A Construction of Traceability Set Systems with Polynomial Tracing Algorithm

Elena Egorova * Marcel Fernandez † Grigory Kabatiansky ‡

***Abstract*—A family F of w-subsets of a finite set X is called** *a set system with the identifiable parent property* **if for any w-subset contained in the union of some t sets, called traitors, of F at least one of these sets can be uniquely determined, i.e.** *traced*. **A** *set system with traceability* **property (TSS, for short) allows to trace at least one traitor by minimal distance decoding of the corresponding binary code, and hence the complexity of tracing procedure is of order $O(M)$, where M is the number of users or the code's cardinality. We propose a new construction of TSS which is based on the old Kautz-Singleton concatenated construction with algebraic-geometry codes as the outer code and Guruswami-Sudan decoding algorithm. The resulting codes (set systems) have exponentially many users (codevectors) M and $polylog(M)$ complexity of code construction and decoding, i.e. tracing traitors. This is the first construction of traceability set systems with such properties.**

I. Introduction

The concept of tracing traitors was introduced in [1] in the context of broadcast encryption. The aim of a tracing traitors scheme is to encrypt data in a way preventing its illegal redistribution. Namely, a distributor encrypts the data blocks with session keys and gives the authorized users personal keys to decrypt them. In order to create unauthorized decryption keys (decoders), some authorized users can form a group (coalition of traitors) and based on their common knowledge (keys/decoders) create a forged key/decoder. Assuming that the cardinality of possible coalition is not greater than t, once a forged key is observed, the distributor should be able to identify at least one traitor from a malicious coalition. Such schemes called tracing traitors schemes as in [1] or schemes with Identifiable Parent Property (IPP schemes, for short) as in [2]. All known such schemes are based on perfect secret sharing schemes (SSS, for short), see [4], [5]. Namely, t-IPP codes introduced in [2] are based on the simplest threshold n-out-of-n SSS.

A general w-out-of-n threshold SSS [4], [5], in which any w (or more) participants can recover the secret key k and any set of less than w participants get no a posteriori information about k, was used for constructing the so-called IPP set systems. Recall, that a family $\mathbf{F} = \{F_1, \ldots, F_M\}$ of w-subsets of a n-set $\{1, ..., n\} = [n]$ is called a t-IPP set system if for any w-subset which belongs to the union of some t (or less) sets of $\mathbf{F}$ at least on of these sets can be uniquely determined, see [6], [7].

Let us call a family of codes C_i of length n_i and cardinality M_i or a family of set systems $\mathbf{F}_i$ of M_i w_i-subsets of a n_i-element set a *good family* if there is a constant $c > 0$ s.t.

$$R_i = n_i^{-1} \log M_i \geq c$$

For a general t-IPP set system, the traitor tracing procedure, i.e., finding at least one of the "involved" sets, has complexity $O(nM^t)$. It was suggested in the original paper [1] to construct IPP codes in which traitor tracing can be done via minimal distance decoding for the corresponding code. This property was called *traceability*. Surely IPP systems with traceability has much smaller complexity, namely of order $O(nM)$. It is proved in [1] that q-ary code has traceability property if its minimal code distance $d > (1 - t^{-2})n$, and hence $q > t^2$ in order to get in this way a good family of IPP codes (an easy consequence of the Plotkin bound). For an IPP set system, which can be considered as a constant weight code of weight w, the analogous condition which guarantees the traceability property is that the minimal code distance $d > 2(1 - t^{-2})w$ [9].

Nevertheless even such smaller complexity is still too big for good families of IPP codes and IPP set systems since their cardinality grows exponentially with the length n and hence the overall decoding complexity is very large for practical applications. A problem of constructing a good family of IPP codes with polynomial complexity, i.e., with the complexity of order $poly(n) = polylog(M)$, was affirmatively solved in [10] and [11]. The main goal of this paper is to construct a good family of IPP set systems with polynomial complexity.

A. Related work

We strongly believe that our work and the works in [3], [8] are different at heart. These papers deal with *collision secure fingerprinting codes*. This name was introduced in [1] as "secret" tracing schemes and later it was coined by Boneh and Shaw [20] under the name "collusion-secure fingerprinting codes". These codes have a probabilistic nature in the sense that the identification of traitors is only performed with high probability. Moreover, a collision secure fingerprinting code is not a single code, but a family of codes from which a particular code is chosen randomly and is not known to traitors.

