On a Class of Traceability 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

Traceability codes are designed to be used in schemes that protect copyrighted digital data against piracy. The main aim of this paper is to give an answer to a Staddon-Stinson-Wei’s problem of the existence of traceability codes with $q < w^2$ and $b > q$. We provide a large class of these codes constructed by using a new general construction method for $q$-ary codes.

1 Introduction

Traceability (TA) codes are designed to be used in schemes that protect copyrighted digital data against piracy. An example of such an application in pay-per-view movies is described in Fiat and Tassa [8]. Different notions of “traceability” have been studied by several researchers in recent years, e.g., [3], [4], [5], [8], [9], [10], [11], [12], [13].

In this paper, notation and definitions of traceability codes are adapted from Staddon, Stinson and Wei’s paper [13].

A code $C$ of length $n$ with $b$ codewords and minimum distance $d$ over an alphabet $Q$ with $|Q| = q$ is called an $(n, b, q; d)$-code. If $d$ is not needed, we call $C$ an $(n, b, q)$-code. A 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 $\text{desc}(C_0)$, is defined by

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

For any $x, y \in Q^n$, define $I(x, y) = \{i : x_i = y_i\}$.

Definition 1.1 Suppose $C$ is an $(n, b, q)$-code and $w \geq 2$ is an integer. $C$ is called a $w$-TA code provided that, for all subsets $C_i \subseteq C$ of size at most $w$ and all $x \in \text{desc}(C_i)$, there is at least one codeword $y \in C_i$ such that $|I(x, y)| > |I(x, z)|$ for any $z \in C \setminus C_i$.

The following result stated in [4], [5], [13] is useful. We present it here with a simple proof.

Theorem 1.1 Any $(n, b, q; d)$ code with $d > n(1 - 1/w^2)$ is an $(n, b, q)$ $w$-TA code.

Proof. Let $C$ be an $(n, b, q; d)$ code with $d > n(1 - 1/w^2)$. Set $\alpha = n(1 - 1/w^2)$. Any two codewords $c_1, c_2 \in C$ agree in at most $\beta = n - (\alpha + 1) = n/w^2 - 1$ positions. Let $C' = \{c'_1, \ldots, c'_v\} \subseteq C$ be a subset of size $v$. For any $u \in \text{desc}(C')$, define $M(u) = \max\{|I(u, c'_i)| : i = 1, \ldots, v\}$. Then $n/v \leq M$. On the
other hand, for any \( c \in C \setminus C' \) we have \( \sum_{c' \in C'} |I(c, c')| \leq v \beta \). Now \( C \) will be a \( v \)-TA code if \( v \beta < n / v \). Thus \( \beta < n / v^2 \), equivalently \( n / v^2 - 1 < n / v^2 \). Hence \( v \leq w \), as desired. \( \square \)

In [13], it is shown that if there exists an \((n, b, q)\) \( w \)-TA code, then \( w < q \). The following theorem [13] is obtained by applying Theorem 1.1 to \( q \)-ary Reed-Solomon codes.

**Theorem 1.2 (Staddon, Stinson and Wei)** Suppose \( n, q \) and \( w \) are given, with \( q \) a prime power and \( n \leq q + 1 \). Then there exists an \((n, b, q)\) \( w \)-TA code in which \( b = q^{\lfloor n / w^2 \rfloor} \).

In Theorem 1.2, if \( q < w^2 \), then \( b = q \). Thus, as an open problem Staddon, Stinson, and Wei [13], ask the following question: Can we construct \( w \)-TA codes with \( q < w^2 \) and \( b > q \)?

Our aim is to give an answer to the Staddon-Stinson-Wei’s problem. Precisely, we present a general construction method for \( q \)-ary codes with large Hamming distance. Using this method we are able to construct a large class of \( w \)-TA codes with \( q < w^2 \) and \( b > q \), and thus obtain a positive answer to the problem.

## 2 A Construction of \((n, b, q; d)\) codes

We depict an \((n, b, q; d)\)-code \( C \) as an \( b \times n \) array \( A(C) \) on \( q \) symbols, where each row of the array corresponds to one of the codewords of \( C \). For any \( a \in Q \), define

\[
m_j(a) = |\{ i : A(C)(i, j) = a \}|.
\]

i.e. \( m_j(a) \) is the frequency of \( a \) on the \( j^{th} \) column of \( A(C) \). Define

\[
m(C) = \max_{1 \leq j \leq n, a \in Q} (m_j(a)).
\]

**Definition 2.1** Let \( C \) be an \((n, b, q; d)\) code. We say that \( C \) has an \( \sigma \)-resolution if the codewords of \( C \) can be partitioned into \( s \) subsets \( A_1, \ldots, A_s \), where \(|A_i| = \sigma\), for \( i = 1, \ldots, s \), in such a way that each \( A_i \) is a code of minimum distance equal to \( n \), i.e. any two codewords of \( A_i \) agree in no position.

