带列表译码的安全码

Yujie Gu、Ilya Vorobyev、Ying Miao

arXiv:2302.02414v1 [cs.IT] 5 Feb 2023

摘要

本文研究用于叛徒追踪的组合安全码，以保护多媒体内容的版权。首先，我们为抗合谋多媒体指纹提出一种新的带列表译码的安全码（SCLD）概念，其中涵盖了许多现有指纹码类型，作为其特例。接下来，我们为具有完全可追踪性的 SCLD 构建高效的识别算法，并确定其最大可能码率的界。与现有指纹码相比，SCLD 不仅具有远高于可分离码的可追踪效率，而且码率也远高于防伪码。此外，我们还确定了二元可分码最大码率的新界。进一步地，我们为动态场景下的多媒体指纹提出了一个两阶段动态叛徒追踪框架，该框架不仅能够高效实现完全可追踪性，还能提供远高于静态场景的容量。

索引词

安全码、列表译码、动态叛徒追踪、版权保护、二进制码、码率

I. 引言

通信网络的发展及其普及，极大地促进了多媒体内容的传播和应用，例如文本、音频、图像、动画和视频；因此，如何确保这些多媒体内容仅由获得授权的合法用户用于预定目的，已成为一项迫在眉睫且充满挑战的任务。为防止多媒体内容遭到恶意再分发，人们提出了数字指纹技术。该技术的优势在于，可以通过水印技术将指纹嵌入多媒体内容中 [13] ；一旦获得授权的用户非法将内容用于非预定目的，便可识别出这些恶意用户 [5], [10]。

正交调制和码调制是将指纹嵌入多媒体内容 [42]的两种典型方法。本文研究后者，即码调制情形；与前者正交调制使用相同数量的正交信号 [42] 相比，它能够容纳更多用户。下面简要回顾码调制。设宿主信号是一个长度为 m的实向量 $\mathbf{x} \in \mathbb{R}^m$ 。在码调制中，有 n 个标准正交基信号 $\{\mathbf{u}_i \in \mathbb{R}^m : \ 1 \le i \le n\}$ ，用户通常不知道这些信号；水印信号（即指纹）$\mathbf{w}_j$ 按如下方式生成：$\mathbf{w}_j = \sum_{i=1}^{n} \mathbf{c}_j(i)\mathbf{u}_i$ ，其中 $\mathbf{c}_j(i) \in \{0,1\}$ ，符合通断键控（OOK）调制。因此，用户 j 将获得 $\mathbf{x}$ 的嵌入指纹的信号副本 $\mathbf{y}_j = \mathbf{x} + \mathbf{w}_j$ 。这是可行的，因为多媒体数据对数据值的微小扰动在感知上并不敏感 [42]。显然，授权用户 j 与指纹 $\mathbf{w}_j$之间存在一一对应关系；等价地，也可表示为与系数向量 $\mathbf{c}_j = (\mathbf{c}_j(1) \ldots, \mathbf{c}_j(n)) \in \{0,1\}^n$的一一对应关系。因此，所有 M 授权用户都与一组指纹 $\{\mathbf{c}_j \in \{0,1\}^n : \ 1 \le j \le M\}$相关联。关于合谋攻击，正如 [42], 所指出的，由于嵌入多媒体中的指纹不同位元可能不易识别，也难以随意操纵，因此，诸如对多个嵌入指纹的信号求平均等线性合谋攻击通常更可行，适用于

Y. Gu 任职于日本福冈九州大学信息科学与电气工程学院。（电子邮箱：gu@inf.kyushu-u.ac.jp）I. Vorobyev 任职于德国慕尼黑慕尼黑工业大学通信工程研究所。（电子邮箱：ilya.vorobyev@tum.ru）Y. Miao 任职于日本茨城筑波大学工程、信息与系统学院。（电子邮箱：miao@sk.tsukuba.ac.jp）本研究获得俄罗斯基础研究基金会（RFBR）和保加利亚国家科学基金会（NSFB）资助（项目编号 20-51-18002）、BMBF-NEWCOM 资助（资助编号 16KIS1005）、日本学术振兴会（JSPS）资助（资助编号 18H01133、21K13830），以及日本学术振兴会与俄罗斯基础研究基金会日俄研究合作计划资助（项目编号 JPJSBP 120204802）。本文的早期版本曾部分发表于 IEEE 信息论国际研讨会（ISIT2022） [29]。

多媒体。在一种线性攻击中，由一组恶意授权用户 $J \subseteq \{1,\dots,M\}$ 组成的联盟通过对其各自的副本 $\mathbf{y}_j$ 进行线性组合，生成一个伪造副本 $\hat{\mathbf{y}}$，即：

$$\hat{\mathbf{y}} = \sum_{j\in J} \lambda_j \mathbf{y}_j = \sum_{j\in J} \lambda_j (\mathbf{x}+\mathbf{w}_j) = \mathbf{x} + \sum_{j\in J} \lambda_j \mathbf{w}_j = \mathbf{x} + \sum_{i=1}^{n} \left(\sum_{j\in J} \lambda_j \mathbf{c}_j(i) \right) \mathbf{u}_i \tag{1}$$

其中 λ_j 为一些实值系数，满足 $0 < \lambda_j < 1$ 和 $\sum_{j\in J} \lambda_j = 1$。特别地，如果对所有$j \in J$都有 $\lambda_j = 1/|J|$，则称其为平均攻击。可追踪性是指，一旦伪造副本 $\hat{\mathbf{y}}$ 被捕获，就能识别出 J 中的部分或全部合谋者/叛徒。特别地，如果能够识别出 J 中的所有合谋者，则称其具有完全可追踪性。在识别阶段，可以通过内积 $\langle \hat{\mathbf{y}}, \mathbf{u}_i \rangle = \sum_{j\in J} \lambda_j \mathbf{c}_j(i)$ 从捕获的 $\hat{\mathbf{y}}$ 中提取有用信息，其中对所有 $1 \le i \le n$ 均成立。容易验证，$\langle \hat{\mathbf{y}}, \mathbf{u}_i \rangle \in [0,1]$ 和

- $\langle \hat{\mathbf{y}}, \mathbf{u}_i \rangle = 0$ 蕴含 $\{\mathbf{c}_j(i) : j \in J\} = \{0\}$；
- $\langle \hat{\mathbf{y}}, \mathbf{u}_i \rangle = 1$ 蕴含 $\{\mathbf{c}_j(i) : j \in J\} = \{1\}$；
- $\langle \hat{\mathbf{y}}, \mathbf{u}_i \rangle \in (0,1)$ 蕴含 $\{\mathbf{c}_j(i) : j \in J\} = \{0,1\}$。

所得向量 $\big(\{\mathbf{c}_j(1) : j \in J\}, \dots, \{\mathbf{c}_j(n) : j \in J\}\big)$ 称为证据向量或 J的后代码，也将在识别/解码算法中使用，其精确定义同样参见第二节。在抗合谋指纹领域，我们希望设计一种多媒体指纹码，即由指纹 $\{\mathbf{c}_j : 1 \le j \le M\}$ 构成、具有高效（部分或完整）可追踪性的集合。这种情形通常称为叛徒追踪静态模型，尤其适用于电子数据分发系统。

另一方面，Fiat 和 Tassa [22], [23] 提出了动态叛徒追踪的概念，该概念在保护宽带多播系统中流媒体数据的知识产权方面有许多实际应用。动态叛徒追踪通常允许分多个阶段识别所有叛徒；在每个阶段，都可以利用上一阶段的反馈，并据此调整追踪策略，参见例如 [2], [32], [37] 。相应地，该方法允许即时使用若干种（不同的）安全码/指纹码，通常能够容纳更多用户，和/或与静态系统相比实现更高效的叛徒追踪。

文献中已提出若干类组合多媒体指纹码，简要回顾如下。1998年，Boneh 和 Shaw [5] 定义了 t抗诬陷码，可在线性时间内追踪到所有叛徒 $O(nM)$，其中 n 和 M 分别表示码的长度和大小。请注意，t抗诬陷码的定义与早得多的文献 [26], [38], [39]中讨论的（t,1）-分离码一致。后来，为了抵御多媒体指纹中的平均攻击，人们先后提出了 AND抗合谋码 [42], 、逻辑抗合谋码（即二元可分码） [8], 和强可分码 [31] ；这些码都能保证完全可追踪性，并已在例如 [24], [28],[45], [46]中得到研究。近期，文献 [17], [19], 研究了用于抗合谋多媒体指纹的、具有完全可追踪性的签名码，并指出它们本质上等价于用于加权二元加法信道通信的唯一可译码本。此外，文献 [17],[18], [19], [44] 也讨论了含噪签名码。

列表译码技术已被应用于叛徒追踪问题（例如参见 [1], [15], [20], [21], [41]），但其中的应用模型与本文相关而又有所不同。在 [16], 中，作者构造了具有高效解码能力但码率相当低的多媒体指纹码，其基础是码级联，以及对具有大距离的 Reed-Solomon码进行快速列表译码。迄今为止，现有叛徒追踪中列表译码的所有应用，都是借助具有较大汉明距离的纠错码和级联构造来构造某些类型的指纹码；相应的解码分析通常依赖于 Reed-Solomon码或代数几何码的高效列表译码算法（例如参见 [30]）。与这些已知结果不同，本文首次直接从基础叛徒追踪模型出发发展列表译码性质，而不是将纠错码作为中介；根据实际模型，本文中的列表译码自然超越了汉明度量。

本文首先针对抗合谋多媒体指纹，引入带列表译码的安全码（SCLD）的概念，将列表译码思想融入抗合谋安全码之中，从而实现高效的识别算法。研究表明，SCLD包含许多现有指纹码作为特例。接着，我们为SCLD 构造了一种两步识别算法，并证明其能够像抗诬陷码一样实现完全可追踪性；然而，与基于现有可分离码的叛徒追踪相比，该算法的效率要高得多。另一方面，我们分别建立了二元和 q 元 SCLD 的码率界，表明 SCLD 的码率可以远高于现有抗诬陷码。作为推论，我们还给出了二元可分离码最大码率的新下界。此外，我们提供了 SCLD 的一些显式构造，以及针对具有代数结构的某些 SCLD 的高效解码算法。进一步地，我们讨论了动态多媒体指纹，并基于列表译码性质建立了两阶段动态叛徒追踪框架；与静态场景相比，该框架不仅能够实现高效解码，还具有大得多的码率（即能够容纳多得多的用户）。

本文组织如下。第 II 节介绍记号和码的定义。第 III 节展示 SCLD 的解码算法。第 IV 节建立列表大小较大的二元 SCLD 和二元可分码的码率界。第 V 节给出固定列表大小的二元 SCLD 最大码率的下界。第 VI 节建立 q 元 SCLD 的界。第 VII 节介绍 SCLD 的一些显式构造及相应的高效解码算法。第 VIII 节讨论动态两阶段叛徒追踪。最后，第 IX 节总结本文。

II. 预备知识

为定义这些码，我们先引入一些记号。设 n、q、M 为正整数，Q、$\{0,1,\ldots,q-1\}$ 为大小为 $|Q|=q$ 的字母表。对于素数幂 q，记 GF(q) 为阶为 q 的有限域。记 $[M]$、$\{1,2,\ldots,M\}$。由 M 个向量 $\mathcal{C}=\{\mathbf{c}_1,\mathbf{c}2,\ldots,\mathbf{c}_M\}\subseteq Q^n$ 构成的集合称为一个 (n,M,q) 码，每个 $\mathbf{c}_i=(\mathbf{c}_i(1),\mathbf{c}_i(2)\ldots,\mathbf{c}_i(n))\in Q^n$ 称为长度为 n 的码字。(n,M,q) 码也称为长度为 n、大小（或基数）为 M 的 q 元码。(n,M,q) 码 $\mathcal{C}$ 的码率为 $R(\mathcal{C})=(\log_q M)/n$。

对于码 $\mathcal{C}\subseteq Q^n$，我们将 $\mathcal{C}$ 的第 i 个投影定义为

$$\mathcal{C}(i)=\{\mathbf{c}(i)\in Q:\mathbf{c}\in\mathcal{C}\},\ 1\le i\le n.$$

