同一密钥,而非对称模型中加密和解密使用不同的密钥。相对于动态模型(dynamic)[4],静态模型中密钥只被分配一次,并在以后的数据传输及追踪过程中不改变。

文献[2,3]中的模型都是盖然论的,而文中的模型是确定性的(deterministic)。确定性是指模型追踪出的一定是叛逆者而非无辜用户。文中提出了防止 N 个用户中至多 k 个叛逆者共谋构造盗版解码器的模型。当一个盗版解码器被获取时,我们的模型确保追踪出至少其中一个叛逆者而不会伤害到其他无辜用户。此模型将整个数据内容分成 $8k\log N$ 个数据块,每一个数据块只需 $2k$ 个密钥。而且它是一个开放模型(open scheme)。

对称追踪模型可进一步分为开放的和秘密的(open or secret)。开放模型是指密钥产生/密钥分配的方法以及用户的解密方法都是公开的,而只有密钥本身是保密的。相反,秘密模型中密钥产生/密钥分配的方法以及密钥本身都是保密的。很明显,秘密模型比开放模型更难破解。但开放模型更简单易用。

文中提出的基于纠错码的开放追踪模型比文献[3]中的开放模型性能更好。文献[3]也提出了一个秘密的盖然论的模型,性能和文中的模型相当,但我们的模型同时是确定性的。

2 基本知识

下面的术语和符号将被用在文中。u 表示可以接受数据的授权用户。U 是所有授权用户的集合,$U = \{u_1,\cdots,u_N\}$,其中 u_i 是 U 集合里的第 i 个用户。整个数据被分成很多数据块,一个数据块是一段数据内容,比如几分钟的视频。每个数据块用不同的数据块密钥(session key)加密。每个授权用户被分配一组密钥,可以用来解密数据块密钥,进而解密实际数据内容。这组密钥被称作用户的个人密钥(personal key)。一些授权用户共谋,帮助未授权用户提供对数据的非法访问。这些用户被称作叛逆者。T 代表共谋者的集合,$T = \{t_1,\cdots,t_k\}$,$T \subset U$,t_i 是其中的一个叛逆者。每个共谋者提供自己的一部分密钥组成一个盗版解码器。

当一个盗版解码器被获取时,可以通过软件或者硬件的方式检测它包含什么密钥。把获取的盗版解码器看作一个黑盒,给予特定的输入检测它的输出。

一个叛逆者追踪模型包含 3 个主要的部分:

(1) 密钥产生/分配算法:数据提供者用该算法产生并分配用户的个人密钥。数据提供者有一个密钥集 α,并定义了一个映射 $P_\alpha: U \mapsto K$,其中 U 是用户集合,K 是所有的个人密钥集合。当一个用户 $u_i \in U$ 加入时,他被提供自己的个人密钥 $P_\alpha(u_i)$。

(2) 加密/解密算法:数据提供者用一个加密算法 E_α 加密数据块密钥,授权用户用解密算法 D_β 解开数据块密钥。很明显,对某个数据块密钥 s,$s = D_\beta(E_\alpha(s))$。

(3) 叛逆者追踪算法:当获取盗版解码器的时候,该算法用于确定叛逆者。

我们用对称加密算法 E 加密数据块。数据提供者把整个数据内容分成一个个大小 E 可接受的数据块。对每一个数据块 M,处理后将会生成两个部分。密文块 B_c 是用从 E 的密钥空间中随机选出的某个密钥 s 加密后的结果。s 被称作数据块密钥。另一个是使能块,它包含了让每个授权用户获取数据块密钥 s 的信息。解密一个数据块的过程可参见图 1。对每一个数据块,每个使能块包含了用一些不同密钥对 s 加密的结果。每个授权用户被分配这些密钥中的一个,以便解密出 s。每个授权用户一共得到 L 个密钥(L 是数据块的个数,后面的模型中被确定)。图 2 描述了上面的叛逆者追踪模型。

