# New Constructions for IPP Codes

Tran van Trung and Sosina Martirosyan
Institute for Experimental Mathematics, University of Essen
Ellernstrasse 29, 45326 Essen, Germany
{trung, sosina}@exp-math.uni-essen.de

**Abstract**

Identifiable parent property (IPP) codes are introduced to provide protection against illegal producing of copyrighted digital material. In this paper we consider explicit construction methods for IPP codes by means of recursion techniques. The first method directly constructs IPP codes, whereas the second constructs perfect hash families that are then used to derive IPP codes. In fact, the first construction provides an infinite class of IPP codes having the best known asymptotic behavior. We also prove that this class has a traitor tracing algorithm with a runtime of $O(M)$ in general, where $M$ is the number of codewords.

**AMS classification:** 05

**Keywords:** Identifiable parent property (IPP) code, traceability(TA) code, perfect hash family, recursive construction.

## 1 Introduction

Codes providing some forms of traceability (TA) to protect copyrighted digital data against piracy have been extensively studied in the recent years. The weak forms of such codes are frameproof codes introduced by Boneh and Shaw [4], and secure frameproof codes [22]. The strong form of codes studied in this paper are identifiable parent property (IPP) codes which have been introduced by Hollmann, van Lint, Linnartz and Tolhuizen [16]. Other strong versions of such codes are TA schemes and TA codes introduced by Chor, Fiat and Naor in [8, 9, 10]. In fact, TA codes turn out to be a subclass of IPP codes [21].

Combinatorial properties of IPP codes and TA codes have been studied by Staddon, Stinson and Wei [21], Sarkar, Stinson [19], Barg, et al. [2], and also in [27]. The question of complexity of traitor tracing algorithms for IPP and TA codes is treated in [20], e.g. certain classes of TA codes are shown to have a faster tracing algorithm than their initially known linear runtime by using the list decoding techniques. New results on bounds of frameproof codes and TA schemes can be found in [18].

Perfect hash families (PHF), due to their significant applications in information retrieval, have undergone considerable investigation, see e.g. [7] for an extensive survey. More recently, perfect hash families have found applications in cryptography, particularly in codes with traceability property [22, 21, 19].

In this paper we focus on explicit construction methods for IPP codes using recursion techniques. We also present a recursive construction for perfect hash families, from which a new class of IPP codes is derived.

Our first construction provides an infinite class of IPP codes with the best known asymptotic behavior. In fact, we are able to construct a class of $w$-IPP codes in which the length $n$ of the codewords is $O((w^2)^{\log^*(M)}(\log(M)))$, where $M$ is the number of codewords. Moreover, we prove that these codes allow a traitor tracing algorith with a runtime of $O(M)$ in general. An even faster tracing algorithm for this class can be achieved. It should be noted that no IPP codes other than TA codes with this property were known before [21].

The rest of this paper is organized as follows. In Section 2 we present some preliminaries. In Section 3 we describe our first construction, in which the concatenation technique and the recursive method are combined. We show the asymptotic behavior of the codes and prove that they have a traitor tracing algorithm with a runtime complexity $O(M)$. In Section 4 we present a double induction method to construct a new class of perfect hash families. This class is then used to derive an infinite class of IPP codes which covers a very large set of parameter values.

## 2 Preliminaries

In this section we give definitions, notation and some basic results for IPP and TA codes and perfect hash families.

Let $Q$ be an alphabet of size $q$ and let $C \subseteq Q^n$. Then $C$ is called a $q$-ary code of length $n$. If $|C| = M$, then we call $C$ an $(n, M, q)$ code. The elements of $C$ are called *codewords* and each codeword will have the form $x = (x_1, \ldots, x_n)$, where $x_i \in Q$, $1 \leq i \leq n$.

For any subset of codewords $C_0 \subseteq C$, the set of *descendants* of $C_0$, denoted $\mathbf{desc}(C_0)$, is defined by

$$\mathbf{desc}(C_0) = \{x \in Q^n : x_i \in \{a_i : a \in C_0\},\ 1 \leq i \leq n\}.$$

Thus $\mathbf{desc}(C_0)$ consists of all $n$-tuples that could be produced by a coalition holding the codewords in $C_0$. If $x \in \mathbf{desc}(C_0)$, then we say that $C_0$ produces $x$.

Let $w$ be an integer. Define the *$w$-descendant code*, denoted $\mathbf{desc}_w(C)$, as follows:

$$\mathbf{desc}_w(C) = \bigcup_{C_0 \subseteq C, |C_0| \leq w} \mathbf{desc}(C_0).$$

Thus $\mathbf{desc}_w(C)$ consists of all $n$-tuples that could be produced by some coalition of size at most $w$.

**Definition 2.1** *Let $\mathcal{C}$ be an $(n, M, q)$ code and let $w \geq 2$ be an integer. $\mathcal{C}$ is called a $(n, M, q, w) - IPP$ code provided that, for all $x \in \mathbf{desc}_w(C)$, it holds that*

$$\bigcap_{\{i : x \in \mathbf{desc}(C_i),\ |C_i| \leq w\}} C_i \neq \emptyset.$$

**Definition 2.2** *Let define $I(x, y) = \{i : x_i = y_i\}$ for any $x, y \in Q^n$. Suppose $C \subseteq Q^n$ is an $(n, b, q)$-code and $w \geq 2$ is an integer. $C$ is called a $w$-TA code provided that, for all $i$ and all $x \in \mathbf{desc}(\mathcal{C}_i)$, there is at least one codeword $y \in C_i$ such that $|I(x, y)| > |I(x, z)|$ for any $z \in C \setminus C_i$.*

In fact, $TA$ codes form a subclass of $IPP$ codes, as pointed out in the following lemma.

**Lemma 2.1 ([21], Lemma 1.3)** *An $(n, M, q, w)$-TA code is an $(n, M, q, w)$-IPP code.*

The converse of Lemma 2.1 is not true, as it can be easily checked with small examples, see e.g [21], [20].