**CONSTRUCTION**

Let \( C_1 \) be an \((n_1, b_1, q_1; d_1)\) code over an alphabet \( Q_1 \). Let \( C_2 \) be an \((n_2, b_2, q_2; d_2)\) code with a \( \sigma \)-resolution \( A_1, \ldots, A_s \). Suppose \( s \geq m(C_1) \). For each \( a \in Q_1 \) denote by \( C_2(a) \) a copy of \( C_2 \) defined over an alphabet \( Q(a) \) such that \( Q(a_1) \cap Q(a_2) = \emptyset \) if \( a_1 \neq a_2 \). Denote by \( A_1(a) \), \( \ldots, A_s(a) \) a \( \sigma \)-resolution of \( C_2(a) \).

Let \( \text{col}_j(a_1, a_2, \ldots, a_{n_1})^T \) be the \( j^{th} \) column of \( A(C_1) \), \( 1 \leq j \leq n_1 \). Let \( a(1), \ldots, a(t) \), say, be \( t \) positions of \( \text{col}_j \) at which symbol \( a \in Q_1 \) appears. Note that \( t \leq m(C_1) \). Now replace \( a \) at position \( a(1) \) by \( A_1(a) \), \( a \) at position \( a(2) \) by \( A_2(a) \), etc., and \( a \) at position \( a(t) \) by \( A_t(a) \). Perform this process for every symbol of \( Q_1 \) and for every column of \( A(C_1) \). The resulting code \( C \) obtained by this replacement has parameters \((n_1 n_2, \sigma b_1, q_1 q_2; n_1 n_2 - (n_1 - d_1)(n_2 - d_2))\).

Obviously, the length and the number of codewords of \( C \) is \( n_1 n_2 \) and \( \sigma b_1 \) respectively. Further, any two codewords \( c_1, c_2 \in C \) agree in at most \((n_1 - d_1)\) positions. After replacement \( c_1 \) and \( c_2 \) correspond to two subsets \( R_1 \) and \( R_2 \) of \( \sigma \) codewords each. Any two
codewords in \( R_1 \) (resp. \( R_2 \)) agree in no position, whereas a codeword from \( R_1 \) and a codeword from \( R_2 \) agree in at most \((n_1 - d_1)(n_2 - d_2)\) positions. Hence the minimum distance of \( C \) is \( n_1 n_2 - (n_1 - d_1)(n_2 - d_2) \), as stated.

Further, if \( q_1 q_2 \geq b_1 \) then \( C \) can be extended to a code \( C^* \) having parameters \((n_1 n_2 + 1, \sigma b_1, q_1 q_2; d)\), where \( d = \min\{n_1 n_2, n_1 n_2 + 1 - (n_1 - d_1)(n_2 - d_2)\}\). Let \( Q = \{a_1, a_2, \ldots, a_{q_1 q_2}\} \) be the alphabet of \( C \) and let \( C_1 = \{c_1, c_2, \ldots, c_{b_1}\} \). By construction, any codeword \( c_i \in C_1 \) corresponds to a subset \( R_i \) of \( \sigma \) codewords. For any \( i = 1, \ldots, b_1 \), we add symbol \( a_i \) to the \((n_1 n_2 + 1)^{th}\) column of each codeword of \( R_i \). This forms a set \( R_i^* \). The collection of all \( R_i^* \) forms an \((n_1 n_2 + 1, \sigma b_1, q_1 q_2; d)\) code \( C^* \) with \( d = \min\{n_1 n_2, n_1 n_2 + 1 - (n_1 - d_1)(n_2 - d_2)\}\). This can be seen as follows. Any two codewords \( x^* \) and \( y^* \) of \( C^* \) belong either to some \( R_i^* \) or to two different \( R_i^* \) and \( R_j^* \). In the first case their distance is \( n_1 n_2 \) because their components agree only at the \((n_1 n_2 + 1)^{th}\) column, and in the second case their distance is at least \( n_1 n_2 + 1 - (n_1 - d_1)(n_2 - d_2) \) because their components at the \((n_1 n_2 + 1)^{th}\) column are distinct.

