Two-Level Fingerprinting Codes: Non-Trivial Constructions

Penying Rochanakul
ROYAL HOLLOWAY, UNIVERSITY OF LONDON
P.Rochanakul@rhul.ac.uk

Abstract

We extend the concept of two-level fingerprinting codes, introduced by Anthapadmanabhan and Barg (2009) in context of traceability (TA) codes [1], to other types of fingerprinting codes, namely identifiable parent property (IPP) codes, secure-frameproof (SFP) codes, and frameproof (FP) codes. We define and propose the first explicit non-trivial construction for two-level IPP, SFP and FP codes.

1 Introduction

Fingerprinting codes have been studied for more than 15 years due to their applications in digital data copyright protection and their combinatorial interest. In order to trace back to the source of pirate piece of digital data, the data distributor hides a unique mark called a fingerprint in each legal copy. The fingerprint can be viewed as a codeword length \( \ell \), which each component is picked from alphabet \( Q \) of cardinality \( q \). Since all fingerprints are unique, if a naive user tries to distribute his or her copy illegally, the fingerprint will lead back to its owner right away. However, if a group of users form a coalition and pool their copies together, it becomes a lot more complicated to trace back to the coalition members. It is assumed that the coalition can only produce data fingerprinted by words such that each component agrees with at least one of the codewords in the coalition: an element of the descendant set of the codewords. We define descendant again clearly in the next section. Once such a descendant is given, TA codes are designed to trace back at least one member of the coalition, provided that the size of the coalition does not exceed a certain parameter \( t \). The user with the codeword that is most similar to the given descendant, is guaranteed to be a member of the coalition. However, there is no guarantee the tracing result is correct if the size of the coalition is greater than \( t \). The first attempt to gain more information about the coalition when its size is bigger than \( t \) was by Anthapadmanabhan and Barg (2009) in [1] who defined the concept of two-level fingerprinting codes; from now on we refer to their codes as two-level traceability codes. In Anthapadmanabhan and Barg’s setting, the users are divided into various groups; for instance, by dividing the distribution area into several geographic regions, and collecting users from the same region into one group. As in traditional
(one-level) codes, when given an illegal copy produced by a coalition of users, the codes can be used to identify one of the guilty users if the coalition size is less than or equal to a certain threshold $t$. However, even when the coalition is of a larger size $T$, where $T > t$, the decoder still provides partial information by tracing one of the groups containing a guilty user. The group that contains the codeword that is closest to the illegal copy must contain at least one member of the coalition. However, they provide no explicit construction of two-level traceability codes apart from using random codes when $t = 1$ or $t = 2$.

IPP, SFP and FP codes have been widely studied as replacements for, or weakenings of, TA codes [5, 10]. In this paper, we aim to explore TA, IPP, SFP and FP codes in the two-level context. Although TA codes have a very efficient algorithm for traitor tracing and very interesting in terms of applications, there is no explicit construction for two-level traceability codes available, except in the trivial case. IPP codes can replace TA codes, since they also have tracing ability when a coalition under a threshold $t$. Moreover, under the same parameters, such as length and alphabet size, IPP codes are much bigger. Therefore, it is important to study two-level IPP codes. We also study SFP and FP codes, which focus on preventing innocent users from being framed, and analyze their interesting combinatorial properties. We define each type of codes properly in the next section. Our main result (Theorems 4, 5, and 6) is a construction for two-level IPP, SFP and FP codes when the number of the groups is small, i.e. less than or equal to the alphabet size. This is actually the first explicit non-trivial constructions for two-level fingerprinting codes.

More details on fingerprinting codes’ applications and combinatorial properties can be found in a survey by Blackburn [3]. We also suggest reading these prior works for more in-depth details: see [6] and [7] for the concept of traitor tracing along with traceability; see [4], [9], and [11] for further combinatorial properties of TA codes; more details about FP and SFP codes are available in [5] and [10], respectively.

The different types of codes are defined in Section 2 and the relationships between the definitions are explored. In Section 3, we provide our proposed construction for two-level codes. Section 4 is a short conclusion containing open problems.

2 Definitions and Relationships

In this section, we restate the definitions of the classical one-level codes we are interested in, then define the corresponding two-level codes.
2.1 One-Level Codes

Before giving the codes’ definitions, it is necessary to introduce some notation. Let $C$ be a code of length $\ell$ on an alphabet $Q$ of (finite) size $q$. Then, we issue data with a hidden different fingerprint from $C$ to each user. The elements of $C$ are called codewords. The hamming distance between codewords $x, y$ will be written as $d_H(x, y)$. For any $X \subseteq Q^\ell$ and $y \in Q^\ell$, let $d(X)$ denote the minimum distance of $X$ and let $d_H(X, y) = \min_{x \in X} d_H(x, y)$.

We call any subset $C \subseteq Q^\ell$, a $q$-ary length $\ell$ code. For each $x \in Q^\ell$, $x_i$ for the $i$th component of $x$. For $X \subseteq Q^\ell$, we define the set of descendants of $X$ to be the subset desc$(X) \subseteq Q^\ell$ given by

$$\text{desc}(X) = \{ d \in Q^\ell : \forall i \in \{1, 2, ..., \ell\}, \exists x \in X \text{ such that } d_i = x_i \}.$$ 

For example, let $P = \{1100, 2102, 1122\}$, then

$$\text{desc}(P) = \{1100, 1102, 1120, 1122, 2100, 2102, 2120, 2122\}.$$ 

Let $t$ be a positive integer. For a code $C$ define the $t$-descendant code of $C$, denoted $\text{desc}_t(C)$, by

$$\text{desc}_t(C) = \bigcup_{X \subseteq C, |X| \leq t} \text{desc}(X).$$

Definition 1. Let $C$ be an $q$-ary length $\ell$ code and let $t$ be a positive integer.

