

数字指纹印中有追踪叛徒能力码的新构造^{*}

蔡庆军, 裴定一

(广州大学数学与信息科学学院, 广东 广州 510405)

摘 要: 如何得到可以追踪盗版者的码是指纹印研究中的一个重要问题。利用 Lovász 引理给出了可认定父元码存在的一个充分必要条件的一个新的证明。然后首次利用码的级联方法构造了一个新的可追踪码。最后给出了几个可追踪码存在的充分条件, 经典的码存在的充分条件可以由前面给出的一个条件推出, 而且当某个码满足某种性质时, 还得到了一个新的码存在的充分条件, 并且通过给出一个例子, 表明了这样的码是存在的。

关键词: 可认定父元码; Lovász 引理; 可追踪码

中图分类号: TP309 **文献标识码:** A **文章编号:** 0529-6579 (2004) S2-0098-03

随着计算机通信技术的迅速发展和多媒体存储及传输技术的进步, 快速复制和传输数字化信息变的越来越容易。然而, 这也使盗版者(称为叛徒)能以低廉的成本复制并出售未经授权的数字产品, 从而谋取暴利。虽然密码学的加密技术是保护数字产品的一种方法, 但它只能够保护数字产品安全传输, 但它却不能保证数字产品解密后的盗版问题。数字水印和数字指纹印就是近年来比较热门的两种数字产权保护技术。数字指纹印技术是由数字产品的版权所有向分发给每个用户的拷贝中引入唯一的标记, 并保存售出的拷贝中对应指纹印的用户身份的数据库, 将来发现盗版产品时, 发行者可以通过检测其中的标记, 即“指纹”, 来追踪该盗版的原始购买者。虽然数字指纹印只能被动地防盗版, 即它们不能防止制作盗版, 但一旦发现某个盗版后至少可以找到一个生成该盗版的用户, 从而可以让盗版者摄于遭到重罚而不敢制造盗版。

指纹印是通过编码的方式嵌入到产品中的。设计好的指纹印编码, 使得在得到某个盗版产品中的假指纹印后都可以追踪到至少一个叛徒就成为一个关键的问题。为了解决这些问题, 文[1]和文[2]中引入了几类有追踪能力的码。可追踪码(traceability code, 当合谋用户个数最多为 t 简记为码 t -TA) 是在文[1]中引入的, 这类码可以高效准确地追踪到至少一个叛徒。有可认定父元性质的码(identifiable parent property code, 当合谋用户个数最多为 t 时简记为码 t -IPP) 是在文[2]中引入的, 顾名思义, 当合谋用户个数小于等于 t 时, 从这些合谋用户生成的盗版中的假“指纹印”一定可

以无误地追踪到某个叛徒。当然这时的追踪算法可能是采取穷搜索的办法。文[2]证明了 t -IPP 码存在当且仅当 $t < q$, 这里 q 是码的字母个数。

若 $C \subseteq \mathbf{Q}^n$, 且 $|C| = M$, 码的最小 Hamming 距离为 d , 则称该码为码 (n, M, d) q -码。 C 中的元素称为码字, 每个码字表为 $x = (x_1, x_2, \dots, x_n)$, 这里 $x_i \in \mathbf{Q}, 1 \leq i \leq n$ 。若 C_0 为 C 的一些码字组成的集合, 则定义 C_0 的导出元集为 C_0 中的码字生成的所有 n -向量。用 $\text{desc}(C_0)$ 来表示这个集合, 即 $\text{desc}(C_0) = \{x \in \mathbf{Q}^n \mid x_i \in \{a_i \mid a \in C_0\}, 1 \leq i \leq n\}$

若 t 为一个正整数, C 的 t 导出元集定义为 C 中所有至多 t 个码字组成的子集合生成的所有 n -向量。即 $\text{desc}_t(C) = \bigvee_{C_0 \subseteq C, |C_0| \leq t} \text{desc}(C_0)$

定义 1 若对 $x \in \text{desc}_t(C)$, 有

$$\{i \mid x \in \text{desc}_t(C_i), |C_i| \leq t\} C_i \neq \emptyset$$

则称 C 为 t -IPP 码, 用符号 $t\text{-IPP}(v, b)$ 表示该码的码长 v , 码字的个数为 b 。

下面给出一个 4 个元素上的 3-IPP 码的例子:

