Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 27 (2024), No. 8, pp. 2473–2482
DOI : 10.47974/JDMSC-2072

Traceable codes formed using certain combinatorial designs

Anu Kathuria
The Technological Institute of Textile and Sciences
Maharshi Dayanand University
Bhiwani 127021
Haryana
India

Abstract

In paper of Staddon and Wei [4] ; there were open problems
(1) " Do there exist w- IPP codes with the property $w < q < [(w + 2)^2 / 4]$? "
(2) "Can we construct interesting w-TA(w-traceable) codes with $n < q$ and $q < w^2$? "

Here in this paper we give that answer and suggest some Combinatorial Designs which meet the above conditions and prove them to be 2-TA and 3-TA Codes. In this paper our spotlight is mainly on Equireplicate Codes and Lexicographic Codes where each element occurs equally often. A few of these codes discussed in (Sinha et. al.2008) are also experimental to be excellent Equidistant Constant Weight Codes

Subject Classification: *12F05, 12F06, 12G15.*

Keywords: *Equireplicate code, Lexicographic code and equidistant constant weight code.*

1. Introduction

This paper discusses challenges in digital rights management (DRM), particularly the need for techniques that go beyond traditional encryption to effectively prevent unauthorized redistribution of digital content. While encryption protects content from unauthorized access, it does not address cases where authorized users share the content illegally. To tackle this issue, researchers have proposed traceability techniques aimed at identifying individuals or groups involved in illegal distribution. These techniques fall into two main categories ;

E-mail: anu_sept24@rediffmail.com

- **Traceability Codes (TA Codes)**: Designed to detect users who redistribute content without permission.
- **Identifiable Parent Property (IPP) Codes**: These codes help pinpoint users or sets of users who may be responsible for the unauthorized sharing.

Together, these are often referred to as **Fingerprinting Codes**, which can be used to trace at least one person involved in illicit distribution. Several types of these codes have been considered for their efficacy in tracing schemes. Different variations of these codes have been developed also time to time .

- **Frameproof Codes**: Originally proposed by Boneh and Shaw [5], these codes are more vulnerable to tampering.
- **IPP Codes**: Introduced by Hollman et.al [7], they offer stronger protection for tracing the origins of shared content.

Further developments include the introduction of traceability codes by Chor et.al [4], and the exploration of combinatorial approaches by Staddon, Stinson, and Wei [9]. A recent approach involves using Combinatorial Designs, Lexicographic codes, and Equireplicate Codes, which have been found to be effective in creating fingerprinting codes that meet certain mathematical criteria for traceability. The concept of frameproof codes, which are considered a more fragile type of fingerprinting code, was first introduced by Boneh and Shaw[5]. Later, a more robust category known as Identifiable Parent Property (IPP) Codes was urbanized in [7]. Another important advancement in this area came from Chor, Fiat, and Naor[4], who introduced traceability codes. Research on the combinatorial properties of these codes has been further expanded by Staddon, Stinson, and Wei, [9] who analyzed the characteristics of both traceability and Frameproof Codes. Additionally, Mathematical Conditions under which equidistant codes can function as 2-TA codes have been established in [1, 2, 3]. These codes represent a new class of extended Reed-Solomon (RS) codes and can be used effectively as 2-TA codes. The current work demonstrates that combinatorial designs based on lexicographic and Equireplicate codes can serve as effective fingerprinting codes, showing that these types of codes exhibit good equidistant properties. Additionally, combinatorial aspects of Fibonacci partial words

and arrays are explored in this context in [10]. The following sections will present the necessary formulae for the subsequent discussions.

2. Some Fundamental Definitions and Terms;

2.1 *Here are some key terms related to fingerprinting codes:*

1. **Fingerprinting Code**: A method used to uniquely identify a copy of a digital object, allowing for tracing the source of illegal distribution. Fingerprinting codes embed specific patterns in digital content to help trace unauthorized sharing.
2. **Pirate**: An individual or entity that illegally shares or redistributes copyrighted digital content without permission.
3. **Traceability**: The ability to identify the source of a digital object, particularly in cases of illegal distribution. Traceability is a fundamental aspect of fingerprinting codes.
4. **Codeword**: An encoded representation of a digital object that is used within fingerprinting codes. Each codeword corresponds to a unique user or copy of the content.
5. **Collusion**: A scenario where multiple pirates cooperate to defeat the fingerprinting system, often by sharing their codewords in a way that obscures the original source.
6. **Traceability Code (TA Code)**: A type of fingerprinting code designed specifically to trace and identify dishonest users who illegally distribute digital content.
7. **Identifiable Parent Property (IPP) Code**: A code that allows for identifying the original source of a digital object, even when a user has attempted to alter or hide their identity.
8. **Hamming Distance**: A measure of the difference between two codewords. In fingerprinting codes, it can be used to quantify how similar or different two codewords are, which is important for tracing purposes.
9. **Weight**: In the context of fingerprinting codes, weight refers to the number of non-zero positions in a codeword. Codes with fixed weight can be particularly useful for maintaining certain properties in the fingerprinting process.