$\mathcal{C}$ 的后代码或证据向量定义为

$$\mathrm{desc}(\mathcal{C})=(\mathcal{C}(1),\ldots,\mathcal{C}(n))\in\mathcal{P}(Q)^n,$$

其中 $\mathcal{P}(Q)^n=\mathcal{P}(Q)\times\cdots\times\mathcal{P}(Q)$，且 $\mathcal{P}(Q)$ 是 Q 的幂集，即 $\mathcal{P}(Q)$ $\{Q_0\colon Q_0\subseteq Q\}$. 例如，若$\mathcal{C}=\{(0,1,0),(0,0,1),(0,1,2)\}$，则 $\mathrm{desc}(\mathcal{C})=(\{0\},\{0,1\},\{0,1,2\})$。特别地，令

$$\mathrm{Desc}_t(\mathcal{C})=\{\mathrm{desc}(\mathcal{C}_0):\mathcal{C}_0\subseteq\mathcal{C},1\le|\mathcal{C}_0|\le t\}.$$

对于任意 $\mathbf{a}\in Q^n$ 和 $\mathbf{b}\in\mathcal{P}(Q)^n$，如果对所有$i\in[n]$，$\mathbf{a}(i)\in\mathbf{b}(i)$，则称 a 被 b 覆盖，记作 $\mathbf{a}\preceq\mathbf{b}$。

A. 带列表译码的安全码

首先，我们介绍带列表译码的安全码这一概念。

定义 1. 设 $\mathcal{C}=\{\mathbf{c}_1,\ldots,\mathbf{c}_M\}\subseteq Q^n$ 是一个 (n,M,q) 码，且 $t\ge2$, $t\le L\le M$是整数。则 $\mathcal{C}$是一个 $\bar{t}$-列表译码安全码，或称为 $\bar{t}$-SCLD$(n,M,q;\ L)$，如果

(1) 对所有不同的 $\mathcal{C}_1$、$\mathcal{C}_2\subseteq\mathcal{C}$，满足 $|\mathcal{C}_1|\le t$ 和 $|\mathcal{C}_2|\le t$，都有 $\mathrm{desc}(\mathcal{C}_1)\ne\mathrm{desc}(\mathcal{C}_2)$;

(2) f 对于任意证据向量 d $\in$Desc$_t(\mathcal{C})$，至多有 L 个码字被 d 覆盖，即

$$Res_{\mathcal{C}}(\mathbf{d}) \triangleq |\{\mathbf{c} \in \mathcal{C} : \mathbf{c} \preceq \mathbf{d}\}| \leq L.$$

我们指出，定义 1 的条件 (1) 保证安全码的完全可追踪性；条件 (2) 表示安全码具有列表译码性质，这可以通过两步译码实现高效的叛徒追踪（参见我们的算法 1）。我们用两个小例子说明这一新概念。

例1. 不难验证，

1) $\mathcal{C}_1 = \{(0, 0, 1), (1, 0, 1), (1, 1, 0)\} \subseteq \{0, 1\}^3$ 是一个 $\bar{2}$-SCLD(3, 3, 2; 3)；以及
2) $\mathcal{C}_2 = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)\} \subseteq \{0, 1\}^3$ 是一个 $\bar{2}$-SCLD(3, 4, 2; 3)。

注意，在 $\mathcal{C}_1$ 中，列表大小为 $L = M = t + 1 = 3$；而在 $\mathcal{C}_2$ 中，列表大小为 $L = M - 1 = t + 1 = 3$。

值得注意的是，从列表大小的角度来看，SCLD 可以视为具有完全可追踪性的多媒体指纹码这一统一概念。回顾一下 [5], [8], [40] 中已有的三类码。

定义 2. 假设 $\mathcal{C} = \{\mathrm{c}_1, \ldots, \mathrm{c}_M\} \subseteq Q^n$ 是一个 (n, M, q) 码，且 $t \geq 2$ 是一个整数。那么

1) 若对任意满足 $|\mathcal{C}_0| = t$ 的 $\mathcal{C}_0 \subseteq \mathcal{C}$ 以及任意 a $\in \mathcal{C} \setminus \mathcal{C}_0$，均有 a $\neq$ desc($\mathcal{C}_0$)，则称 $\mathcal{C}$ 为 t-抗诬陷码，或称为 t-FPC(n, M, q)；2) 若对所有满足 $|\mathcal{C}_1| \leq t$ 和 $|\mathcal{C}_2| \leq t$ 的不同 $\mathcal{C}_1$、$\mathcal{C}_2 \subseteq \mathcal{C}$，均有 desc($\mathcal{C}_1$) $\neq$desc($\mathcal{C}_2$)，则称 $\mathcal{C}$ 为 $\bar{t}$-可分离码，或称为 $\bar{t}$-SC(n, M, q)。3) 若对任意满足 $|\mathcal{C}_0| = t$ 的 $\mathcal{C}_0 \subseteq \mathcal{C}$，以及 d $=$desc($\mathcal{C}_0$)，均有下式成立，则称 $\mathcal{C}$ 为具有列表译码的 t-超码，或称为 t-HLD(n, M, q; L)。

$$Res_{\mathcal{C}}(\mathbf{d}) = |\{\mathbf{c} \in \mathcal{C} : \mathbf{c} \preceq \mathbf{d}\}| \leq L.$$

根据上述定义可直接验证，对于任意 $t' \leq t$，$\bar{t}$-SCLD、t-FPC 和 t-SC 实际上都是 t-SCLD 的特例，且 t-HLD(n, M, q; L) 蕴含 t'-HLD(n, M, q; L)。更具体地说，以下结论可立即得出。

引理 1. 设 t, n, q, M, L 为正整数，且 $M \geq L \geq t$，则以下结论成立。

1) t-FPC(n, M, q) 等价于 $\bar{t}$-SCLD(n, M, q; t)。2) $\bar{t}$-SC(n, M, q) 等价于 $\bar{t}$-SCLD(n, M, q; M)。3) 设 $\mathcal{C}$ 为一个 (n, M, q) 码。那么，$\mathcal{C}$ 是 $\bar{t}$-SCLD(n, M, q; L)，当且仅当 $\mathcal{C}$ 同时是 $\bar{t}$-SC(n, M, q) 和 t-HLD(n, M, q; L)。4) 对任意 $L \leq L' \leq M$，$\bar{t}$-SCLD(n, M, q; L) 都是 $\bar{t}$-SCLD(n, M, q; L')。

B. 码率

分别用 $M_{\mathrm{FPC}}(t, n, q)$、$M_{\mathrm{SCLD}}(\bar{t}, n, q; L)$、$M_{\mathrm{HLD}}(t, n, q; L)$、$M_{\mathrm{SC}}(\bar{t}, n, q)$ 表示长度为 n 的 q 元 t-FPC、$\bar{t}$-SCLD、t-HLD、t-SC 的最大基数。

根据实际应用需求，通常需要 [42]，二进制指纹码。众所周知，通过 Forney 级联 [25]，也可以由一般的 q 元码导出二进制码。因此，接下来我们将考虑以下两种典型情形。

- 对于二进制码，将其最大的渐近码率记为

$$R_{\mathrm{FPC}}(t) = \limsup_{n\to\infty} \frac{\log_2 M_{\mathrm{FPC}}(t, n, 2)}{n},$$
$$R_{\mathrm{SCLD}}(\bar{t}; L) = \limsup_{n\to\infty} \frac{\log_2 M_{\mathrm{SCLD}}(\bar{t}, n, 2; L)}{n},$$
$$R_{\mathrm{HLD}}(t; L) = \limsup_{n\to\infty} \frac{\log_2 M_{\mathrm{HLD}}(t, n, 2; L)}{n},$$

$$R_{\mathrm{SC}}(\bar{t}) = \limsup_{n\to\infty} \frac{\log_2 M_{\mathrm{SC}}(\bar{t}, n, 2)}{n}.$$

- 对于长度为 n 的 q 元码，将其最大的渐近码率记为

$$R_{\mathrm{FPC}}(t, n) = \limsup_{q\to\infty} \frac{\log_q M_{\mathrm{FPC}}(t, n, q)}{n},$$
$$R_{\mathrm{SCLD}}(\bar{t}, n; L) = \limsup_{q\to\infty} \frac{\log_q M_{\mathrm{SCLD}}(\bar{t}, n, q; L)}{n},$$
$$R_{\mathrm{HLD}}(t, n; L) = \limsup_{q\to\infty} \frac{\log_q M_{\mathrm{HLD}}(t, n, q; L)}{n},$$
$$R_{\mathrm{SC}}(\bar{t}, n) = \limsup_{q\to\infty} \frac{\log_q M_{\mathrm{SC}}(\bar{t}, n, q)}{n}.$$

根据引理 1，我们立即得到以下推论。

推论 1。设 t、n、L、L' 为整数，且满足 $2 \le t \le L \le L' \le M$，其中 M 为码的基数。则有

1) $R_{\mathrm{FPC}}(t) \le R_{\mathrm{SCLD}}(\bar{t};\ L) \le \min\{R_{\mathrm{SC}}(\bar{t})\ R_{\mathrm{HLD}}(t;\ L)$ }. 2) $R_{\mathrm{FPC}}(t, n) \le R_{\mathrm{SCLD}}(\bar{t}, n;\ L) \le R_{\mathrm{SC}}(\bar{t}, n)$. 3) $R_{\mathrm{SCLD}}(\bar{t}, n;\ L) \le R_{\mathrm{SCLD}}(\bar{t}, n;\ L')$.

我们在表 I 中总结了 FPC、HLD、SC 的最新界，以及 SCLD 的新界，适用于字母表大小 q 足够大的情况。

表 I：当 $q \to \infty$ 时，q 元 FPC、SCLD、HLD 和 SC 的渐近码率

	$R_{\mathrm{FPC}}(t, n)$	$R_{\mathrm{SCLD}}(\bar{t}, n;\ L)$	$R_{\mathrm{HLD}}(t, n;\ L)$	$R_{\mathrm{SC}}(\bar{t}, n)$
码率	$= \frac{\lceil n/t \rceil}{n}$	$\ge \begin{cases} 2/3 & \text{若 } t=2,\ L \ge 3 \\ 1/(t-1) & \text{若 } t>2,\ L \ge t+1 \end{cases}$	$\ge \frac{L-t+1}{L}$	$\ge \begin{cases} 2/3 & \text{若 } t=2 \\ 1/(t-1) & \text{如果 } t>2 \end{cases}$
		$\le \begin{cases} \lceil 2n/3 \rceil / n & \text{如果 } t=2 \\ \lceil n/(t-1) \rceil / n & \text{如果 } t>2 \end{cases}$	$\le \frac{(L-t+1)\lceil n/L \rceil}{n}$	$\le \begin{cases} \lceil 2n/3 \rceil / n & \text{如果 } t=2 \\ \lceil n/(t-1) \rceil / n & \text{如果 } t>2 \end{cases}$
参考文献	[3], [11]	定理 5, [4], [27]	[14], [40]	[4], [27]

III. SCLD 识别算法

本节介绍一种用于 SCLD 的两步识别算法，并将其性能与现有 FPC 和 SC 进行比较。类似的思路也曾用于非自适应群组测试中的并集自由码 [43]。

定理 1. $\bar{t}$-SCLD$(n, M, q;\ L)$ 存在一个具有完全可追踪性的识别算法，其时间复杂度为 $O\big(\max\{nM, nL^t\}\big)$。

证明：$\bar{t}$-SCLD$(n, M, q;\ L)$ 的识别算法可按算法 1 执行，其中

- 步骤 1 的时间开销为 $O(nM)$，且 $|\mathcal{W}| \le L$;
- 步骤 2 的时间开销为 $O(n|\mathcal{W}|^t) = O(nL^t)$。

因此，算法 1 的总时间开销为 $O(\max\{nM, nL^t\})$。此外，根据 SCLD 的定义 1，步骤 1 之后的真实合谋集合 $\mathcal{T}$ 是 $\mathcal{W}$ 的子集，并可在步骤 2 之后被准确识别。也就是说，SCLD 的算法 1 具有完全可追踪性，证毕。 ■