(i) $C$ has the $t$-frameproof property (or is $t$-FP) if for all $X \subseteq C$ such that $|X| \leq t$, it holds that

$$\text{desc}(X) \cap C \subseteq X$$

(ii) $C$ has the $t$-secure-frameproof property (or is $t$-SFP) if for all $X_0, X_1 \subseteq C$ of size at most $t$, if $\text{desc}(X_0) \cap \text{desc}(X_1) \neq \emptyset$, then $X_0 \cap X_1 \neq \emptyset$.

(iii) $C$ has the $t$-identifiable parent property (or is $t$-IPP) if for all $x \in \text{desc}_t(C)$, it holds that

$$\bigcap_{X \subseteq C, |X| \leq t, x \in \text{desc}(X)} X \neq \emptyset.$$ 

(iv) $C$ has the $t$-traceability property (or is $t$-TA) if for all $X \subseteq C$ that $|X| \leq t$ and for all $x \in \text{desc}(X)$, for all $z \in C$ with $d_H(x, z)$ minimal, $z \in X$. 

3
Presuming that the coalition size is at most $t$. None of the coalition under FP codes can produce a codeword that is not belong to the coalition. In SFP codes, the descendant sets of codewords from two disjoint coalitions are always disjoints. All the coalition of IPP codes that can produce the same words, own at least one codeword in common. Lastly, in TA codes, given a descendant of any coalition under TA code, a codeword that is most similar to the given descendant, is always a member of the coalition.

It is not difficult to check from the definitions that the relationships among different types of codes are as follows; $t$-TA codes are $t$-IPP codes, $t$-IPP codes are $t$-SFP codes and $t$-SFP codes are $t$-FP codes.

### 2.2 Two-Level Codes

In this section, we extend the concept of two-level fingerprinting codes, introduced by Anthapadmanabhan and Barg (2009) in context of traceability (TA) codes [1], to IPP, SFP and FP codes. Then, we present an overview of relationships between different types of fingerprinting codes.

The motivation behind the definitions of two-level fingerprinting codes is to gain more information from a word produced by a coalition that is bigger than the first threshold $t$. Consider the following scenario. The content distributor distributes the same number of distinct digital copies to each company. Then those companies assign each copy in their hands to an individual employer. Here each company is acting as a group in our two-level model. Once a piracy has occurred, apart from tracing back to an individual traitor or protecting an innocent user from being framed as can be achieved by one-level codes, one might be interested in just tracing back to a company that employs one of the coalition members and sue the whole company (two-level TA and two-level IPP), or to make sure that none of those companies can cooperate and frame the other innocent companies without getting one of their employee involved (two-level SFP and two-level FP).

In traditional one-level codes, we assign a different fingerprint from $C$ to each user. Here $C$ is a code of length $\ell$ over an alphabet $Q$ of (finite) size $q$. In two-level codes, we partition a one-level code $C$ into $g$ disjoint sets of $p$ elements each, denoted by $C_1, C_2, ..., C_g$. Then we call $C$ a $q$-ary length $\ell$ two-level code, containing $g$ groups of size $p$. Hence, $C = C_1 \cup C_2 \cup ... \cup C_g$, we have $|C_i| = p$ for all $i, j \in \{1, 2, ..., g\}$, and $C_i \cap C_j = \emptyset$ when $i \neq j$. For a codeword $c \in C_i$, let $G(c)$ represent its group index $i$.

**Definition 2.** Let $C = C_1 \cup C_2 \cup ... \cup C_g$ be a two-level $q$-ary length $\ell$ code and let $T, t$ be positive integers where $T \geq t$.

(i) $C$ has the $(T, t)$-frameproof property (or is $(T, t)$-FP) if
(a) $C$ is $t$-FP when viewed as a $q$-ary length $\ell$ code, and
(b) for all $X \subseteq C$ such that $|X| \leq T$ and for all $x \in \text{desc}(X) \cap C$, $\mathcal{G}(x) \in \mathcal{G}(X)$.

(ii) $C$ has the $(T, t)$-secure-frameproof property (or is $(T, t)$-SFP) if

(a) $C$ is $t$-SFP when viewed as a $q$-ary length $\ell$ code, and
(b) for all $X_0, X_1 \subseteq C$ of size at most $T$, if $\text{desc}(X_0) \cap \text{desc}(X_1) \neq \emptyset$, then $\mathcal{G}(X_0) \cap \mathcal{G}(X_1) \neq \emptyset$.

(iii) $C$ has the $(T, t)$-identifiable parent property (or is $(T, t)$-IPP) if

(a) $C$ is $t$-IPP when viewed as a $q$-ary length $\ell$ code, and
(b) for all $x \in \text{desc}_T(C)$,

\[
\bigcap_{X \subseteq C:|X| \leq T, x \in \text{desc}(X)} \mathcal{G}(X) \neq \emptyset.
\]

(iv) $C$ has the $(T, t)$-traceability property (or is $(T, t)$-TA) if

(a) $C$ is $t$-TA when viewed as a $q$-ary length $\ell$ code, and
(b) for all $X \subseteq C$ that $|X| \leq T$ and for all $x \in \text{desc}(X)$, for all $z \in C$ with $d_H(x, z)$ minimal (i.e. $d_H(x, z) = d_H(C, x)$), $\mathcal{G}(z) \in \mathcal{G}(X)$.