The following result is useful, which states that error-correcting codes with "sufficiently large" minimum distance are necessarily $TA$ codes and $IPP$ codes, [21].

**Theorem 2.2 ([21], Theorem 4.4)** *Any $(n, M, q)$ code $C$ having minimum distance $d > n(1 - 1/w^2)$ is an $(n, M, q, w)$-TA code. In particular, $C$ is an $(n, M, q, w)$-IPP code.*

A finite set $\mathcal{H}$ of $n$ functions $h : A \longrightarrow B$, where $|A| = M \geq |B| = m$, is called an *(n,M,m)-hash family*, denoted by $(n, M, m) - HF$.

**Definition 2.3** *Let $M$, $m$, $w$ be integers such that $M \geq m \geq w \geq 2$. An $(n, M, m)$-hash family $\mathcal{H}$ is called an $(n, M, m, w)$-perfect hash family, denoted $(n, M, m, w) - PHF$, if for any subset $X \subseteq A$ with $|X| = w$, there is at least one function $h \in \mathcal{H}$ such that $h$ is injective on $X$.*

An $(n, M, q)$-code $\mathcal{C}$ can be depicted as an $M \times n$ matrix $C$ on $q$ symbols, where each row of the matrix corresponds to one of the codewords. Similarly, an $(n, M, m) - HF$, $\mathcal{H}$, can be presented as an $M \times n$ matrix on $m$ symbols, where each column of the matrix corresponds to one of the function in $\mathcal{H}$.

A direct connection between error-correcting codes and perfect hash families, due to Alon, is as follows.

**Lemma 2.3** [1] *Suppose there is an $(n, M, q)$ code $\mathcal{C}$ with minimum distance $d$. Then there is an $(n, M, q, w) - PHF$, where*

$$(n-d)\binom{w}{2} < n.$$

*Proof.* Let $C$ be the matrix representation of $\mathcal{C}$. Then $C$ is an $M \times n$ matrix, whose entries are from a set of $q$ symbols. The condition $(M-d)\binom{w}{2} < N$ asserts that for any given $n$ rows, say $i_1, \ldots, i_w$, of $C$ there is at least one column whose $w$ entries in the rows $i_1, \ldots, i_w$ are pairwise distinct. Thus $C$ is an $(n, M, q, w) - PHF$, as desired. ■

The following theorem, due to Staddon, Stinson and Wei [21], is useful for our discussion in the sequel.

**Theorem 2.4 ( [21], Theorem 2.8)** *Let $\mathcal{C}$ be an $(n, M, q)$-code whose matrix representation is $C$. If $C$ is an $(n, M, q, \lfloor (w+2)^2/4 \rfloor) - PHF$, then $\mathcal{C}$ is an $(n, M, q, w) - IPP$ code.*

## 3 A construction of IPP codes

Our first construction of IPP codes is carried out in two steps. In the first step we prove Theorem 3.4 and the crucial Theorem 3.5. In the second step we describe the construction by making use of Theorem 3.4 and 3.5, and the result is presented in Theorem 3.6. The asymptotic behavior of these codes is shown in Theorem 3.7. Using the same method a more general result is obtained, which is formulated in Theorem 3.8. Finally, Theorem 3.9 shows that the codes of Theorem 3.6 have a traitor tracing algorithm with a runtime of $O(M)$.

We first describe a simple construction for $q$-ary codes which has been presented by Bush (1952) [5] for orthogonal arrays.

Let $A \subseteq Q_1^n$ be an $(n, M_1, q_1)$ code with minimum distance $d_1$ and $|Q_1| = q_1$, and let $B \subseteq Q_2^n$ be an $(n, M_2, q_2)$ code with minimum distance $d_2$ and $|Q_2| = q_2$. Let $Q = Q_1 \times Q_2$. We define a code $C$ over alphabet $Q$ as follows. For any pair of codewords $\mathbf{a} = (a_1, \ldots, a_n) \in A$ and $\mathbf{b} = (b_1, \ldots, b_n) \in B$ we construct a vector

$\mathbf{c}(\mathbf{a},\mathbf{b}) = ((a_1,b_1),\dots,(a_n,b_n)) \in Q^n$. Then it is easy to verify that $C = \{\mathbf{c}(\mathbf{a},\mathbf{b}) : \mathbf{a} \in A, \mathbf{b} \in B\} \subseteq Q^n$ is an $(n, M_1M_2, q_1q_2)$ code with minimum distance $d = \min\{d_1, d_2\}$.

Thus we have the following result.

**Theorem 3.1** *Suppose there exist $(n, M_1, q_1)$ code and $(n, M_2, q_2)$ code with minimum distance $d_1$ and $d_2$, respectively. Then there exists an $(n, M_1M_2, q_1q_2)$ code with minimum distance $d = \min\{d_1, d_2\}$.*

Theorem 3.1 can be used to construct $q$-ary codes achieving Singleton bound with equality, namely $MDS$ codes (*maximum distance separable*), for which $q$ is not a prime power. In fact, in the language of orthogonal arrays an $(n, M, q)$ $MDS$ code with minimum distance $d$ is an $OA_1(n-d+1, n, q)$; here we have $M = q^{n-d+1}$. We record this special case of Bush construction in the following theorem.

**Theorem 3.2** *The existence of $(n, q_1^t, q_1)$ and $(n, q_2^t, q_2)$ $MDS$ codes having the same minimum distance $d = n-t+1$ implies the existence of an $(n, (q_1q_2)^t, q_1q_2)$ $MDS$ code with minimum distance $d$.*

As a consequence of Theorem 3.2, we have the following corollary.

