A note on an upper bound of traceability codes *

S. -L. Ng†, S. Owen‡

February 25, 2015

Abstract

Blackburn, Etzion and Ng showed in a paper in 2010 that there exist 2-traceability codes of length l of size $cq^{\lceil l/4 \rceil}$, where the constant c depends only on l. The question remains as to what the best possible c may be. A well-known construction using error-correcting codes with high minimum distance gives 2-traceability codes of size $cq^{\lceil l/4 \rceil}$ with $c \geq 1$. However, in the same paper, an example of a 2-traceability code of length 3 with size $\frac{3}{2}(q-1)$ was given, which shows that $c > 1$ in some situations, and that there are traceability codes that are bigger than the construction using error-correcting codes. Here we give an upper bound $4q-3$ for 2-traceability codes of length 4 and give an example of $(l-1)$-traceability codes of length l with size $\frac{l}{l-1}(q-1)$. This example also gives a 2-traceability code of length 4 larger than any codes constructed using the error-correcting code construction.

1 Introduction

Traceability codes are combinatorial objects introduced by Chor, Fiat and Naor [3] in 1994 for the construction of traitor tracing schemes to protect digital content. We will only introduce the necessary notation and definitions for describing our results here and refer the reader to [1, 2, 4] for further references and other important results in the area.

Let F be a finite set of cardinality q. A code $\mathcal{C}$ of length l over F is a set $\mathcal{C} \subseteq F^l$. We write $d(\mathbf{u}, \mathbf{v})$ for the Hamming distance between two words $\mathbf{u}$, $\mathbf{v} \in F^l$, and for a code $\mathcal{C}$, we write $d(\mathcal{C})$ for the minimum distance of $\mathcal{C}$. For $\mathbf{x} \in F^l$ we write $\mathbf{x} = (x_1, x_2, \ldots, x_l)$ or $\mathbf{x} = x_1 x_2 \ldots x_l$ for convenience.

Let $\mathcal{C}$ be a code of length l over F. If $X \subseteq \mathcal{C}$ then a *descendant* of the set X is a word $\mathbf{d} \in F^l$ such that for each $i \in \{1, \ldots, l\}$, there is an $\mathbf{x} \in X$ such that $d_i = x_i$. The *set of descendants of* X is

$$desc(X) = \{\mathbf{d} \in F^l : \mathbf{d} \text{ is a descendant of } X\}.$$

*This work was part of S. Owen's MSc project at the School of Mathematics and Information Security, Royal Holloway, University of London.

†(corresponding author) s.ng@rhul.ac.uk, Information Security Group, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom

‡steph.louise.owen@gmail.com, Telefonica, 260 Bath Road, Slough, Berkshire SL1 4DX

Moreover, the *k-descendant set of* $\mathcal{C}$ is the union of all descendant sets $desc(X)$ for all subsets $X \subseteq \mathcal{C}$ with $|X| \leq k$, and is denoted $desc_k(\mathcal{C})$.

Now, let $\mathbf{d}$ be a word in F^l and let $k \geq 2$. A codeword $\mathbf{x} \in \mathcal{C}$ is a *(possible) parent* of $\mathbf{d}$ if there exists a set $P \subseteq \mathcal{C}$, $P \leq k$, such that $\mathbf{x} \in P$ and $\mathbf{d} \in desc(P)$. A *parent set of* $\mathbf{d}$ is a set $P \subseteq \mathcal{C}$ such that $\mathbf{d} \in desc(P)$ and $|P| \leq k$. We use $P_{k,\mathcal{C}}(\mathbf{d})$ to denote the set of all possible parent sets for $\mathbf{d}$. For any word $\mathbf{d} \in F^l$, let $P_{\mathbf{d}}$ be the set of all codewords of $\mathcal{C}$ at minimum distance to $\mathbf{d}$. If $P_{\mathbf{d}}$ appears in every possible parent set of $\mathbf{d}$ of size at most k, then $\mathcal{C}$ is a *k-traceability (k-TA) code.*

In [2], Blackburn, Etzion and Ng showed that there exist 2-TA codes of length l of size $cq^{\lceil l/4 \rceil}$, where the constant c depends only on l [2, Theorem 3]. The question remains open as to what the best possible c may be. A well-known construction using error-correcting codes with high minimum distance gives 2-TA codes of size $cq^{\lceil l/4 \rceil}$ with $c \geq 1$:

Theorem 1.1 *[3] Let $\mathcal{C}$ be an error-correcting code of length l, over a q-ary alphabet F. If $d(\mathcal{C}) > l - \left\lceil \frac{l}{k^2} \right\rceil$ then $\mathcal{C}$ is a k-TA code.*

For example, a Reed-Solomon code would give a k-TA code of size $q^{\lceil l/4 \rceil}$. In the same paper, however, an example of a 2-TA code of length 3 with size $\frac{3}{2}(q-1)$ was given:

Example 1.2 *[2, Example 1]. Let $q = 2r+1$, where r is a positive integer. Let $A = \{0, 1, \ldots, 2r\}$. Define $C = C_1 \cup C_2 \cup C_3$, where*

$$C_1 = \{(0, i, i) \colon 1 \leq i \leq r\}$$
$$C_2 = \{(i, 0, r+i) \colon 1 \leq i \leq r\}$$
$$C_3 = \{(r+i, r+i, 0) \colon 1 \leq i \leq r\}.$$

Then C is a q-ary 2-TA code of length 3, containing $3r = \frac{3}{2}(q-1)$ codewords.

This example shows that the constant c in the upper bound of [2] can be greater than 1 in some situations, and that there are traceability codes that are bigger than the construction of Theorem 1.1. It is not known what the best constant c is. From the proof of [2, Theorem 3], the constant is exponential in l. Here we obtain an upper bound $4q - 3$ for 2-TA codes of length 4 and give an example of $(l-1)$-TA codes of length l with size $\frac{l}{l-1}(q-1)$. This example also gives a 2-TA code of length 4 larger than any codes constructed using the error-correcting code construction.

2 A bound on 2-TA codes of length 4

We consider the cases of $\mathcal{C}$ having minimum distance 2 and 3 separately.

Lemma 2.1 *Let $\mathcal{C}$ be a q-ary 2-TA code of length 4, $q \geq 2$, with minimum distance $d(\mathcal{C}) = 2$. Without loss of generality, suppose that $\mathbf{x} = 0000$, $\mathbf{y} = 0011 \in \mathcal{C}$. Then any other codeword $\mathbf{z} \in \mathcal{C} \setminus \{\mathbf{x}, \mathbf{y}\}$ must have either $z_1 = 0$ or $z_2 = 0$, and z_3, $z_4 \notin \{0, 1\}$.*

Proof: Let $\mathbf{z} \in \mathcal{C} \setminus \{\mathbf{x}, \mathbf{y}\}$, $\mathbf{z} = z_1 z_2 z_3 z_4$. Let $\mathbf{d} = 00 z_3 z_4$. Then

$$\mathbf{d} \in \mathit{desc}(\mathbf{x}, \mathbf{z}) \cap \mathit{desc}(\mathbf{y}, \mathbf{z}),$$

and since $P_{\mathbf{d}}$ must be contained in the parent sets, we must have $P_{\mathbf{d}} = \{\mathbf{z}\}$. Since $\mathcal{C}$ is a 2-TA code, we must have $d(\mathbf{d}, \mathbf{z}) < d(\mathbf{d}, \mathbf{x})$.

Now $d(\mathbf{d}, \mathbf{x}) \leq 2$. Since $d(\mathbf{d}, \mathbf{z}) < d(\mathbf{d}, \mathbf{x})$, we can't have $d(\mathbf{d}, \mathbf{x}) = 0$. If $d(\mathbf{d}, \mathbf{x}) = 1$, then either $z_3 = 0$, $z_4 \neq 0$ or $z_3 \neq 0$, $z_4 = 0$. In the first case, since we must have $d(\mathbf{d}, \mathbf{z}) < d(\mathbf{d}, \mathbf{x})$, we have $\mathbf{d} = \mathbf{z} = 000 z_4$. This contradicts the fact that $\mathcal{C}$ has minimum distance 2. Similarly for the second case.

Hence we must have $d(\mathbf{d}, \mathbf{x}) = 2$ and so $z_3 \neq 0$, $z_4 \neq 0$. Similarly we have $z_3 \neq 1$, $z_4 \neq 1$.