We refer to $(T, t)$-FP codes, $(T, t)$-SFP codes, $(T, t)$-IPP codes and $(T, t)$-TA codes as two-level codes, and refer to all codes in Definition 1 as one-level codes.

The relationships among different types of two-level codes and one-level codes can be illustrated as in the following diagram.

\[
\begin{array}{ccc}
T-\text{TA} & \implies & (T, t)-\text{TA} & \implies & t-\text{TA} \\
\downarrow & & \downarrow & & \downarrow \\
T-\text{IPP} & \implies & (T, t)-\text{IPP} & \implies & t-\text{IPP} \\
\downarrow & & \downarrow & & \downarrow \\
T-\text{SFP} & \implies & (T, t)-\text{SFP} & \implies & t-\text{SFP} \\
\downarrow & & \downarrow & & \downarrow \\
T-\text{FP} & \implies & (T, t)-\text{FP} & \implies & t-\text{FP}.
\end{array}
\]

The proofs of the implications are straightforward from the codes’ definitions. We state and prove the next theorem as an example.
Theorem 1. A $(T,t)$-TA code is a $(T,t)$-IPP code.

Proof. Let $C$ be a $(T,t)$-TA code. Accordingly, $C$ is $t$-TA when viewed as a one-level code. That is, $C$ is a $t$-IPP code.

Let $x \in \text{desc}_T(C)$ and let $D_x$ be a set of all elements $z$ in $C$ that give the minimum value of $d_H(z,x)$. Let $U \subseteq C$ be any subset of size at most $T$ such that $x \in \text{desc}(U)$. By the definition of two-level traceability code, we know that for all $z \in G D_x$, $G(z) \in G(U)$. Hence $G(D_x) \subseteq G(U)$. Then, we have

$$G(D_x) \subseteq \bigcap_{X \subseteq C, |X| \leq T \atop x \in \text{desc}(X)} G(X).$$

Consequently,

$$\bigcap_{X \subseteq C, |X| \leq T \atop x \in \text{desc}(X)} G(X) \neq \emptyset.$$

Therefore, $C$ is a $(T,t)$-IPP code. \hfill \Box

The proof of the rest of the relationships between the different type of codes can be found in [8].

As a consequence of the definitions of codes, all two-level fingerprinting codes possess the properties of one-level codes. So, it is natural to construct two-level fingerprinting codes from existing one-level fingerprinting codes. We propose our explicit construction in the next section.

3 Code Constructions

In this section, we aim to construct two-level fingerprinting codes from existing one-level fingerprinting codes. Our construction works provided the number of groups is at most the alphabet size. Our construction begins with a one-level code, removes some codewords, groups and modifies the remaining codewords. The results are two-level codes with are guaranteed to be at least half the size of their original codes.

Straight from the definitions, it is easy to see that the upper bounds on the size of one-level codes are also relevant to two-level codes. In fact, rather than using our construction, the following example is better for some special cases. The construction here is simple and provides two-level codes that meet the lower bounds of one-level codes.
Example 1. Let $\ell$ be a positive integer greater than 2, and let $Q$ a finite field of cardinality at least $\ell$. Let $\alpha_1, \alpha_2, \ldots, \alpha_\ell$ be distinct elements of $Q$. Consider the following $t$-FP code $C = \{ (f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_\ell) : f \in Q[X] \text{ and } \deg f < \lceil \ell/t \rceil \}$ (see [2]). Then, for any integer $T > t$, there exists a $(T, t)$-FP code $C'$ of the same cardinality as $C$ with $q$ groups, where $q = |Q|$.  

Proof. Partition $C$ into $q$ groups, $C_1, C_2, \ldots, C_q$, by letting $C_i = \{ x \in C : x_1 = i \}$. Since $|C| = q^{\lceil \ell/t \rceil}$ and $C' = \bigcup_{i=1}^{q} C_i$, now we have a code $C' = C$ containing $q$ groups of cardinality $q^{\lceil \ell/t \rceil - 1}$. It is easy to see that none of the codewords in $C'$ can frame the codeword outside its own group, since any pair of codewords from the different groups have different symbols in the first co-ordinate. Together with $t$-FP property that $C'$ inherits directly from $C$, $C'$ is a $(T, t)$-FP code of the same cardinality as $C$. \hfill $\square$

Example 1 has shown that there exist $(T, t)$-FP codes as big as $(T, t)$-FP codes in some cases. In fact, for any $t$-IPP, $t$-SFP or $t$-FP code $C$, if there exists an $i \in \{1, 2, \ldots, \ell\}$ that only $g \leq q$ symbols from $Q$ are used in the $i$th co-ordinate and they are uniformly distributed, we can construct two-level codes by partitioning its codewords into $g$ groups according to the $i$th co-ordinate, then obtain a two-level code $C'$ with the same cardinality as $C$. We exclude full proof from this paper: see [8] for details.

There are one-level codes that the distribution of alphabet symbols in any co-ordinate is non-uniform (and there are examples that are the largest known for some parameters) In this case, the above simple two-level construction cannot be used. The construction we propose in the next subsection is general enough to work in these cases, though at a cost of reducing the size of the code by a factor of up to 2.