在定理 1 中取 $L = \lfloor M^{1/t} \rfloor$，可立即得到一个推论。

算法 1 {v1}的识别算法 $\bar{t}$-SCLD

输入：$\bar{t}$-SCLD码 $\mathcal{C}$；证据向量 d $\in$Desc$_t(\mathcal{C})$
输出：所有叛徒的集合 $\mathcal{T}$

```
T = ∅, W = ∅。                                   ▷ 初始化候选集
对每个 c ∈ C 执行                                 ▷步骤 1
    如果 c ⪯ d，则
        W = W ∪ {c};
    结束 if
</style}结束循环
对每个大小至多为 t 的子集 S ⊆ W 执行               ▷步骤 2
    如果 desc(S) ==d，则
        T = S;
        输出 T;
    结束 if
结束 for
```

推论 2. 一个 $\bar{t}$-SCLD$\left(n, M, q; \lfloor M^{1/t} \rfloor\right)$ 具有一种具备完全可追踪性的识别算法，其时间复杂度为 $O(nM)$。

回顾 [8]中的 FPC 和 SC 的叛徒追踪。

引理 2（[8]）。1） t-FPC（n， M， q）具有识别算法，可实现完全可追踪性，时间复杂度为 O（nM）。2） $\bar{t}$-SC（n， M， q）具有识别算法，可实现完全可追踪性，时间复杂度为 O（nM^t）。

从推论 2 和引理 2 可以很容易看出，新提出的 SCLD 可以具有与 FPC 相同的可追踪性，而且效率远高于 SC。下一节将展示，SCLD 的码率可能远高于 FPC。

IV. Bounds for binary $\bar{t}$-SCLDs

本节分别建立二元 SC、HLD 和 SCLD 的最大渐近码率的界。

A. 二元 SC、SCLD、HLD 的下界

根据定理 1，值得考虑列表大小为 $L = M^{\alpha}$ 且 $\alpha \in [0,1]$的情形。为简化记号，我们用 $M^{(\alpha)}{}_{\mathrm{SCLD}}(\bar{t}, n, q)$ 和 $M^{(\alpha)}{}_{\mathrm{HLD}}(t, n, q)$ 分别表示长度为 n 、列表大小为 $L = M^{\alpha}$的 q 元 $\bar{t}$-SCLD 和 t-HLD 的最大基数，其中 M 是相应的码大小。据此，我们将它们最大的渐近二元码率记为

$$R^{(\alpha)}_{\mathrm{SCLD}}(\bar{t}) = \limsup_{n\to\infty} \frac{\log_2 M^{(\alpha)}_{\mathrm{SCLD}}(\bar{t}, n, 2)}{n}, \tag{2}$$

$$R^{(\alpha)}_{\mathrm{HLD}}(t) = \limsup_{n\to\infty} \frac{\log_2 M^{(\alpha)}_{\mathrm{HLD}}(t, n, 2)}{n}. \tag{3}$$

现在，我们通过带删余的随机编码，建立 SC、HLD 和 SCLD 的以下下界。设 $h(x)$ $-x\log_2(x)$ $-(1-x)$ $\log_2(1-x)$ 为二元熵函数。

定理 2. 1) $R_{\mathrm{SC}}(\bar{t})$ $\geq \max_{p\in(0,1)}$ $\underline{R}_{\mathrm{SC}}(t, p)$，其中

$$\underline{R}_{\mathrm{SC}}(t,p) \triangleq \min_{\substack{1\le t_1\le t_2\le t,\\ 0\le m\le t_1,\\ m\neq t_2}} \left(\frac{-\log_2(1-P_g(t_1,t_2,m))}{t_1+t_2-m-1} \right), \tag{4}$$

and

$$P_g(t_1,t_2,m) \triangleq p^{t_1}+p^{t_2}+(1-p)^{t_1}+(1-p)^{t_2}-2p^{t_1+t_2-m}-2(1-p)^{t_1+t_2-m} \quad \textit{for } m>0,$$
$$P_g(t_1,t_2,0) \triangleq 1-p^{t_1+t_2}-(1-p)^{t_1+t_2}-(1-p^{t_1}-(1-p)^{t_1})(1-p^{t_2}-(1-p)^{t_2}).$$

2) 对任意 $\alpha \in (0,1)$，我们有 $R^{(\alpha)}{}_{\mathrm{HLD}}(t) \geq \max\limits_{p\in(0,1)} \underline{R}^{(\alpha)}{}_{\mathrm{HLD}}(t,p)$，其中

$$\underline{R}^{(\alpha)}_{\mathrm{HLD}}(t,p) \triangleq \frac{h(p)-tph(1/t)}{1-\alpha}. \tag{5}$$

3) 对于任意 $\alpha \in (0,1)$，我们有 $R^{(\alpha)}{}_{\mathrm{SCLD}}(\bar{t}) \geq \max_{p\in(0,1)} \min\{\underline{R}_{\mathrm{SC}}(t,p)\ \underline{R}^{(\alpha)}{}_{\mathrm{HLD}}(t,p)\}$。

证明：设 $\mathcal{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_M\}$ 是由 M 个长度为 n 的二进制向量组成的集合，其中每个坐标 $\boldsymbol{c}_i(j)$ 都独立随机选取，且以概率 p 取值为 1，其中 $p\in(0,1)$。在以下关于这三个命题的论证中，我们将使用此随机系综。

1) 考虑 $\bar{t}$-安全集合码的要求。两个不同的指标集 $I_1\subseteq[M]$, $I_2\subseteq[M]$, $|I_1|=t_1$, $|I_2|=t_2$,$|I_1\cap I_2|=m$, $1\leq t_1,t_2\leq t$ 称为一个坏的 (t_1,t_2,m)-对，若 $\mathrm{desc}(\{\boldsymbol{c}_i\colon\ i\in I_1\}) = \mathrm{desc}(\{\boldsymbol{c}_i\colon\ i\in I_2\})$。为简化记号，对于任意 $I\subseteq[M]$，我们记 $\mathcal{C}_I=\{\boldsymbol{c}_i\colon\ i\in I\}\subseteq\mathcal{C}$。

接下来，我们估计所有 $(t_1,t_2,m)\in\Xi$ 中坏 (t_1,t_2,m)-对数量的期望 E_1，其中

$$\Xi \triangleq \{(t_1,t_2,m): 1\leq t_1\leq t_2\leq t, 0\leq m\leq t_1, m\neq t_2\}. \tag{6}$$

为此，记

$$\mathcal{B}(t_1,t_2,m) \triangleq \{(I_1,I_2): I_1\subseteq[M], I_2\subseteq[M], |I_1|=t_1, |I_2|=t_2, |I_1\cap I_2|=m\}.$$

显然，$|\mathcal{B}(t_1,t_2,m)|\leq M^{t_1+t_2-m}$。令 $P_g(t_1,t_2,m)$ 表示对于任意一行 i，$\mathcal{C}_{I_1}(i)$ 和 $\mathcal{C}_{I_2}(i)$ 不同的概率，其中 $(I_1,I_2)\in\mathcal{B}(t_1,t_2,m)$。那么，对于任意 $(I_1,I_2)\in\mathcal{B}(t_1,t_2,m)$，$(I_1,I_2)$ 是坏的概率为 $(1-P_g(t_1,t_2,m))^n$。现在我们通过讨论以下两种情况来计算 $P_g(t_1,t_2,m)$。

情况 1。考虑交集 $I_1\cap I_2$ 非空的情形，即 $m>0$。若 $\mathcal{C}_{I_1}(i)\neq\mathcal{C}_{I_2}(i)$，则有四种可能的选项：

- $\mathcal{C}_{I_1}(i)=\{1\}$, $\mathcal{C}_{I_2}(i)=\{0,1\}$;
- $\mathcal{C}_{I_1}(i)=\{0\}$, $\mathcal{C}_{I_2}(i)=\{0,1\}$;
- $\mathcal{C}_{I_1}(i)=\{0,1\}$, $\mathcal{C}_{I_2}(i)=\{1\}$;
- $\mathcal{C}_{I_1}(i)=\{0,1\}$, $\mathcal{C}_{I_2}(i)=\{0\}$.

这四种情况对应的概率为：

- $p^{t_1}(1-p^{t_2-m})$;
- $(1-p)^{t_1}(1-(1-p)^{t_2-m})$;
- $p^{t_2}(1-p^{t_1-m})$;
- $(1-p)^{t_2}(1-(1-p)^{t_1-m})$.

$m>0$ 的总概率 $P_g(t_1,t_2,m)$ 是这些概率之和

$$P_g(t_1,t_2,m)=p^{t_1}+p^{t_2}+(1-p)^{t_1}+(1-p)^{t_2}-2p^{t_1+t_2-m}-2(1-p)^{t_1+t_2-m}.$$

情况 2。现在考虑 $I_1\cap I_2=\emptyset$ 的情况，即 $m=0$。我们有

$$P_g(t_1,t_2,0)=1-P_{\mathrm{bad}}(t_1,t_2)$$

其中，$P_{\mathrm{bad}}(t_1,t_2)$ 表示 $\mathcal{C}_{I_1}(i)$ 和 $\mathcal{C}_{I_2}(i)$ 在坐标 i 处重合的概率。与情况 1 类似，我们得到

$$P_{\mathrm{bad}}(t_1,t_2)=p^{t_1+t_2}+(1-p)^{t_1+t_2}+(1-p^{t_1}-(1-p)^{t_1})(1-p^{t_2}-(1-p)^{t_2}).$$

根据以上内容，期望 E_1 为

$$\begin{aligned} E_1 &= \sum_{(t_1,t_2,m)\in\Xi} |\mathcal{B}(t_1,t_2,m)|(1-P_g(t_1,t_2,m))^n \\ &\le \sum_{(t_1,t_2,m)\in\Xi} M^{t_1+t_2-m}(1-P_g(t_1,t_2,m))^n. \end{aligned}$$

接下来，我们选择参数 $p\in(0,1)$ 和 $M=2^{nR}$，使得对于任意常数 $c>0$ 以及充分大的 n，都有 $E_1<cM/(n+1)$。实际上，当 $p\in(0,1)$ 且 $R<\underline{R}_{\mathrm{SC}}(t,p)$ 时即可满足这一条件，其中 $\underline{R}_{\mathrm{SC}}(t,p)$ 定义于 (4) 式。取 $c=1/4$。然后从每个坏对中移除一个元素。注意，剩余向量中没有重复项，因为对应的索引已作为坏 $(1,1,0)$-对被移除。因此可以得出，所得的码确实是一个 $\bar{t}$-安全集合码，其大小大于 $M/2$，码率至少为 $\log_2(M/2)/n=R-o(1)$，其中 $n\to\infty$。

2) 现在我们考虑一个 HLD。回想一下随机生成的 $\mathcal{C}$。

- 如果一个向量 $\boldsymbol{c}_i$ 的重量不等于 $\lfloor p(n+1)\rfloor$，我们就称它为坏向量；否则称它为好向量。将所有好向量组成的集合 $\boldsymbol{c}_i\in\mathcal{C}$ 记为 $\mathcal{C}_0$。令 E_2 表示 $\mathcal{C}$ 中坏向量数量的数学期望。
- 对于满足 $|\mathcal{C}_0'|=t$ 的 $\mathcal{C}_0'\subseteq\mathcal{C}_0$，若 $\mathrm{Res}_{\mathcal{C}_0}(\mathrm{desc}(\mathcal{C}_0'))$ 的大小大于 L'，则称子集 $\mathcal{C}_0'$ 为坏集合，其中 $L'=(\frac{M}{2n+2})^\alpha$。令 E_3 表示 $\mathcal{C}_0$ 中坏集合数量的数学期望。

接下来，我们如下估计 E_2 和 E_3。

(i) 考虑 E_2。向量 $\boldsymbol{c}\in\mathcal{C}$ 的重量可视为服从参数为 n 和 p 的二项分布的随机变量，即 $B(n,p)$。已知 $B(n,p)$ 的众数（即出现频率最高的值）为 $\lfloor p(n+1)\rfloor$。由于总共有 $n+1$ 种可能的重量，因此向量 $\boldsymbol{c}$ 为好向量（即重量为 $\lfloor p(n+1)\rfloor$）的概率不低于平均值 $1/(n+1)$。因此，坏向量数量的数学期望 E_2 至多为 $Mn/(n+1)$。