**Corollary 3.3** *For any integer $n \geq 2$ and $s$ with a prime factorization $s = p_1^{e_1}\dots p_r^{e_r}$ such that $n \leq p_i^{e_i}$, $i = 1,\dots,r$, there is an $(n, s^t, s)$ $MDS$ code, for all $2 \leq t \leq n$.*

*Proof.* The corollary follows from the existence of $(n, (p_i^{e_i})^t, (p_i^{e_i}))$ $MDS$ (Reed-Solomon) codes for $i = 1,\dots,r$. ∎

By combining Corollary 3.3 and Theorem 2.2 we obtain the following theorem.

**Theorem 3.4** *Let $w \geq 2$ be any given integer. For any integer $n > w^2$ and $s$ having $s = p_1^{e_1}\dots p_k^{e_k}$ as its prime factorization with $n \leq p_i^{e_i}$ for all $i = 1,\dots,k$ there exists an $(n, M, s, w) - IPP$ code, where $M = s^{\lceil n/w^2 \rceil}$.*

Let $A$ be an $(n_2, M_2, q_2)$ code over an alphabet $Q_2$ with $|Q_2| = q_2$ and let $B$ be an $(n_1, q_2, q_1)$ code over an alphabet $Q_1$ with $|Q_1| = q_1$. Let $Q_2 = \{a_1,\dots,a_{q_2}\}$ and let $B = \{\mathbf{b_1},\dots,\mathbf{b_{q_2}}\}$. Let $\theta : Q_2 \longrightarrow B$ be the one-to-one mapping defined by $\theta(a_i) = \mathbf{b_i}$ for $1 \leq i \leq q_2$. For any codeword $\mathbf{a} = (a_1,\dots,a_{n_2}) \in A$ we denote by $\tilde{\mathbf{a}} = (\theta(a_1),\dots,\theta(a_{n_2})) = (\mathbf{b_1},\dots\mathbf{b_{n_2}})$ the $q_1$-ary sequence of length $n_1n_2$ obtained from $\mathbf{a}$ by using $\theta$. The set $C = \{\tilde{\mathbf{a}} = (\mathbf{b_1},\dots,\mathbf{b_{n_2}})/\ \mathbf{a} = (a_1,\dots,a_{n_2}) \in A\}$ is an $(n_1n_2, M_2, q_1)$ code, called the concatenated code of $A$ and $B$.

Our next important theorem shows that the concatenation technique works for IPP codes.

**Theorem 3.5** *Let $A$ be an $(n_2, M_2, q_2, w) - IPP$ code and let $B$ be an $(n_1, q_2, q_1, w) - IPP$ code. Then the concatenated code $C$ of $A$ and $B$ is an $(n_1 n_2, M_2, q_1, w) - IPP$ code.*

*Proof.* Let $\mathbf{x} = (x_1, \ldots, x_{n_1 n_2}) \in Q_1^{n_1 n_2}$. We partition $\mathbf{x}$ into $n_2$ blocks $\mathbf{x}_1, \ldots, \mathbf{x}_{n_2}$ with $\mathbf{x}_i = (x_{(i-1)n_1+1}, \ldots, x_{in_1}) \in Q_1^{n_1}$, $1 \leq i \leq n_2$. We will write $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_{n_2})$. Specially, if $\mathbf{x} = \mathbf{c} = (\mathbf{b}_1, \ldots, \mathbf{b}_{n_2}) \in C$, then $\mathbf{b}_i$'s are themselves blocks of the partition of $\mathbf{c}$.

Suppose $\mathbf{x} \in \mathbf{desc}(C_i)$, $1 \leq i \leq r$, where $C_i \subseteq C$ with $|C_i| = \alpha_i \leq w$. We prove that $\bigcap_{1 \leq i \leq r}(C_i) \neq \emptyset$, i.e. $C$ is a $w - IPP$ code.

Let $C_i = \{\mathbf{c}_1^{(i)}, \ldots, \mathbf{c}_{\alpha_i}^{(i)}\} \subseteq C$, where $\mathbf{c}_j^{(i)} = (\mathbf{b}_{j1}^{(i)}, \ldots, \mathbf{b}_{jn_2}^{(i)})$. For any $1 \leq i \leq r$ and any $1 \leq \ell \leq n_2$ define $D_\ell^{(i)} = \{\mathbf{b}_{1\ell}^{(i)}, \ldots, \mathbf{b}_{\alpha_i \ell}^{(i)}\}$, i.e. $D_\ell^{(i)}$ is the collection of all $\ell^{th}$ blocks of the codewords of $C_i$. In other words $D_\ell^{(i)} \subseteq B$ is a subset of $\alpha_i$ codewords. As $\mathbf{x} \in \mathbf{desc}(C_i)$ by the assumption, we have $\mathbf{x}_\ell \in \mathbf{desc}(D_\ell^{(i)})$ for $1 \leq i \leq r$ and $1 \leq \ell \leq n_2$. Since $B$ is a $w - IPP$ code, we have

$$\bigcap_{1 \leq i \leq r} D_\ell^{(i)} \neq \emptyset.$$

Let $\mathbf{b}_\ell \in \bigcap_{1 \leq i \leq r} D_\ell^{(i)}$ be an arbitrary but fixed codeword, i.e. $\mathbf{b}_\ell$ is a parent of $\mathbf{x}_\ell$ in code $B$. Set $\mathbf{y} = (\mathbf{b}_1, \ldots, \mathbf{b}_{n_2})$. Let $\bar{\mathbf{y}} = (a_1, \ldots, a_{n_2}) \in Q^{n_2}$ be the corresponding sequence obtained from $\mathbf{y}$ using $\theta$, i.e. $a_i = \theta^{-1}(\mathbf{b}_i)$. In the same way let $\bar{C}_i = \{\bar{\mathbf{c}}_1^{(i)}, \ldots, \bar{\mathbf{c}}_{\alpha_i}^{(i)}\} \subseteq A$ denote the corresponding subset of $C_i$.

