

A new class of traceability schemes

Elena Egorova¹, Marcel Fernandez², and Grigory Kabatiansky³

¹ Skolkovo Institute of Science and Technology (Skoltech)
egorvahelene@gmail.com

² Universitat Politècnica de Catalunya, Barcelona, Spain
marcel@entel.upc.edu

³ Skoltech
g.kabatiansky@skoltech.ru

Abstract. A new class of tracing traitors schemes with traceability property which combines ideas of nonbinary IPP-codes and IPP set systems is proposed. A detailed comparison of the proposed scheme with previously known traceability schemes is provided.

Keywords: IPP codes; IPP set systems; tracing traitors; traceability; perfect secret sharing schemes; constant-weight codes

1 Introduction

A modern statement of a fingerprinting problem for digital content was first stated in [1]. Ten years later Chor, Fiat and Naor introduced and developed in [2] a model of digital fingerprinting in the frame of broadcast encryption. A distributor has some digital content to broadcast and sells the access (decoder) to this content. To prevent unauthorized users from accessing the data, the distributor encrypts the data blocks with session keys and gives each authorized user the corresponding personal decoder, i.e., the personal set of keys to decipher data. The main problem of such type of data distribution is to make it collusion resistant, i.e., for a given unauthorized decoder (pirate version), the distributor should be able to identify at least one of the sources of the leakage even if this unauthorized copy was produced by a group of malicious users.

Indeed, in order to hide their identities, some authorized users can form a group (coalition of traitors) and, basing on their common knowledge (keys/decoders), create a forged decoder. Assuming that the cardinality of a possible coalition is not greater than some integer t , the desired property is that once a forged decoder is observed, the distributor can trace it back to at least one traitor from the corresponding malicious coalition.

The problem of data protection against such collusion attacks has given rise to the well known concept of *tracing traitors* [2], and its two particular cases known as *codes with the identifiable parent property (IPP codes)* [3] and *set systems with the identifiable parent property (IPP set systems)* [5], [4]. The IPP

codes were extensively studied, see e.g. [6], [7], [8], also a detailed overview can be found in [9]. As for the IPP set systems, it started with the papers [5], [4] and the most recent results can be found in [10], [11].

Note that the original idea of [2] was based on usage of the perfect secret sharing schemes (SSS, for short), which were discovered in [12] [13]. From this point of view t -IPP codes are based on the simplest (n, n) -threshold SSS. Extension of this idea to arbitrary (w, n) -threshold SSS was proposed in [4] [5] under the name of family of IPP schemes or, equivalently, IPP set systems, and further developed in [11], [10]. A new class of tracing traitors scheme, called q -ary (or “*colored*”) IPP set system, has been recently proposed in [14], [15]. This new class contains nonbinary IPP-codes and IPP set systems as particular cases.

In this paper we extend the “*colored*” approach of [14], [15] to traceability property. In fact, first tracing traitors schemes constructed in [2] have the traceability property, namely, the nearest (in Hamming distance) codevector to the forged vector (decoder) belongs to the malicious coalition. Systematic study of traceability codes has been started from [8], see also [16]-[22]. An original approach to construction of traceability set systems via constant-weight codes was proposed in [23]. Unfortunately there are some mistakes in evaluation of error-correcting codes parameters, which led to wrong results as it was remarked in [24]. The correct version of constructing traceability set systems via *binary* constant-weight codes was recently given in [10]. The current paper could be considered as an extension of the approach and results of [10] from ordinary Johnson scheme to nonbinary Johnson scheme (see [25]) in order to create a new class of traceability schemes which generalizes IPP-codes and IPP set systems with traceability property.

This paper is organized in the following way. In section 2 we describe basic facts about IPP-type schemes, in particular, we explain how they are based on underlying threshold secret sharing schemes. In section 3 we give the definitions of q -ary IPP set systems and q -ary set systems with traceability property and prove a simple sufficient condition for possessing the traceability property. In section 4 we derive an analog of Gilbert-Varshamov bound for traceability schemes (codes). Finally in section 5 we make a comparison with previously known IPP-type schemes and give a concluding remark.