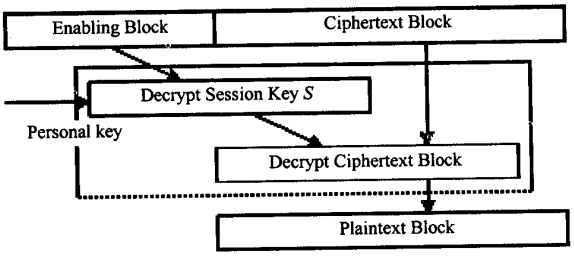

图 1 一个数据块的解密

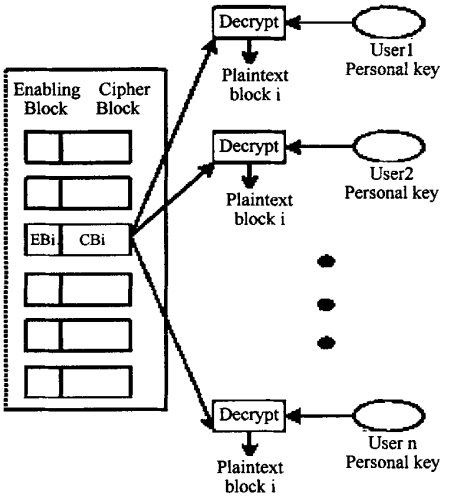

图 2　整个数据内容的解密过程

叛逆者共谋给出自己密钥中的一部分,组成盗版解码器,让未授权用户可以计算出各个数据块密钥。所以叛逆者追踪模型设计的目标是:在最多有 k 个共谋者的条件下,当一个盗版解码器被抓获时,模型至少能确定其中的一个叛逆者(不能指望可以追踪出所有的叛逆者,因为盗版解码器可能完全只由某个叛逆者提供的密钥组成,而其他叛逆者什么也不提供)。

对叛逆者追踪模型一个必需的要求是对数据块密钥的加密算法是计算上安全的。也就是说,当没有密钥的时候是不能通过其他方式破解出数据块密钥的。加密算法在计算上安全性在对称加密系统中依赖所使用密钥的长度,可以假设密钥够长,加密算法是安全的。

3　具体模型

和前文中提到的一样,我们的模型把整个数据内容划分为 L 个数据块,每个数据块密钥 s 用密钥集中的 m 个密钥加密(见图 3),L,m 的具体值在后面给出。

$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,m}$
$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_{L,1}$	$a_{L,2}$	$\cdots$	$a_{L,m}$

图 3　模型的密钥

当用户 $u_i \in U$ 加入时,他将被分配他的个人密钥 $P_a(u_i) = \{a_{1,j_{i,1}}, a_{2,j_{i,2}}, \cdots a_{L,j_{i,L}}\}$。

$a_{h,j_{i,h}}$ 是用户 u_i 的第 h 个数据块密钥的解密密钥($1 \le h \le L, 1 \le j_{i,h} \le m$)。显然,每一个个人密钥就是一个合法解码器。

为了解出数据,对每个数据块一个盗版解码器必须包含一个解密密钥。叛逆者共谋构造一个盗版解码器,对每个数据块 h,必须要有某个叛逆者 t_i 提供他的该数据块密钥解密密钥 $a_{h,j_{i,h}}$。

先给出几个定义。

定义 1　一个解码器的集合 $\Gamma = \{P_a(u_1), P_a(u_2), \cdots, P_a(u_N)\}$ 被称作一个 (L,N) - 解码器集,其中 $P_a(u_i) = \{a_{1,j_{i,1}}, a_{2,j_{i,2}}, \cdots, a_{L,j_{i,L}}\}$。解码器 $P_a(u_i)(1 \le i \le N)$ 将被分配给用户 u_i 作为他的个人密钥。用 $Y(P_a(u_i), h)$ 代表用户 u_i 的第 h 个数据块密钥的解密密钥,也即 $Y(P_a(u_i), h) = a_{h,j_{i,h}}$。

我们的目标是构造一个共谋安全的叛逆者追踪模型,T 表示共谋叛逆者的集合。首先给出所有 T 能够构造的盗版解码器。

定义 2　$\Gamma = \{P_a(u_1), P_a(u_2), \cdots, P_a(u_N)\}$,是一个 (L,N) - 解码器集,$T = \{t_1, \cdots, t_k\}$,是一个共谋叛逆者的集合。$Y(T,h) = \bigcup_{i=1}^{k} \{Y(P_a(t_i), h)\}$,表示 T 的对第 h 个数据块密钥的解密密钥集合。定义 T 构造的盗版解码器集为 $F(T,\Gamma) = \bigodot_{i=1}^{L} Y(T,i)$,其中 $\odot$ 表示笛卡儿集。