例 1 $c_1 = (000), c_2 = (111), c_3 = (222), c_4 = (333)$ 。由于每个码字在第一位都不相同, 则这个码是 3-IPP 码。

1 可认定父元码的码率的一个下界

t -IPP 码的存在问题一直是大家关心的一个重要问题。文[2]中证明了存在 t -IPP 码当且仅当 $q \geq t+1$ 。利用 Lovász 局部引理对 IPP 码的码率进行讨论, 并用 Lovász 局部引理简化了文[2]中的证

^{*} 收稿日期: 2004-09-11

作者简介: 蔡庆军 (1976 年生), 男, 讲师, E-mail: qingjuncai@sohu.com

明过程。

首先引入一些定义和记号。

令 $R = R(C) = \log_q M/n$ 表示码 C 的码率。
 $R_q(t) = \lim_{n \rightarrow \infty} \inf \max R(C_n)$, 这里最大值取自所有的长为 n 的 t -IPP 码 C_n 。 $H(x) = -x \log_2 x - (1-x) \log_2 (1-x)$ 表示熵函数。令 $A_1, \dots, A_n$ 为 n 个事件。

定义 2 事件 A 与事件 B 称为是独立的, 若 $\Pr(A \mid B) = \Pr(B)$ 。

定义 3 事件 A 与事件 $B_1, \dots, B_m$ 事件称为是相互独立的, 若有 $\Pr(A \mid \bigcap_{i \in I} B_i) = \Pr(A)$ 成立, 对任意 $I \subseteq \{1, \dots, m\}$ 。

注意事件 A 可以与事件 $B_1, \dots, B_m$ 中的每个都独立, 而与整个 n 个事件不相互独立。

定义 4 结点集 $V = \{1, \dots, n\}$ 上的一个图 $G = (V, E)$ 称为依靠图, 若对图中的结点 i , 有结点 j 在 G 中不与结点 i 邻接 (即 $\{i, j\} \notin E$), 则事件 A_i 与 A_j 事件是独立的。显然由 n 个事件可得到一个依靠图。

下面就可以引入一个重要的引理。

引理 1 Lovász 局部引理 若图 G 是 n 个事件上的依靠图, 且 G 的每个结点的度至多为 d , 且 $\Pr(A_i) \leq p$ (对所有的 i)。当 $4pd \leq 1$ 时, 就有 $\Pr(\overline{A_1} \circ \overline{A_2} \circ \dots \circ \overline{A_n}) > 0$ 。

此时我们可以证明如下的定理了。

定理 1 令 $u = \lceil (t/2 + 1)^2 \rceil$, 则有

$$R_q(t) \geq \frac{1}{u-1} \log_q \frac{(q-t)! q^u}{(q-t)! q^u - q! (q-t)^{u-t}}$$

特别的, 若 $q \geq t+1$, 则 $R_q(t) > 0$ 。

证明 对 M 个码字的 u -子集 U , 令 A_U 表示 U

不是 t -IPP 码的事件, 则 $\Pr(A_U) = \left\{ \begin{matrix} u \\ t \end{matrix} \right\} P_{T,U,C}$, 其中

$$P_{T,U,C} = \left[1 - \frac{q! (q-t)^{u-t}}{(q-t)! q^u} \right]^n$$

。令 G 为一个图, 其中每个结点 i 对应一个 u -子集 U_i , 满足 $U_i \cap U_j \neq \emptyset$ 当且仅当两个结点 i 和 j 是相连的。这个图的度为

$$d = \left\{ \begin{matrix} M \\ u \end{matrix} \right\} - \left\{ \begin{matrix} M-u \\ u \end{matrix} \right\}$$

易见 G 是 A_U 对应的相关图, 则由引理 1 知当 $4pd \leq 1$ 时, 存在 t -IPP 码。而 $4pd \leq 1$ 可如下变化:

$$\left\{ \begin{matrix} M \\ u \end{matrix} \right\} - \left\{ \begin{matrix} M-u \\ u \end{matrix} \right\} < \left[4 \left\{ \begin{matrix} u \\ t \end{matrix} \right\} \left(1 - \frac{q! (q-t)^{u-t}}{(q-t)! q^u} \right)^n \right]^{-1}$$

不等式的两边取以 q 为底的对数, 然后两边同除

以 n 有

$$n^{-1} \log_q [u^2 M^{u-1} + O(M^{u-2})] < -\log_q \left[1 - \frac{q! (q-t)^{u-t}}{(q-t)! q^u} \right]$$