Since $\mathbf{y} \in \mathbf{desc}(C_i)$ by the construction, we have $\bar{\mathbf{y}} \in \mathbf{desc}(\bar{C}_i)$ for $1 \leq i \leq r$. Hence

$$\bar{\mathbf{y}} \in \bigcap_{1 \leq i \leq r} \mathbf{desc}(\bar{C}_i).$$

Since $A$ is a $w - IPP$ code, we have

$$\bigcap_{1 \leq i \leq r} \bar{C}_i \neq \emptyset.$$

Let $\bar{\mathbf{z}}' = (a_1', \ldots, a_{n_2}') \in \bigcap_{1 \leq i \leq r}(\bar{C}_i)$ be a parent of $\bar{\mathbf{y}}$ in $A$. Then $\mathbf{z}' = (\mathbf{b}_1', \ldots, \mathbf{b}_{n_2}') \in C_i$ for $1 \leq i \leq r$, where $\mathbf{z}'$ the codeword of $C$ corresponding to $\bar{\mathbf{z}}'$. Therefore

$$\bigcap_{1 \leq i \leq r} C_i \neq \emptyset.$$

Thus $C$ is an $w - IPP$ code. ■

**Remark 3.1** Note that the proof of Theorem 3.5 describes how to identify a traitor. This fact is used for the proof of Theorem 3.9.

**Remark 3.2** Observe that the minimum distance of $C$ in Theorem 3.5 does not satisfy the condition of Theorem 2.2 and $C$ is not a $w$-TA code even when $A$ and $B$ are $w$-TA codes, in general. Thus Theorem 3.5 gives a construction of "proper" $IPP$ codes in the sense that they are not $TA$ codes.

## 3.1 An infinite class of w-IPP codes with efficient traitor tracing algorithm

We are now in a position to describe our first construction.

The construction is carried out by induction on the number of iterations.

Let $w \geq 2$ be any integer. Let $n_0 > w^2$ be integer and let $s_0$ be an integer with the prime factorization $s_0 = p_1^{e_1} \dots p_k^{e_k}$ such that $n_0 \leq p_i^{e_i}$ for all $i = 1, \dots, k$.

For the $1^{st}$ iteration we choose two codes $C_0$ and $C_1^*$ using Theorem 3.4:

$C_0$ is an $(n_0, M_0, s_0, w) - IPP$ code with $M_0 = s_0^{\lceil \frac{n_0}{w^2} \rceil}$;

$C_1^*$ is an $(n_0^*, M_1, M_0, w) - IPP$ code with $n_0^* = n_0^{\lceil \frac{n_0}{w^2} \rceil}$ and $M_1 = M_0^{\lceil n_0^*/w^2 \rceil}$.

Applying Theorem 3.5 with $A$ replaced by $C_1^*$ and $B$ by $C_0$ we obtain an $(n_1, M_1, s_0, w) - IPP$ code $C_1$ with $n_1 = n_0 * n_0^* = n_0 * n_0^{\lceil \frac{n_0}{w^2} \rceil}$.

Now an $(n_{i-1}, M_{i-1}, s_0, w) - IPP$ code $C_{i-1}$ exists by induction for the $(i-1)^{th}$ iteration. Choose an $(n_{i-1}^*, M_i, M_{i-1}, w) - IPP$ code $C_i^*$ from Theorem 3.4 with

$$n_{i-1}^* = n_{i-2}^{*\lceil \frac{n_{i-2}^*}{w^2} \rceil} \text{ and } M_i = M_{i-1}^{\lceil \frac{n_{i-1}^*}{w^2} \rceil}.$$

Applying Theorem 3.5 with $A = C_i^*$ and $B = C_{i-1}$, we then get an $(n_i, M_i, s_0, w) - IPP$ code $C_i$ with

$$n_i = n_{i-1} * n_{i-2}^{*\lceil \frac{n_{i-2}^*}{w^2} \rceil}.$$

Thus we obtain the following result.

**Theorem 3.6** *Let $w \geq 2$ be any integer. Let $n_0 > w^2$ be integer and let $s_0$ be an integer with the prime factorization $s_0 = p_1^{e_1} \dots p_k^{e_k}$ such that $n_0 \leq p_i^{e_i}$ for all $i = 1, \dots, k$. Then, for all $h \geq 0$ there exists an $(n_h, M_h, s_0, w) - IPP$ code, where*

$$n_h = n_{h-1} * n_{h-1}^*, \quad M_h = M_{h-1}^{\lceil \frac{n_{h-1}^*}{w^2} \rceil}, \quad n_{h-1}^* = n_{h-2}^{*\lceil \frac{n_{h-2}^*}{w^2} \rceil},$$

$$M_0 = s_0^{\lceil \frac{n_0}{w^2} \rceil}, \quad and \quad n_0^* = n_0^{\lceil \frac{n_0}{w^2} \rceil}.$$

The asymptotic behavior of the parameters of the codes produced by Theorem 3.6 can be examined by a similar argument, which is demonstrated in [24], pp. 196-197. In fact, we can show that

$$n_h \leq \alpha.(w^2)^{\log^*(M_h)}(\log M_h),$$

for all sufficiently large $h$, where $\alpha$ is some constant and the function $\log^* : \mathbb{Z}^+ \longrightarrow \mathbb{Z}^+$ is defined recursively by

$$\begin{aligned} \log^*(1) &= 1 \\ \log^*(n) &= \log^*(\lceil \log n \rceil) + 1, \quad \text{if } n > 1. \end{aligned}$$

Note that the function $\log^*(n)$ grows very slowly, e.g. $\log^*(n) \leq 7$ for $n \leq 2^{2^{65536}}$.

We have the following result.