10. **Frameproof Code**: A specific type of fingerprinting code designed to resist certain types of attacks, particularly those involving collusion among pirates.
11. **Robustness**: The ability of a fingerprinting code to maintain its functionality and integrity in the presence of attacks or attempts to subvert its traceability features.
12. **Optimal Code**: A fingerprinting code that achieves the best possible performance in terms of traceability, robustness, and efficiency, often defined by parameters like minimum distance and dimension. These terms form the foundation for understanding the mechanics and challenges associated with fingerprinting codes and their application in digital rights management.

3. "Equidistant Codes as Traceable Codes "

In this part we present the results established in [1] regarding the situation under which an equidistant constant weight code qualifies as a **2-Traceable (2-TA) code**.

Theorem 3.1 [1]: "*An Equidistant Constant Weight Code carrying distance end to end n with distance d > 2n /3 is for all time a two-Traceable Code'. (The condition d>2n/3; implies that the minimum distance between any two code words is sufficiently large. This huge minimum distance enhances the code's ability to distinguish between code words, making it harder for pirates to collude without being detected).*

Theorem 3.2 [6, 9]: "*if D is a code of length n and minimum distance d such that certain conditions are met, we can derive specific properties and implications about the code i.e.* $d > \left(1-\frac{1}{m^2}\right)n$, *where* $m \geq 2$". *Then D is m-Traceable Code.*

Lemma 3.3[1]: "*A Linear Maximum Distance Separable Code D with parameters [p+1, 2, p] possesses several properties that classifies it as an Equidistant Constant Weight Code". In summary, the linear MDS code D with parameters [p+1, 2, p] is indeed an equidistant constant weight code and is classified as an optimal code due to its maximized minimum distance.*

Definition 3.4: (Plotkin Bound [1]). "Let D be a code with the following characteristics;

- **Length** N**:** The number of symbols in each codeword.
- **Size** n: The total number of codewords in the code.

- **Minimum Distance** d: The minimum distance between any two distinct codewords**.**
- **Alphabet Size** m: The number of elements in the underlying alphabet Q. if for these parameters $d \leq \frac{nN(m-1)}{(n-1)m}$. if $d = \frac{nN(m-1)}{(n-1)m}$ (i.e. d is equal to that numeric value) then D is Optimized code also."

Definition 3.5: (Deslarte Bound [9]). "$B_q(n,d) \leq (p-1) \quad n+1$"; for a particular case $n = p+1;$ we have $B_q(p+1,d) \leq (p-1)\ (p+1)\ +1$ i.e. $= p^2$.

4. "Equireplicate Code as 2-TA Code "

Here in Section 4, how Equireplicate codes can be used as different form of fingerprinting codes.

Definition 4.1: [11] In this definition, we are describing a structured combinatorial system known as a *set system* or *block design*, where we have:

1. **Set M**: This is a set containing m distinct elements (points or symbols). Each point can represent a unique entity, symbol or value.
2. **Set of Blocks B**: This is a collection of b subsets (blocks) of M. Each block Bi ∈ B is a subset of M with its *block size* defined by the number of elements it contains, denoted as |Bi|
3. **Uniformity (k-uniform)**: If each block in B contains exactly k elements, then the set system is called *k-uniform*.
4. **Replication Number n**: For each element m ∈M , the *replication number* n represents the number of blocks in B that include m. This is effectively the frequency of each point across the blocks.
5. **n-equireplicate**: A set system is called *n-equireplicate* if every element in M appears in exactly n blocks. This implies uniform replication, where each point's appearance is distributed evenly across the blocks in B.

In short:

- (M,B) forms a (m,b) - set scheme.
- The system is *k-uniform* if all blocks have size k.
- The system is *n-equireplicate* if each point appears in exactly n blocks.