2 Threshold Secret Sharing Schemes as a base for IPP-type schemes

Consider the following broadcasting scenario where the distributor delivers some digital content x to M users. In order to prevent illegal redistribution, the distributor transmits the content x in an encrypted form $y = \varphi(x, k)$ obtained by using some secret key $k \in K$, which serves as a session key and should be changed for distributing another portion of digital content. Firstly the key k is “splitted” into

shares s_1, \dots, s_n according to a chosen Secret Sharing Scheme (SSS). Then each share s_i is transmitted to all users as some blocks of information $e_{i,1}, \dots, e_{i,q}$, where $e_{i,l} = \psi(s_i, f_{i,l})$, ψ is some encryption map, $\mathcal{F}^i = \{f_{i,1}, \dots, f_{i,q}\}$ is the set of corresponding encryption keys, $l = 1, \dots, q$ and $q \geq 1$ - integer number. As the result of encryption the distributor has $N = nq$ encrypted shares $\{e_{11}, \dots, e_{1q}, \dots, e_{n1}, \dots, e_{nq}\}$ which are transmitted along with y , i.e. together with the encrypted digital content x . The j -th user receives (during the initialization phase) a decoder consisting of the corresponding set of decryption keys \mathcal{D}_j , which allows the user to find k and hence to reveal x . A secret sharing scheme is called a *perfect* (w, n) -*threshold* secret sharing scheme if any w users of n can recover the secret and any less number of users gains from their shares no a posteriori information about the secret.

Below we make this definition more precise for two particular and most popular tracing traitors schemes. Let us make this description more precise for two particular types of tracing traitors schemes, namely t -IPP codes and t -IPP set systems.

IPP codes are based on the simplest (n, n) -threshold SSS. We assume that the session key k belongs to a q -ary alphabet, which can be considered as an Abelian group G of size q , for instance, the group \mathbf{Z}_q of residues modulo q . The corresponding shares s_1, \dots, s_n are random uniformly distributed variables on G with the following property

$$s_1 + \dots + s_n = k \tag{1}$$

where summation is taken in the group G . The distributor encrypts every share s_i on q different keys from the set \mathcal{F}^i . Denote by \mathcal{D}^i the set of the corresponding decryption keys and enumerate them by symbols of q -ary alphabet. The set \mathcal{D}_j of decryption keys for j -th user should contain one key from every \mathcal{D}^i , since then the user can find each share s_i for $i = 1, \dots, n$ and hence recover the key k by (1). Let $c_i^{(j)} = \mathcal{D}^i \cap \mathcal{D}_j$ be the corresponding element of the q -ary alphabet. Consider q -ary code $C = \{c^{(1)}, \dots, c^{(M)}\}$ which we call a *fingerprinting code*, where $c^{(j)} = (c_1^{(j)}, \dots, c_n^{(j)})$ is an n -dimensional q -ary vector corresponding to the j -th user.

If a coalition of malicious users (traitors) $U \subset \{1, \dots, M\}$ wants to create a “device” (“decoder”) which will be able to decrypt every transmitted encrypted portion of digital content, then the coalition have to create a new set $\mathcal{Y} = \{y_1, \dots, y_n\}$ of decryption keys with the property that $y_i \in \mathcal{D}^i$ for all $i \in \{1, 2, \dots, n\}$ and $y_i \in P_i(U)$, where $P_i(U) = \cup_{u \in U} c_i^{(u)}$, i.e., the coalition can choose keys only from the set of the coalition’s keys. Hence, the resulting problem can be formulated in the language of codes as it was done in [3]. Namely, denote by $\langle U \rangle$ the set of all false fingerprints (also called descendants [3]) that the coalition U can create, i.e.,

$$\langle U \rangle = \{\mathbf{x} = (x_1, \dots, x_n) \in GF_q^n : \forall j x_j \in P_j(U)\} \tag{2}$$

Let for a fingerprinting code C denote by U a coalition of traitors as well as the set of codevectors corresponding to them.

Definition 1. [3]. A code C has the identifiable parent property of order t , or C is t -IPP code for short, if for all $z \in GF(q)^n$ either

$$\bigcap_{U: \mathbf{z} \in \langle U \rangle, |U| \leq t} U \neq \emptyset \quad (3)$$

or there is no coalition that can produce z .

Hence, if the fingerprints form a code possessing the *Identifiable Parent Property (IPP)*, then from any false fingerprint \mathbf{z} created by a coalition U , $|U| \leq t$, at least one user from U will be identified without any doubt.