(ii) 考虑 E_3。首先，我们计算一个好向量 $\boldsymbol{c}\in\mathcal{C}_0$ 被固定证据向量$\boldsymbol{d}=\mathrm{desc}(\mathcal{C}_0')$ 覆盖的概率，其中 $\mathcal{C}_0'\subseteq\mathcal{C}_0$、$|\mathcal{C}_0'|=t$ 和 $\boldsymbol{c}\notin\mathcal{C}_0'$。注意，满足 $1\in\boldsymbol{d}(i)$ 的坐标 i 的数量至多为$t\lfloor p(n+1)\rfloor$。因此，$\boldsymbol{c}\preceq\boldsymbol{d}$ 的概率可以上界为

$$\Pr(\boldsymbol{c}\preceq\boldsymbol{d})\le\frac{\binom{t\lfloor p(n+1)\rfloor}{\lfloor p(n+1)\rfloor}}{\binom{n}{\lfloor p(n+1)\rfloor}}=2^{n(tph(1/t)-h(p)+o(1))}$$

因为 $n\to\infty$。如果 $\mathcal{C}_0$ 中存在多于 $L'=(\frac{M}{2n+2})^\alpha$ 个向量被 $\boldsymbol{d}$ 覆盖，那么 $\mathcal{C}_0\setminus\mathcal{C}_0'$ 中就存在一个恰含$L'-t$ 个向量的集合，这些向量均被 $\boldsymbol{d}$ 覆盖。因此，$\boldsymbol{d}$ 在 $\mathcal{C}_0$ 中覆盖多于 L' 个向量的概率小于

$$\begin{aligned} \binom{|\mathcal{C}_0\setminus\mathcal{C}_0'|}{L'-t}2^{n(tph(1/t)-h(p)+o(1))(L'-t)} &\le \binom{M}{L'-t}2^{n(tph(1/t)-h(p)+o(1))(L'-t)} \\ &\le \frac{M^{L'-t}}{(L'-t)!}\left(M^{R^{-1}(tph(1/t)-h(p)+o(1))}\right)^{L'} \\ &\le \frac{M^{L'-t}}{(L')^{-t}(L')!}\left(M^{R^{-1}(tph(1/t)-h(p)+o(1))}\right)^{L'} \\ &\le \frac{e^{L'}M^{L'-t}}{(L')^{L'-t}}\left(M^{R^{-1}(tph(1/t)-h(p)+o(1))}\right)^{L'} \\ &= M^{L'(1-\log_M L'+R^{-1}(tph(1/t)-h(p)+o(1)))}. \end{aligned}$$

取 $R<\underline{R}^{(\alpha)}{}_{\mathrm{HLD}}(t,p)$，其中 $\underline{R}^{(\alpha)}{}_{\mathrm{HLD}}(t,p)$ 在 (5) 中定义。对于任意这样的 R，$\boldsymbol{d}$ 覆盖 $\mathcal{C}_0$ 中超过 L'个向量的概率至多为 $M^{L'(-\varepsilon+o(1))}$，其中某个正数 ε。

那么，数学期望 E_3 至多为

$$E_3 < \binom{|\mathcal{C}_0|}{t} M^{L'(-\varepsilon+o(1))} < M/(4n+4)$$

当 n 足够大时。

删除所有坏向量，并从每个坏集合中删除一个向量。得到的码是一个 t-HLD，列表大小为 L'，基数为 $\geq \frac{M}{2n+2}$。再删除一些向量，使码的大小恰好为 $M' = \frac{M}{2n+2}$。最终得到的码是一个 t-HLD，列表大小为 $L' = (M')^{\alpha}$，码率为 $\frac{1}{n}\log_2 \frac{M}{2n+2} = R - o(1)$，当 $n \to \infty$ 时。

3）现在，基于前面对安全集合码和 HLD 的讨论，我们来考察 SCLD 的要求。回顾同一个随机系综 $\mathcal{C}$。根据前两个命题中的论证，如果我们取

$$R \leq \min\{\underline{R}_{\mathrm{SC}}(t,p), \underline{R}^{(\alpha)}_{\mathrm{HLD}}(t,p)\}$$

那么有

$$E_1 < M/(4n+4);\ E_2 \leq Mn/(n+1);\ E_3 < M/(4n+4),$$

这意味着 $E_1 + E_2 + E_3 < M - M/(2n+2)$。

删除所有坏向量、每个坏对中的一个向量，以及每个坏集合中的一个向量。得到的码既是一个 $\bar{t}$-安全集合码，也是一个 t-HLD；根据引理 1，它实际上是一个 $\bar{t}$-SCLD，码大小为 $M' = M/(2n+2)$，列表大小为 $L' = (M')^{\alpha}$。码率至少为 $n^{-1}\log_2 \frac{M}{2n+2} = R - o(1)$，当 $n \to \infty$ 时。定理得证。 ■

B根据上述定理 2，我们还可以得到 HLD 的如下下界 s.

定理 3. 设 $t \geq 2$ 为正整数， $0 < \alpha < 1$ 为常数。则

$$R^{(\alpha)}_{\mathrm{HLD}}(t) \geq \frac{1}{1-\alpha}\left[h\left(\frac{1}{2^{th(1/t)}+1}\right) - \frac{th(1/t)}{2^{th(1/t)}+1}\right]. \tag{7}$$

此外，如果 t 足够大，则有

$$R^{(\alpha)}_{\mathrm{HLD}}(t) \geq \frac{\log_2 e}{et(1-\alpha)}(1+o(1)) \approx \frac{0.530738}{t(1-\alpha)}(1+o(1)), \quad (t \to \infty). \tag{8}$$

证明：回顾定理 2，可知

$$R^{(\alpha)}_{\mathrm{HLD}}(t) \geq \max_{p\in(0,1)} \frac{h(p) - tph(1/t)}{1-\alpha}. \tag{9}$$

设 $F(p,t)\ h(p)\ - tph(1/t)$。则

$$\frac{\partial F(p,t)}{\partial p} = \log_2\left(\frac{1-p}{p}\right) - th(1/t).$$

令 $\frac{\partial F(p,t)}{\partial p} = 0$，得到 $p^* = \frac{1}{2^{th(1/t)}+1}$。显然 $0 < p^* < 1$。然后将 $p = p^*$ 代入 (9)，得到界 (7)。接下来考虑 t 足够大的情况。将 $p = \frac{1}{et}$ 代入 $F(p,t)$，其中 $e \approx 2.71828$ 为欧拉数，可得

$$\begin{aligned}
F\left(p=\frac{1}{et},t\right) &= h\left(\frac{1}{et}\right) - \frac{1}{e}h\left(\frac{1}{t}\right) \\
&= -\frac{1}{et}\log_2\left(\frac{1}{et}\right) - \left(1-\frac{1}{et}\right)\log_2\left(1-\frac{1}{et}\right) + \frac{1}{et}\log_2\left(\frac{1}{t}\right) + \frac{1}{e}\left(1-\frac{1}{t}\right)\log_2\left(1-\frac{1}{t}\right) \\
&= \frac{1}{et}\log_2 e - \left(1-\frac{1}{et}\right)\log_2\left(1-\frac{1}{et}\right) + \frac{1}{e}\left(1-\frac{1}{t}\right)\log_2\left(1-\frac{1}{t}\right) \\
&= \frac{1}{et}\log_2 e + \log_2 e \cdot \left(1-\frac{1}{et}\right)\left[\frac{1}{et} + \frac{1}{2(et)^2} + \cdots\right] - \frac{\log_2 e}{e}\cdot\left(1-\frac{1}{t}\right)\left[\frac{1}{t} + \frac{1}{2t^2} + \cdots\right]
\end{aligned}$$

$$= \frac{1}{et}\log_2 e + \frac{\log_2 e}{et} - \frac{\log_2 e}{e}\cdot\frac{1}{t} + o\left(\frac{1}{t}\right)$$
$$= \frac{1}{et}\log_2 e \cdot (1 + o(1)),$$

其中第四个等号由泰勒级数得出：对于任意$x \in [-1,1]$，$\ln(1-x) = -\sum_{n=1}^{\infty}\frac{x^n}{n} = -x - \frac{x^2}{2} - \cdots$。因此我们得出

$$R^{(\alpha)}_{\mathrm{HLD}}(t) \geq \max_{p\in(0,1)}\frac{F(p,t)}{1-\alpha} \geq \frac{F(p=\frac{1}{et},t)}{1-\alpha} = \frac{\log_2 e}{et(1-\alpha)}\cdot(1+o(1)) \approx \frac{0.530738}{t(1-\alpha)}(1+o(1)), \quad (t\to\infty)$$

如所期望的。 ■

值得注意的是，当 $\alpha \to 0$ 和 t 趋于无穷时，列表大小 L 也趋于无穷，而列表大小 $L \ll M$，其中 M 是相应的码基数。对于这一特定情形，上述定理 3表明 $\lim_{\alpha\to 0} R^{(\alpha)}{}_{\mathrm{HLD}}(t) \geq \frac{0.530738}{t}(1+o(1))$，这与 [40, 定理 3]中推导出的界一致。

B. 备注

表 II 展示了二元 SC、FPC、SCLD 的最大码率最新下界的数值，其中粗体数值由定理 2 推导得出；$R_{\mathrm{SC}}(\bar{2})$ 来自 [34]；$R_{\mathrm{FPC}}(2)$ 来自 [36]；而对于 $t>2$，$R_{\mathrm{FPC}}(t)$ 来自 [40]。

表 II：定理 2 和 [34], [36], [40] 得出的下界

t	2	3	4	5	6
$R_{\mathrm{SC}}(\bar{t}) \geq$	0.5	**0.13834**	**0.06198**	**0.03138**	**0.02003**
$R^{(1/t)}{}_{\mathrm{SCLD}}(t) \geq$	**0.44452**	**0.13205**	**0.05770**	**0.03105**	**0.01997**
$R_{\mathrm{FPC}}(t) \geq$	0.20756	0.07999	0.04392	0.02794	0.01936

值得注意的是，$R^{(1/t)}{}_{\mathrm{SCLD}}(\bar{t})$ 和 $R_{\mathrm{SC}}(\bar{t})$ 的界相当接近，且远大于 $R_{\mathrm{FPC}}(t)$（尤其是当 t 较小时）。结合第 III 节中的讨论，我们得出结论：与 SC 相比，SCLD 不仅具有高得多的可追踪性，而且码率也远高于 FPC。

V. 具有固定列表大小的二元 t-SCLD 的界

本节考察具有固定列表大小的二元 SCLD 的存在性。具体而言，我们为具有固定列表大小的 SCLD 的最大渐近码率建立一个下界。与前一节（第 IV 节）讨论列表大小非常大的情形不同，在此分析固定列表大小的情形需要更加谨慎。下面将运用马尔可夫不等式和霍夫丁不等式。

引理 3（例如 [9]）。设 Y为随机变量，使得 $Y \geq 0$，且 $Z_1, \ldots, Z_N$为相互独立的随机变量，使得对所有 $i \in [N]$，$a_i \leq Z_i \leq b_i$。令 $S_N = Z_1 + \cdots + Z_N$。对于任意 $\gamma > 0$，有
1) $\Pr(Y \geq \gamma) \leq \frac{\mathbb{E}[Y]}{\gamma}$（马尔可夫不等式）；2) $\Pr(|S_N - \mathbb{E}[S_N]| \geq \gamma) \leq 2\cdot \exp\left(-\frac{2\gamma^2}{\sum_{i=1}^{N}(b_i-a_i)^2}\right)$（霍夫丁不等式），
其中 $\mathbb{E}[Y]$ 是 Y的期望。

定理 4。设 $t \geq 2$ 为正整数。则

$$R_{\mathrm{SCLD}}(\bar{t};L) \geq \max_{p\in(0,1/2]} \min\left\{\underline{R}_{\mathrm{SC}}(t,p), \underline{R}_{\mathrm{HLD}}(t,p,L)\right\}, \tag{10}$$