Table 1.

0	0	0	0
2	1	0	1
1	2	0	2
1	1	1	0
0	2	1	1
2	0	1	2
2	2	2	0
1	0	2	1
0	1	2	2

Table 1, gives an example of (4,9,3) - code consisting of 9 codewords each of length 4,distance 3 defined over the set of 3 elements. It is simple to confirm that each point or element appears for 3 times in every row and in every column. Here in this example, moreover if we check the distance and apply the property of traceability condition over equidistant code, then this code come out to be 2-traceable code. We even here show that the above code C satisfies the unlock problem discussed in [4]. "*Can we design w-TA codes such that* $q < w^2$ *and* $n > q$? " it is very clear that here length of each codeword is 4 and distance d between every two codewords is 3. in this context even we represent one more example.

Table 2

0	0	0	0	0
0	1	1	1	1
1	0	1	2	2
1	2	2	0	1
2	1	2	2	0
2	2	0	1	2

In Table 2, we also find that the distance d between every two codewords is same and each element appears for same number of times in every column and in every row. Here the above code C also is an equidistant code and if we apply the property of traceability over there then it comes out to be 2-TA code (d = 4, i.e. 4 > 2.5/3).Construction and Terminology of Equireplicate Codes has been suggested in [8].

Proposition 4.1 : If there exists an equireplicate (n_1, M, d_1)- 2-TA code and (n_2 ,M, d_2)-2-TA code then concatenation of 2-TA i.e. given as ($n_1 + n_2$, M , d_1+d_2) code is also 2-TA code.

Proof: As if the given equireplicate is 2-TA code then it is obvious that it will meet the condition of traceability as $d_1 > \frac{2n_1}{3}$ and for second type of code it is also clear that $d_2 > \frac{2n_2}{3}$. if we concatenate both codes then also we get that 2-TA code as it satisfies the condition that $d_1 + d_2 > \frac{2(n_1+n_2)}{3}$. It completes our result. Here in this context we represent an example.

Table 3

0	0	0	0	0	0	0	0
2	1	0	1	0	1	1	2
1	2	0	2	0	2	2	1
1	1	1	0	1	0	1	1
0	2	1	1	1	1	2	0
2	0	1	2	1	2	0	2
2	2	2	0	2	0	2	2
1	0	2	1	2	1	0	1
0	1	2	2	2	2	1	0

For that example it is easy to verify that the above code in Table 3 meets the properties of being 2-TA code and it also satisfies the condition that it is w-TA satisfying $n > q$ and $q < w^2$ discussed in [4]. In the same way we can design 3-TA codes satisfying the desired condition too. No such types of 3-TA codes are available in Literature till now.

5. Lexicographic Codes as m-TA Codes

Here in this part we show that how Lexicographic codes be able to be used as 2-TA and 3- TA codes. Lexicographic codes consisting of length n and Hamming distance d are derived by taking into consideration, p – ary vectors with weight w in lexicographic order and and adding together up them to the code if they are at a distance exactly d from the words they have been added earlier. The notion of **Lexicographic Codes** was introduced by Conway and Sloane in [8,12]. These codes serve as a heuristic approach to achieving optimal codes, often providing superior parameters and coding representations. In their study, a number of general bounds for equidistant and constant weight codes are established. Additionally, the

construction of constant weight equidistant codes, utilizing the computer package QPlus, is elaborated upon in [9].

Table 4

0	0	0	0	0
0	1	1	1	1
1	0	1	2	2
1	2	2	0	1
2	1	2	2	0
2	2	0	1	2

Here we can easily found that the above code in Table 4 consists of the codewords at fixed distance 4 and constant weight 4. Length of each codeword is 4. On applying the condition of traceability over this code it is easy to verify that the above code is 2-TA code. That above code also meets the condition mentioned in open problem "*Can we construct w –TA codes with* $n > q$ and $q < w^2$? " . In above example it is easy to confirm that the above code C consists of each codeword with length 4 over the field size 3, i.e. $n > q$ and $q < w^2$ comes out to be true over this code. It is also straightforward to obtain a code of length 6 and distance 6, with a weight of 4, using the computer package QPlus. There will be a set of 3 codewords as we discuss with an example.