It is easy to see that nontrivial q -ary t -IPP codes do not exist for $t \geq q$. In particular, there are no nontrivial binary t -IPP codes even for coalitions of size 2. On the other hand, for $t < q$ there exist families of t -IPP codes with non-vanishing rate, i.e., with a number of codewords growing exponential in n , see [6, 7].

The above described tracing traitors scheme based on the simplest n -out-of- n threshold perfect secret sharing scheme. General case of w -out-of- n threshold perfect SSS ([12],[13]) was used for constructing the following tracing traitors scheme, called the t -IPP family of sets [4, 5]. Each share is encrypted on only single key and the dealer distributes to j -th user the corresponding set of decryption keys \mathcal{D}_j , consisting of w keys, what allows the user to recover w shares and hence reveal the secret k and finally decipher the transmitted digital content. A malicious coalition U can create a fraud “decoder” by arranging together at least w different keys \tilde{E}_j which belong to members of U . Thus the set of descendants of the coalition U equals to

$$\langle U \rangle_w = \{B \subset \{1, \dots, n\} : B \subset \bigcup_{u \in U} \mathcal{D}_u, |B| \geq w\} \quad (4)$$

Now the Identifiable Parent Property can be reformulated for a family of sets in the following way.

Definition 2. [4, 5] Family \mathbf{F} of w -subsets of an n -set $\{1, \dots, n\}$ has the *Identifiable Parent Property of order t* , or \mathbf{F} is t -IPP set system for short, if for any set $A \subset \{1, \dots, n\}$ such that $|A| \geq w$ either

$$\bigcap_{U: A \in \langle U \rangle_w, |U| \leq t} U \neq \emptyset, \quad (5)$$

or there is no U such that $|U| \leq t$ and $A \in \langle U \rangle_w$

Informally, a family \mathbf{F} of w -subsets of a n -set $\{1, \dots, n\}$ is a t -IPP set system if for any w -subset which belongs to the union of some t sets of \mathbf{F} at least one of these sets can be uniquely determined. In particular, it means that no one set of \mathbf{F} belongs to the union of t other sets of \mathbf{F} . Such families of sets are called nowadays as cover-free families, thanks to [26], but they appeared in coding theory twenty years before as superimposed codes [27].

Now let us define a new, more general type of IPP schemes introduced in [14],[15]. Consider w -out-of- n SSS. In the original model of IPP set systems each share is encrypted using only one key, i.e. $q = 1$ and each user receives the w -subset of such keys. In order to define a q -ary (q -“colored”) IPP set system let us encrypt each share via q keys. Then the set \mathcal{D}_j of decryption keys for j -th user should contain w (at least) keys from the union of all \mathcal{D}^i , since then the user can find w shares and hence can recover the key k . It leads to the following definitions [14],[15].

Let $P_i^*(U) := P_i(U) \cup \{0\}$. A coalition U can produce the following set of forged decoders (vectors):

$$\langle U \rangle_w^* := \{y \in P_1^*(U) \times \dots \times P_n^*(U) \mid wt(y) \geq w\}$$

Definition 3. [14],[15].

A $(q + 1)$ -ary code $C \subset \{0, 1, \dots, q\}^n$ of constant weight w has q -ary IPP property of order t , or, for short, C is q -ary t -IPP set system, if for any $y \in \{0, 1, \dots, q\}^n$ and $wt(y) \geq w$ either

$$\bigcap_{U: |U| \leq t, y \in \langle U \rangle_w^*} U \neq \emptyset, \tag{6}$$

or there is no U such that $|U| \leq t$ and $y \in \langle U \rangle_w^*$.

Remark 1. For $q = 1$ this definition coincides with the definition of ordinary t -IPP set systems, and for $w = n$ coincides with the definition of q -ary t -IPP codes (symbol 0 disappears because of the demand that a forgery vector \mathbf{y} should have Hamming weight at least n).

3 “Colored” IPP set systems with traceability

A notion of traceability was firstly considered in [2] in order to construct traitor tracing schemes (in the context of IPP codes), especially with efficient tracing algorithm. The traceability property is a particular case of IPP property that makes the search of malicious users easier. Traceability makes IPP property stronger in the following way: given the forged vector (“decoder”) the distributor can find a malicious user by searching for the closest (in some “metric”, to be precised later) decoder (vector).