将 $M = q^{Rn}$ 代入上式并令 $n \rightarrow \infty$ 有

$$R(u-1) \leq \log_q \frac{(q-t)! q^u}{(q-t)! q^u - q! (q-t)^{u-t}}$$

此时由于 $R_q(t) = \lim_{n \rightarrow \infty} \inf \max R(C_n)$, 则有

$$R_q(t) \geq \frac{1}{u-1} \log_q \frac{(q-t)! q^u}{(q-t)! q^u - q! (q-t)^{u-t}}$$

特别的, 若 $q \geq t+1$, 则 $R_q(t) > 0$ 。

Lovász 局部引理虽然在文 [2] 中用过, 但只是对 $t=2$ 的情况。虽然这个定理与文 [3] 的定理 4 结果一样, 但这个证明更简单。

2 t -可追踪码

定义 5 若 $x, y \in Q^n$, 令

$$I(x, y) = \{i: x_i = y_i\}$$

$$I(x, y, z) = \{i: x_i = y_i = z_i\}$$

若 $v \in \text{desc}_t(C)$, 且有某个 i , 使得当 $v \in \text{desc}(C_i)$ 时, 其中 $C_i \subseteq C$, $|C_i| \leq t$, 最少存在一个码字 $y \in C$, 使得对所有 $z \in C \setminus C_i$ 都有 $|I(v, y)| > |I(v, z)|$ 。那么 C 称为 t -可追踪码。用符号 $t\text{-TA}(v, b)$ 表示该码的码长为 v , 码字的个数为 b 。有时简称 $t\text{-TTA}$ 码。

$t\text{-TA}$ 码与上一节给出的 $t\text{-IPP}$ 码实际有如下的关系。

定理 2 $t\text{-TA}(N, M)$ 码是 $t\text{-IPP}(N, M)$ 码, 但反之则不然。反例见例 2。

例 2 $2\text{-IPP}(3, 4)$ 码但不是 $2\text{-TA}(3, 4)$ 码。

$C = \{(122), (230), (322), (003)\}$, 由于这 4 个码字的第一位都不相同, 则由它们生成的“假指纹”一定可以找到某个叛徒。即这个码是 $4\text{-IPP}(3, 4)$ 码, 由 $t\text{-IPP}$ 码的定义知这个码还是 $2\text{-IPP}(3, 4)$ 码。但这个码不是 $2\text{-TA}(3, 4)$ 。令某个“假指纹”为 $v = (222)$, 则 v 是 $(230), (122)$ 的导出元。但是

$$|I(v, (230))| = 1,$$

$$|I(v, (122))| = |I(v, (322))| = 2$$

这样这个码不满足 2-TA 码的定义。

3 存在 t -可追踪码的几个充分条件

在所知的文献中, 级联方法被用来构造 IPP 码, 但却没有用这种方法来构造 TA 码。这是因为一般 2 个 $t\text{-TA}$ 码的级联构造不一定是 $t\text{-TA}$ 码。但我们却有下面的定理。

引理 2^[1] 若一个 (N, n) 码 C 满足 $d > N(1 -$

$\frac{1}{t^2}$), 则 C 是 t -TA 码。

定理 3 若级联码的内码 C_I 为 $(N, n, d_I)_q$, 且 $d_I > N\left(1 - \frac{1}{2t^2}\right)$, 外码 C_o 为 $(l, m, d_o)_n$, 且 $d_o > l\left(1 - \frac{1}{2t^2}\right)$, 则级联码 $C_{I,o}$ 是 t TA (lN, m) 码。

证明 由于级联码的最小距离 $d \geq d_I d_o > l\left(1 - \frac{1}{2t^2}\right) \cdot N\left(1 - \frac{1}{2t^2}\right) > lN\left(1 - \frac{1}{t^2}\right)$, 则由引理 2 可知级联码 $C_{I,o}$ 是 t TA (lN, m) 码。

这样的码确实是存在的。内码是 (N, q') 的 RS 码, 其中 $t = \left\lfloor \frac{N}{2t^2} \right\rfloor$, 外码是 $(l, (q')')$ 的 RS 码, 其中 $t' = \left\lfloor \frac{l}{2t^2} \right\rfloor$ 。