其中 $\underline{R}_{\mathrm{SC}}(t,p)$ 在式 (4) 中定义，且

$$\underline{R}_{\mathrm{HLD}}(t,p,L) \triangleq h(p) + \frac{B(t,p,L)}{L} \tag{11}$$

$$B(t,p,L)=p\log_2\left(\frac{q_1(t,L,1-z)}{q_1(t,L,1-z)+q_2(t,L,1-z)}\right)+(1-p)\log_2\left(\frac{q_1(t,L,z)}{q_1(t,L,z)+q_2(t,L,z)}\right) \tag{12}$$

$$q_1(t,L,z)=z^t(z-z^t)^{L-t+1} \tag{13}$$

$$q_2(t,L,z)=(z-z^t)(1-z^t-(1-z)^t)^{L-t+1} \tag{14}$$

且 $z\in(0,1)$ 是该方程的唯一根

$$p(q_1(t,L,z)+q_2(t,L,z))=(1-p)(q_1(t,L,1-z)+q_2(t,L,1-z)). \tag{15}$$

证明：令 $\mathcal{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_M\}$ 为由 $M=2^{nR}$ 个长度为 n的二进制向量组成的集合，其中每个坐标$\boldsymbol{c}_i(j)$ 都从 $\{0,1\}$ 中独立随机选取，且取值为 1 的概率为 p，其中 $p\in(0,1/2]$。现在，我们要从 $\mathcal{C}$中移除一些不符合 SCLD 定义的向量。为此，我们将坏项定义如下。

- 称一对不同的指标集 (I_1,I_2)（其中 $I_1\subseteq[M]$, $I_2\subseteq[M]$, $|I_1|=t_1$, $|I_2|=t_2$, $|I_1\cap I_2|=m$,$1\le t_1,t_2\le t$）为坏的 (t_1,t_2,m)-对，若 $\mathrm{desc}(\{\boldsymbol{c}_i\colon i\in I_1\})=\mathrm{desc}(\{\boldsymbol{c}_i\colon i\in I_2\})$。记所有 $(t_1,t_2,m)\in\Xi$ 对应的坏 (t_1,t_2,m)-对的数量为 X_1，其中 Ξ 定义见式 (6)。
- 若向量 $\boldsymbol{c}_i$ 的重量不等于 $\lfloor p(n+1)\rfloor$，则称其为坏向量；否则称其为好向量。将 $\boldsymbol{c}_i$ 中所有好向量组成的集合记为 $\mathcal{C}_0$。显然 $\mathcal{C}_0\subseteq\mathcal{C}$。将 $\mathcal{C}$ 中的好向量数量记为 X_2 和 $X_2=|\mathcal{C}_0|$。
- 若一对不同的集合 $(\mathcal{C}_1,\mathcal{C}_2)$ 满足 $\mathcal{C}_1,\mathcal{C}_2\subseteq\mathcal{C}_0$, $|\mathcal{C}_1|=t$, $|\mathcal{C}_2|=L+1$，且 $\mathrm{desc}(\mathcal{C}_1)$ 覆盖了来自 $\mathcal{C}_2$的所有码字，则称其为坏对。将 $\mathcal{C}_0$ 中的集合坏对数量记为 X_3。

现在我们要证明以下三个命题。
断言1. $\Pr\left(X_1<\frac{M}{16(n+1)}\right)>\frac{3}{4}$。

事实上，码 $\mathcal{C}$ 与定理 2 的证明中所用的随机码相同。在该证明中，我们估计了坏 (t_1,t_2,m)-对数量的数学期望，在 $R<\underline{R}_{\mathrm{SC}}(t,p)$ 的情况下，对于任意常数 c 且 n足够大时，该期望为 $\mathbb{E}[X_1]<cM/(n+1)$。这里我们取 $c=1/64$。于是得到

$$\Pr\left(X_1<\frac{M}{16(n+1)}\right)=1-\Pr\left(X_1\ge\frac{M}{16(n+1)}\right)\ge1-\frac{\mathbb{E}[X_1]}{\frac{M}{16(n+1)}}>1-\frac{\frac{M}{64(n+1)}}{\frac{M}{16(n+1)}}=\frac{3}{4}$$

其中，第一个不等式由引理 3 中的马尔可夫不等式得出。由此证明了断言1。
断言 2. $\Pr\left(X_2\ge\frac{M}{2(n+1)}\right)=1-o(1)$ 当 $n\to\infty$。

事实上，回想一下，在定理 2 的证明中，我们证明了一个向量是好向量的概率至少为 $1/(n+1)$。这意味着好向量数量的期望为 $\mathbb{E}[X_2]\ge M/(n+1)$。因此

$$\Pr\left(X_2<\frac{M}{2(n+1)}\right)\le\Pr\left(|X_2-\mathbb{E}[X_2]|\ge\frac{M}{2(n+1)}\right)\le2\exp\left(-\frac{M}{2(n+1)^2}\right) \tag{16}$$

其中 第二个不等式由引理 3 中的霍夫丁不等式得出。现在我们有

$$\Pr\left(X_2\ge\frac{M}{2(n+1)}\right)=1-\Pr\left(X_2<\frac{M}{2(n+1)}\right)\ge1-2\exp\left(-\frac{M}{2(n+1)^2}\right)=1-o(1)\quad(n\to\infty)$$

其中，该不等式由式 (16) 得出。由此证明了断言 2。
断言 3。$\Pr\left(X_3<\frac{3X_2}{4}\right)\ge\frac{1}{3}$。

事实上，注意到由 $\mathcal{C}$ 中所有好向量组成的码 $\mathcal{C}_0$ 可以看作一个固定重量为 $\lfloor p(n+1)\rfloor$、基数为 $|\mathcal{C}_0|=X_2$ 的随机码，其分布与每个码字都从所有重量为 $\lfloor p(n+1)\rfloor$的向量集合中独立且等概率选取时的分布相同。这一事实与 [40, 定理 3]，证明中的假设一致，其中同样考虑了固定重量系综。[40, 定理 3] 表明，如果码 $\mathcal{C}_0$ 的码率 $R_{\mathcal{C}_0}=\frac{1}{n}\log_2$({v16) 小于 $\underline{R}_{\mathrm{HLD}}(t,p,L)$，其中 $\underline{R}_{\mathrm{HLD}}(t,p,L)$ 是

由 (11)-(15) 定义的，那么当 n 足够大时，坏集合对的数量 X_3 的数学期望满足 $\mathbb{E}[X_3] \le X_2/2$。因此，我们有

$$\Pr\Big(X_3 < \frac{3X_2}{4}\Big) = 1 - \Pr\Big(X_3 \ge \frac{3X_2}{4}\Big) \ge 1 - \frac{\mathbb{E}[X_3]}{\frac{3X_2}{4}} \ge 1 - \frac{\frac{X_2}{2}}{\frac{3X_2}{4}} = \frac{1}{3}$$

其中，第一个不等式由引理 3 中的马尔可夫不等式得出。由此证明了断言 3。

根据断言 1–3，我们得出结论：若 $R < \min\{\underline{R}_{\mathrm{SC}}(t,p),\ \underline{R}_{\mathrm{HLD}}(t,p,L)\}$，则以下三个条件有正概率同时成立：

$$X_1 < \frac{M}{16(n+1)} \quad \text{and} \quad X_2 \ge \frac{M}{2(n+1)} \quad \text{and} \quad X_3 < \frac{3X_2}{4}.$$

据此，从码 $\mathcal{C}_0$ 中删除每一对坏集合中的一个向量，并对所有 $(t_1,t_2,m) \in \Xi$，从每个坏 (t_1,t_2,m)-对中删除一个向量。容易验证，所得码是一个 $\bar{t}$-SCLD$(n, M_0, 2;\ L)$，其码大小为

$$M_0 \ge X_2 - X_1 - X_3 > \frac{X_2}{4} - X_1 > \frac{M}{8(n+1)} - \frac{M}{16(n+1)} = \frac{M}{16(n+1)}$$

且码率为 $\frac{1}{n}\log_2 M_0 \ge R + o(1)$，当 $n \to \infty$ 时成立。定理 4 的证明至此完成。 ■

表 III 展示了定理 4 中 $t = 2,3$ 的数值计算结果。值得注意的是，从情形 $t = 3$ 可以看出，即使列表大小不够大，二元 SCLD 的码率已经相当出色，与表 II 中涉及的渐近大列表大小情形相比更是如此。在下一节第 VI 节中，我们将展示，对于字母表大小 q 足够大的 q-元 SCLD，这一有趣现象更加明显。

表 III：通过定理 4 得到的固定列表大小二元 SCLD 下界

(t, L)	(2, 3)	(2, 4)	(2, 5)	(2, 6)	(2, 7)
$R_{\mathrm{SCLD}}(\ t, L)\ \ge$	0.245655	0.263492	0.274428	0.281927	0.287402
(t, L)	(3, 4)	(3, 5)	(3, 6)	(3, 7)	(3, 8)
$R_{\mathrm{SCLD}}(\ t, L)\ \ge$	0.115118	0.126598	0.129504	0.130385	0.130601

VI. q元 SCLD 的界

本节将给出 q 元 SCLD 在 q 较大时，其最大可能码率的下界。有趣的是，我们针对 q 元 SCLD 得到的界与 q 元 SC 的（几乎最优）界相匹配，这意味着已构造出的 q 元 SCLD 也几乎是最优的。具体而言，我们证明以下定理。

定理 5. 设 t、n、L 为正整数。则

$$R_{\mathrm{SCLD}}(\bar{t}, n; L) \ge \begin{cases} 2/3 & \text{if } t = 2 \text{ and } L \ge 3, \\ 1/(t-1) & \text{if } t > 2 \text{ and } L \ge t+1. \end{cases}$$

证明：设 $\mathcal{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_M\}$ 是由 $M = \varepsilon$ 个 $(n,t)q^{nR_\delta}$ q 元、长度为 n 的向量组成的集合，其中每个坐标 $\boldsymbol{c}_i(j)$ 都独立地随机选取，并且对每个 $k \in Q$，以概率 $1/q$ 取值 k；$\varepsilon(n,t)$ 是一个仅依赖于 n 和 t、小于 1 的充分小正数常量，并且

$$R_\delta = \begin{cases} 2/3 - \delta & \text{if } t = 2, \\ 1/(t-1) - \delta & \text{if } t > 2 \end{cases} \tag{17}$$

对于任意小的 $\delta > 0$。接下来我们讨论定义 1 中 SCLD 的两个要求。

(I) 考虑可分性质的第一项要求。我们使用 [4] 中的以下结果。

断言4（[4, 定理4和 5]）。对于任意小的 $\delta > 0$ 以及充分小的 $0 < \varepsilon\ (n,t)$，存在一个随机集合 $\mathcal{C}' \subseteq \mathcal{C}$，使得

- 对于所有满足 $|\mathcal{C}_1| \le t$ 和 $|\mathcal{C}_2| \le t$的不同 $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathcal{C}'$，均有 $\mathrm{desc}(\mathcal{C}_1) \neq \mathrm{desc}(\mathcal{C}_2)$；并且
- $|\mathcal{C}'|$ 的期望至少为 κq^{nR_δ}，其中 κ 是小于 $\varepsilon - \varepsilon^2\kappa'$的任意正数常量，而 κ' 是仅取决于 n 和 t的常数。

换句话说，可以从 $\mathcal{C}$ 中移除少量向量（数量为 $\mathcal{C} \setminus \mathcal{C}'$），使剩余向量的集合 $\mathcal{C}' \subseteq \mathcal{C}$ 满足定义 1 中 SCLD 的第一个要求。