Let us give formal definitions in a historical order, i.e., we start with IPP codes.

Definition 4. A q -ary code C is said to have t -traceability property if for any coalition $U \subset C$, $|U| \leq t$, any $y \in \langle U \rangle$ and any $v \in C \setminus U$, it holds

$$d_H(y, v) > \min_{u \in U} d_H(y, u),$$

where $d_H(x, y) = |\{i \mid x_i \neq y_i\}|$ is the Hamming distance.

This definition means that for a given a forged vector (decoder) y the distributor calculates the distance between y and all code vectors and then takes as a traitor the vector that deliver the minimum (of the Hamming distance). It was shown in [2] that if the minimal distance $d_H(C)$ of a code C is bigger than $n(1 - t^{-2})$ then C has the t -traceability property.

For the IPP set systems there is the following analogous definition:

Definition 5. A family $\mathbf{F} = \{F_1, \dots, F_M\}$ of w -subsets of $\{1, \dots, n\}$ is called a t -traceability set system (t -TSS) if for any coalition $U \subset \mathbf{F}$, $|U| \leq t$ and any $S \in \langle U \rangle_{set}$, it holds $|S \cap F| < \max_{u \in U} |S \cap u|$ for any $F \in \mathbf{F} \setminus U$.

In this case t -traceability property means that the search of malicious user(s) reduces to the search of “closest” sets which in this case means the maximum cardinality of intersection, or, since we deal with constant weight code, the minimal Hamming distance.

The notion of t -TSS was introduced in [4] and further studied in [5], [11]. These papers provide results about upper bounds as well as about lower bounds on the size of traceability set systems. The most recent result concerning lower bound can be found in [10].

In order to formulate the traceability concept for the new type of tracing traitors schemes, i.e., q -ary IPP set systems, we need the following notion. Define function $s(a, b)$ in the following way: $s(a, b) = 1$ if $a = b \neq 0$ and $s(a, b) = 0$ otherwise. And define

$$S(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n s(x_i, y_i) = |\{i \mid x_i = y_i \neq 0\}|,$$

i.e., $S(\mathbf{x}, \mathbf{y})$ is the number of coinciding *non-zero* coordinates.

Then, the traceability property can be formulated as follows:

Definition 6. A $(q + 1)$ -ary code C is said to have t -traceability property of order t if for any coalition $U \subset C$, $|U| \leq t$ and any $y \in \langle U \rangle_w^*$, it holds $S(y, v) < \max_{u \in U} S(y, u)$ for any $v \in C \setminus U$.

The following lemma establishes a sufficient condition on a q -ary set system to have t -traseability property which is similar to the original approach of [2]:

Lemma 1. If for a $(q+1)$ -ary code C holds that $S(u, v) < w/t^2$ for any $u, v \in C$, then C has t -traceability property of order t .

Proof. Consider any coalition $U \subset C$, $|U| \leq t$ and any $y \in \langle U \rangle_w^*$. Then, $\max_{u \in U} S(u, y) \geq w/t$ since $wt(y) \geq w$. On the other hand, for any $v \in C \setminus U$,

$$S(v, y) < \sum_{u \in U} S(v, u) < t \cdot \frac{w}{t^2} = \frac{w}{t}.$$

Proved lemma gives us a hint how to estimate the maximum number of codewords in a code with t -traceability property. In next section we will establish the Gilbert-Varshamov type of bound for such estimation.

4 Gilbert-Varshamov bound for nonbinary IPP set systems

Let $A(n, q, w, d)$ denote the maximum number of codewords in a $(q + 1)$ -ary constant weight w code C of length n with $S(x, y) < d$ for any $x, y \in C$. To establish the lower bound for $A(n, q, w, d)$ we propose to derive Gilbert-Varshamov type bound. Define the ‘‘ball’’ $B_v(n, d, w)$ with center at v as the all constant weight vectors $B_v(n, d, w) = \{x : S(x, v) \geq d, wt(x) = w\}$. Then the standard Gilbert-type arguments show that

$$\begin{aligned} A(n, q, w, d) &\geq \frac{\binom{n}{w} q^w}{|B_v(n, d, w)|} = \\ &= \frac{\binom{n}{w} q^w}{\sum_{s,u} \binom{w}{s} \binom{w-s}{u} \binom{n-w}{w-(s+u)} (q-1)^u q^{w-s-u}} \geq \\ &\geq \frac{\binom{n}{w} q^w}{n^2 \max_{s,u} \left[\binom{w}{s} \binom{w-s}{u} \binom{n-w}{w-(s+u)} (q-1)^u q^{w-s-u} \right]} \end{aligned}$$