We record the result of the construction in the following theorem.

**Theorem 2.1** Suppose there is an \((n_1, b_1, q_1; d_1)\) code \( C_1 \) and there is an \((n_2, b_2, q_2; d_2)\) code \( C_2 \) with a \( \sigma \)-resolution \( \{A_1, \ldots, A_{b_2}\} \) such that \( s \geq m(C_1) \). Then the following hold.

(i) There is an \((n_1 n_2, \sigma b_1, q_1 q_2; n_1 n_2 - (n_1 - d_1)(n_2 - d_2))\) code \( C \).

(ii) Further, if \( q_1 q_2 \geq b_1 \), then \( C \) can be extended to a code \( C^* \) having parameters \((n_1 n_2 + 1, \sigma b_1, q_1 q_2; d)\), where \( d = \min\{n_1 n_2, n_1 n_2 + 1 - (n_1 - d_1)(n_2 - d_2)\}\).

We illustrate the construction in Theorem 2.1 by the following example.

**Example 2.1** Let \( C_1 \) be a \((3, 4, 2; 2)\) code over the alphabet \( Q_1 = \{0, 1\} \) given by

\[
C_1 = \begin{pmatrix}
0 & 0 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 0
\end{pmatrix}
\]

Let \( C_2(0) \) be a \((3, 6, 3; 2)\) code on the alphabet \( \{1, 2, 3\} \) having a 3-resolution \( A_1(0) \) and \( A_2(0) \):

\[
A_1(0) = \begin{pmatrix}
1 & 2 & 3 \\
2 & 3 & 1 \\
3 & 1 & 2
\end{pmatrix}, \quad A_2(0) = \begin{pmatrix}
1 & 3 & 2 \\
2 & 1 & 3 \\
3 & 2 & 1
\end{pmatrix}
\]

Let \( C_2(1) \) be a copy of \( C_2(0) \) on the alphabet \( \{4, 5, 6\} \) with the corresponding 3-resolution

\[
A_1(1) = \begin{pmatrix}
4 & 5 & 6 \\
5 & 6 & 4 \\
6 & 4 & 5
\end{pmatrix}, \quad A_2(1) = \begin{pmatrix}
4 & 6 & 5 \\
5 & 4 & 6 \\
6 & 5 & 4
\end{pmatrix}
\]

Replacing entries of \( A(C_1) \) by \( A_j(\bar{j}) \) gives

\[
A_1(0) \quad A_1(0) \quad A_1(0) \\
A_2(0) \quad A_1(1) \quad A_1(1) \\
A_1(1) \quad A_2(0) \quad A_2(1) \\
A_2(1) \quad A_2(1) \quad A_2(0)
\]
Thus, we obtain a $(9, 12, 6; 8)$ code $C$. Now, since the condition $q_1 q_2 > b_1$ is satisfied, $C$ can be extended to a $(10, 12, 6; 9)$ code $C^*$.

$$
\begin{array}{cccccc}
1 & 2 & 3 & 1 & 2 & 3 \\
2 & 3 & 1 & 2 & 3 & 1 \\
3 & 1 & 2 & 3 & 1 & 2 \\
1 & 3 & 2 & 4 & 5 & 6 \\
2 & 1 & 3 & 5 & 6 & 4 \\
3 & 2 & 1 & 6 & 4 & 5 \\
\end{array}
\quad
\begin{array}{cccccc}
1 & 2 & 3 & 1 & 2 & 3 \\
2 & 3 & 1 & 2 & 3 & 1 \\
3 & 1 & 2 & 3 & 1 & 2 \\
1 & 3 & 2 & 4 & 5 & 6 \\
2 & 1 & 3 & 5 & 6 & 4 \\
3 & 2 & 1 & 6 & 4 & 5 \\
\end{array}
$$

$C - C^*$

3  Construction of $(n, b, q)$ $w$-TA codes with $q < w^2$ and $b > q$

In this section we discuss a concrete application of the above construction. We see that the method is suitable for constructing $q$-ary codes with large distance, and therefore, by Theorem 1.1, for constructing $w$-TA codes with large $w$. The following theorem shows this fact.