如果一个盗版解码器包含用户 u_i 的很多解密密钥,将会怀疑 u_i 是一个叛逆者。但是,u_i 能够声称他是被诬陷的,共谋者构造了一个包含很多他的密钥的解码器。因此,我们的模型就是要构造一个 (L,N) - 解码器集,并且满足下面的特性:当一个盗版解码器被获取时,模型一定能够确定其中一个叛逆者,而不会诬陷其他用户,同时限制共谋用户的个数最大为 k,称它为 k - ECC (L,N) - 解码器集。

下面将具体介绍如何构造一个 k - $ECC(L,N)$ - 解码器集。主要思想是把 $(1,k)$ - 解码器集和一个纠错码相结合构造一个解码器集,此解码集在一定的参数条件下,可以实现叛逆者追踪。先回顾一下纠错码的基本定义。

定义 3　δ 是包含 N 个长度为 L 的字符串的集合,其中字符串是基于有 p 个字符的字符集上,如果 δ 中的任两个字符串之间的汉明距都至少是 D,则称 δ 是一个 $(L,N,D)_p$ - 纠错码,简称为 $(L,N,D)_p$ - ECC[2,8]。可以给含 p 个字符的字符集中的每个字符 r 编一个顺序号,用 $SN(r)$ 表示。

上面已经提到了,文中的模型把整个数据内容划分为 L 个数据块,每个数据块密钥 s 用密钥集中的 m 个密钥加密(见图 3)。很显然,每一行彼此之间是独立的,并且可以看作是一个 $(1,m)$ - 解码器集 Γ_i =

$(a_{i,1},a_{i,2},\cdots,a_{i,m})(1\leq i\leq L)$。

定义 4 δ 是一个 $(L,N,D)_m-ECC$，$\Gamma_i=\{a_{i,1},a_{i,2},\cdots,a_{i,m}\}(1\leq i\leq L)$ 和 δ 按下面的方式组合成的解码器集 Γ'：对每个字符串 $v_j=v_{j1},v_{j2},\cdots,v_{jL}\in\delta$，定义：

$P_a(v_j)=\{a_{1,SN(v_{j1})},a_{2,SN(v_{j2})},\cdots,a_{L,SN(v_{jL})}\}$，$SN(v_{ji})$ 是字符 v_{ji} 在字符集中的顺序号$(1\leq j\leq N)$。其中 $SN(v_j)=\{SN(v_{j1}),SN(v_{j2}),\cdots,SN(v_{jL})\}$ 组成了 $P_a(v_j)$ 的 L 个连续标记。Γ' 是 δ 中所有字符串按以上方式组合成的解码器集，即 $\Gamma'=\{P_a(v_j)\mid v_j\in\delta\}$。$\Gamma'$ 就被称作 $ECC(L,N)$ - 解码器集。

和文献[5]中的定理 4 类似，下面给出定理 1 和证明。定理 1 描述了一个 $k-ECC(L,N)$ - 解码器集能追踪出至少一个叛逆者需要满足的参数条件。

定理 1 $\Gamma_i=\{a_{i,1},a_{i,2},\cdots,a_{i,m}\}(1\leq i\leq L)$ 是一个 $(1,m)$ - 解码器集，δ 是一个 $(L,N,D)_m-ECC$，Γ' 是 $\Gamma_i(1\leq i\leq L)$ and δ 组合成的 $ECC(L,N)$ - 解码器集。如果 $D>L(1-\frac{1}{k})$，那么当最多只有 $k(k\leq m)$ 用户共谋时，一定可以根据获取的盗版解码器追踪出至少其中一个共谋者，即 Γ' 是个 $k-ECC(L,N)$ - 解码器集。

