Noname manuscript No.
(will be inserted by the editor)

On non-binary traceability set systems

Elena Egorova · Marcel Fernandez · Grigory Kabatiansky

Received: date / Accepted: date

Abstract We introduce non-binary IPP set systems with traceability properties that have IPP codes and binary IPP set systems with traceability capabilities as particular cases. We prove an analogue of the Gilbert-Varshamov bound for such systems.

Keywords IPP schems · IPP codes · IPP set system · constant-weight codes

1 Introduction

Consider a distribution model where a dealer uses a broadcast channel to transmit some digital content to a wide audience. In order to restrict the access to the distributed content only for the authorized users (who paid for the access) the distributor should use broadcast encryption schemes. For the first time such schemes were considered in [14]. In what follows we will be interested in broadcast encryption schemes resistant to the so-called *collusion attacks* [6]. Such type of attacks can be described as follows.

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, consisting of the personal set of keys needed to decrypt the data. Note that different users receive different decoders. Malicious users, who want to resell the access to the distributed content without revealing their identities, can form a group (coalition of traitors) and, based on their

Elena Egorova, Grigory Kabatiansky
Skolkovo Institute of Science and Technology (Skoltech), Moscow region, Russia
E-mail: egorovahelene@gmail.com, g.kabatyansky@skoltech.ru. The work of E.Egorova and G.Kabatiansky has been supported by the RFBR grants 20-07-00652

Marcel Fernandez
Universitat Politècnica de Catalunya, Barcelona, Spain
E-mail: marcel@entel.upc.edu The work of M. Fernández has been supported by the Spanish Government grant TEC2015-68734-R and Catalan Government grant SGR 782

common knowledge (present keys and decoders), create a forged decoder. This type of forgery constitutes the main idea of a collusion attack. So, assuming that the cardinality of a possible coalition is not grater than some integer t, the main problem is to construct such set of decoders (for authorized users) that for a given unauthorized decoder (pirate version), the distributor will be able to identify at least one of the sources of the leakage even if this unauthorized copy was produced by a coalition.

The problem of data protection against such collusion attacks has given rise to the well known concept of *tracing traitors* (TT) [6]. As a base of TT-schemes, in [6] it was proposed to use different types of perfect secret sharing schemes (SSS, for short), which were discovered in [5] [23]. For the moment three main tracing traitor schemes are known. Historically the first scheme is known as *codes with the identifiable parent property (IPP codes)*. Such scheme is based on the simplest (n, n)-threshold SSS and was proposed in [6], then, it was further developed in [17,2,1,24]; interested reader may address to the detailed overviews [3,19,18]. Another known scheme, based on arbitrary (w, n)-threshold SSS, was proposed in [25,7] and is known under the name of *set systems with the identifiable parent property (IPP set systems)*. The most recent results can be found in [11,16,10,15,12]. The generalization of these two schemes was proposed in [9]. It is also based on (w, n)-threshold SSS as for IPP set systems but uses an encryption process similar to one used for IPP codes. In this paper we shall call this generalized scheme as *non-binary IPP set systems*.

In this paper we investigate the particular case of IPP-type schemes, known as *tracing traitors schemes with traceability property* or *traceability schemes*, for short. The main idea of traceability schemes is to create such set of decoders that a malicious user (participant of the coalition) can be found as the "nearest" decoder to the forged one. In fact, the first tracing traitors schemes constructed in [6] have the traceability property, namely, the malicious users can be recovered as the nearest in Hamming metric codevector to the forged vector (decoder). They were further studied in [20,4]. The systematic study of traceability set systems has been started in [25,26]. An original approach to construction of traceability set systems via constant-weight codes was proposed in [22]. Unfortunately there were some mistakes in evaluation of error-correcting codes parameters, which led to wrong results as it was remarked in [21]. The correct version of constructing traceability set systems via *binary* constant-weight codes was given in [11].