3.1 Main Theorem

In this section, we aim to construct two-level fingerprinting codes from existing one-level fingerprinting codes. Our construction works provided the number of groups is at most the alphabet size. Our construction begins with a one-level code, removes some codewords, groups and modifies the remaining codewords. The results are two-level codes with are guaranteed to be at least half the size of their original codes.

The next theorem is the core of our construction.
Theorem 2. Let $q, g$ and $\ell$ be integers greater than 1, where $g \leq q$. Let $C$ be an $q$-ary length $\ell$ code. Then there exists a $q$-ary length $\ell$ code $C'$ of cardinality at least $\frac{|C|}{2}$, where $C'$ possesses the following properties;

1. there exists an injection from $C'$ to $C$ with changes occur only in the first co-ordinate of the codewords,
2. $C'$ can be partitioned into $g$ groups of the same size, each at least $\left\lceil \frac{|C|}{2g} \right\rceil$,
3. the first co-ordinate of codewords in each group of $C'$ are distinct from those of any other group.

The explicit construction of two-level codes is embedded in the proof of Thorem 2. Before showing the detailed proof, we provide the following example to give a rough idea about how to construct $C'$ from a given code $C$ satisfying Theorem 2.

Example 2. Let $n = 91$, $q = 11$, $g = 9$ and $p = \left\lceil \frac{n}{2g} \right\rceil = 6$. Let $g_1, g_2, \ldots, g_q$ be number of codewords of $C$ beginning with each symbol. Suppose we have:

\[
\begin{align*}
g_1 &= 4 & g_2 &= 5 & g_3 &= 10 & g_4 &= 11 \\
g_5 &= 17 & g_6 &= 5 & g_7 &= 2 & g_8 &= 4 \\
g_9 &= 18 & g_{10} &= 10 & g_{11} &= 5.
\end{align*}
\]

Our aim is to form new 9 groups of size 6, where the first co-ordinate of codewords in each new group are distinct from any other groups. Our method can be divided in to 3 main steps: splitting, amalgamating and replacing. We illustrate the construction of $C'$ in Figure 1.

**Step 1: Splitting**

The purpose of this step is to split all big groups, which contain $p$ codewords or more, into one or more smaller groups of size at least $p$. Observe that $g_3, g_4$ and $g_{10}$ provide 1 group each, while $g_5$ and $g_9$ give 2 and 3 groups, respectively. We now have 8 groups with only 5 different first co-ordinates. Remove any 3 other groups to obtain 3 more first co-ordinates, let them be $g_2, g_6$ and $g_{11}$. Hence, 2, 6 and 11 have become unused first co-ordinates.

**Step 2: Amalgamating**

In this step, we aim to create more groups, of size at least $p$, by merging at least 2 smaller groups together. So, we merge the remaining groups together to obtain the last group of size $g_1 + g_7 + g_8 = 10$.

**Step 3: Replacing**

Since some first co-ordinates of the groups we have constructed are repeated, we replace them with those unused first co-ordinates in this step. Then,
reduce the number of codewords in each group to \( p \). Here the repeated first co-ordinates are 5 and 9, replace the first co-ordinate of codewords in those “repeated” groups (1 group from \( g_5 \) and 2 groups from \( g_9 \)) by the unused first co-ordinates; 2, 6, 11.

The result after completing these 3 steps and removing any extra elements is 9 disjoint groups of size 6, where the first co-ordinate of codewords in each group are distinct from any other groups. The following table illustrates what we did earlier. We abuse the notation and use \( k_i \) for the \( i \)th codeword in group \( t \), i.e. the \( i \)th codeword beginning with symbol \( t \).

Now, we prove Theorem 2 using the similar approach as in Example 2.

**Proof of Theorem 2.** Let \( q \) and \( \ell \) be integers greater than 1. Let \( C \) be a one-level code length \( \ell \) over an alphabet \( Q \), where \( |Q| = q \). Let \( g \) be a positive integer less than or equal to \( q \). Denote \( \lceil \frac{|C|}{2g} \rceil \) by \( p \). For each symbol \( a \in Q \), let \( G_a = \{ x \in C : x_1 = a \} \) and denote the size of \( G_a \) by \( g_a \). And let \( g_a = \alpha_a p + \beta_a \), where \( \alpha_a, \beta_a \) are integers and \( 0 \leq \beta_a < p \). Let \( Q_1 \) be \( \{ a \in Q : \alpha_a > 0 \} \), \( q_1 = |Q_1| \) and \( v = \sum_{a \in Q} \alpha_a = \sum_{a \in Q_1} \alpha_a \). We are now trying to construct \( g \) groups of \( p \) codewords which the first co-ordinate of each group is different from the others. As in Example 2 we proceed through the three main steps: splitting, amalgamating and replacing.

**Step 1: Splitting**

Now, for each \( a \) in \( Q_1 \), we pick \( \alpha_a p \) codewords from each \( G_a \), then divide these codewords into \( \alpha_a \) sets of \( p \) codewords. At this stage, we obtain \( v \) disjoint sets of \( p \) codewords, namely \( C_1, C_2, ..., C_v \), with the property that all the codewords within the same set have the same symbol in the first co-ordinate. However, some of the symbols are still being used by more than one group.

If \( v \geq g \) we are done: for \( i \in \{ 1, 2, ..., g \} \), we replace the first co-ordinate of the codewords in \( C_i \) by symbol \( i \) to form subset \( C'_i \), and define \( C' = \bigcup_{i=1}^{g} C'_i \).