Since $d(\mathbf{d}, \mathbf{z}) < d(\mathbf{d}, \mathbf{x})$, we must have $d(\mathbf{d}, \mathbf{z}) \leq 1$, so either $z_1 = 0$ or $z_2 = 0$. Hence all codewords of $\mathcal{C}$ are of the form $00 z_3 z_4$, $z_1 0 z_3 z_4$ or $0 z_2 z_3 z_4$, with z_3, $z_4 \notin \{0, 1\}$.

■

From the proof of Lemma 2.1, we see that in a code of length 4 with minimum distance 2, if two codewords agree in the first two positions then the symbols in the third position of these codewords do not appear again in the third position of another codeword. Hence we have the following result:

Lemma 2.2 *Let $\mathcal{C}$ be a q-ary 2-TA code of length 4, $q \geq 2$, with minimum distance $d(\mathcal{C}) = 2$. Define $X \subseteq \mathcal{C}$ as follows:*

$$X = \{\mathbf{u} \in \mathcal{C} \colon \exists \mathbf{v} \in \mathcal{C}, \mathbf{u} \neq \mathbf{v}, \text{ with } u_1 = v_1 \text{ and } u_2 = v_2\}.$$

Then $|X| \leq q$.

We can now prove a bound on $\mathcal{C}$:

Theorem 2.3 *Let $\mathcal{C}$ be a q-ary 2-TA code of length 4, with $d(\mathcal{C}) = 2$ and $q \geq 2$. Then $|\mathcal{C}| \leq 3q - 2$.*

Proof: Again, without loss of generality, suppose that $\mathbf{x} = 0000$, $\mathbf{y} = 0011 \in \mathcal{C}$. Define the set $X \subseteq \mathcal{C}$ as in Lemma 2.2. Clearly $\mathbf{x}$, $\mathbf{y} \in X$. We further define the sets $U, V \subseteq \mathcal{C}$:

$$U = \{u_1 0 u_3 u_4 \colon u_1 \text{ does not occur in the 1st position of another codeword}\},$$
$$V = \{0 v_2 v_3 v_4 \colon v_2 \text{ does not occur in the 2nd position of another codeword}\}.$$

By Lemma 2.1, all codewords of $\mathcal{C}$ are of the form $00 z_3 z_4$, $z_1 0 z_3 z_4$ or $0 z_2 z_3 z_4$. Hence every codeword belongs to X, U or V. By definition X is disjoint from U and from V. Clearly U and V are also disjoint, since if a codeword agrees with another codeword in positions one and two then, by definition, they are contained in X. Hence $\mathcal{C} = X \cup U \cup V$ and $|\mathcal{C}| = |X| + |U| + |V|$.

Consider the set U. No two codewords of U can agree in position one and they cannot contain 0 in position one. Hence U can contain at most $(q-1)$ codewords. By the same argument, there are at most $(q-1)$ codewords in V. By Lemma 2.2, X contains at most q codewords. Hence $|\mathcal{C}| = |X| + |U| + |V| \leq q + 2(q-1) = 3q - 2$, as required.

■

Now we consider a q-ary 2-TA code $\mathcal{C}$ of length 4 with minimum distance $d(\mathcal{C}) = 3$. We prove the following theorem.

Theorem 2.4 *Let $\mathcal{C}$ be a q-ary 2-TA code of length 4, $q \geq 3$, with minimum distance $d(\mathcal{C}) = 3$. Then $|\mathcal{C}| \leq 4q - 3$.*

Proof: We define two disjoint subsets X, Y of $\mathcal{C}$ as follows:

$$X = \{\mathbf{x} \in \mathcal{C} : \text{ at least 4 codewords in } \mathcal{C} \text{ contain } x_1 \text{ in their first position}\},$$
$$Y = \{\mathbf{x} \in \mathcal{C} : \text{ at most 3 codewords in } \mathcal{C} \text{ contain } x_1 \text{ in their first position}\}.$$

Clearly $\mathcal{C} = X \cup Y$ and since X and Y are disjoint, we have that $|\mathcal{C}| = |X| + |Y|$. Therefore we will obtain a bound on the size of $\mathcal{C}$ by finding bounds on the size of the subsets X and Y.