To apply Lemma we set $d = nt^{-2}$ and hence we are interested in the following value

$$M(n, q, t) = \max_w \min_{s,u} \frac{1}{n^2} \frac{\binom{n}{w} q^w}{\binom{w}{s} \binom{w-s}{u} \binom{n-w}{w-(s+u)} (q-1)^u q^{w-s-u}} \quad (7)$$

where $q, t \geq 2$ -integers, $w > \frac{n}{t^2}$, $s, u \geq 0$ and $s + u \leq w$. For fixed value w we can find the maximum of

$$G(n, q, t; w) = \binom{w}{s} \binom{w-s}{u} \binom{n-w}{w-(s+u)} (q-1)^u q^{-s-u}$$

or, asymptotically it means

$$\begin{aligned} \max_{s,u} xH(y) + x(1-y)H\left(\frac{z}{1-y}\right) + (1-x)H\left(\frac{x(1-y-z)}{1-x}\right) + \\ + xz \log_2(q-1) - x(y+z) \log_2 q \end{aligned}$$

where $H(x) = -(x \log_2 x + (1-x) \log_2 (1-x))$ is binary entropy function. Let $w = xn, s = yw, u = zw$. For the most interesting case when $q = 2$ (and easy to analyse) we get the following optimization problem

$$R_t = \max_x \min_{y,z} H(x) + x(y+z) - \left(xH(y) + x(1-y)H\left(\frac{z}{1-y}\right) + (1-x)H\left(\frac{x(1-y-z)}{1-x}\right) \right)$$

subject to $x, z \geq 0, y > t^{-2}, q, t$ are integers greater than 1, where R_t is a rate of corresponding code, i.e., $R = n^{-1} \log_2 M(n, 2, t)$.

The corresponding calculations give for $q = 2$ that for $t = 2$

$$R_2 \geq 0.0360178851,$$

which is achieved for $x = 0.1156$, i.e. for $w/n = 0.1156$, and for $t = 3$

$$R_3 \geq 0.0063140344$$

which is achieved for $x = 0.0.048$.

5 How to compare tracing traitors schemes?

In order to compare different tracing traitors schemes we need to return to the origin of this subject, namely to [2], where it was suggested to consider the total number $N = nq$ of transmitted “blocks” containing encrypted shares, i.e., consider N as a “block length” and correspondingly calculate the *effective* code rate as $R^* = N^{-1} \log_2 M$. It means that for q -ary IPP-codes and for q -ary IPP-set systems we need to multiple their ordinary rate on $\frac{\log_2 q}{q}$. In the case of IPP set systems the effective rate equals to the ordinary rate.

Let us compare numerically the new traceability scheme with the known ones in the particular case of coalitions of size 2. Our binary 2-IPP set systems with traceability have the effective rate $R_{2-set}^* = 0.0180$ what outperforms significantly the best known traceability ternary codes with their effective rate $R_3^* = \frac{0.007}{3}$, see [21], and slightly worse than the effective rate of the best known IPP-set systems $R_{1-set}^* = 0.0181$, see [10].

How the effective rate of the best t -IPP systems with traceability behaves for $t \rightarrow \infty$ is the open question and the subject of our future work.