In our present work we deal with zero error identification, i.e. with "open" tracing traitors schemes (according to the

*Elena Egorova is with Skolkovo Institute of Science and Technology (Skoltech), Russia. Email egorovahelene@gmail.com

† M.Fernandez is with Universitat Politecnica de Catalunya, Barcelona, Spain. Email marcelf@entel.upc.edu

‡G. Kabatiansky is with Skoltech, Russia. Email g.kabatyansky@skoltech.ru.

language of [1]), which later became known as codes with Identifiable Parent Property (IPP). So in this sense they are incomparable. Moreover, set systems can be viewed as a different paradigm since we are not dealing with ordered vectors as fingerprinting marks but with unordered sets. It is only our approach that seems to make both paradigms comparable. Let us note also that, to our knowledge, all known identification algorithms for Tardos codes have "decoding times which are sublinear in the total number of users", i.e. exponential in the code length, and our algorithms have polynomial complexity.

II. Concatenated construction of good IPP set systems with polynomial complexity based on algebraic-geometric codes

We start from some definitions and previous results on IPP set systems.

Definition 1: A family $\mathbf{F} = \{F_1, \ldots, F_M\}$ of w-subsets of $\{1, \ldots, n\}$ is called a (t, w)-*traceability set system* ((t, w)-TSS) if for any coalition $U \subset [M]$, $|U| \leq t$ and any set S s.t. $S \subset \cup_{u \in U} F_u$ and $|S| \geq w$, it holds

$$|S \cap F_j| < \max_{u \in U} |S \cap F_u| \text{ for all } j \ \in [M] \setminus U$$

It will be more convenient to consider instead of subsets F_i their corresponding characteristic vectors c_i, instead of a family $\mathbf{F} = \{F_1, \ldots, F_M\}$ the corresponding constant weight code $C = \{c_1, \ldots, c_M\}$ of weight w and length n. We shall say that a binary vector $a = (a_1, \ldots, a_n)$ is covered by a vector $b = (b_1, \ldots, b_n)$ and denote $a \prec b$ if $a_i \leq b_i$ for all i. Then, Definition 1 can be reformulated as follows

Definition 2: A binary constant weight code C of weight w is called a (t, w)-*traceability code* if for any subset $U \subset C$, $|U| \leq t$ and any vector s s.t. $s \prec \bigvee_{u \in U} u$ and $wt(s) \geq w$, it holds

$$d(s, c) > \min_{u \in U} d(s, u) \text{ for all } c \ \in C \setminus U$$

We need the following statement which can be proved along the line of analogous fact for IPP codes [1]. We prove the following Proposition for completeness.

Proposition 1: [9] A binary constant weight code C of weight w is a (t, w)-*traceability code* if for its minimal code distance $d(C)$ the following inequality holds

$$d(C) > 2(1 - t^{-2})w \tag{1}$$

Proof. Let define for two binary vectors $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ the "intersection" function

$$I(a, b) = |\{i : a_i = b_i = 1\}|$$

Then the Hamming distance

$$d(a, b) = wt(a) + wt(b) - 2I(a, b),$$

where $wt(a) = |\{i : a_i = 1\}|$ is the Hamming weight of a. Hence the equation (1) is equivalent to $I(C) < t^{-2}w$, where $I(C) = \max_{c \neq c' \in C} I(c, c')$. Consider any subset (coalition) $U \subset C$, $|U| \leq t$ and any vector s s.t. $s \prec \bigvee_{u \in U} u$ and $wt(s) \geq w$. Then $\sum_{u \in U} I(u, s) \geq wt(s)$ and hence

$$\max_{u \in U} I(u, s) \geq \frac{wt(s)}{t} \geq \frac{w}{t}$$

Since $s \prec \bigvee_{u \in U} u$ then for any $c \in C \setminus U$ one has that $I(c, s) \leq \sum_{u \in U} I(u, c) < t \times w/t^2 = w/t$. ∎

Note, that constant weight codes satisfying (1) possess a stronger property than the traceability property. Namely, consider for a given vector s generated by a coalition U, i.e., $s \prec \bigvee_{u \in U} u$ and $wt(s) \geq w$, the following list of codewords

$$L(s) = \{c \in C : \ I(s, c) \geq wt^{-1}\}$$

Then it follows from the proof that the list is nonempty and belongs to U.