The non-binary IPP set systems with traceability constitutes the subject matter of this paper. Our main result is the existence of such schemes with non-vanishing rate. This paper is organized in the following way. In section 2 we propose a short reminder of the basics of non-binary IPP set systems, namely, we show how (w, n)-SSS is incorporated in it and explain the traceability paradigm for such scheme. In section 3 we prove GV-bound for non-binary IPP set systems with traceability. In section 4 we define the effective rate of IPP-schemes what allows to compare different schemes with traceability property. In the conclusion we formulate an open problem.

2 Non-binary IPP set systems

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 $z = \varphi(x, \sigma)$ obtained by using some secret key $\sigma \in K$, which serves as a session key and should be changed for distributing another portion of digital content. Firstly the key σ is matched with the set of shares $s_1, \ldots, s_n$ according to perfect (w, n)-threshold Secret Sharing Scheme [5,23]. Let us recall that a secret sharing scheme is called a *perfect* (w, n)*-threshold* secret sharing scheme if any w shares out of n are enough to recover the secret σ and any less number of shares provides no a posteriori information about the secret. Initially, in [6] authors proposed to use perfect (n, n)-SSS and then encrypt each share on q different keys. Different shares are encrypted on different sets consisting of q keys, i.e., overall there are nq encrypted shares and q encrypted versions of each share. This idea gave rise to a notion of IPP codes [17]. Then, general case of w-out-of-n threshold perfect SSS was used in [25,7] for constructing IPP set systems. For such model each share is encrypted using only a single key. Different shares are encrypted on different keys, so, overall, there are n encrypted shares.

In [9] it was proposed to combine the main ideas of these two schemes. More precisely, it was proposed to use w-out-of-n threshold perfect SSS and encrypt each share on q different keys as it was done in [6]. Formally, the share s_i is encrypted q times on the keys from the set $\mathcal{A}_i = \{\alpha_i^1, ..., \alpha_i^q\}$. Encrypted version of shares are transmitted along with the encrypted content z. During the initial stage (before the transmission) the j-th user receives the set consisting of w decryption keys that are then used to decrypt w shares and, so, to decrypt the secret key σ (according to the chosen SSS). Formally, j-th user receives the subset $\mathcal{D}_j \subset \bigcup_{i\in[n]} \mathcal{A}_i$ consisting of w different keys needed to decrypt w different shares, i.e., $|\mathcal{D}_j| = w$ and $|\mathcal{D}_j \cap \mathcal{A}_i| \leq 1$ for all $i \in [n]$.

In what follows we will move from subset representation of users' decoders to vector representation. Indeed, consider some ordering of keys for each set $\mathcal{A}_i$ and map each key to a symbol of q-ary alphabet $\mathbf{A}_q^* = \{1, 2, ..., q\}$, for example by mapping α_i^k to $k \in \mathbf{A}_q^*$ for all $i \in [n]$. Define also the $(q+1)$-ary alphabet $\mathbf{A}_q = \{0, 1, ..., q\}$. Then, instead of considering the subset $\mathcal{D}_j$ we will consider the corresponding characteristic vector $\mathbf{c}^{(j)} \in \mathbf{A}_q^n$ such that its i-th coordinate $c_i^{(j)} = k$ if $\alpha_i^k \in \mathcal{D}_j$ and $c_i^{(j)} = 0$ if $\mathcal{D}_j \cap \mathcal{A}_i = \emptyset$ (absence of the key for i-th share). Note that the resulting vector $\mathbf{c}^{(j)}$ has exactly w non-zero coordinates, i.e., it has weight $wt(\mathbf{c}^{(j)}) = w$ over $q+1$-ary alphabet $\mathbf{A}_q$.

For this model the collusion attack proceeds in the following way. A malicious coalition $U \subset \mathbf{A}_q^n$ in order to create a working forged "decoder" has to collect at least w different keys that can decrypt w different shares. The participant of the coalition can do so by taking at least w different keys among those keys that belong to them. Thus, the set of all forged decoders that the coalition U can create equals to

$$\langle U\rangle_w = \{\mathbf{y} \in P_1^*(U) \times ... \times P_n^*(U) \ : \ wt(\mathbf{y}) \geq w\}, \tag{1}$$

where

$$P_i^*(U) = \{u_i \ : \ \mathbf{u} \in U\} \cup \{0\} \tag{2}$$

is the i-th "projection" of the coalition U. Informally, it means that the participants of the coalition can take one of the keys among those that they have for any given share. If no one has a key for a particular share, then we assume that they cannot guess the possible key.