If X is non-empty then it contains at least four codewords. Without loss of generality we assume that $\mathbf{x} = 0000$, $\mathbf{y} = 0111$, $\mathbf{z} = 0222$, $\mathbf{w} = 0333 \in X$.

If all the codewords in x coincide in the first position, then there can be at most q codewords in X since $d(\mathcal{C}) = 3$.

Suppose then that not all codewords in X coincide in the first position, and so X contains at least two groups of four codewords which agree in position one. Without loss of generality we assume that four codewords of this second group are $\mathbf{u}$, $\mathbf{v}$, $\mathbf{s}$, $\mathbf{t} \in X$:

$$\mathbf{u} = 1u_2u_3u_4,$$
$$\mathbf{v} = 1v_2v_3v_4,$$
$$\mathbf{s} = 1s_2s_3s_4,$$
$$\mathbf{t} = 1t_2t_3t_4,$$

with distinct symbols at position i for each $i = 2, 3, 4$.

We will show that none of the u_i, v_i, s_i, t_i can be 0, 1, 2 or 3. This show that the symbols in the second (third, and fourth, respectively) position must be distinct and therefore $|X| \leq q$.

Suppose that $u_2 = 0$, so $\mathbf{u} = 10u_3u_4$. Since $d(\mathcal{C}) = 3$, u_3, $u_4 \neq 0$ and v_2, s_2, $t_2 \neq 0$.

Consider the word $\mathbf{d} = 10v_30 \in desc(\mathbf{x}, \mathbf{v})$. Since $d(\mathcal{C}) = 3$ $\mathbf{d}$ is not a codeword. Since $u_4 \neq 0$, $d(\mathbf{d}, \mathbf{u}) = 2$. By the traceability property, either $\mathbf{x}$ or $\mathbf{v}$ (or both) must belong to $P_{\mathbf{d}}$, the set of all codewords of $\mathcal{C}$ at minimum distance to $\mathbf{d}$. So either $d(\mathbf{d}, \mathbf{x}) = 1$ or $d(\mathbf{d}, \mathbf{v}) = 1$.

Suppose $d(\mathbf{d}, \mathbf{x}) = 1$. This means that $v_3 = 0$ and $\mathbf{v} = 1v_20v_4$ and v_4, u_3, s_3, $t_3 \neq 0$.

Now consider the word $\mathbf{d}' = 10s_30 \in desc(\mathbf{x}, \mathbf{s})$. Again $\mathbf{d}'$ is not a codeword, and $d(\mathbf{d}', \mathbf{u}) = d(\mathbf{d}', \mathbf{x}) = 2$. So we must have $\mathbf{s} \in P_{\mathbf{d}'}$ and therefore we must have $d(\mathbf{d}', \mathbf{s}) = 1$ and $s_4 = 0$.

To recapitulate, we now have:

$$\begin{aligned}\mathbf{x} &= 0000,\\ \mathbf{u} &= 10u_3u_4,\\ \mathbf{v} &= 1v_20v_4,\\ \mathbf{s} &= 1s_2s_30,\\ \mathbf{t} &= 1t_2t_3t_4,\end{aligned}$$

with u_i, v_i, s_i, $t_i \neq 0$.

Finally, consider the word $\mathbf{d}'' = 1t_200 \in desc(\mathbf{x}, \mathbf{t})$. Again $\mathbf{d}''$ is not a codeword, and $d(\mathbf{d}'', \mathbf{x}) = d(\mathbf{d}'', \mathbf{t}) = d(\mathbf{d}'', \mathbf{v}) = 2$.

This contradicts the traceability property that says that codewords nearest to the descendant must be contained in the parent set.

Following the same argument, assuming $d(\mathbf{d}, \mathbf{v}) = 1$ also gives a constradiction. Therefore our initial assumption that there are two codewords in X which agree in a position other than the first position, must be incorrect. Hence all codewords in X are distinct in each of their final three positions, and X can contain at most q distinct codewords.

We now have a bound on the size of X, namely that $|X| \leq q$. A bound on the size of the set Y is straightforward. If X is not empty, then at least one symbol has already appeared in the first position of at least four codewords, and so cannot appear in the first position of any codeword in Y, that is, we have only $q-1$ choices for x_1 in Y, and $|Y| \leq 3(q-1)$, and so $|\mathcal{C}| \leq q + 3(q-1) = 4q - 3$. If X is empty, $|\mathcal{C}| \leq 3q \leq 4q - 3$ for all $q \geq 3$, and hence for any 2-TA code of length 4 with $d(\mathcal{C}) = 3$ and $q \geq 3$, we have $|\mathcal{C}| \leq 4q - 3$.