**Theorem 3.7** *For any integer $w \geq 2$ and any integer $s$ having the prime factorization $s = p_1^{e_1} \dots p_k^{e_k}$ with $w^2 < p_i^{e_i}$ for all $i = 1, \dots, k$, there exists an infinite class of $(n, M, s, w) - IPP$ codes for which $n$ is $O((w^2)^{log^*(M)}(log(M))$.*

As we want to show that the constructed codes in Theorem 3.6 having an efficient tracing algorithm, we have chosen the starter code as an $MDS$ code. In fact, the construction works for any starter code. For instance, for given $M, q, w \geq 2$, the probabilistic method in [2] shows the existence of $(n', M, q, w) - IPP$ codes with $q > w$ and some $n'$. Thus, if we take this $(n', M, q, w) - IPP$ code as a starter code and carry out the same recursive construction, then we get a more general result as follows.

**Theorem 3.8** *For any integer $w \geq 2$ and any integer $q \geq w$, there exists an infinite class of $(n, M, q, w) - IPP$ codes for which $n$ is $O((w^2)^{log^*(M)}(log(M))$.*

To our knowledge Theorem 3.7 and 3.8 yield a class of explicit constructed codes with the best known asymptotic behavior. In fact, Stinson, Wei and Zhu [24] recently give an explicit construction for an infinite class of perfect hash families $(n, M, q, w) - PHF$, in which $n$ is $O((w^2)^{\log^*(M)}(\log(M))$. This class is asymptotically among the best explicit constructed perfect hash families known in the literature. On the other hand, an $(n, M, q, w) - IPP$ code is an $(n, M, q, w+1) - PHF$, see e.g. [21], and therefore an $(n, M, q, w) - PHF$. But the converse is not true: an $(n, M, q, w+1) - PHF$ is not an $(n, M, q, w) - IPP$ code in general. This is to say that an $(n, M, q, w) - IPP$ code is a much stronger structure than an $(n, M, q, w) - PHF$. Even though, our constructed $IPP$ codes have the same asymptotic size as that of the best known explicitly constructed classes of $PHF$.

**Remark 3.3** It is worth noting that in a recent paper [19], Sarkar and Stinson construct an infinite class of $(n, M, q, w)$-IPP codes for which $n$ is $O((w^3)^{\log^*(M)}(\log(M))$, for integers $q > w \geq 2$ in terms of strong separating hash families.

### 3.2 An efficient traitor tracing algorithm

For $w$-IPP codes, a traitor tracing algorithm (TTA) will have a runtime complexity of size $O(\binom{M}{w})$, in general. For $w$-TA codes, however, the runtime of a TTA will be $O(M)$, (see e.g.[20]) for more information. Therefore, the question of the existence of $w$-IPP codes in general with an improved runtime for a TTA was raised in [21].

Here, we show that our constructed $w$-IPP codes have a TTA with a runtime $O(M)$, thereby answering the above question affirmatively.

The recursive process of concatenation used to construct $w-IPP$ codes in Theorem 3.6 provides a way to build a TTA for code $C_i$ based on the TTA's of codes $C_{i-1}$ and $C_i^*$. In fact, the proof of Theorem 3.5 describes precisely how a traitor can be traced back for the code $C_i$. In doing so we assume that the TTA's for codes $C_{i-1}$ and $C_i^*$ are known. Let $L_{i-1}$ and $L_i^*$ be the runtime complexity of such a TTA for $C_{i-1}$ and $C_i^*$, respectively. Let assume $\mathbf{x} \in \mathbf{desc}(K_j)$ , for $j = 1, \ldots, r$, and $K_j \subseteq C_i$ with $|K_j| \leq w$, i.e., $\mathbf{x}$ is a pirate word of length $n_i = n_{i-1} * n_{i-1}^*$ created by $r$ possible coalitions $K_j$. From the proof of Theorem 3.5 we see that the runtime $L_i$ of a TTA for code $C_i$ is given by $L_i = L_{i-1} * n_{i-1}^* + L_i^*$. If we start with $C_0$ and $C_1^*$ as $w$-TA codes, for which the runtime of their TTAs are $O(M_0)$ and $O(M_1)$, then we have $L_1 = O(M_1)$, as $|M_0| << |M_1|$. Therefore, if $C_i^*$ is a $w$-TA code for each step of the recursion, then we have $L_i = O(M_i)$. Now the codes $C_0$ and $C_i^*$ in Theorem 3.6 are in fact $w$-TA codes, so we have the following result.

**Theorem 3.9** *For any integer $w \geq 2$ and any integer $s$ having the prime factorization $s = p_1^{e_1} \ldots p_k^{e_k}$ with $w^2 < p_i^{e_i}$ for all $i = 1, \ldots, k$, there exists an infinite class of $(n, M, s, w) - IPP$ codes with $n$ is $O((w^2)^{log^*(M)}(log(M))$, which have a traitor tracing algorithm of linear runtime $O(M)$.*

In [25, 26], Sudan develops methods for list decoding for certain class of error-correcting codes. The method has been improved since then. For example, in [13, 14, 15] Guruswami and Sudan present efficient list decoding algorithms for Reed-Solomon codes, algebraic-geometric, and certain concatenated codes. It turns out that the method of list decoding can be applied to traitor tracing algorithms, when the mentioned codes are used as $TA$-codes. This fact is discussed by Silverberg, Staddon and Walker in [21] . For instance, the $TA$ codes based on Reed-Solomon codes will have traitor tracing algorithms of runtime $poly(\log M)$, where $M$ is the size of the codes. This, in turn, implies that the method can be applied to our constructed $IPP$ codes.

Consequently, if $s = q$ is a prime power and the ingredients of the recursion are Reed-Solomon codes, then the $IPP$ codes of Theorem 3.9 allow a traitor tracing algorithm which can run in $poly(\log M)$ time.

# 4 A construction for perfect hash families and IPP codes

The main result in this section is the construction of an infinite class of perfect hash families by means of a double recursive method. The resulting perfect hash families are then used to derive IPP codes in view of Theorem 2.4. But we emphasize that our construction method for perfect hash families presented here is of independent interest. This construction is a generalization of a construction given in [17]. Moreover, we would remark that the construction in this section appears to be rather complex, even though we have attempted to give a clear concise explanation.