2.1 Set systems with identifiable parent property

Now we are ready to formulate the identifiable parent property for such scheme.

Definition 1 [9] A $(q+1)$-ary constant-weight code $C \subset \mathbf{A}_q^n$ of weight w is (t, w, q)-IPP code if for any vector $\mathbf{y} \in \mathbf{A}_q^n$ s.t. $wt(\mathbf{y}) \geq w$ either

$$\bigcap_{U \subset C:\, \mathbf{y} \in \langle U\rangle_w,\ |U| \leq t} U \neq \emptyset, \tag{3}$$

or there is no $U \subset C$ such that $|U| \leq t$ and $\mathbf{y} \in \langle U\rangle_w$.

Such property guarantees that at least one malicious user will be identified correctly. Note that if $w = n$ then the definition 1 transforms to a definition of t-IPP codes [17], and for the case $q = 1$ it transforms to (t, w)-IPP set systems [7].

2.2 Set systems with traceability property

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 "proximity measure" $S(\mathbf{x}, \mathbf{y})$ between two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{A}_q^n$ defined as

$$S(\mathbf{x}, \mathbf{y}) = |\{i \mid \mathbf{x}(i) = \mathbf{y}(i) \neq 0\}|, \tag{4}$$

i.e., $S(\mathbf{x}, \mathbf{y})$ is the number of coinciding *non-zero* coordinates. The function $S(\mathbf{x}, \mathbf{y})$ is obviously related to the Hamming distance $d_H(\mathbf{x}, \mathbf{y})$, namely,

$$d_H(\mathbf{x}, \mathbf{y}) = wt(\mathbf{x}) + wt(\mathbf{y}) - 2S(\mathbf{x}, \mathbf{y}) - J(\mathbf{x}, \mathbf{y}), \tag{5}$$

where $J(\mathbf{x}, \mathbf{y}) = \{l \ : \ x_l \neq 0, y_l \neq 0, x_l \neq y_l\}$.

The traceability property can be formulated as follows.

Definition 2 A $(q+1)$-ary constant weight code $C \subset \mathbf{A}_q^n$ of weight w is called a (t, w, q)-traceability set system ((t, w, q)-TSS code, for short) if for any coalition $U \subset C$, $|U| \leq t$ and any $\mathbf{y} \in \langle U\rangle_w$, it holds

$$S(\mathbf{y}, \mathbf{v}) < \max_{\mathbf{u} \in U} S(\mathbf{y}, \mathbf{u}) \tag{6}$$

for any $\mathbf{v} \in C \setminus U$.

Remark 1 Note that for the case $w = n$ this definition is equivalent to the definition of t-IPP codes with the traceability property. For the case $q = 1$ this definition is equivalent to the definition of t-IPP set systems with traceability property. In the general case the given definition is more convenient than a similar one based on the Hamming distance as we can see from the next lemma.

The following lemma establishes a sufficient condition on a (t, w, q)-set system to have t-traceability property, which is similar to the original approach of [6]:

Lemma 1 *A $(q+1)$-ary constant-weight code $C \subset \mathbf{A}_q^n$ of weight w is a (t, w, q)-TSS code if for any $\mathbf{u}, \mathbf{v} \in C$ it holds*

$$S(\mathbf{u}, \mathbf{v}) < w/t^2. \tag{7}$$

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

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

which concludes the proof.

According to Remark 1, Lemma 1 gives for IPP codes the same results as in [6], namely, a q-ary code C with the minimal code distance $d_H(C) > (1 - t^{-2})n$ has the t-traceability property. As for t-IPP set systems, Lemma 1 coincides with Lemma 61 from [26].