So without loss of generality, we assume \( v < g \). To construct the first \( v \) groups from \( C_1, C_2, ..., C_v \), we need \( v \) different symbols to replace the first co-ordinate of each group. Besides the \( q_1 \) symbols in \( Q_1 \), we need \( v - q_1 \) extra symbols. Let \( Q_2 \) be any subset of \( Q \) of cardinality \( v \) containing \( Q_1 \). Throw away all codewords in \( G_a \) where \( a \in Q_1 \setminus Q_2 \).

**Step 2: Amalgamating**

We merge some of the remaining \( G_a \), where \( a \in Q \setminus Q_1 \), into \( g - v \) groups of size between \( p \) and \( 2p - 2 \), where the first co-ordinates of each group are different from the other groups. This can be done as follows.
| Original No. | No. of group obtained | New No. | No. of removed codewords |
|-------------|-----------------------|---------|--------------------------|
| $g_3 = 10$  | 1 group               | $g'_3 = 6$ | $3_1, ..., 3_{10}$ |
| $4_{11}, ..., 4_{12}$ |               | $g'_4 = 6$ | $4_1, ..., 4_6$ |
| $5_{13}, ..., 5_{17}$ |               | $g'_5, g'_6 = 6$ | $5_1, ..., 5_6$ |
| $9_{13}, ..., 9_{18}$ |               | $g'_9, g'_{10}, g'_{11} = 6$ | $9_1, ..., 9_6$ |
| $g_{10} = 10$ | 1 group               | $g'_{10} = 6$ | $10_{11}, ..., 10_6$ |
| $g_2 = 5$  | Removed               |         | $2_{11}, ..., 2_5$ |
| $g_6 = 5$  | Removed               |         | $6_{11}, ..., 6_5$ |
| $g_{11} = 5$ | Removed             |         | $11_{11}, ..., 11_5$ |
| $g_4 = 4$  | 1 Group after amalgamated $g_1, g_7$ and $g_8$ | $g'_{1,7,8} = 6$ | $1_{1}, 1_{2}, 1_{3}, 1_{4}, 7_1, 8_1, 8_2$ |
| $g_7 = 2$  |                       |         | $1_{4}, 7_2, 8_3, 8_4$ |
| $g_8 = 4$  |                       |         |                           |

| Total Removed Remaining codewords | 38 | 53 |

Figure 1: Groups Dividing in Example 2

10
Consider
\[ \sum_{a \in Q \setminus Q_2} g_a = |C| - \sum_{a \in Q_2} g_a \]
\[ = |C| - \sum_{a \in Q_2} (\alpha_a p + \beta_a) \]
\[ = |C| - \left( \sum_{a \in Q_1} \alpha_a p + \sum_{a \in Q_2} \beta_a \right) \]
\[ \geq |C| - (vp + v(p - 1)) \]
\[ \geq 2gp - vp - v(p - 1) \]
\[ = 2(g - v)p + v. \]

Hence, apart from \( \cup_{a \in Q_2} G_a \), we have at least \( 2(g - v)p + v \) codewords left in the system. And, since \( g_a = \alpha_a + \beta_a = 0 + \beta_a = \beta_a \leq p - 1 \) for all \( a \in Q \setminus Q_2 \), we can group \( g_a, a \in Q \setminus Q_2 \) in a greedy fashion into \( g - v \) sets of size between \( p \) and \( 2p - 2 \), where each \( a \in Q \setminus Q_2 \) is allowed to merge into at most one set. Name the sets \( C_{v+1}, C_{v+2}, \ldots, C_g \). Observe that the first co-ordinate of codewords in each set varies, but differs from any other groups.

**Step 3: Replacing**
Here we construct the groups of \( C' \) as follows: Let \( Q_2 = \{a_1, a_2, \ldots, a_g\} \).

1. for \( i = 1 \) to \( v \), let \( C'_i \) be a set of codewords obtained from \( C_i \) by replacing the first co-ordinate by the symbol \( a_i \in Q_2 \),
2. for \( i = v + 1 \) to \( g \), let \( C'_i \) be a set of any \( p \) codewords from \( C_i \),
3. let \( C' = \bigcup_{i=1}^{g} C'_i \).

Now, we need to show that our constructed code \( C' \) satisfies Theorem 2.
Let the mapping \( \varphi : C' \rightarrow C \) map each codeword of \( C' \) to its corresponding codeword in \( C \). It is not difficult to see that \( \varphi \) is an injection that makes changes in only the first co-ordinate of any codeword.

Hence, we are now obtain a code \( C' \) contains \( g \) groups of the same size, each at least \( \lceil \frac{|C|}{2g} \rceil \) with the property that the first co-ordinate of codewords in each group are distinct from any other groups, and an injection \( \varphi \) from \( C' \) to \( C \) with changes occur only in the first co-ordinate of the codewords.

Note that to construct \( C' \), we have eliminated \( \left( \sum_{a \in Q} \beta_a \right) - (g - v)p \) codewords from \( C \). Here the first term represents all the remainders, and
the second term is derived form \((g - v)\) amalgamated groups that were taken back from the thrown away remainders. Now \(\left(\sum_{a \in Q} \beta_a \right) - (g - v)p = (|C| - vp) - (g - v)p = |C| - gp \leq |C| - \frac{|C|}{2} = \frac{|C|}{2}\). Hence we eliminate at most \(\frac{|C|}{2}\) codewords from \(|C|\).

\[\square\]

For any one-level FP, SFP and IPP code, the two-level code satisfying Theorem 2 has the corresponding two-level fingerprinting property. To make it more convenient for us to show this in the next subsection, we define some mappings and a lemma we need here.