证明(概要) Γ' 是由 $\Gamma_i(1\leq i\leq L)$ and δ 组合成的 $ECC(L,N)$ - 解码器集。将 Γ' 中的每一个解码器分配给一个用户(Γ' 总共有 N 个解码器)。假设某个用户的解码器是由 δ 中的字符串 v_j 构造而来，$SN(v_j)$ 是他的 L 个连续标记。不妨假设获取的盗版解码器是 $\{a_{1,j_1},a_{1,j_2},\cdots,a_{1,j_L}\}$，其中 $\{j_1,j_2,\cdots,j_L\}$ 也组成了 L 个连续标记。在最多只有 k 个共谋者的条件下，容易看出当 $L\geq k(L-D+1)$ 时，一定有某个用户的 L 个连续标记与 $\{j_1,j_2,\cdots,j_L\}$ 至少在 $(L-D+1)$ 个位置上相同。此时这个用户就可认定是一个叛逆者。

所以为了追踪出一个叛逆者，解码器的长度必须满足 $L\geq k(L-D+1)$。而定理中的 $D>L(1-\frac{1}{k})$ 确保了该条件成立。

按照定理 1，为了能构造 $k-ECC(L,N)$ - 解码器集，必须构造出满足 $D>L(1-\frac{1}{k})$ 的 $(L,N,D)_m-ECC$。这样的纠错码是否存在，该如何构造，定理 2 给出了答案。

定理 2 对任何整数 k 和 N，让 $L=8k\log N$，那么一定存在一个 $(L,N,D)_{2k}-ECC$ 并且满足 $D>L(1-\frac{1}{k})$(见文献[2]中引理 3.3 和文献[8])。

下面结合定理 2 给出如何构造 $k-ECC(L,N)$ - 解码器集。

定理 3 对任何整数 k 和 N，让 $L=8k\log N$，那么一定存在一个 $k-ECC(L,N)$ - 解码器集。

证明 把整个数据内容分成 L 个数据块，$L=8k\log N$，并且每个数据块 s 用 $m=2k$ 个密钥加密(见图 4)。显然，每一行是一个 $(1,k)$ - 解码器集 $\Gamma_i=\{a_{i,1},a_{i,2},\cdots,a_{i,m}\}\quad(1\leq i\leq L)$

$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,2k}$
$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,2k}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_{L,1}$	$a_{L,2}$	$\cdots$	$a_{L,2k}$

图 4 所构造模型的密钥

由定理 2，一定存在一个 $(L,N,D)_{2k}-ECC$ 并且满足 $D>L(1-\frac{1}{k})$。结合定理 1 可构造一个 $k-ECC(L,N)$ - 解码器集。

通过上面的分析，不难看出我们的模型是开放的确定性的。$2kL$ 个密钥是唯一需要保密的信息。当至多有 k 个叛逆者共谋时，我们的模型可以通过和用户的个人密钥比照从而发现叛逆者。确定叛逆者的算法性能很高，因为它只是序列的匹配。

我们的模型把数据内容分成个 $8k\log N$ 数据块，每个数据块密钥 s 用 $m=2k$ 个密钥来加密。对每个用户而言需要的存储空间是 $L=8k\log N$ 个密钥的大小。每个数据块的使能块区域需要存储 $2k$ 个对数据块密钥 s 加密的结果。表 1 中将文献[3]中提供的几种模型与我们的模型进行了性能对比，容易看出我们模型的性能更好(表中 w - 叛逆者不能被追踪出的概率；L - 数据内容需要划分的数据块个数；m - 每个数据块密钥需要加密的次数)。

表 1 几种模型的性能对比

模型	L	m
open one - level	$o(k^2\log N)$	$o(k^2)$
open two - level	$o(k^2\log^2 k\log(N/k))$	$o(k\log^2 k)$
secret one - level	$o(k\log(N/w))$	$o(k)$
ECC - Based Scheme	$o(k\log N)$	$o(k)$

4 总 结

叛逆者追踪模型用于在盗版解码器被发现时追踪出共谋的叛逆者。文中提供了一种基于纠错码追踪模型。它是由纠错码构造而来，是开放的、确定性的、性能高的，能很好地防止最多 k 个叛逆者共谋。

(下转第 203 页)