Let $M_q(n, t, w)$ denote the maximal possible cardinality of (t, w, q)-TSS code of length n. Define the lower asymptotic bound on the rate of best (t, w, q)-TSS code as

$$R_t(\omega, q) = \liminf_{n \to \infty} n^{-1} M_q(n, t, \lfloor n\omega \rfloor). \tag{8}$$

We will be interested in the maximal possible rate of q-ary t-IPP set systems with traceability as

$$R_t(q) = \max_{\omega} R_t(\omega, q). \tag{9}$$

In the next section we will establish the Gilbert-Varshamov type bound on the size of (t, w, q)-TSS codes.

3 Gilbert-Varshamov bound for non-binary IPP set systems

Let $L_q(n,w,T)$ denote the maximum possible number of codewords in a $(q+1)$-ary code C of length n and constant weight w with $S(\mathbf{u},\mathbf{v}) < T$ for any $\mathbf{u},\mathbf{v} \in C$. To establish the lower bound for $L_q(n,w,T)$ we employ Gilbert-Varshamov type bound similar to GV-bound for constant weight codes.

Define the "ball" $B_z(n,w,T)$ of radius T with the center at $\mathbf{z}$ as the set of all vectors $\mathbf{x}$ of weight w such that $S(\mathbf{x},\mathbf{z}) \geq T$. Let us denote the "size" of the ball as $B(n,w,T)$ since it is the same for all $\mathbf{z}$ s.t. $wt(\mathbf{z}) = w$. It is easy to see that

$$B(n,w,T) = \sum_{s,u: s\geq T, s+u\leq w} \binom{w}{s}\binom{w-s}{u}\binom{n-w}{w-(s+u)}(q-1)^u q^{w-(s+u)}, \tag{10}$$

where $s = S(\mathbf{x},\mathbf{z})$ and $u = |\{l : x_l \neq 0, z_l \neq 0, x_l \neq z_l\}|$. The standard Gilbert-type arguments show that

$$L_q(n,w,T) \geq \frac{\binom{n}{w}q^w}{B(n,w,T)}, \tag{11}$$

From Lemma 1 and the equation (**??**) we have the following theorem

Theorem 1

$$M_q(n,t,w) \geq \frac{\binom{n}{w}q^w}{B(n,w,wt^{-2})} \tag{12}$$

We shall use the following simple upper bound on the size of the ball $B_z(n,w,T)$

$$B(n,w,T) \leq n^2 \max_{s,u:\ s\geq T, s+u\leq w}\left[\binom{w}{s}\binom{w-s}{u}\binom{n-w}{w-(s+u)}(q-1)^u q^{w-s-u}\right] \tag{13}$$

and the well known approximation of binomial coefficient

$$\binom{n}{k} = 2^{n(H(k/n)+o(1))} \text{ for } k \leq n/2,$$

where $H(x) = -(x\log_2 x + (1-x)\log_2(1-x))$ is the binary entropy function. Then from (12), by substituting $w = \omega n, s = yw, u = zw$, next corollary follows:

Corollary 1

$$R_t(\omega,q) \geq H(\omega) - \max_{y,z:\ y\geq t^{-2}, y+z\leq 1, z\geq 0} F_q(\omega,y,z), \tag{14}$$

where

$$F_q(\omega,y,z) = \omega H(y) + \omega(1-y)H\left(\frac{z}{1-y}\right) + (1-\omega)H\left(\frac{\omega(1-y-z)}{1-\omega}\right) + \\ +\omega z\log_2(q-1) - \omega(y+z)\log_2 q. \tag{15}$$

Remark 2 It is easy to see from (5), (10) and (11) that in the case of t-IPP codes, which corresponds to $w = n$, $s + u = w$, the GV-type bound (12) coincides with the result of [6]. In the case $q = 1$ we have t-IPP set systems and the bound (14) was obtained in [11].

For the next simple case $q = 2$ the optimization problem (14) transforms to

$$R_t(2) = \max_{\omega} \min_{y,z} H(\omega) + \omega(y+z) -$$

$$- \left(\omega H(y) + \omega(1-y) H\left(\frac{z}{1-y}\right) + (1-\omega) H\left(\frac{\omega(1-y-z)}{1-\omega}\right)\right) \qquad (16)$$

subject to $\omega, z \geq 0$, $y \geq t^{-2}, y + z \leq 1$, and t is integer greater than 1. The corresponding numerical optimization gives that for $t = 2$

$$R_2(2) \geq 0.03602,$$

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

$$R_3(2) \geq 0.006314$$

which is achieved for $\omega = 0.048$.