Table 5

0	0	1	2	1	2
2	1	0	0	2	1
1	2	2	1	0	0

It is straightforward to verify that the code in Table 5 is a 3-Traceable (3-TA) code. If we assign these three codewords as A, B and C suppose these three colluders conspire and produce a new codeword D then (1 0 0 0 1 0), then $d(A, D) = 4$, $d(B, D) = 4$ and $d(C, D) = 4$. The above example is an example of 3- TA code , no such example is available in literature. That code even itself gives the answer of open problem (i) also.

Proposition 5.1: If a code (n_1, M, d_1)– Lexicographic code exists as m – *TA* code and there also exists an (n_2, M, d_2) – Lexicographic code as m – *TA* code; then concatenation of such *m-TA* codes is also m – *TA* code.

Proof: As it is obvious that if the given Lexicographic code with parameters (n_1, M, d_1) is m-TA code then it will satisfy the condition of traceability i.e. $d_1 > (1 - 1/m^2)$. $n_1$and for the second code if we apply the condition of traceability i.e. $d_2 > (1 - 1/m^2)$. n_2 is also true over there. On concatenating the above codes now we will be getting the code with the length $n_1 + n_2$ and it will also satisfy the condition of traceability as $d_1 + d_2 > (1 - 1/m^2)$. $(n_1 + n_2)$

So the concatenation of 2-TA codes is also 2-TA code.

6. Conclusions

(i) In this paper we show that how combinatorial designs like Equireplicate codes and Lexicographic codes can be used as fingerprinting Codes ?

(ii) Both types of Combinatorial Designs can be used in designing $w - TA$ codes specifying the conditions $n > q$ and $q < w^2$ in open problem of Stinson and Wei [4].

7. Acknowledgements

This acknowledgment expresses the authors' appreciation to the anonymous referees who reviewed the work and provided insightful feedback, which likely contributed to refining the research.

References

[1] A. Kathuria, S. Batra, and S. K. Arora, "On traceability property of equidistant codes," *Discrete Math.*, vol. 340, no. 4, pp. 713–721, Apr. (2017).

[2] Anu Kathuria, Sudhir Batra and S.K. Arora " A class of 2-FP Codes" *Journal of Information and Optimization Sciences*, Taylor and Francis, vol.38, issue 8, pg.1311-1324, December (2017).

[3] Anu Kathuria, Sudhir Batra " On Algebraic Conditions of Equidistant Codes " *International Journal of Science and Research Archive*", vol.8 Issue 2, pg. 575-588, April (2023).

[4] B. Chor, A. Fiat and M. Naor , "Tracing Traitors", in Advances in Cryptology – CRYPTO 94 (Lecture Notes in Computer Science) Berlin, Germany, Springer Verlag, vol. 839, pp. 257-270 (1994).

[5] D. Boneh and J. Shaw, "Collusion –Secure fingerprinting for Digital Data", *IEEE Transactions on Information Theory*, vol. 44, pp. 1897-1905 (1998).

[6] Fang-Wei Fu, Torleiv Klove ,Yuan Luo "On Equidistant Constant Weight Codes" *Discrete Applied Mathematics*, 128, pg. 157-164 (2003).

[7] H. D. L. Hollman, J. H. Van Lint, and J.-P. Linnartz, "On Codes with the identifiable parent property," *J. Combin. Theory, Ser. A*, vol. 82, pp. 121–133 (1998).

[8] J.H. Conway and N.J.A Sloane, "Lexicographic Codes : Error – Correcting codes from game Theory" *IEEE Transactions on Information Theory*, vol. 32, pp. 337- 348 (1986).

[9] J.N. Staddon, D.R. Stinson and R. Wei ," Combinatorial Properties of Frameproof and Traceable Codes" *IEEE Transactions on Information Theory*, vol.47, pp. 1042-1049 (2001).

[10] R. Krishna Kumari, R. Arulprakasam, and V. R. Dare, "Combinatorial properties of Fibonacci partial words and arrays," *J. Discrete Math. Sci. Cryptogr.*, vol. 24, no. 4, pp. 1007–1020 (2021), doi: 10.1080/09720529.2020.179452.

[11] N. Alon, C. J. Colbourn, A. C. Ling, and M. Tompa, "Equireplicate balanced binary codes for oligo arrays," *SIAM J. Discrete Math.*, vol. 14, no. 4, pp. 481–497 (2001).

[12] T. Todorov, G. Bogdanova, and T. Yorgova, "Lexicographic constant weight equidistant codes over the alphabet of three, four and five elements," *J. Intell. Inf. Manag.*, vol. 2, pp. 183–187 (2010).

Received January, 2024