(II) 现在我们考虑　　定义 1 中 SCLD 的第二个要求。对于 $\mathcal{C}_0 \subseteq \mathcal{C}$、$|\mathcal{C}_0| = t$，如果 $\mathcal{C}_0$ 满足

$$|\mathrm{Res}_{\mathcal{C}}(\mathrm{desc}(\mathcal{C}_0))| > L.$$

令 E 表示 $\mathcal{C}$中坏集合数量的期望。我们将估计 E。

考虑由持有 $\mathcal{C}$中指纹的 t 名合谋者组成的联盟生成的固定证据向量 d。该 d 覆盖 $\mathcal{C}$ 中超过 L 个向量的概率至多为

$$\begin{aligned}\binom{M}{L+1-t}(t/q)^{n(L+1-t)} &\le \frac{M^{L+1-t}}{(L+1-t)!}q^{n(\log_q t-1)(L+1-t)}\\ &\le \frac{M^{L+1-t}}{(L+1-t)!}\Big(\frac{M}{\varepsilon}\Big)^{R_\delta^{-1}(\log_q t-1)(L+1-t)}\\ &\le \Big(\frac{M}{\varepsilon}\Big)^{(L+1-t)(1+R_\delta^{-1}(\log_q t-1))}/(L+1-t)!.\end{aligned}$$

根据期望的线性性，坏集合数量的期望 E 至多为

$$E < M^t\Big(\frac{M}{\varepsilon}\Big)^{(L+1-t)(1+R_\delta^{-1}(\log_q t-1))}/(L+1-t)!$$

具体而言，对于任意小的 $\delta > 0$，我们有：

- 对于 $t = 2$、$R_\delta = 2/3 - \delta$、$L = 3$ 和 $q > 2^{\frac{2}{3\delta}}$，我们得到

$$E < \frac{1}{2}\Big(\frac{M}{\varepsilon}\Big)^{4+2\frac{\log_q 2-1}{2/3-\delta}} = o(M) = o(q^{nR_\delta}).$$

- 对于 $t > 2$、$R_\delta = \frac{1}{t-1} - \delta$、$L = t+1$ 和 $q > t^{t-3+\frac{2t-2}{(t^2-t)\delta}}$，我们有

$$E < \frac{1}{2}\Big(\frac{M}{\varepsilon}\Big)^{t+2+2\frac{\log_q t-1}{1/(t-1)-\delta}} = o(M) = o(q^{nR_\delta}).$$

根据 (I) 和 (II)，我们移除 $\mathcal{C} \setminus \mathcal{C}'$ 中的向量，并从 $\mathcal{C}$ 中的每个坏集合移除一个向量。容易验证，所得码是一个 $\bar{t}$-SCLD，且码大小的期望至少为 $\kappa q^{nR_\delta}(1 - o(1))$。因此，结合引理 1，我们得出存在一个 $\bar{t}$-SCLD$(n, M', q; L)$，满足 $L \ge t+1$ 和 $M' \ge \kappa' q^{nR_\delta}$，其中 κ' 是一个仅依赖于 n 和 t 的常数。换言之，我们证明了对于任意 $\delta > 0$，q 元 t-SCLD 码的码率至少为 R_δ。令 $\delta \to 0$ 趋于极限，即完成证明。

■

值得注意的是，定理 5 中 SCLD 的码率达到了已知最佳（且几乎最优）的 SC 码率[4], [27]。我们还注意到一个有趣的性质：t-SCLD 的列表大小只需略大于 t，即 $L \ge t+1$。结合第 III 节，我们得出结论：对于较大的 q，SCLD 不仅可以达到与 SC 相同的码率，而且解码效率也远高于 SC。

VII. SCLD 的构造

在本节中，我们提　　供 SCLD 的显式构造，并讨论相应的识别算法。

A. 基于广义打包的 SCLD

由于 t-SCLD 的码长对应多媒体内容中的正交归一信号数量，而码大小对应授权用户的数量，因此，我们希望构造码大小较大的 t-SCLD，

同时保持其长度较短。本小节将从一种称为广义打包 [12]的组合结构出发，构造 $\bar{2}$-SCLD(2, M, q'; 3)。

定义3. 设 K 是非负整数的一个子集， v、 b 是两个正整数。广义$(v, b, K, 1)$打包是一个集合系统$(X, \mathcal{B})$，其中 X 是一个含有 v 个元素的集合， $\mathcal{B}$ 是由 X 的 b 个子集组成的集合，称为区组，并满足

1) 对任意 $B \in \mathcal{B}$， $|B| \in K$； 2) X 中的任意一对不同元素至多出现
在 $\mathcal{B}$的一个区组中。

在广义$(v, b, K, 1)$打包中，若对某个 $k > 1$ 有 $K = \{k\}$，且 X中的每一对不同元素恰好出现在一个区组中，则通常称其为平衡不完全区组设计，简称$(v, k, 1)$-BIBD。满足 $q \geq 2$ 的$(q^2+q+1, q+1, 1)$-BIBD对应于阶为 q的射影平面。

构造1。设 $(X, \mathcal{B})$ 是一个广义 $(v, v, K, 1)$ 覆盖，其中 $X = \{0, 1, \ldots, v-1\}$ 且 $\mathcal{B} = \{B_0, B_1, \ldots, B_{v-1}\}$。则 $\mathcal{C} = C_0 \cup C_1 \cup \cdots \cup C_{v-1}$（其中 $C_i = \{(i, b) : b \in B_i\}$ 若 $B_i \neq \emptyset$，且 $C_i = \emptyset$ 若 $B_i = \emptyset$）是定义在 X 上的一个$\bar{2}$-SCLD(2, M, v; 3)，并且满足 $M = |B_0| + |B_1| + \cdots + |B_{v-1}|$。

证明：文献 [7] 已证明， $\mathcal{C}$ 是定义在 X 上的一个 $\bar{2}$-安全集合码(2, M, v)，并且满足 $M = |B_0| + |B_1| + \cdots + |B_{v-1}|$。为证明其列表大小为 3，注意到当 $i = j$ 时，desc($\{(i, a), (j, b)\}$) 不可能覆盖其他任何码字；当 $i \neq j$ 时，它至多覆盖一个码字 (i, b) 或 (j, a)，但不能同时覆盖二者。否则，desc($\{(i, a), (j, b)\}$) 将等于 desc($\{(i, b), (j, a)\}$)，这与 $\mathcal{C}$ 是一个 $\bar{2}$- 安全集合码 $(2, M, v)$ 矛盾。证明完毕。 ■

由构造1可立即得到以下结论。

推论3. 对任意素数幂 q，存在

1) 一个 $\bar{2}$-SCLD(2, $(q+1)(q^2+q+1)$ q^2+q+1;
3)； 2) 一个 $\bar{2}$-SCLD(2, q^3+2q^2, q^2+q;3)。

证明：断言1) 由一个众所周知的事实得出：当 q 为素数幂 [12]时，存在一个阶为 q 的射影平面。从一个射影平面中删除一个区组，并从该区组中删除一个元素，即得到一个广义 $(q^2+q, q^2+q, \{q, q+1\}, 1)$ 码组，其中有 q 个大小为 q 的区组，以及 q^2 个大小为 $q+1$ 的区组。这就得到了一个 $\bar{2}$-SCLD(2, q^3+2q^2, q^2+q; 3)，证毕。

■

根据引理 1， t-SCLD 的最大码大小可由 t-SC 的最大码大小上界限定。由 [7], 我们得出给定字母表大小时，上述两个长度为 2 的 2-SCLD 分别具有可能的最大码大小。

B. 级联构造

本小节给出一种级联构造，该构造可以根据字母表大小较大的 SCLD（例如第 VII-A 节中的构造）推导出字母表大小较小的 SCLD。

构造 2. 设 $2 \leq q \leq q'$ 为两个整数。设 $\mathcal{A}$ 为字母表 Q 上的 t-FPC(n_1, q', q)，且 $\mathcal{B}$ 为字母表 Q'上的 $\bar{t}$-SCLD(n_2, M, q'; L)。定义一个双射 ϕ: $Q' \to \mathcal{A}$。设 $\mathcal{C}$为如下定义的码

$$\Phi : \mathcal{B} \to \mathcal{C}$$
$$\mathbf{b} = (\mathbf{b}(1), \ldots, \mathbf{b}(n_1)) \mapsto \Phi(\mathbf{b}) = (\phi(\mathbf{b}(1)), \ldots, \phi(\mathbf{b}(n_1))).$$

那么 $\mathcal{C}$是 Q上的 $\bar{t}$-SCLD($n_1 n_2, M, q$; L)。

证明：显然， $\mathcal{C}$是定义在 Q 上的 $(n_1 n_2, M, q)$ 码。由于任意 t-FPC 都是一个 $\bar{t}$-安全集合码，根据 [8, 引理 5.4], 可知， $\mathcal{C}$是一个 $\bar{t}$-安全集合码($n_1 n_2, M, q$)。只需证明 $\mathcal{C}$也是一个 t-HLD($n_1 n_2, M, q$; L)。注意到 $\mathcal{A}$ 是一个 t- FPC(n_1, q', q)，因此联盟的后代码无法覆盖联盟之外的任何码字。由此可知， $\mathcal{C}$中的任意联盟 $\{\Phi(\mathbf{b}_1)$ $\ldots, \Phi(\mathbf{b}_s)\}$, $s \leq t$，都恰好对应于 $\mathcal{B}$ 中的联盟 $\{\mathbf{b}_1, \ldots, \mathbf{b}_s\}$。这意味着

$\mathcal{C}$中被 desc({Φ(b$_1$) .. ., Φ(b$_s$)}) 覆盖的码字至多有 L 个，即 $\mathcal{C}$是一个 t-HLD($n_1n_2, M, q; L$)。根据引理 1，结论得证。 ■

为了利用构造 2 推导出 SCLD，我们回顾如下两个已知的抗诬陷码构造。

构造3（[3]）。设 $l \geq 2$ 和 $t \geq 2$ 是两个整数，$q \geq l$ 是一个素数幂。设 α_1、α_2、……、$\alpha_l \in$ GF(q) 互不相同。在 GF(q) 上定义一个码 $\mathcal{C}$ 如下：

$$\mathcal{C} = \{(f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_l)) : f \in GF(q)[x] \text{ and } \deg(f) < \lceil l/t \rceil\}.$$

那么，$\mathcal{C}$ 是一个 t-FPC($l, q^{\lceil l/t \rceil}, q$)。

构造 4（[3]）。设 $l \geq 4$ 是一个偶整数。设 $m \geq l+1$ 是一个素数幂，且 $q = m^2 + 1$。设β_0、β_1、α_1、α_2、……、$\alpha_{l-1} \in$ GF(m) 互不相同。定义

$$\mathcal{C}_1 = \left\{(\infty, (f(\alpha_1), g(\alpha_1)), \ldots, (f(\alpha_{l-1}), g(\alpha_{l-1})) : f, g \in GF(m)[x],\ \deg(f) = \frac{l}{2} - 1, \deg(g) \leq \frac{l}{2} - 1\right\},$$

$$\mathcal{C}_2 = \left\{\big((t(\beta_0), t(\beta_1)), (s(\alpha_1), t(\alpha_1)), \ldots, (s(\alpha_{l-1}), t(\alpha_{l-1}))\big) : s, t \in GF(m)[x],\ \deg(s) \leq \frac{l}{2} - 2, \deg(t) \leq \frac{l}{2}\right\}.$$

那么，码 $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$ 是一个定义在 $F = \{\infty\} \cup$ (GF(m))^2 上的 2-FPC(l, 2($q-1$)$^{l/2}$(1 − 1/(2$\sqrt{q-1}$) q)。

应用构造 2、推论3以及构造 3、4，我们立即得到以下结果。

推论4. 设 n 为素数幂。

1）对于任意整数 $l \geq 2$，存在一个 $\bar{2}$-SCLD(2l，($n+1$)(n^2+n+1) q；3）对于任意满足 $q \geq l$和 $q^{\lceil l/2 \rceil} \geq n^2+n+1$的素数幂 q。2）对于任意偶数 $l \geq 4$，存在一个 $\bar{2}$-SCLD(2l，($n+1$)(n^2+n+1) q；3）对于任意 $q = m^2+1$，其中 $m \geq l+1$ 为满足2($q-1$)$^{l/2}$(1 − 1/(2$\sqrt{q-1}$)) $\geq n^2+n+1$的素数幂。3）对于任意整数 $l \geq 2$，存在一个 $\bar{2}$-SCLD(2l，n^3+2n^2，q；3）对于任意满足 $q \geq l$和$q^{\lceil l/2 \rceil} \geq n^2+n$的素数幂 q。4）对于任意偶数 $l \geq 4$，存在一个 $\bar{2}$-SCLD(2l，n^3+2n^2，q；3）对于任意 $q = m^2+1$，其中 $m \geq l+1$ 为满足2($q-1$)$^{l/2}$(1 − 1/(2$\sqrt{q-1}$)) $\geq n^2+n$的素数幂。