特征对于分类器的区别能力要比其他视觉特征强,直观上用于实验的各类图像在边缘上的特征的区分度较强,因此用形状特征能够比较容易地区别各类。

图 2　实验数据库中的图像

在对待标注图像进行语义分类以后,其所属语义类的语义标签就自动地被图像所继承。然后再利用统计模型在语义类内部进行基于统计的图像标注,这样,图像语义标注的效率就大大地得到了提高。在所做的实验中,基于形状特征的实验达到了 81.5% 的较高的标注准确率,证明了算法的有效性和在文物图像上的适用性。图 3 对于算法在形状特征和纹理特征上各自的标注精度(标注准确率)做了相应的对比。

测试数据	纹理特征	形状特征
训练图像集	52.8%	81.5%
测试图像集	47.4%	71.4%
所有图像集	50.1%	75.0%

图 3　实验结果数据

4　结束语

图像语义标注作为图像检索领域一个较新的研究方向,是减小图像视觉特征和图像语义鸿沟的一种有效手段。基于语义分类的文物图像标注方法在实验中表现出了较好的标注准确率和效率。但图像语义标注在文物图像中的应用还有很多尚未解决的问题,需要做进一步深入的研究。

参考文献:

[1] 张鸿斌,陈　豫.连接基于内容图像检索技术中的语义鸿沟[J].情报理论与实践,2004(2):196－198.

[2] 沈青松.图像语义标注与检索及在数字图书馆中的应用[D].杭州:浙江大学,2006.

[3] Mori Y, Takahashi H, Oka R. Image－to－word transformation based on dividing and vector quantizing images with words [C]// In MISRM'99 First International workshop on multimedia intelligent storage and retrieval management. [s. l.]: [s. n.], 1999.

[4] Duygulu P, Bamard K, de Freitas N, et al. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary[C]// The 7th European Conf on Computer Vision. Copenhagen, Denmark:[s. n.], 2002.

[5] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross－media relevance models[C]// In: Proc of the 26th Annual Int'l ACM SIGIR Conf. New York: ACM Press, 2003:119－126.

[6] 彭青松.多媒体交叉参照检索和语义自动标注[D].杭州:浙江大学,2005.

[7] Fu Y, Wang W, Gao W. Content－Based Natural Image Classification and Retrieval Using SVM[J]. Chinese Journal of Computers, 2003, 26(10):1260－1265.

[8] 许天兵.一个用语义分类实现的图象检索框架[J].计算机工程与应用,2003(2):106－107.

(上接第 166 页)

尽管基于纠错码的叛逆者追踪模型有较高的性能,但提出划分更少数据块个数和需要更少加密密钥的新模型仍需要进一步研究。

参考文献:

[1] Fiat A, Naor M. Broadcast encryption[C]//In Advances in Cryptology－CRYPTO'93 Lecturé Notes in Computer Science. [s. l.]:[s. n.], 1994:480－491.

[2] Boneh D, Shaw J. Collusion－secure fingerprinting for digital data[J]. IEEE Trans Inform Theory, 1998, 44:1897－1905.

[3] Chor B, Fiat A, Naor M. Tracing traitors[C]//in Proc. Advances in Cryptology－Crypto '94. Santa Barbara, California: Springer－Verlag, 1994:257－270.

[4] Fiat A, Tassa T. Dynamic traitor tracing[C]//in Proc. Advances in Cryptology－Crypto '99. Santa Barbara, California: Springer－Verlag, 1999:388－397.

[5] Safavi－Naini R, Wang Y. Sequential Traitor Tracing[C]// Proc Crypto 2000. Santa Barbara, California: Springer－Verlag, 2000.

[6] Dwork C, Lotspiech J, Naor M. Digital signets: Self－enforcing protection of digital information[C]//Proceeding of the 28th Annual Symposium on Theory of Computing. [s. l.]: ACM, 1996.

[7] Boneh D, Franklin M. An efficient public key tracing scheme [C]//in Proc Advances in Cryptology－Crypto'99. Santa Barbara, California: Springer－Verlag, 1999:338－353.

[8] van Lint. Introduction to coding theory[M]. Berlin: Springer－Verlag, 1982.