A. KS-EZ construction

To construct a family of good (t, w)-traceability codes with polynomial complexity of encoding and decoding (tracing traitors) procedures we employ the same idea as in Ericson and Zinoviev paper [12] where constant weight codes asymptotically better than the corresponding GV-bound were constructed based on algebraic-geometry codes [13] as outer codes and the old concatenation construction of Kautz and Singleton paper [14].

Consider a $Q = q^2 = p^{2m}$-ary one-point algebraic-geometry (AG) code W, see [15], of length N, dimension K, code rate $R = K/N$ and the minimal code distance

$$D = N(1 - R - (q - 1)^{-1} + o(1)) \tag{2}$$

We choose the following binary code C of length Q and cardinality Q, which consists of codevectors $a^1, \ldots, a^Q$ each of the Hamming weight 1, where the vector a^i has the only one in the i-th coordinate and other coordinates equal to zero. Let us denote the elements of the finite field $GF(Q) = \{\alpha_1, \ldots, \alpha_Q\}$. Now, construct the concatenated code V by replacing coordinates of a code vector of AG code on the corresponding binary code vectors of the code C. The resulting *binary* code V has the following parameters: length $n = NQ$, distance $d(V) = 2D$, cardinality $M = Q^K$ and hence its rate is

$$R(V) = n^{-1} \log_2 M = R \frac{\log_2 Q}{Q}$$

Moreover, the code V is a constant weight code of weight $w = N$.

Theorem 1: For any rate

$$\mathcal{R} \leq R(t) = \frac{1 + \log_2 t}{8(t^3 + t)^2} \tag{3}$$

there exists a family of good (t, w)-traceability codes of rate at least $\mathcal{R}$.

Proof. By (2) and Proposition 1 the code V is a (t, w)-traceability code for all enough large N if

$$1 - R - (q - 1)^{-1} > 1 - t^{-2} \tag{4}$$

Let $R=\frac{1}{2t^2}$ and q be the minimal prime number or power of prime s.t. $q\geq 2t^2+2$. Then

$$1-R-(q-1)^{-1}\geq 1-(\frac{1}{2t^2}+\frac{1}{2t^2+1})=1-t^{-2}+\frac{1}{2t^2(2t^2+1)}$$

and hence the inequality (4) holds for all N such that $o(1)$ in (2) is smaller than $(2t^2(2t^2+1))^{-1}$. Now, let us estimate the rate $R(V)$ of the code V. By Bertrand's postulate $q<2(2t^2+1)$, and, hence, $q<4(t^2+1)$. Therefore

$$R(V)=R\frac{\log_2 Q}{Q}=\frac{1}{2t^2}\frac{\log_2 q^2}{q^2}>\frac{2+\log_2(t^2+1)}{16(t^3+t)^2}>R(t)$$

■

B. Decoding (tracing) for KS-EZ construction

In this subsection we design tracing (decoding) algorithms with decoding complexity *polynomial* in the code length for the above constructed asymptotically good (t,w)-traceability codes. Traditional decoding algorithms realizing bounded distance decoding of a concatenated code, like Forney's algorithm [16], do not fit as we need to correct the number of errors much higher than the half of the code distance (moreover, we need to correct the number of errors bigger than half of the code length). This takes us to the concept of *list decoding* [17], [19]. Instead of trying to deliver a single codeword, a list decoder outputs a list of all codewords within distance larger than half of the code distance of the received word, thus offering a potential way to recover from errors beyond the error correction bound of the code.

In *hard-decision* decoding the *decoder* estimates the sent codeword symbols from the received word symbols. On the other hand, *soft-decision* decoding applies to the cases where the decoding process takes advantage of "side information" generated by the receiver and instead of using the received word symbols, the decoder uses probabilistic reliability information about these received symbols.

Hard decision and soft decision list decoding algorithms for AG codes are given in [17], [19]. The idea of the hard decision decoding algorithm is to interpolate and find a bivariate polynomial. The factorization this polynomial gives the estimated sent codewords. In the soft decision decoding case the polynomial is forced to pass through i different points a different number of r_i times, where this r_i is related to the soft information gien by the decoder. See equation (5) below.