References

1. G. R. Blakley, C. Meadows, and G. Purdy, Fingerprinting long forgiving messages, in Proc. CRYPTO, 1985, pp. 180189.

2. B.Chor, A.Fiat, and M.Naor,“Tracing traitors”. in *Advances in Cryptology-Crypto’94, LNCS*, vol. 839, pp. 480–491, 1994.
3. H. D. Hollmann, J.H. Van Lint, J. P. Linnartz, L. M. Tolhuizen. On codes with the identifiable parent property. *Journal of Combinatorial Theory, Series A*, 82(2), 121-133, 1998.
4. D. R. Stinson, R. Wei. Combinatorial properties and constructions of traceability schemes and frameproof codes, *SIAM Journal on Discrete Mathematics*, 11(1), 41-53, 1998.
5. M. J. Collins. Upper bounds for parent-identifying set systems. *Designs, Codes and Cryptography*, 51(2), 167-173, 2009.
6. A. Barg, G. Cohen, S. Encheva, G. Kabatiansky, G. Zémor.. A hypergraph approach to the identifying parent property: the case of multiple parents. *SIAM Journal on Discrete Mathematics*, 14(3), 423-431, 2001.
7. N. Alon, G. Cohen, M. Krivelevich, S. Litsyn. Generalized hashing and parent-identifying codes. *J. Comb. Theory, Ser. A* 104(1): 207-215, 2003.
8. J. N. Staddon, D. R. Stinson, R. Wei, Combinatorial properties of frameproof and traceability codes, *IEEE Trans. Inform. Theory*, 47, 1042-1049, 2001.
9. S. R. Blackburn. Combinatorial schemes for protecting digital content. *Surveys in combinatorics*, 307, 43-78, 2003.
10. E. Egorova, G. Kabatiansky, Analysis of two tracing traitor schemes via coding theory. *Coding Theory and Applications, LNCS*, vol. 10495, pp. 84-92, 2017.
11. Y. Gu, Y. Miao. Bounds on Traceability Schemes. *IEEE Transactions on Information Theory*, vol. 64, N 5, 3450-3460, 2018.
12. G. R. Blakley. Safeguarding cryptographic keys. *Proceedings of the National Computer Conference* 48: 313-317, 1979.
13. A. Shamir. How to share a secret, *Communications of the ACM* 22 (11): 612-613, 1979.
14. E. Egorova, How to combine IPP codes and IPP sets systems - proof of the concept. 16th International Workshop on Algebraic and Combinatorial Coding Theory, Svetlogorsk,Russia, September 2017.
15. E. Egorova, Colored IPP set systems. 3rd International Conference C2SI-2019 on Codes, Cryptology and Information Security, April 2019 (submitted).
16. T. Lindkvist, J. Lfvenberg, M. Svanstrm, A class of traceability codes, *IEEE Trans. Inform. Theory* 48 (2002) 20942096
17. T. van Trung, S. Martirosyan, On a class of traceability codes, *Des. Codes Cryptogr.* 31 (2004) 125132.
18. M. Fernandez, J. Cotrina, M. Sorario, N. Domingo, A note about the traceability properties of linear codes, in: K.-H. Nam, G. Rhee (Eds.), *Information Security, Cryptology ICISC 2007*, in: *Lecture Notes in Comput. Sci.*, vol. 4817, Springer-Verlag, Berlin, 2007, pp. 251258.
19. H. Jin, M. Blaum, Combinatorial properties for traceability codes using error correcting codes, *IEEE Trans. Inform. Theory* 53 (2007) 804808.
20. G.A. Kabatiansky, Good ternary 2-traceability codes exist, in: *Proc. IEEE Symp. Inform. Theory*, Chicago, IL, 2004, p. 203.
21. G.A. Kabatiansky, Codes for copyright protection: the case of two pirates, *Probl. Inf. Transm.* 41 (2005) 182186.
22. Simon R Blackburn, TuvI Etzion, Siaw-Lynn Ng, Traceablity codes
23. Safavi-Naini ,R and Wang, Y, New results on frame-proof codes and traceability schemes, *IEEE Transactions on Information Theory*, November 2001, 47(7) 3029-3033.

24. J. Lfvenberg and J-A. Larsson, Comments on "New Results on Frame-Proof Codes and Traceability Schemes" IEEE Transactions on Information Theory, December 2010
25. H. Tarnanen, M.J. Aaltonen and J-M Goethals, On the Nonbinary Johnson Scheme, Europ. J. Combinatorics, v. 6, pp. 279-285, 1985
26. Z. Furedi, P. Erdos, P. Frankl. Families of finite sets in which no set is covered by the union of r others, Israel J. Math. 51(1), 79-89, 1985.
27. W. Kautz, R. Singleton. Nonrandom binary superimposed codes, IEEE Transactions on Information Theory, vol. 10, N 4, 363-377, 1964.