Let the mapping \(\pi : Q \rightarrow Q\) be defined as follows. Let \(\pi(a) = a\) when \(a\) does not appear as the first co-ordinate of any codeword of \(C'\); otherwise let \(\pi(a) = b \in Q\) when there exists a codeword \(c' \in C'\) with \(c'_1 = a\) that was derived from \(c \in C\) with \(c_1 = b\).

Let the mapping \(\psi : Q^\ell \rightarrow Q^\ell\) be defined by mapping \(x \in Q^\ell\) to \(\psi(x) \in Q^\ell\) where

\[
\psi(x)_i = \begin{cases} 
\pi(x_i) & \text{if } i = 1; \\
x_i & \text{otherwise.}
\end{cases}
\]

It is not difficult to see that \(\psi\) is a well-defined function and \(\varphi\) from Theorem 2 is actually \(\psi\) when restricted to \(C'\), i.e. \(\varphi = \psi|_{C'}\).

Observe that for any \(i \in \{1, 2, ..., \ell\}\) and any codewords \(y, z \in C'\), if \(y_i = z_i\), then \(\varphi(y)_i = \varphi(z)_i\). Moreover, \(\varphi(y)_i = y_i = z_i = \varphi(z)_i\) when \(i \neq 1\).

**Lemma 3.** Let \(C\) and \(C'\) be codes length \(\ell\) over \(Q\) satisfying Theorem 2. Let \(X\) be a subset of \(C'\). Then

\(\psi(\text{desc}(X)) \subseteq \text{desc}(\psi(X))\).

**Proof.** Let \(X\) be a subset of \(C'\) and let \(y\) be a codeword in \(\psi(\text{desc}(X))\). Then, there exists a codeword \(x\) in \(\text{desc}(X)\) such that \(\psi(x) = y\). For any component \(i\) in \(\{1, 2, ..., \ell\}\), there exists a codeword \(x^i\) in \(X\), where \(x_i = x^i_i\). Hence \(\psi(x)_i = \psi(x^i)_i\) for all \(i \in \{1, 2, ..., \ell\}\). Which implies \(\psi(x) \in \text{desc}(\{\psi(x^1), \psi(x^2), ..., \psi(x^\ell)\}) \subseteq \text{desc}(\psi(X))\). Therefore \(y \in \text{desc}(\psi(X))\), which implies \(\psi(\text{desc}(X)) \subseteq \text{desc}(\psi(X))\). \(\square\)

### 3.2 Existence of Codes

Here we demonstrate that the codes \(C'\) satisfying Theorem 2 are two-level FP, SFP or IPP code if the original codes \(C\) are respecting FP, SFP or
IPP codes. Also, we provide an example showing that the two-level code constructed from a TA code using Theorem 2 does not always possess two-level TA property.

**Theorem 4.** Let \( t, q, g \) and \( \ell \) be integers greater than 1, where \( g \leq q \). And let \( T \) be any integer greater than \( t \). Suppose there exists a \( q \)-ary length \( \ell \) one-level \( t \)-FP code \( C \). Then there exists a \( q \)-ary length \( \ell \) two-level \(( T, t )\)-FP code \( C' \) of cardinality at least \( \left\lceil \frac{|C|}{2} \right\rceil \), where \( C' \) contains \( g \) groups of the same size.

**Proof.** Let \( C' \) be a code obtained from the \( t \)-FP code \( C \) as in Theorem 2. It is easy to see that none of the codewords in \( C' \) can frame the codeword outside its own group, since any pair of codewords from the different groups have different symbols in the first co-ordinate. So, only \( t \)-FP property of \( C' \) needs to be proved.

Let \( U \) be any subset of \( C' \) containing at most \( t \) codewords. Let \( x \in \text{desc}(U) \cap C' \). We will show that \( x \in U \). Since \( x \in \text{desc}(U) \cap C' \), we have \( \varphi(U) \subseteq C \) and \( |\varphi(U)| \leq |U| \leq t \).

Since \( x \in \text{desc}(U) \cap C' \), then \( x \in \text{desc}(U) \) and \( x \in C' \). By Lemma 3, \( \varphi(x) \in \text{desc}(\varphi(X)) \). Also it is easy to see that \( \varphi(C') \subseteq C \), hence \( \varphi(x) \in \text{desc}(\varphi(X)) \cap C \). This makes \( \varphi(x) \in \varphi(U) \) by the \( t \)-FP property of \( C \). Hence \( x \in U \), which implies \( C' \) has \( t \)-FP property, i.e. \( C' \) is a \(( T, t )\)-FP code.

Then we can conclude that there exists a \( q \)-ary length \( \ell \) two-level \(( T, t )\)-FP code \( C' \) of size at least \( \left\lceil \frac{|C|}{2} \right\rceil \), containing \( g \) groups (each of size at least \( \left\lceil \frac{|C|}{2g} \right\rceil \)).

**Theorem 5.** Let \( t, q, g \) and \( \ell \) be integers greater than 1, where \( g \leq q \). And let \( T \) be any integer greater than \( t \). Suppose there exists a \( q \)-ary length \( \ell \) one-level \( t \)-SFP code \( C \). Then there exists a \( q \)-ary length \( \ell \) two-level \(( T, t )\)-SFP code \( C' \) of cardinality at least \( \left\lceil \frac{|C|}{2g} \right\rceil \), where \( C' \) contains \( g \) groups of the same size.