We shall use Guruswami-Sudan list decoding algorithm [17] in a way somewhat similar as it was employed in [18] for the family of good digital fingerprinting codes, see also [11] where an analogous tracing algorithm was proposed for IPP codes. Recall that for arbitrary non-negative integer "weights" r_{ij} assigned to symbols of $GF(Q)$ the Guruswami-Sudan (GS) list soft-decoding algorithm [17], [19] returns all codewords $\mathbf{w}\in W$ that satisfy the following inequality

$$r(\mathbf{w}):=\sum_{i=1}^{N} r_{i,w_i}\geq\sqrt{(N-D)\sum_{i,j} r_{ij}^2}, \qquad (5)$$

Let $U\subset V$ be a coalition of at most t users and let $U^Q=\{u=(u_1^Q,\dots,u_N^Q\}\subset W$ be the corresponding set of Q-ary vectors. Denote by $U_i^Q=\{u_i^Q: u^Q\in U^Q\}$ the i-th projection of the set U^Q.

Consider a binary vector

$$s=(s_{11},\dots,s_{1Q},s_{21},\dots,s_{2Q},\dots,s_{N1},\dots,s_{NQ})$$

generated by a coalition U, i.e., $s\prec\bigvee_{u\in U}u$ and $wt(s)=w'\geq w$.

Substitute to each subblock $s^{(i)}=(s_{i1},\dots,s_{iQ})$ the corresponding set $H_i=\{\alpha_j: j\in supp(s^{(i)})\}\subset U_i^Q\subset GF(Q)$. Let us define $r_{i,j}:=1$ if $\alpha_j\in H_i$ and zero otherwise. Then

$$r(\mathbf{w}):=\sum_{i=1}^{N} r_{i,w_i}=|\{i: w_i\in H_i\}|\leq|\{i: w_i\in U_i^Q\}|$$

since $H_i\subset U_i^Q$. For the choosen values of $R=(2t^2)^{-1}$ and Q the minimal code distance D of the outer code W is at least $N(1-t^{-2})$ for enough large N. Then, on the one hand, for any $\mathbf{w}\notin U$

$$r(\mathbf{w})\leq\sum_{u\in U}|\{i: w_i=u_i\}|<t\times N/t^2=N/t \qquad (6)$$

On the other hand,

$$\sum_{u\in U} r(\mathbf{u})\geq\sum_{i=1}^{N}|H_i|=wt(s)=w'=\lambda N, \qquad (7)$$

where $1\leq\lambda\leq t$, and hence

$$\max_{u\in U} r(\mathbf{u})\geq t^{-1}N\lambda$$

It follows from (5) that the GS decoding algorithm outputs the following list

$$L=\{\mathbf{w}\in W: r(\mathbf{w})\geq t^{-1}N\sqrt{\lambda}\},$$

which does not contain any $\mathbf{w}\notin U$, see (6), i.e., $L\subset U$, and L is non empty because of (7).

Remark: Our contribution presents a "structured" model of set systems just so we can apply well established ideas of coding theory in the decoding process, in particular list decoding techniques. In [21] hard decision list decoding is used as the underlying routine in the tracing process.

In our case we have to deal with a concatenated construction that demands to use soft-decision list decoding with the challenge of having to assign the appropriate weights to each symbol so as to be able to discern between guilty and innocent users.

III. Examples

Consider KS-EZ construction for small t.

For $t=2$ consider an AG-code W over $GF(64)$, i.e., $q=8$, and with rate $R<\frac{1}{4}-\frac{1}{7}=\frac{3}{28}$. Then according to (2) and Proposition 1 the corresponding concatenated code C is a 2-traceability code with total rate $R(C)=R\frac{\log_2 64}{64}$. Hence

for any rate $\mathcal{R} < \frac{3}{28}\frac{6}{64} = 0.010...$ there exists a family of 2-traceability codes with rate $\mathcal{R}$ and a tracing algorithm of polynomial complexity. For comparison the best known 2-traceability codes have rate 0.018 [9] but their tracing complexity is exponential in the code length.

For $t = 3$ consider an AG-code W over $GF(256)$, i.e., $q = 16$, and with rate $R < \frac{1}{9} - \frac{1}{15} = \frac{2}{45}$. Then according to (2) and Proposition 1 the corresponding concatenated code C is a 3-traceability code with total rate $R(C) = R\frac{\log_2 256}{256}$. Hence for any rate $\mathcal{R} < \frac{2}{45}\frac{1}{32} \approx 0.0014$ there exists a family of 3-traceability codes with rate $\mathcal{R}$ and tracing algorithm of polynomial complexity. For comparison the best known 3-traceability codes have rate 0.003 [9] but their tracing complexity is exponential in the code length.