## 4.1 A recursive construction of perfect hash families

We first prove Lemma 4.1 below, which is essential for our purpose.

From now on let $q$ be a prime power. We begin with a description of a collection of matrices derived from mutually orthogonal latin squares (MOLS) whose symbols are elements in the finite field $F_q = \{0, 1, \ldots, q-1\}$. For basic facts on MOLS, we refer to [6].

Let $M_1, \ldots, M_{q-1}$ be a set of $q-1$ MOLS, of which the first column is the vector $(0, 1, \ldots, q-1)^T$. Let $M_0$ be the $q \times q$ matrix whose all $q$ columns are equal to the vector $(0, 1, \ldots, q-1)^T$ (i.e. each row of $M_0$ consists of a $q$ time repeating of a symbol). The collection of $M_0, \ldots, M_{q-1}$ is equivalent to an orthogonal array $OA_1(2, q, q)$ (see, for example [6, p. 130]) and hence to a Reed-Solomon code $\mathcal{RS}$ with parameters $(q, q^2, q, d = q-1)$.

For $2 \leq m \leq q$, set

$$\mathcal{A} = \{A_{0,m}, \ldots, A_{q-1,m}\},$$

where each matrix $A_{h,m}$ is obtained from $M_h$ by deleting its $q - m$ rightmost columns.

Consider the $q^2 \times (m+1)$ array $\mathcal{A}^E$ obtained from $\mathcal{A}$ by extending each matrix $A_{i,m}$ with the $(m+1)^{th}$ column $(i, i, \ldots, i)^T$. Then $\mathcal{A}^E$ is equivalent to the Reed-Solomon code $(m+1, q^2, q, d = m) - \mathcal{RS}$. By Lemma 2.3 $\mathcal{A}^E$ is an $(m+1, q^2, q, w) - PHF$ with $\binom{w}{2} < m + 1$.

Conversely, if $w$ is given, we set $m = \binom{w}{2}$. Then the collection $\mathcal{A}$ has the following crucial property: every subset $\mathcal{B}$ of $w'$ distinct matrices $A_{i_1,m}, \ldots, A_{i_{w'},m}$ of $\mathcal{A}$, where $1 \leq w' \leq w-1$, forms an $(m, qw', q, w) - PHF$.

This can be easily seen as follows.

Consider $\mathcal{B}$ as part of $\mathcal{A}^E$. Note that $\mathcal{A}^E$ has exactly one column more than $\mathcal{B}$, the $(m+1)^{th}$ column. For any given set $W$ of $w$ rows of $\mathcal{B}$, there is a column $\mathbf{c}$ in $\mathcal{A}^E$, such that the symbols of $\mathbf{c}$ at the given $w$ rows are pairwise distinct, because $\mathcal{A}^E$ is a $(m+1, q^2, q, w) - PHF$. Further, since $\mathcal{B}$ is a collection of $w'$ matrices $A_{h,m}$, there are at least two rows of $W$ belonging to the same matrix in $\mathcal{B}$. This implies that the column $\mathbf{c}$ is not the $(m+1)^{th}$ column of $\mathcal{A}^E$, hence $\mathbf{c}$ must be a column of $\mathcal{B}$, as desired.

Thus, we have proved the following result.

**Lemma 4.1** *Let $\mathcal{A}$ be the collection of $q$ matrices $\{A_{0,m}, \ldots, A_{q-1,m}\}$ just described above, where each $A_{h,m}$ is a $q \times m$ matrix, whose entries are elements of $F_q$. Let $m = \binom{w}{2}$. Then, any subset $\mathcal{B}$ of $w'$ distinct matrices $A_{i_1,m}, \ldots, A_{i_{w'},m}$ of $\mathcal{A}$, where $1 \leq w' \leq w-1$, forms an $(m, q.w', q, w) - PHF$.*

We are now ready to prove the following theorem.

**Theorem 4.2** *For any positive integers $i \geq 1$, $w \geq 2$ and any prime power $q \geq \binom{w}{2}$ there exists an $(O((i+1)^{w-1}), q^{i+1}, q, w) - PHF$.*

*Proof.* The proof is by induction on $w$ and $i$.

In the following we use $n_i(w)$ as an abbreviation for $O((i+1)^{w-1})$ and $C_i^w$ for $(n_i(w), q^{i+1}, q, w) - PHF$.

Note that the vector space $F_q^{i+1}$ is an $(n_i(2), q^{i+1}, q, 2) - PHF$, where $n_i(2) = i+1$. Thus $C_i^2$ exists for all $i \geq 1$. In other words the statement is valid for $w = 2$.

Assume that the statement is valid for $w - 1 > 2$. That means that for every $2 \leq u \leq w-1$ there exists an $C_i^u = (n_i(u), q^{i+1}, q, u) - PHF$ for all $i$. We prove that the statement is true for $w$, i.e. there is an $C_i^w = (n_i(w), q^{i+1}, q, w) - PHF$ for every $i$.

This is done by induction on $i$.

For $i = 1$ there is a $C_1^w = (n_1(w), q^2, q, w) - PHF$, where $n_1(w) = \binom{w}{2} + 1$ and $q \geq n_1(w) - 1$. In fact, $C_1^w$ is obtained from the Reed-Solomon code $(n_1(w), q^2, q) - \mathcal{RS}$ by using Lemma 2.3. Assume that $C_j^w$ exists for all $j \leq i-1$.

Let

$$\tilde{C_i^w} = (D_{i-1}^w, E_{i-1}^{w-1})$$

denote the concatenation of $D_{i-1}^w$ and $E_{i-1}^{w-1}$, which are defined as follows.

$D_{i-1}^w$ is obtained from $C_{i-1}^w$ by repeating each of its rows $q$ times.