**Proof.** Let \( C' \) be a code obtained from the \( t \)-SFP code \( C \) as in Theorem 2.

1. Let \( X_1, X_2 \) be subsets of \( C \) of size at most \( t \), where \( \text{desc}(X_1) \cap \text{desc}(X_2) \neq \emptyset \). We will show that \( X_1 \cap X_2 \neq \emptyset \).

Let \( x \in \text{desc}(X_0) \cap \text{desc}(X_1) \). Then \( \varphi(x) = \psi(x) \in \psi(\text{desc}(X_0) \cap \text{desc}(X_1)) \). Now

\[
\varphi(x) \in \psi(\text{desc}(X_1) \cap \text{desc}(X_2)) \\
\subseteq \psi(\text{desc}(X_1)) \cap \psi(\text{desc}(X_2)) \\
\subseteq \text{desc}(\psi(X_1)) \cap \text{desc}(\psi(X_2)) \text{ by Lemma 3} \\
= \text{desc}(\varphi(X_1)) \cap \text{desc}(\varphi(X_2)).
\]
Therefore desc(\(\varphi(X_1)\)) \(\cap\) desc(\(\varphi(X_2)\)) \(\neq\) \(\emptyset\). By the \(t\)-SFP property of \(C\), we deduce that \(\varphi(X_1) \cap \varphi(X_2) \neq \emptyset\). Since \(\varphi\) is an injection, \(\varphi(X_1) \cap \varphi(X_2) = \varphi(X_1 \cap X_2)\). Therefore \(X_1 \cap X_2 \neq \emptyset\), which implies \(C'\) has the \(t\)-SFP property.

2. Let \(Y_1, Y_2\) be subsets of \(C\) of size at most \(T\), where desc(\(Y_1\)) \(\cap\) desc(\(Y_2\)) \(\neq\) \(\emptyset\). We will show that \(G(Y_0) \cap G(Y_1) \neq \emptyset\).

Let \(x \in\) desc(\(Y_1\)) \(\cap\) desc(\(Y_2\)). Then there exist codewords \(a\) in \(Y_1\) and \(b\) in \(Y_2\), where \(a_1 = x_1 = b_1\). Since the first co-ordinate of each group is different from the others, we can conclude that \(G(a) = G(b)\). Therefore \(G(a) \in G(Y_1) \cap G(Y_2) \neq \emptyset\), so \(C'\) is a \((T, t)\)-SFP code.

\(\square\)

**Theorem 6.** Let \(t, q, g\) and \(\ell\) be integers greater than 1, where \(g \leq q\). And let \(T\) be any integer greater than \(t\). Suppose there exists a \(q\)-ary length \(\ell\) one-level \(t\)-IPP code \(C\). Then there exists a \(q\)-ary length \(\ell\) two-level \((T, t)\)-IPP code \(C'\) of cardinality at least \(|C|^2\), where \(C'\) contains \(g\) groups of the same size.

**Proof.** Let \(C'\) be a code obtained from the \(t\)-IPP code \(C\) as in Theorem 2.

1. Let \(x \in\) desc\(_t\)(\(C'\)). Then, there exists \(U \subseteq C'\) such that \(|U| \leq t\) and \(x \in\) desc\(_t\)(\(U\)). By Lemma 3, we know that \(\psi(x) \in\) desc\(_t\)(\(\varphi(U)\)) and \(\varphi(U) \subseteq C\). Observe that \(|\varphi(U)| \leq |U| \leq t\). Hence \(\psi(x) \in\) desc\(_t\)(\(C\)). Since \(C\) is an IPP code,

\[
\bigcap_{X \in C: |X| \leq t \atop \psi(x) \in\text{desc}(X)} X \neq \emptyset.
\]

Also, for any \(X \subseteq C'\), \(x \in\) desc\(_t\)(\(X\)) implies \(\psi(x) \in\) desc\(_t\)(\(\varphi(X)\)) and \(|X| = |\varphi(X)|\). Hence

\[
\bigcap_{X \in C: |X| \leq t \atop \psi(x) \in\text{desc}(X)} X \subseteq \bigcap_{X \in C': |X| \leq t \atop x \in\text{desc}(X)} \varphi(X).
\]

Since \(\varphi\) is injective, we have

\[
\bigcap_{X \in C': |X| \leq t \atop x \in\text{desc}(X)} \varphi(X) = \varphi(\bigcap_{X \in C: |X| \leq t \atop x \in\text{desc}(X)} X).
\]
Hence
\[ \varphi(\bigcap_{\substack{X \subseteq C' : |X| \leq t \\text{and} \\exists x \in \text{desc}(X) \\text{where} \ x \in X}} X) \neq \emptyset. \]

Therefore
\[ \bigcap_{\substack{X \subseteq C' : |X| \leq t \\text{and} \\exists x \in \text{desc}(X) \\text{where} \ x \in X}} X \neq \emptyset, \]
which shows that \( C' \) is a \( t \)-IPP code.