Consider also the case $q = 3$. The corresponding numerical optimization gives that for $t = 2$

$$R_2(3) \geq 0.05369,$$

which is achieved for $\omega = 0.172$, and for $t = 3$

$$R_3(3) \geq 0.00946$$

which is achieved for $\omega = 0.073$. Note, that numerical results for the case $q = 1$, i.e., the case of t-IPP set systems, can be found in [11, 12].

4 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 [6], where it was suggested to consider the total number $N = nq$ of transmitted "blocks" containing encrypted shares, i.e., consider N as the "block length" and correspondingly calculate the *effective* rate of (t, w, q)-TSS code C as

$$R^{\text{eff}} = N^{-1} \log_2 |C|.$$

In the case of IPP set systems ($q = 1$) the effective rate equals to the ordinary code rate, since $q = 1$ and $N = n$.

Define the *maximal possible effective rate* of (t, w, q)-TSS codes as

$$R_t^{\text{eff}} = \max_{q} R_t^{\text{eff}}(q),$$

where $R_t^{\text{eff}}(q) = q^{-1}R_t(q)$.

Let us compare numerically the new traceability scheme with the known ones in the particular case of coalitions of size two and three. For $t = 2$ and $q = 1$ in [11] it was proved that $R_2^{\text{eff}}(q = 1) = 0.0181$, this bound was later improved in [12] using combinatorial methods, and the best known bound for today is $R_2^{\text{eff}}(q = 1) = 0.0219$. For the case $t = 3$ from [12] we have $R_3^{\text{eff}}(q = 1) = 0.00365$. As for the new scheme from (16) we have $R_2^{\text{eff}}(q = 2) = 0.018, R_2^{\text{eff}}(q = 3) = 0.0179$, it can be shown that $R_2^{\text{eff}}(q)$ decreases with the growth of q. If we consider 2-IPP traceability codes ($w = n$) the corresponding effective rate achieves its maximum at $q = 18$ and is equal to 0.0162, and for the case $t = 3$ the maximum is at $q = 43$ and is equal to 0.00301. So, we can conclude that for now the best effective rate R_t^{eff} for $t = 2$ is achieved at $q = 1$ and is equal to 0.0219 which is due to binary 2-IPP set systems with traceability property [12]. The same can be said for the case $t = 3$, the best effective rate is also due to binary 3-IPP set systems with traceability and is equal to 0.00365.

5 Conclusion

In this paper we introduced generalized IPP-schemes with the traceability property that allow to investigate uniformly t-IPP codes and t-IPP set systems with the traceability property as two marginal cases of non-binary IPP set systems.

How the effective rate of the best general t-IPP schemes with traceability behaves for $t \to \infty$ is an open question. It is known that the effective rate of t-IPP set systems with traceability $R_t^{\text{eff}}(q = 1) = t^{-4+o(1)}$. Indeed, it was proved in [16] that t-traceability set systems is a t^2-cover-free family [13], therefore, it follows from the known upper bound on the cardinality of t-cover-free families, see [13], [8], that $R_t(1) = O(t^{-4+o(1)})$. On the other hand, the GV-bound shows that $R_t(1) \geq c_1 t^{-4}$, where $c_1 > 0$ is some constant.