$E_{i-1}^{w-1}$ is obtained from $C_{i-1}^{w-1}$ by replacing each symbol $j$ by matrix $A_{j,w}$, described in Lemma 4.1.

We depict $\tilde{C_i^w}$ as an $q.q^i \times (n_{i-1}(w) + n_{i-1}(w-1).\binom{w}{2}$ array, where the first $n_{i-1}(w)$ columns correspond to $D_{i-1}^w$ and the remaining $n_{i-1}(w-1).\binom{w}{2}$ columns correspond to $E_{i-1}^{w-1}$. And we partition the rows of the array $\tilde{C_i^w}$ into $q^i$ consecutive blocks, say $B_1, \ldots, B_{q^i}$, each block $B_i$ has $q$ rows.

| | $D^{w}_{i-1}$ | $E^{w-1}_{i-1}$ | | | |
|---|---|---|---|---|---|
| $B_1$ | 1st row of $C^{w}_{i-1}$ repeated $q$ times | $A_{(1,1),w}$ | $A_{(1,2),w}$ | $\dots$ | $A_{(1,n_{i-1}(w)),w}$ |
| $B_2$ | 2nd row of $C^{w}_{i-1}$ repeated $q$ times | $A_{(2,1),w}$ | $A_{(2,2),w}$ | $\dots$ | $A_{(2,n_{i-1}(w)),w}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $B_{q^i}$ | $q^i$th row of $C^{w}_{i-1}$ repeated $q$ times | $A_{(q^i,1),w}$ | $A_{(q^i,2),w}$ | $\dots$ | $A_{(q^i,n_{i-1}(w)),w}$ |

Array $\tilde{C_i^w}$

Remark that the matrix $A_{(j,k),w}$ in the table corresponds to the symbol at the entry $(j,k)$ of the array $C^{w-1}_{i-1}$.

Next, we prove that $\tilde{C_i^w}$ is a $w-PHF$.

Let $r_1,\dots,r_w$ be any given $w$ rows of $\tilde{C_i^w}$. If $r_1,\dots,r_w$ belong to $w$ different blocks, say $B_{i_1},\dots,B_{i_w}$, then from the definition of $D^w_{i-1}$ there is at least one column in $D^w_{i-1}$ containing pairwise distinct symbols in the rows $r_1,\dots,r_w$. Assume that $r_1,\dots,r_w$ belong to $w'$ blocks, say $B_{i_1},\dots,B_{i_{w'}}$, where $w' \le w-1$. As $C^{w-1}_{i-1}$ is an $(w-1)-PHF$, there exists a column, say $\mathbf{c}$, whose symbols, say $j_1,\dots,j_{w'}$, in the rows $i_1,\dots,i_{w'}$ are pairwise distinct. From the definition of $E^{w-1}_{i-1}$, the symbols $j_1,\dots,j_{w'}$ are replaced by matrices $A_{j_1,w},\dots,A_{j_{w'},w}$ (notice that $A_{j_1,w},\dots,A_{j_{w'},w}$ together form a set of $\binom{w}{2}$ consecutive columns of the blocks $B_{i_1},\dots,B_{i_{w'}}$ in $E^{w-1}_{i-1}$). By Lemma 4.1 $A_{j_1,w},\dots,A_{j_{w'},w}$ is an $(\binom{w}{2}, q.w', q, w)-PHF$, so there is a column in $E^{w-1}_{i-1}$ having different symbols in the rows $r_1,\dots,r_w$. Thus $\tilde{C_i^w}$ is a $w-PHF$.

Now recall that $C^{w-1}_{i-1}$ is an $q^i \times n_{i-1}(w-1)$ array and that $E^{w-1}_{i-1}$ is obtained from $C^{w-1}_{i-1}$ by replacing each entry $j \in \{0,\dots,q-1\}$ of $C^{w-1}_{i-1}$ by the $(q \times \binom{w}{2})$-matrix $A_{j,w}$.

Since the first column of each matrix $A_{j,w}$ is always the vector $(0,\dots,q-1)^T$, there are $n_{i-1}(w-1)$ identical columns in $\tilde{C_i^w}$.

Now let $C_i^w$ denote the array obtained from $\tilde{C_i^w}$ by deleting $n_{i-1}(w-1)-1$ of these identical columns. Then $C_i^w$ is an $q^{i+1} \times n_i(w)$ array, where

$$n_i(w) \quad = \quad n_{i-1}(w) + n_{i-1}(w-1) \times \binom{w}{2} - (n_{i-1}(w-1)-1)$$

$$= \quad n_{i-1}(w) + n_{i-1}(w-1)(\binom{w}{2} - 1) + 1.$$

It is obvious that $C_i^w$ is an $w - PHF$, just as $\tilde{C_i^w}$.

As $n_{i-1}(w) = O(i^{w-1})$ and $n_{i-1}(w-1) = O(i^{w-2})$, we have

$$n_i(w) = O(i^{w-1}) + O(i^{w-2})[\binom{w}{2} - 1].$$

Consequently

$$n_i(w) = O((i+1)^{w-1}).$$

Hence $C_i^w$ is an $(O((i+1)^{w-1}, q^{i+1}, q, w) - PHF$, as desired. ■

**Remark 4.1** We remark that the case $w = 3$ and $q = 3$ in Theorem 4.2 has been studied by S. S. Martirosyan and S. S. Martirosyan in [17], wherein a recursive algorithm is presented, which constructs an infinite class of $(j^2, 3^j, 3, 3) - PHF$, for every integer $j \geq 1$.

## 4.2 A new class of $w$-IPP codes

Using Theorem 2.4 and Theorem 4.2 we immediately obtain the following new class of IPP codes.