C. 一种具有高效解码能力的代数构造

在本小节中，我们给出一种二元 2-SCLD 的代数构造，它兼具较高的码率和运行时间为 O(polylog(M)) 的高效识别算法。以下构造最初见于 [34]。

构造 5（[34]）。设 $l \geq 2$ 为一个整数。对于每个元素 x ∈GF(2^l)，x 都可以表示为一个长度为 l *[33]*的二进制向量。定义二进制码 $\mathcal{C}$ 如下：

$$\mathcal{C} = \{(\mathbf{x}, \mathbf{x}^{\mathbf{3}}) : \mathbf{x} \in GF(2^l)\}.$$

于是，$\mathcal{C}$是一个 $\bar{2}$-SCLD(2l, 2^l, 2; 2^l)。

很容易看出，上述列表大小为 $L = M = 2^l$ 的 $\bar{2}$-SCLD 的码率为 1/2，高于定理 2 中随机编码方法所得的码率。值得注意的是，我们发现，通过构造 5 得到的 2-SCLD 可以高效地进行识别，如下所示。

定理 6. 对于由构造 5 得到的 $\bar{2}$-SCLD(n, M, 2; M)，存在一个识别算法，其时间复杂度为 O(poly(n)) $= O$(polylog(M))。

证明：设 $\mathcal{C}$ = {(x, x$^{\mathbf{3}}$) : x ∈GF(2^l)} 是由构造 5 导出的 $\bar{2}$-SCLD(n, M, 2; M)，其中 $n = 2l$ 和 $M = 2^l$。对于任意一个大小为二的联盟，例如 (x, x$^{\mathbf{3}}$) 和 (y, y$^{\mathbf{3}}$)，其中 x ≠ y ∈ GF(2^l)，其证据向量 $\boldsymbol{d}$ 意味着

算法 2 GF(2^l) 上二次方程 (19) 的一个解 [6]

输入：GF(2^l)的一个本原元 ω，使得 ω 的迹等于 1，即 $\sum_{i=0}^{l-1}\omega^{2^i}=1$;

the constant term $k=\mathbf{1}+\mathbf{v}/\mathbf{u^3}\in \mathrm{GF}(2^l)$

如果 $l \pmod 2 == 1$，则 ▷ Case 1

$\mathbf{z}=\sum_{i=0}^{(l-1)/2} k^{2^{2i}}$

否则

$T(k)=\sum_{i=0}^{(l-2)/2} k^{2^{2i}}$

结束 if

if $l \pmod 4 == 2$ 且 $T(k)==0$ 则 ▷ Case 2

$\mathbf{z}=\sum_{i=0}^{(l-6)/4} (k+k^2)^{2^{2+4i}}$

否则 if $l \pmod 4 == 2$ 且 $T(k)==1$ 则 ▷ Case 3

$\mathbf{z}=\omega^{(2^l-1)/3}+\sum_{i=0}^{(l-6)/4} (k+k^2)^{2^{2+4i}}$

否则 if $l \pmod 4 == 0$ 且 $T(k)==1$ 则 ▷ Case 4

$S=\sum_{j=1}^{l/4-1}\sum_{i=j}^{l/4-1} k^{2^{2i-1+l/2}+2^{2j-2}}$

$\mathbf{z}=S+S^2+k^{2^{l-1}}\left(1+\sum_{i=0}^{l/4-1} k^{2^{2i+l/2}}\right)$

否则如果 l（模 4）$==0$ 且 T（k）$==0$ 则 ▷ Case 5

$S_1=\sum_{j=1}^{l/4-1}\sum_{i=j}^{l/4-1} (\omega+\omega^2+k)^{2^{2i-1+l/2}+2^{2j-2}}$

$\mathbf{z}=\omega+S_1+S_1^2+(\omega+\omega^2+k)^{2^{l-1}}\left(1+\sum_{i=0}^{l/4-1} (\omega+\omega^2+k)^{2^{2i+l/2}}\right)$

结束条件判断

输出：方程（19）在 ∈GF(2^l) 中的一个解 z

(u, v)，其中 u、v ∈ GF(2^l)，满足

$$\mathbf{x}+\mathbf{y}=\mathbf{u} \quad \text{and} \quad \mathbf{x^3}+\mathbf{y^3}=\mathbf{v}, \tag{18}$$

其中，当 d(i) $=\{0,1\}$ 时 u(i) $=1$，当 d(i) $=\{0\}$ 或 $\{1\}$ 时 u(i) $=0$；当 d($l+i$) $=\{0,1\}$ 时 v(i) $=1$，当 d($l+i$) $=\{0\}$ 或 $\{1\}$ 时，v(i) $=0$，对于 $1\le i\le l$。显然，u $\neq 0$。识别的目标是利用（18）中的 u、v 找到 x 和 y 的解。为此，我们首先注意到，（18）可以转化为 $\mathbf{x^2}+\mathbf{u}\cdot\mathbf{x}+\mathbf{u^2}+\mathbf{v}/\mathbf{u}=\mathbf{0}$. 令 z $=$x/u。于是我们有

$$\mathbf{z^2}+\mathbf{z}+\mathbf{1}+\mathbf{v}/\mathbf{u^3}=0. \tag{19}$$

接下来我们证明，可以借助算法 2 确定 GF(2^l) 中方程 (19) 的一个解；其中，我们还用到了如下结果：根据 [35, 定理 2]，GF(2^l) 包含一个 $trace$ 等于 1 的本原元。

事实上，算法 2 可以在 O(poly(l)) 时间内确定二次方程 (19) 的一个解 z。结合关系 x $=$u $\cdot$ z 以及 x 和 y 的对称性，即可确定方程 (18) 的解。换言之，基于算法 2 的 $\mathcal{C}$ 解码可在 O(poly(l)) $=O$(poly(n)) $=O$ (polylog(M)) 时间内完成，符合要求。 ■

VIII. 两阶段动态叛徒追踪

本节讨论动态场景下的两阶段叛徒追踪。特别地，我们建立了一种基于 HLD 和 SCLD 的两阶段动态叛徒追踪框架。结果表明，该框架不仅能够提供更高效的

解码算法，但 al　　码率（即容纳更多用户），而且远高于静态场景下的 SCLD。

定理 7. 设 t 为正整数，且叛徒数量不超过 t。存在一种两阶段动态叛徒追踪方案，可容纳 $M = \Omega(2^{nR_{\mathrm{TDTT}}(\alpha,\beta,t)})$ 个用户，并在时间 $O(nM^{\max\{1,\alpha\beta t\}})$ 内实现完全可追踪性，其中 $R_{\mathrm{TDTT}}(\alpha,\beta,t) = \frac{1}{2}R^{(\alpha)}{}_{\mathrm{HLD}}(t)$，且 $0 \le \alpha, \beta \le 1$ 满足 $\alpha \cdot R^{(\alpha)}{}_{\mathrm{HLD}}(t) \le R^{(\beta)}{}_{\mathrm{SCLD}}(\bar{t})$。

证明：假设授权用户总数为 M，叛徒总数至多为 t，其中 $t \le M$。我们按照算法 3 构造一个两阶段叛徒追踪方案。

算法 3 两阶段动态叛徒追踪算法

输入： t-HLD 码 $\mathcal{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M\}$，列表大小为 $L^{(1)} = M^{\alpha}$；证据向量 $\mathbf{d}^{(1)} \in \mathcal{P}(Q)^n$　　▷阶段 1

1: $\mathcal{T} = \emptyset$，$\mathcal{W} = \emptyset$。　　▷ 初始化候选集

2: 对于每个 $j \in [M]$，执行

3: 　　如果 $\mathbf{c}_j \preceq \mathbf{d}^{(1)}$，则

4: 　　$\mathcal{W} = \mathcal{W} \cup \{j\}$;

5: 　　结束 if

6: 结束 for

输出：索引集合 $\mathcal{W}$

输入： $\bar{t}$-SCLD码 $\mathcal{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{|\mathcal{W}|}\}$，列表大小为 $L^{(2)} = |\mathcal{W}|^{\beta}$；证据向量 $\mathbf{d}^{(2)} \in \mathcal{P}(Q)^n$　　▷第二阶段

1：执行与 SCLD 算法 1 中相同的两个步骤

输出：所有叛徒的集合 $\mathcal{T}$

在第一阶段，我们利用一个二进制 t-HLD$(n_0, M, 2; L^{(1)})$，其列表大小为 $L^{(1)} = M^{\alpha}$，其中 $0 \le \alpha \le 1$ 是一个常数，稍后将确定。相应地，码大小为 $M = 2^{n_0 R^{(\alpha)}{}_{\mathrm{HLD}}(t)}$，其中 $R^{(\alpha)}{}_{\mathrm{HLD}}(t)$ 的定义见式 (3)。根据算法 3 的前半部分（即第一阶段），给定合谋攻击生成的证据向量 $\mathbf{d}^{(1)}$，可以在 $O(n_0 M)$ 时间内识别出 $\mathcal{W}$ 索引的一组用户。根据 t-HLD 的定义，$\mathcal{W}$的大小上界为 $L^{(1)} = M^{\alpha}$，且所有叛徒都在 $\mathcal{W}$中。换言之，在这一阶段，我们排除了一些无辜用户，并缩小了下一阶段的搜索空间，使得所有叛徒都位于 $\mathcal{W}$中，从而有助于进一步高效地识别所有确切的叛徒。

在第二阶段，我们采用列表大小为 $L^{(2)} = |\mathcal{W}|^{\beta}$ 的二进制 $\bar{t}$-SCLD$(n_0, |\mathcal{W}|, 2; L^{(2)})$，其中 $0 \le \beta \le 1$ 是一个常数，稍后确定。根据算法 3（等价地，算法 1）的后半部分以及定理 1，给定由合谋攻击生成的证据向量 $\mathbf{d}^{(2)}$，所有叛徒都能在时间 $O\big(\max\{n_0|\mathcal{W}|, n_0|\mathcal{W}|^{\beta t}\}\big) = O\big(n_0 M^{\max\{\alpha,\alpha\beta t\}}\big)$ 内被追踪出来。

令 $n = 2n_0$。根据上述内容，算法 3 的总时间开销为 $O\big(nM^{\max\{1,\alpha\beta t\}}\big)$，而这种两阶段动态叛徒追踪的码率（相应地，用户容量）为

$$R_{\mathrm{TDTT}}(\alpha,\beta,t) = \limsup_{n\to\infty} \frac{\log_2 M}{n} = \limsup_{n_0\to\infty} \frac{\log_2 2^{n_0 R^{(\alpha)}_{\mathrm{HLD}}(t)}}{2n_0} = \frac{1}{2} R^{(\alpha)}_{\mathrm{HLD}}(t). \tag{20}$$

注意到，为了保证第二阶段中相应的 t-SCLD存在，需要满足

$$|\mathcal{W}| \le M^{\alpha} = 2^{\alpha \cdot n_0 R^{(\alpha)}_{\mathrm{HLD}}(t)} \le 2^{n_0 R^{(\beta)}_{\mathrm{SCLD}}(\bar{t})}.$$

换言之，α 和 β 需要满足 $\alpha \cdot R^{(\alpha)}{}_{\mathrm{HLD}}(t) \le R^{(\beta)}{}_{\mathrm{SCLD}}(\bar{t})$。因此定理得证。 ■

值得注意的是，从定理 7 可以看出，α和 β 的选择在码率（即用户容量）与两阶段动态叛徒追踪的追踪/识别时间复杂度之间实现权衡的过程中发挥着重要作用。粗略而言，如果 α 和 β 较大（较小），则码率 R_{TDTT} 会较大（较小），而识别时间复杂度会较高（较低）。为了准确了解它们的性能，我们将在下文讨论两个有趣的情形。