这样就得到了期望的 t TA 码。

接下来给出 t TA 码存在的一些充分条件。

定理 4 若对任意 t 个码字 $C_o = \{c_1, c_2, \dots, c_t\}$, $c \in C \setminus C_o$, 有 $\sum_{i=1}^t d(c_i, c) > Nt - \frac{N}{t}$, 则 C 是 t TA 码。

证明 设 $\{c_1, c_2, \dots, c_t\}$ 生成的某个向量为 v 。由于 $\{c_1, c_2, \dots, c_t\}$ 生成 v , 且 C 的码长为 N , 则必存在某个码字, 不妨设为 c_1 , 使得 $|I(c_1, v)| \geq \left\lfloor \frac{N}{t} \right\rfloor$,

由于 $|I(c, v)| \leq \left| \sum_{i=1}^t (c, c_i) \right|$ 和 $\sum_{i=1}^t d(c, c_i) > Nt - \frac{N}{t}$ 知 $|I(c, v)| < \left\lfloor \frac{N}{t} \right\rfloor$, 即 $|I(c, v)| < |I(c_1, v)|$, 可见 C 是 t TA 码。

推论 1 若 C 是线性码, 且 C 的任何 t 个码字的重量和 $\sum_{i=1}^t w\left(c_i > Nt - \frac{N}{t}\right)$, 则 C 是 t TA 码。

定理 5 若对任何 $x, y, z \in C$, $\max_{x,y} \{ |I(x, y)| \} = l$, 且 $|I(x, y, z)| = l' < l$, 则当

$l < l' + \left\lfloor \frac{N-l}{2} \right\rfloor$ 时, C 是 2 TA 码。

证明 若 c_1, c_2 生成向量 v , 则 v 必与的某个码字, 设为 c_1 使得 $|I(v, c_1)| > l + \frac{N-l}{2}$, 对 $C \setminus \{c_1, c_2\}$ 中的码字 c' , 都有

$$|I(v, c')| \leq |I(c_1, c_2)| +$$

$$|I(c_2, c')| - |I(c_1, c_2, c')| < 2l - l'$$

由于 $l < l' + \frac{N-l}{2}$, 则 $|I(v, c')| < |I(v, c_1)|$, 于是 C 是 2 TA 码。

满足定理 5 条件的码确实是存在的。

例 3 若码 C 是 $\{(111222333), (122233311), \{133211322)\}$ 。则由于 $N=9, l=3, l'=3$, 由于 $l < l' + \frac{N-l}{2}$, 则由定理 5 知 C 是 2 TA 码。

若码的最小距离 $d = N - l > 3N/4$, 则 $\max_{x,y} \{ |I(x, y)| \} = l < N/4$, 于是 $\frac{N-l}{2} \geq \frac{3Nl}{8} \geq \frac{3N}{8} > \frac{N}{4}$

$> l$, 这样就有了 $l < l' + \left\lfloor \frac{N-l}{2} \right\rfloor$, 由定理 5 知 C 是

2 TA 码。由于 $d = N - l > \frac{3N}{4}$, 则引理 2 当 $t=2$ 时,

实际就是定理 5 的一个推论。

参考文献:

- [1] STADDON J N, STINSON D R, WEI R. Combinatorial properties of frameproof codes and traceability codes [J]. IEEE Trans Information Theory, 2001, 47(3): 1042-1049.
- [2] HOLLMANN H D L, van LINT J H, LINNARIZ J P, et al. On codes with the identifiable parent property [J]. J Combinatorial Theory, Series A, 1998, 82(2): 121-133.
- [3] BARG A, COHEN G, ENCHEVA S, et al. A hypergraph approach to the identifying parent property, the case of multiple parents [J]. SIAM J Disc Math, 2001, 14(3): 423-432.

The New Results of the Codes Having Traceability in Fingerprinting

CAI Qing jun, PEI Ding yi

(School of Mathematics and Information Sciences, Guangzhou University, Guangzhou 510405, China)

Abstract: How to get good frameproof codes and secure frameproof codes is an important problem in the fingerprinting. Using Lovász lemma, a new proof of the sufficient and necessary condition of the existence of t -IPP code is given. Firstly, the concatenation construction of t -TA code is given, which had not talked about before. Several sufficient conditions of the existence of t -TA code is given. The classical existence condition of t -TA code can be deduced by one of sufficient conditions of the existence of t -TA code given in former subsection. Moreover, a new sufficient condition of existence of 2 TA is given when the code has some property.

Key words: identifiable parent property code; Lovász lemma; traceability code