IV. Conclusion

For the first time, for any fixed number t of traitors, we constructed tracing traitor schemes of set systems type with non-vanishing rate and polynomial complexity in the tracing process. One more advantage of such schemes is that they are *binary* as opposed to IPP-codes that have to be nonbinary [2].

Acknowledgments

The work of M. Fernandez has been supported by the Spanish Government Grant TEC2015-68734-R (MINECO/FEDER) "ANFORA" and Catalan Government Grant 2017 SGR 782. The work of E. Egorova and G. Kabatiansky has been supported by the RFBR Grants 18-07-01427.

References

[1] B.Chor, A.Fiat, and M.Naor. "Tracing traitors". *Advances in Cryptology-Crypto'94, LNCS*, 839, pp. 480–491, 1994.
[2] H. D. L. Hollmann, J. H. van Lint, J.-P. Linnartz, and L. M. G. M. Tolhuizen, "On codes with the Identifiable Parent Property," *J. Combinatorial Theory, Ser. A*, vol. 82, no. 2, pp. 121–133, May 1998.
[3] G. Tardos. "Optimal Probabilistic Fingerprint Codes". *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing, ACM*, pp. 116–125, 2003.
[4] G. R. Blakley. Safeguarding cryptographic keys. Proceedings of the National Computer Conference 48: 313-317, 1979.
[5] A. Shamir. How to share a secret, Communications of the ACM 22 (11), 612-613, 1979.
[6] 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.
[7] M. J. Collins. Upper bounds for parent-identifying set systems. Designs, Codes and Cryptography, 51(2), 167-173, 2009.
[8] K. Nuida, S. Fujitsu, M. Hagiwara, T. Kitagawa, H. Watanabe, K. Ogawa, H. Imai. An improvement of discrete Tardos fingerprinting codes. Designs, Codes and Cryptography, 52(3), 339-362, 2009.
[9] Egorova, E., Kabatiansky, G. Analysis of two tracing traitor schemes via coding theory. Coding Theory and Applications, *LNCS*, vol. 10495, pp. 84-92, 2017.
[10] A. Silverberg ; J. Staddon ; J.L. Walker, Applications of list decoding to tracing traitors, IEEE Trans. Information Theory vol. 49, no 5, pp 1312-1318, 2003.
[11] A.Barg and G. Kabatiansky, Class of i.p.p codes with effective tracing algorithm , Journal of Complexity, vol. 20, no 2-3, pp.137-147, 2004.
[12] T.H.E. Ericson, V.A. Zinoviev. An improvement of the Gilbert bound for constant weight codes. IEEE Trans. Information Theory vol. 33, no 5, pp 721-723, 1987.
[13] M.A. Tsfasman, S.G. Vladuts, T. Zink. Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound, Mathematische Nachrichten 109 (1), 21-28, 1982.
[14] W. Kautz, R. Singleton. Nonrandom binary superimposed codes, IEEE Transactions on Information Theory 10(4), 363-377, 1964.
[15] M.A. Tsfasman, S.G. Vladuts, T. Zink. Algebraic geometric codes: basic notions, American Mathematical Soc., vol. 139, 2007
[16] G.D. Forney,Jr., *Concatenated Codes*. Cambridge,MA: MITPress, 1966.
[17] V. Guruswami and M. Sudan, "Improved decoding of Reed-Solomon and algebraic-geometry codes", *IEEE Trans. Inform. Theory*, vol. 45, pp. 1757-1767, 1999.
[18] A. Barg, G. R. Blakley, and G. Kabatiansky, "Digital fingerprinting codes: Problem statements, constructions, identification of traitors," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 852–865, 2003.
[19] V. Guruswami, "List decoding of error-correcting codes," Ph.D. dissertation, MIT, Cambridge, MA, 2001.
[20] D. Boneh and J. Shaw, *Collusion-secure fingerprinting for digital data*, IEEE Trans. Inform. Theory vol. 44, no. 5, pp. 1897–1905, 1998.
[21] A. Silverberg, J. Staddon, J.L. Walker, *Applications of list decoding to tracing traitors*, IEEE Trans. Inform. Theory vol. 49, no. 5, pp. 1312–1318, 2003.