We conjecture that for large t

$$R_t^{\text{eff}} = t^{-4+o(1)}.$$

References

1. Alon, N., Cohen, G., Krivelevich, M., Litsyn, S.: Generalized hashing and parent-identifying codes. Journal of Combinatorial Theory, Series A **104**(1), 207–215 (2003)
2. Barg, A., Cohen, G., Encheva, S., Kabatiansky, G., Zémor, G.: A hypergraph approach to the identifying parent property: the case of multiple parents. SIAM Journal on Discrete Mathematics **14**(3), 423–431 (2001)
3. Blackburn, S.: Combinatorial schemes for protecting digital content. Surveys in combinatorics **307**, 43–78 (2003)
4. Blackburn, S., Etzion, T., Ng, S.L.: Traceability codes. Journal of Combinatorial Theory, Series A **117**(8), 1049–1057 (2010)

5. Blakley, G.R.: Safeguarding cryptographic keys. Proceedings of the national computer conference **48**(313) (1979)
6. Chor, B., Fiat, A., Naor, M.: Tracing traitors. Annual International Cryptology Conference pp. 257–270 (1994)
7. Collins, M.J.: Upper bounds for parent-identifying set systems. Designs, Codes and Cryptography **51**(2), 167–173 (2009)
8. D'yachkov, A., Rykov, V.: Bounds on the length of disjunctive codes. Problemy Peredachi Informatsii **18**(3), 7–13 (1982)
9. Egorova, E.: On generalization of ipp codes and ipp set systems. Problems of Information Transmission **55**(3), 241–253 (2019)
10. Egorova, E., Fernandez M.and Kabatiansky, G.: A construction of traceability set systems with polynomial tracing algorithm. International Symposium on Information Theory, 2019. ISIT 2019. Proceedings. (2019)
11. Egorova, E., Kabatiansky, G.: Analysis of two tracing traitor schemes via coding theory. International Castle Meeting on Coding Theory and Applications pp. 84–92 (2017)
12. Egorova, E., Vorobyev, I.: New lower bound on the rate of traceability set systems. XVI International Symposium Problems of Redundancy in Information and Control Systems pp. 93–98 (2019)
13. Erdös, P., Frankl, P., Füredi, Z.: Families of finite sets in which no set is covered by the union of r others. Israel Journal of Mathematics **51**(1), 79–89 (1985)
14. Fiat, A., Naor, M.: Broadcast encryption. Annual International Cryptology Conference pp. 480–491 (1993)
15. Gu, Y., Cheng, M., Kabatiansky, G., Miao, Y.: Probabilistic existence results for parent-identifying schemes. IEEE Transactions on Information Theory (2019). DOI 10.1109/TIT.2019.2927020
16. Gu, Y., Miao, Y.: Bounds on traceability schemes. IEEE Transactions on Information Theory **64**(5), 3450–3460 (2017)
17. Hollmann, H.D., Van Lint, J., Linnartz, J.P., Tolhuizen, L.: On codes with the identifiable parent property. Journal of Combinatorial Theory, Series A **82**(2), 121–133 (1998)
18. Kabatiansky, G.: On the tracing traitors math. In: International Conference on Codes, Cryptology, and Information Security, pp. 371–380. Springer (2019)
19. Kabatiansky, G.: Traceability codes and their generalizations. Problems of Information Transmission **55**(3), 283–294 (2019)
20. Lindkvist, T., Lofvenberg, J., Svanstrom, M.: A class of traceability codes. IEEE Transactions on Information Theory **48**(7), 2094–2096 (2002)
21. Lofvenberg, J., Larsson, J.: Comments on "new results on frame-proof codes and traceability schemes". IEEE Transactions on Information Theory **56**(11), 5888–5889 (2010)
22. Safavi-Naini, R., Wang, Y.: New results on frame-proof codes and traceability schemes. IEEE Transactions on Information Theory **47**(7), 3029–3033 (2001)
23. Shamir, A.: How to share a secret. Communications of the ACM **22**(11), 612–613 (1979)
24. Staddon, J., Stinson, D., Wei, R.: Combinatorial properties of frameproof and traceability codes. IEEE transactions on information theory **47**(3), 1042–1049 (2001)
25. Stinson, D., Wei, R.: Combinatorial properties and constructions of traceability schemes and frameproof codes. SIAM Journal on Discrete Mathematics **11**(1), 41–53 (1998)
26. Stinson, D., Wei, R.: Key preassigned traceability schemes for broadcast encryption. International Workshop on Selected Areas in Cryptography pp. 144–156 (1998)