**Theorem 3.1**  
(i) Let $q_0$ be a prime power. If there is a set of at least $(q_0 - 1)$ mutually orthogonal latin squares (MOLS) of order $\sigma$, then there is an $(n, b, q; d)$ code with

$$
\begin{align*}
  n &= (q_0 + 1)\sigma^m \\
  b &= q_0^2\sigma^m \\
  q &= q_0\sigma^m \\
  d &= (q_0 + 1)\sigma^m - 1,
\end{align*}
$$

for any positive integer $m$.

(ii) There is an $(n, b, q; d)$ code with

$$
\begin{align*}
  n &= \underbrace{(\cdots((q_0 + 1) q_1 + 1)q_1 + 1)\cdots q_1 + 1)}_m \\
  b &= q_0^2 q_1^m \\
  q &= q_0 q_1^m \\
  d &= n - 1,
\end{align*}
$$

where $q_1 \geq q_0$ are prime powers and $m \geq 1$ is an integer.

**Proof.** Take $C_0$ to be an $OA_1(2, q_0 + 1, q_0)$ orthogonal array $A$, (see e.g., [6]), i.e. $C_0$ is a $(q_0 + 1, q_0^2, q_0; q_0)$ extended Reed-Solomon code. The array $A$ has the property that any symbol appears exactly $q_0$ times in each column. A remark upon MOLS, which are used
here, needs to be made. It is known that any given set of $u$ MOLS $M_1, \ldots, M_u$ can be transformed in such a way that any two rows from different $M_i$ and $M_j$ agree in at most one column. Here, we assume that our MOLS have this property.

(i) Now suppose we have a set of $q_0$ MOLS $M_1, \ldots, M_{q_0}$ of order $\sigma$. In the case that we only have $(q_0 - 1)$ MOLS $M_1, \ldots, M_{q_0-1}$, we will take $M_0$ to be the $\sigma \times \sigma$ matrix with entries from the $\sigma$ symbols of the Latin squares such that each symbol appears $\sigma$ times in exactly one row. In either cases, $M_0, M_1, \ldots, M_{q_0-1}$ together form a $\sigma$ resolution of a $(\sigma, q_0, \sigma; \sigma - 1)$ code $\mathcal{C}$. Applying Theorem 2.1 to $\mathcal{C}$ and $\mathcal{C}$ gives a $((q_0 + 1) \sigma, q_0^2 \sigma, q_0 \sigma; (q_0 + 1) \sigma - 1)$ code $\mathcal{C}_1$. As each symbol of the alphabet appears in each column of $\mathcal{A}(\mathcal{C}_1)$ $q_0$ times, Theorem 2.1 can be applied to $\mathcal{C}_1$ and $\mathcal{C}$ again. This recursive procedure gives rise to codes in (i).

(ii) If $\sigma = q_1$ ($\geq q_0$) is a prime power, then there are $q_1 - 1$ MOLS $M_1, \ldots, M_{q_1-1}$ of order $q_1$. $M_1, \ldots, M_{q_1-1}$ and $M_0$ together form a code $\mathcal{C}$ with a $q_1$ resolution. Extend $\mathcal{C}_1$ in (i) to a code $\mathcal{C}_1^*$ by adding one more column, as shown in Theorem 2.1. Observe that in $\mathcal{C}_1^*$ a symbol appears $q_1$ or $q_0$ times in each column. Thus, we can apply Theorem 2.1 to $\mathcal{C}_1^*$ and $\mathcal{C}$. Therefore, if at each step the obtained code is extended before applying Theorem 2.1, the resulting code after $m$ steps will have parameters given in (ii).

The following theorem shows that codes constructed in Theorem 3.1, in fact, provide a large class of $w$-TA codes with $q < w^2$ and $b > q$.

**Theorem 3.2** Let $q_0$ and $q_1$ be prime powers such that $q_1 \geq q_0$.

(i) Suppose $\sqrt{q_0 q_1^2} + 1 < \lfloor \sqrt{q_0 q_1} + q_1 + 1 \rfloor$. Then for any integer $n$ with

$$\sqrt{q_0 q_1^2} + 1 < \lfloor \sqrt{n} \rfloor \leq \lfloor \sqrt{q_0 q_1} + q_1 + 1 \rfloor$$

there exists an $(n, b, q)$ $w$-TA code with $q < w^2$ and $b > q$, where

$$b = q_0^2 q_1$$
$$q = q_0 q_1$$
$$w = \lfloor \sqrt{n} \rfloor - 1.$$ 

(ii) For any integer $m \geq 2$ and for any integer $n$ with

$$\sqrt{q_0 q_1^m} + 1 < \lfloor \sqrt{n} \rfloor \leq \lfloor \sqrt{q_0 q_1^m} + q_1^m + \cdots + q_1 + 1 \rfloor$$