2. Let \( y \in \text{desc}_T(C') \). Then, there exists \( V \subseteq C' \) such that \( |V| \leq T \) and \( y \in \text{desc}(V) \). Let \( v \) be a codeword in \( V \) such that \( v_1 = y_1 \), and let \( i \in \{1, 2, ..., g\} \) such that \( v \in C'_i \). Hence \( G(v) = i \). For any \( X \subseteq C' \) of cardinality at most \( T \) with \( \text{desc}(X) \) containing \( y \), there exists a codeword \( y^X \) such that \( y_1 = y^X_1 \). Since the group index of a codeword can be determined from its first co-ordinate, we have \( G(y^X) = i \). That implies
\[ i \in \bigcap_{\substack{X \subseteq C' : |X| \leq T \\text{and} \\exists y \in \text{desc}(X) \\text{where} \ y \in X}} G(X). \]
Consequently,
\[ \bigcap_{\substack{X \subseteq C' : |X| \leq T \\text{and} \\exists y \in \text{desc}(X) \\text{where} \ y \in X}} G(X) \neq \emptyset. \]

Therefore, \( C' \) is a \((T, t)\)-IPP code. \( \square \)

The two-level codes satisfying Theorem 2 preserve the fingerprinting property from their corresponding one-level codes for IPP, SFP and FP codes. However, this is not always true in the case of TA codes as can be seen in the following example.

**Example 3.** Let \( C = \{011, 022, 033, 044, 105, 206, 307, 408, 550, 660, 770, 880\} \subseteq \{0, 1, 2, ..., 8\}^3 \). It is not difficult to check that \( C \) is a 2-TA code. Let \( g = 4 \), then \( p = \left\lceil \frac{12}{8} \right\rceil = 2 \). Then Theorem 2 does not guarantee two-level traceability code from \( C \).
Proof. Here we have $g_1 = 4$, $g_2 = g_3 = \ldots = g_8 = 1$. Consider $C_1 = \{011, 022\}$, $C_2 = \{033, 044\}$, $C_3 = \{105, 550\}$ and $C_4 = \{206, 660\}$, which leads to $C_1' = \{011, 022\}$, $C_2' = \{833, 844\}$, $C_3' = \{105, 550\}$ and $C_4' = \{206, 660\}$ by Theorem 2. Let $U = \{011, 105, 550\}$, then $000 \in \text{desc}(U)$ and $G(U) = \{1, 3\}$. Observe that 206 is a codeword of $C'$ with $d_H(000, 206)$ minimal, but $G(206) = 4 \notin G(U)$. Therefore $C'$ is not a $(K, 2)$-TA code for any integer $T$ greater than 2.

4 Conclusion and Open problems

Theorem 2 ensures that we can always construct two-level IPP, SFP and FP codes, with $g \leq q$, of size at least half of the size of one-level codes. When one-level code of exponential size, throwing away half of the codewords would not effect the codes’ size significantly. However, we do not have the same result for TA codes. Hence the following question comes up naturally.

**Question 1:** Let $g \leq q$, and let $C$ be a $q$-ary $t$-TA code of length $\ell$, does there always exist a $q$-ary $(T, t)$-TA code $C'$ of length $\ell$ of cardinality at least a half of the original code $C$, containing $g$ groups?

We believe the answer to this question is yes.

The results in this paper require the number of groups to be small.

**Question 2:** Are there any good constructions of two-level fingerprinting codes when the number of groups is greater than the alphabet size?

And the next question follows.

**Question 3:** Are there any good upper bounds on the size of two-level codes that are significantly better than the one-level case?

We believe that, in general cases, the bounds will be significantly better than the one-level case and will begin to depend on $T$ when $g$ grows bigger.

References

[1] N. P. Anthapadmanabhan and A. Barg. Two-level fingerprinting codes. In *Proceedings of the 2009 IEEE international conference on Sympo-*
System on Information Theory, volume 4 of ISIT’09, pages 2261–2265, Piscataway, NJ, USA, 2009. IEEE Press.

[2] S. R. Blackburn. Frameproof codes. *SIAM Journal on Discrete Mathematics*, 16(3):499–510, 2003.

[3] S. R. Blackburn. Combinatorial schemes for protecting digital content. *Surveys in combinatorics 2003 (ed. C.D. Wensley)*, 307:43–78, Cambridge University Press, Cambridge, 2003.

[4] S. R. Blackburn, T. Etzion, and S Ng. Traceability codes. *Journal of Combinatorial Theory Series A*, 117:1049–1057, 2010.

[5] D. Boneh and J. Shaw. Collusion-secure fingerprinting for digital data. *IEEE Transactions on Information Theory*, 47:1897–1905, 1998.

[6] B. Chor, A. Fiat, and M. Naor. Tracing traitors. In *CRYPTO ’94: Proceedings of the 14th Annual International Cryptology Conference on Advances in Cryptology (ed. Y. G. Desmedt)*, pages 257–270, London, UK, 1994. Springer-Verlag.

[7] B. Chor, A. Fiat, M. Naor, and B. Pinkas. Tracing traitors. *IEEE Transactions on Information Theory*, 46:893–910, 1994.

[8] P. Rochanakul. PhD Thesis, in preparation. Royal Holloway, University of London, London, UK.

[9] J. N. Staddon, D. R. Stinson, and R. Wei. Combinatorial properties of frameproof and traceability codes. *IEEE Transactions on Information Theory*, 47:1042–1049, 2000.

[10] D. R. Stinson, T. van Trung, and R. Wei. Secure frameproof codes, key distribution patterns, group testing algorithms and related structures. *Journal of Statistical Planning and Inference*, 86:595–617, 1997.

[11] D. R. Stinson and R. Wei. Combinatorial properties and constructions of traceability schemes and frameproof codes. *SIAM Journal on Discrete Mathematics*, 11:41–53, 1998.