■

It is not hard to show that for the remaining cases of $q = 2$, $d(\mathcal{C}) = 1$ and $d(\mathcal{C}) = 4$, $|\mathcal{C}| \leq q$. Hence the following theorem is now proved:

Theorem 2.5 *Let $\mathcal{C}$ be a 2-TA code of length 4, over a q-ary alphabet. Then $|\mathcal{C}| \leq 4q - 3$.*

Having proved the bound, the question of whether this bound is tight once again arises. Furthermore, can we find a construction method for a code of maximum size? This remains an open question.

For a 2-TA code of length 4, Theorem 1.1 requires a code $\mathcal{C}$ with $d(\mathcal{C}) = 4$. However a code of length 4 with minimum distance 4 can contain at most q codewords. In the next section we show that it is possible to constuct a 2-TA code of length 4 larger than any codes constructed using Theorem 1.1.

3 An $(l-1)$-TA code of length l

We will show that a q-ary $(l-1)$-TA code of length l and size $\frac{l}{l-1}(q-1)$ exists. The following example is an extension of Example 1.2:

Example 3.1 *Let $q = (l-1)r + 1$, where r is a positive integer. Let $A = \{0, 1, ..., (l-1)r\}$.*

Define $C = C_1 \cup C_2 \cup ... \cup C_l$, where

$$\begin{aligned} C_1 &= \{(0, i, ..., i) \colon 1 \leq i \leq r\}, \\ C_2 &= \{(i, 0, r+i, ..., r+i) \colon 1 \leq i \leq r\}, \\ C_3 &= \{(r+i, r+i, 0, 2r+i, ..., 2r+i) \colon 1 \leq i \leq r\}, \\ &\vdots \\ C_l &= \{((l-2)r+i, ..., (l-2)r+i, 0) \colon 1 \leq i \leq r\}. \end{aligned}$$

Each subset C_j contains r codewords, and hence $|C| = lr = \frac{l}{l-1}(q-1)$.

Let $\mathbf{d} \in desc(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^{l-1})$, $\mathbf{x}^i \in C$ for all $i = 1, \ldots, l-1$. Let $\mathbf{x}$ be a nearest codeword to $\mathbf{d}$. Then $\mathbf{x}$ must agree with $\mathbf{d}$ in at least two positions. But any two positions will identify the only codeword that can contribute to those positions in $\mathbf{d}$. Hence $\mathbf{x}$ must be in every possible parent set for $\mathbf{d}$ and C is an $(l-1)$-TA code.

If we take $l = 4$ (so $q = 3r + 1$) this gives us a 3-TA code of length 4 of size $\frac{4}{3}(q-1)$. Now, since a k-TA code is also a $(k-1)$-TA code, we have the following result:

Theorem 3.2 *There is a q-ary 2-TA code $\mathcal{C}$ of length 4 with $|\mathcal{C}| = \frac{4}{3}(q-1)$.*

This shows that it is possible to constuct a 2-TA code of length 4 larger than any codes constructed using Theorem 1.1. Unfortunately, our results only confirm that c can be greater than 1, but do not indicate how c varies with l. The question remains open as to what the best possible c may be.

References

[1] Simon R Blackburn. Combinatorial schemes for protecting digital content. *Surveys in combinatorics*, 307:43–78, 2003.

[2] Simon R. Blackburn, Tuvi Etzion, and Siaw-Lynn Ng. Traceability codes. *J. Comb. Theory, Ser. A*, 117(8):1049–1057, 2010.

[3] Benny Chor, Amos Fiat, and Moni Naor. Tracing traitors. In Yvo Desmedt, editor, *CRYPTO*, volume 839 of *Lecture Notes in Computer Science*, pages 257–270. Springer, 1994.

[4] Jessica Staddon, Douglas R. Stinson, and Ruizhong Wei. Combinatorial properties of frameproof and traceability codes. *IEEE Transactions on Information Theory*, 47(3):1042–1049, 2001.