there exists an $(n, b, q)$ $w$-TA code with $q < w^2$ and $b > q$, where

$$b = q_0^2 q_1^m$$
$$q = q_0 q_1^m$$
$$w = \lfloor \sqrt{n} \rfloor - 1.$$ 

**Proof.** First, recall that the parameters $(N, b, q; d)$ of a code $\mathcal{C}^*$ in Theorem 3.1 (ii) are $N = q_0 q_1^m + q_1^m + q_1^{m-1} + \cdots + q_1 + 1$, $b = q_0^2 q_1^m$, $q = q_0 q_1^m$, and $d = N - 1$, where $m \geq 1$ is an integer. We remark that if $\mathcal{C}^*$ is shortened, the resulting code with length $n \leq N$ always have minimum distance $d = n - 1$.

Let $(n, b, q; n - 1)$ be the parameters of a shortened code $\mathcal{C}$ of $\mathcal{C}^*$ (the case $\mathcal{C} = \mathcal{C}^*$ is also included). So, $n \leq N$. Let $w = \lfloor \sqrt{n} \rfloor - 1$. By Theorem 1.1, $\mathcal{C}$ is a $w$-TA code. The condition $q < w^2$, i.e., $\sqrt{q} < w$, thus becomes $\sqrt{q} < \lfloor \sqrt{n} \rfloor - 1$, equivalently $\sqrt{q} + 1 < \lfloor \sqrt{n} \rfloor$. 

5
As $n \leq N$, we have $\sqrt{q} + 1 < \sqrt{n} \leq \sqrt{N}$. Now $q = q_0 q_1^m$, so if $m = 1$, we have the condition $\sqrt{q_0 q_1} + 1 < \sqrt{n} \leq \sqrt{q_0 q_1 + q_1 + 1}$. Thus (i) follows. If $m \geq 2$, we see that the condition $\sqrt{q} + 1 < \sqrt{n}$ is always satisfied. In fact, we only need to verify that $\sqrt{q_0 q_1} + 1 < \sqrt{N}$, i.e., $(\sqrt{q_0 q_1} + 1)^2 < q_0 q_1^m + q_1^m + q_1^{m-1} + \cdots + q_1 + 1$. Simplifying the last inequality yields $4q_0 q_1^{m-2} < (q_1^{m-1} + \cdots + q_1 + 1)^2$, which is satisfied for all integers $q_1 \geq q_0 \geq 2$ and $m \geq 2$. Thus we have (ii). The proof is complete.

**Remark 3.1** In the proof of Theorem 3.2 above, we do not use the approximation $\sqrt{q} + 1 < \sqrt{n}$ to show $\sqrt{q} + 1 < \sqrt{N}$ for case $m = 1$. If we used it, we would get an inequality $4q_0 < q_1$. And therefore, we would miss a large number of $w$-TA codes. In fact, the condition $\sqrt{q_0 q_1} + 1 < \sqrt{q_0 q_1 + q_1 + 1}$, as stated in the theorem, is much stronger.

**Example 3.1** Some small $w$-TA codes of Theorem 3.2 (i) are as follows. A $(10, 12, 6)$ $3$-TA code corresponds to $q_0 = 2$ and $q_1 = 3$. This code is also displayed in Example 2.1. For $q_0 = 3$ and $q_1 = 4$ we have a $(17, 36, 12)$ $4$-TA code, and for $q_0 = 4$ and $q_1 = 5$ we have a $(26, 80, 20)$ $5$-TA code.

**Remark 3.2** It is worth to note that the construction method in Theorem 2.1 can produce good $q$-ary codes. Recall that for any $(n, b, q; d)$ code the Plotkin bound is given by $b(b - 1)d \leq 2n \sum_{i=0}^{b-1} \frac{q^i}{q+1} b_i$, where $b_i = \lceil (b+i)/q \rceil$, see, e.g., [1]. Now consider, for example, the codes in Theorem 3.1 (ii). It is easy to check that if $q_0 = q_1$, these codes meet the Plotkin bound with equality. Moreover, for the three codes mentioned in Example 3.1 we have the following. The $(10, 12, 6; 9)$ code is optimal. The $(17, 36, 12; 16)$ and $(26, 80, 20; 25)$ codes are ‘quasi’ optimal because the maximum value for $b$ derived from the Plotkin bound is 37 in the first case and 81 in the second case.

**References**