情况 1：在不考虑追踪时间复杂度的情况下，寻找最大的码率 R_{TDTT}。为此，我们可以设定 $\beta = 1$，因为 $R_{\mathrm{TDTT}}(\alpha,\beta,t)$ 是 $\beta \in [0,1]$ 的非递减函数。因此，对于每个 t，我们旨在探索

的最优值

$$\max \ \frac{1}{2}R^{(\alpha)}_{\mathrm{HLD}}(t) \ \ \text{subject to} \ \ \alpha \cdot R^{(\alpha)}_{\mathrm{HLD}}(t) \le R^{(\beta)}_{\mathrm{SCLD}}(\bar{t}), \ \ \beta = 1, \ \ 0 < \alpha < 1. \tag{21}$$

表 IV 展示了通过 (21) 与定理 2 得到的 R_{TDTT} 数值下界，并将其与解码复杂度相同情况下 SCLD 的码率进行了比较。显然，在相同的完全可追踪性要求下，两阶段动态叛徒追踪的码率可能远高于 SCLD（即能够容纳更多用户）。

表 IV：TDTT 和 SCLD 的下界，其中 $\beta = 1$

(α, β, t)	(0.40606, 1, 3)	(0.29257, 1, 4)	(0.21608, 1, 5)
$R_{\mathrm{TDTT}}(\alpha, \beta, t)$	0.16778	0.10224	0.07245
$R^{(\alpha)}_{\mathrm{SCLD}}(\bar{t})$	0.13258	0.05783	0.03106
解码开销	$O(nM^{1.21818})$	$O(nM^{1.17028})$	$O(nM^{1.08040})$

情况 2。寻找在追踪时间复杂度为 $O(nM)$ 时的最大码率 R_{TDTT}。由定理 7 可知，两阶段叛徒追踪的时间开销为 $O(nM^{\max\{1,\alpha\beta t\}})$；当 $\alpha\beta \le 1/t$ 时，该开销达到最小值 $O(nM)$。由于 $R_{\mathrm{TDTT}}(\alpha, \beta, t)$ 是关于 α, $\beta \in [0, 1]$ 的非递减函数，因此我们可以考虑 $\alpha\beta = 1/t$ 的情况，以探寻可能达到的最大码率。相应地，对于每个 t，我们旨在寻找其最优值

$$\max \ \frac{1}{2}R^{(\alpha)}_{\mathrm{HLD}}(t) \ \ \text{subject to} \ \ \alpha \cdot R^{(\alpha)}_{\mathrm{HLD}}(t) \le R^{(\beta)}_{\mathrm{SCLD}}(\bar{t}), \ \ \alpha\beta = 1/t, \ \ 0 < \alpha, \beta < 1. \tag{22}$$

表 V 展示了根据式 (22) 和定理 2 得到的 R_{TDTT} 数值下界。表 V 中，在相同解码开销 $O(nM)$ 下，对 SCLD 与两阶段动态叛徒追踪进行比较可见，两阶段动态叛徒追踪的码率可能高得多（即能够容纳更多用户）。

表 V：当 $\alpha\beta = 1/t$ 时 TDTT 和 SCLD 的下界

(α, β, t)	(0.40406, 0.82496, 3)	(0.29107, 0.85890, 4)	(0.21508, 0.92989, 5)
$R_{\mathrm{TDTT}}(\alpha, \beta, t)$	0.16722	0.10202	0.07236
$R^{(1/t)}_{\mathrm{SCLD}}(\bar{t})$	0.13205	0.05770	0.03105
解码开销	$O(nM)$	$O(nM)$	$O(nM)$

IX. 结论

本文研究了用于叛徒追踪的组合安全码。我们首先将列表译码的思想直接融入多媒体指纹技术中的实际叛徒追踪模型，并提出了带列表译码的安全码（SCLD）这一概念。研究表明，SCLD 可视为一种统一的概念，因为许多现有指纹码都可作为其特例。我们建立了 SCLD 的高效解码与识别算法，并给出了其最大可能码率的界。这些结果表明，在解码效率和／或码率方面，SCLD 可能优于现有指纹码。此外，我们还针对动态场景提出了两阶段叛徒追踪框架，并证明该框架不仅解码速度快，而且码率远高于静态场景。未来，进一步提高码率并探索二元 SCLD 更明确的构造方法，将是值得研究的方向。

致谢

Y. Gu 谨向高木剛教授致谢，感谢他在早期研究带列表译码的叛徒追踪时提出的启发性建议；感谢栗林稔教授对第八节提出的深刻意见；并感谢河野修一教授进行的富有启发性的讨论。

参考文献

[1] A. Barg 和 G. Kabatiansky，“一类具有高效识别能力的识别概率码”，复杂性杂志，第20卷，第2-3期，第137–147页，2004年。[2] O. Berkman、M. Parnas 和 J. Sgall，“高效的动态叛徒追踪”，SIAM计算杂志，第30卷，第6期，第1802–1828页，2001年。[3] S. R. Blackburn，“防伪码”，SIAM离散数学杂志，第16卷，第3期，第499–510页，2003年。[4] ——，“可分离码的概率存在性结果”，IEEE信息论汇刊，第61卷，第11期，第5822–5827页，2015年。[5] D. Boneh 和 J. Shaw，“数字数据的抗合谋指纹”，IEEE信息论汇刊，第44卷，第5期，第1897–1905页，1998年。[6] C.-L. Chen，“GF(2^m)上二次方程解的公式”，IEEE信息论汇刊，第IT-28卷，第5期，第792–794页，1982年。[7] M. Cheng、L. Ji 和 Y. Miao，“可分离码”，IEEE信息论汇刊，第58卷，第3期，第1791–1803页，2012年。[8] M. Cheng 和 Y. Miao，“关于多媒体指纹的抗合谋码与检测算法”，IEEE信息论汇刊，第57卷，第7期，第4843–4851页，2011年。[9] H. Chernoff，“基于观测值之和的假设检验的渐近效率度量”，《数学统计年刊》，第493–507页，1952年。[10] B. Chor、A. Fiat 和 M. Naor，“追踪叛徒”，载于国际密码学年会（CRYPTO）。施普林格，1994年，第257–270页。[11] G. D. Cohen 和 H. G. Schaathun，《分离码渐近综述》。CiteSeer，2003年，第248号。[12] C. J. Colbourn，《组合设计手册》。CRC出版社，2010年。[13] I. J. Cox、M. L. Miller、J. A. Bloom 和 C. Honsinger，《数字水印》。施普林格，2002年，第53卷。[14] A. D’yachkov、N. Polyanskii、V. Shchukin 和 I. Vorobyev，“对称多址信道的可分离码”，IEEE信息论汇刊，第65卷，第6期，第3738–3750页，2019年。[15] E. Egorova、M. Fernandez 和 G. Kabatiansky，“一种具有多项式时间追踪算法的可追踪集合系统构造”，载于 IEEE 信息论国际研讨会（ISIT），2019年，第2739–2742页。[16] E. Egorova、M. Fernandez、G. Kabatiansky 和 M. H. Lee，“a信道的签名码与抗合谋多媒体指纹码”，载于 IEEE 信息论国际研讨会（ISIT），2016年，第3043–3047页。[17] ——，“加权噪声加法信道、多媒体指纹和压缩感知的签名码”，《设计、编码与密码学》，第87卷，第2期，第455–462页，2019年。[18] E. E. Egorova、M. Fernandez、G. A. Kabatiansky 和 Y. Miao，“能够抵抗平均攻击和对抗噪声的完全可追踪多媒体指纹码的存在性与构造”，《信息传输问题》，第56卷，第4期，第388–398页，2020年。[19] J. Fan、Y. Gu、M. Hachimori 和 Y. Miao，“加权二进制加法信道与多媒体指纹的签名码”，IEEE信息论汇刊，第67卷，第1期，第200–216页，2020年。[20] M. Fernandez、J. Moreira 和 M. Soriano，“使用 Koetter-Vardy算法识别叛徒”，IEEE信息论汇刊，第57卷，第2期，第692–704页，2011年。[21] M. Fernandez 和 M. Soriano，“代数几何可追踪码中的叛徒识别”，IEEE信号处理汇刊，第52卷，第10期，第3073–3077页，2004年。[22] A. Fiat 和 T. Tassa，“动态叛徒追踪”，载于国际密码学年会（CRYPTO）。施普林格，1999年，第354–371页。[23] ——，“动态叛徒追踪”，《密码学杂志》，第14卷，第3期，第211–223页，2001年。[24] S. D. Fiore 和 M. Dalai，“关于 $\bar{2}$-可分离码和B_2 码的注记”，《离散数学》，第345卷，第3期，第112751页，2022年。[25] G. D. Forney，“级联码”。1965年。[26] A. D. Friedman、R. L. Graham 和 J. D. Ullman，“通用单跃迁时间异步状态赋值”，IEEE计算机汇刊，第100卷，第6期，第541–547页，1969年。[27] F. Gao 和 G. Ge，“多媒体指纹可分离码的新界”，IEEE信息论汇刊，第60卷，第9期，第5257–5262页，2014年。[28] Y. Gu、J. Fan 和 Y. Miao，“可分离码和B_2 码的改进界”，IEEE通信快报，第24卷，第15–19页，2020年。[29] Y. Gu、I. Vorobyev 和 Y. Miao，“带列表译码的安全码”，载于 IEEE 信息论国际研讨会（ISIT），2022年，第2350–2355页。[30] V. Guruswami 和 M. Sudan，“Reed-Solomon码和代数几何码的改进解码”，载于第39届计算机科学基础年会论文集（FOCS），1998年，第28–37页。[31] J. Jiang、M. Cheng 和 Y. Miao，“强可分码”，《设计、编码与密码学》，第79卷，第2期，第303–318页，2016年。[32] T. Laarhoven、J. Doumen、P. Roelse、B. ˇSkori´c 和 B. de Weger，“动态Tardos叛徒追踪方案”，IEEE信息论汇刊，第59卷，第7期，第4230–4242页，2013年。[33] R. Lidl 和 H. Niederreiter，《有限域》。剑桥大学出版社，2003年。[34] B. Lindstr¨om，“由和确定两个向量”，《组合理论杂志》，第6卷，第4期，第402–407页，1969年。[35] O. Moreno，“GF(2^m)中迹等于1的本原元”，《离散数学》，第41卷，第53–56页，1982年。[36] H. Randriambololona，“超越概率界的(2, 1)-分离系统”，《以色列数学杂志》，第195卷，第171–186页，2013年。[37] R. Safavi-Naini 和 Y. Wang，“序贯叛徒追踪”，IEEE信息论汇刊，第49卷，第5期，第1319–1326页，2003年。[38] Y. L. Sagalovich，“提高有限自动机可靠性的一种方法”，《信息传输问题》，第1卷，第2期，第27–35页，1965年。[39] ——，“分离系统”，《信息传输问题》，第30卷，第2期，第14–35页，1994年。[40] V. Y. Shchukin，“多址超信道的列表译码”，《信息传输问题》，第52卷，第4期，第329–343页，2016年。[41] A. Silverberg、J. Staddon 和 J. L. Walker，“利用列表译码的高效叛徒追踪算法”，载于国际密码学与信息安全理论及应用会议（ASIACRYPT）。施普林格，2001年，第175–192页。

[42] W. Trappe、M. Wu、Z. J. Wang 和 K. R. Liu，《用于多媒体的抗合谋指纹》，IEEE信号处理汇刊，第 51 卷，第 4 期，第 1069–1087 页，2003 年。[43] I. Vorobyev，《并集自由码的快速译码》，载于 2021 年第 XVII 届“信息与控制系统中的冗余问题”国际研讨会（REDUNDANCY）。IEEE，2021 年，第 106–109 页。[44] ——，《具有最优码率、可抵抗平均攻击和对抗噪声的完全可追踪多媒体指纹码》，《设计、编码与密码学》，第 1–9 页，2022 年。[45] X. Wang，《父集识别集合系统和可分离码的改进上界》，《设计、编码与密码学》，第 89 卷，第 1 期，第 91–104 页，2021 年。[46] Y. Yang、Y. Zhang 和 G. Ge，《安全码及相关哈希族的新下界：一种超图理论方法》，IEEE信息论汇刊，第 63 卷，第 4 期，第 2446–2453 页，2017 年。