**Theorem 4.3** *For any positive integers $i$ and $w \geq 2$ and any prime power $q$ with $q \geq \binom{\lfloor (w+2)^2/4 \rfloor}{2}$ there exists an $(O((i+1)^{\lfloor (w+2)^2/4 \rfloor - 1}), q^{i+1}, q, w) - IPP$ code.*

*Proof.* By Theorem 4.2 there is an $(O((i+1)^{\lfloor (w+2)^2/4 \rfloor - 1}), q^{i+1}, q, \lfloor (w+2)^2/4 \rfloor) - PHF$. The theorem then follows from Theorem 2.4. ■

It is worth noting that at each recursion step the size of the constructed code in Theorem 4.3 increases much slower than that in Theorem 3.6. Actually, Theorem 4.3 roughly states that $w - IPP$ codes of certain codeword length can be constructed for any given $w$ and any given code size $q^i$. Thus, Theorem 4.3 gives an explicit construction of IPP codes for a very large set of parameter values.

# References

[1] N. Alon, Explicit construction of exponential sized families of k-independent sets, *Discrete Math.* **58** (1986), 191–193.

[2] A. Barg, G. Cohen, S. Encheva, G. Kabatiansky and G. Zémor, A hypergraph approach to the identifying parent property: the case of multiple parents, *SIAM J. Discrete. Math.,* **14** (2001), pp. 423–431.

[3] D. Boneh and M. Franklin, An efficient public key traitor tracing schemes, in *Advances in Cryptology - Crypto'94 (Lecture Notes in Computer Science)*, Springer-Verlag **839** (1994), 257–270.

[4] D. Boneh and J. Shaw, Collusion-secure fingerprinting for digital data, *IEEE Trans. Inform. Theory* **44** (1998), 1897–1905.

[5] K. A. Bush, A generalization of a theorem due to MacNeish, *Ann. Math. Stat.,* **23** (1952), 293–295.

[6] C. J. Colbourn and J. H. Dinitz, CRC Handbook of Combinatorial Designs, *CRC Press, Inc.,* 1996.

[7] Z. J. Czech, G. Havas, and B. S. Majewski, Perfect hashing, *Theor. Comp. Sci.* **182** (1997), 1–143.

[8] B. Chor, A. Fiat and M. Naor, Tracing traitors, in *Advances in Cryptology - Crypto'94 (Lecture Notes in Computer Science)*, Springer-Verlag, **839** (1994), 257–270.

[9] B. Chor, A. Fiat, M. Naor, and B. Pinkas, Tracing traitors, *IEEE Trans. Inform. Theory* **46** (2000), 480–491.

[10] A. Fiat and T. Tassa, Dynamic traitor tracing, in *Advances in Cryptology Crypto'99 (Lecture Notes in Computer Science)*, Springer-Verlag, **1666** (1999), 354–371.

[11] A. Fiat and T. Tassa, Dynamic traitor tracing, *J. Cryptology* **14** (2001), 211–223.

[12] kE. Gafni, J. Staddon, and Y. L. Yin, Efficient methods for integrating traceability and broadcast encryption, in *Advances in Cryptology–Crypto'99 (Lecture Notes in Computer Science)*, Springer-Verlag, **1666** (1999), 372–387.

[13] V. Guruswami and M. Sudan, Improved decoding of Reed-Solomon and algebraic-geometry codes, *IEEE Trans. Inform. Theory* **45** (1999), 1757–1767.

[14] V. Guruswami and M. Sudan, List decoding algorithms for certain concatenated codes, *Proc. 32nd ACM Symposium on Theory of Computing (STOC 2000)*, 181–190.

[15] V. Guruswami, List decoding of error-correcting codes, PhD Thesis, MIT (2001).

[16] H. D. L. Hollmann, J.H. van Lint, J- P. Linnartz and L. M. G. M Tolhuizen, On codes with identifiable parent property, *J. Comb. Theory A*, **82** (1998), 121–133.

[17] S. S. Martirosyan and S. S. Martirosyan, New Upper Bound on the Cardinality of a K-separated Set or Perfect Hash Family and a Near Optimal Construction for It, *Transactions of IPIA of NAN RA & YSU "Mathematical Problems of Computer Science"*, vol. XXI (2000), 104–115.

[18] R. Safavi-Naini and Yejing Wang, New results on frameproof codes and traceability schemes, *IEEE Trans. Inform. Theory* **47** (2001), 3029–3033.

[19] P. Sarkar, D. R. Stinson, Frameproof and IPP codes, Preprint.

[20] A. Silverberg, J. N. Staddon, and J. L. Walker, Efficient Traitor Tracing Algorithms using List Decoding, *ASIACRYPT 2001*, Lect. Notes Comput. Sci. **2248** (2001), pp. 175–192.

[21] J. N. Staddon, D. R. Stinson and R. Wei, Combinatorial properties of frameproof and traceability codes, *IEEE Trans. Inform. Theory* **47** (2001), 1042–1049.

[22] D. R. Stinson, Tran van Trung and R. Wei, Secure frameproof codes, key distribution patterns, group testing algorithms and related structures, *J. Statist. Planning Inference* **86** (2000), 595–617.

[23] D. R. Stinson and R. Wei, Combinatorial properties and constructions of traceability schemes and frameproof codes, *SIAM J. Discrete Math.* **11** (1998), 41–53.

[24] D. R. Stinson and R. Wei and L. Zhu, New constructions for perfect hash families and related structures using combinatorial designs and codes, *J. Combin. Designs* **8** (2000), 189–200.

[25] M. Sudan, Decoding of Reed-Solomon codes beyond the error-correction bound. *Journal of Complexity* **13** (1997), 180–193.

[26] M. Sudan, Decoding of Reed-Solomon codes beyond the error-correction diameter. *Proc. 35th Annual Allerton Conference on Communication, Control and Computing* (1997), 215–224.

[27] Tran van Trung and Sosina Martirosyan, On a Class of Traceability Codes, to appear in *Designs, Codes and Cryptography.*