DIMACS Technical Report 2002-36
September 2002

A class of i.p.p. codes with efficient identification

by

Alexander Barg[1,2]
Math. Research
Bell Labs, Lucent Technologies
700 Mountain Ave.,
Murray Hil,. NJ 07974, USA
and IPPI RAN, Moscow, Russia

Gregory Kabatiansky[3]
IPPI RAN, Bol'shoj Karetnyj 19, Moscow 101447, Russia

[1]Permanent Member
[2]Supported in part by the Binational Science Foundation (USA-Israel), grant no. 1999099
[3]Research done while visiting DIMACS Center, Rutgers University, Piscataway, NJ, in May 2002

ABSTRACT

Let $\mathcal{C}$ be a code of length n over a q-ary alphabet. An n-word y is called a descendant of a set of t codewords $x^1, \dots, x^t$ if $y_i \in \{x_i^1, \dots, x_i^t\}$ for all $i = 1, \dots, n$. A code is said to have the *t-identifying parent property* (t-i.p.p.) if for any n-word y that is a descendant of at most t parents it is possible to identify at least one of them.

An explicit construction is presented of t-i.p.p. codes of rate bounded away from zero, for which identification can be accomplished with complexity $\mathsf{poly}(n)$.

1 Introduction: IPP and traceability codes

Consider the set of n-words over an alphabet $\mathcal{Q}$ of size q. Let $\mathrm{d}(a,b)$ denote the Hamming distance and $s(a,b)=n-\mathrm{d}(a,b)$ denote the number of agreements between the vectors a and b. A subset $\mathcal{C}$ of $\mathcal{Q}^n$ is called a code, and we write $\mathcal{C}(n,M)$ to denote a code of length n and size M. By $R=R(\mathcal{C})$ we denote the code rate:

$$R=(1/n)\log_q M.$$

For a subset $\mathcal{X}=\{x^1,\dots,x^t\}\subset\mathcal{Q}^n$ define its *envelope* $e(\mathcal{X})$ by

$$e(\mathcal{X})=\{x\in\mathcal{Q}^n : x_i\in\{x_i^1,\dots,x_i^t\}\}.$$

Elements of the envelope $e(\mathcal{X})$ will be called descendants of $\mathcal{X}$; elements of $\mathcal{X}$ will be called parents with respect to any of their descendants.

The union of the envelopes of all the subsets of $\mathcal{C}$ of size $\le t$ is called the *t-envelope* of the code:

$$E_t(\mathcal{C})=\bigcup_{\mathcal{X}\subset\mathcal{C},\,|\mathcal{X}|\le t} e(\mathcal{X}).$$

The parent-descendant relation gives rise to several classes of codes that have recently gained considerable attention in the literature.

Definition. An (n,M) code $\mathcal{C}$ has the *t-identifying parent property* if for any $y\in E_t(\mathcal{C})$,

$$P_{\mathcal{C}}(y):=\bigcap_{\substack{\mathcal{X}\subset\mathcal{C}\\ y\in e(\mathcal{X}),\,|\mathcal{X}|\le t}}\mathcal{X}\neq\emptyset. \qquad (1)$$

For brevity we call such codes t-i.p.p. codes. Informally, given a t-i.p.p. code it is possible for every vector $y\in E_t(\mathcal{C})$ to identify with certainty at least one of its parents.

t-i.p.p. codes and codes for related cryptographic problems received considerable attention in the recent years [1]-[8], [10] (paper [10] introduced i.p.p. codes for the case $t=2$), [13].

For $q\ge 3$, existence of 2-i.p.p. codes with rate R bounded away from 0 for $n\to\infty$ was proved in [10]. For the case of general t, existence of i.p.p. codes of nonzero rate for any $q\ge t+1$ was first proved in [4]. These papers also provided a characterization of 2- and 3-i.p.p. codes, respectively.

Let $R_q(t)$ denote the maximum asymptotically attainable rate of q-ary t-i.p.p. codes. The following results are known about it.

Theorem 1 [4] *Let $u=\lfloor (t/2+1)^2\rfloor$. Then for any $\varepsilon>0$,*

$$R_q(t)\ge\frac{1}{u-1}\log_q\frac{(q-t)!q^u}{(q-t)!q^u-q!(q-t)^{u-t}}-\varepsilon.$$

For $q=4,t=3$ a better estimate $R_4(3)\ge\frac{1}{5}\log_4(1024/1018)-\varepsilon$ is valid.

This result was proved by introducing (t,u) hashing codes (functions) and using the probabilistic method to derive an existence bound on their number. In particular, for $t=2$ Theorem 1 coincides with an earlier result of [10]. Generally, we see that for t-i.p.p. codes to exist it is sufficient that $t\le q-1$. This is also a necessary condition because for small q there are only trivial identifying codes: the maximal size of a t-i.p.p. code, $t\ge q$, is $M=t$.

The probabilistic argument for (t,u) hashing was refined in [1] resulting in the following estimate which improves the result of Theorem 1 for all t except $t=2,3$.

Theorem 2 [1]

$$R_{t+1}(t) \geq \frac{t!(u-t)^{u-t}}{u^u(u-1)\ln(t+1)} - \varepsilon.$$

The procedure of finding a codeword x in an i.p.p. code $\mathcal{C}$ which necessarily is a parent of a given word y, *i.e.,* , $x \in P_{\mathcal{C}}(y)$, is called *identification* (or *tracing traitors*) [7]. It is clear that for any $y \in E_t(\mathcal{C})$, we can identify at least one of its parents by examining all the $\binom{M}{t}$ subsets of $\mathcal{C}$ of size t. For codes of large size this requires prohibitively long computation. Therefore, the following open problem was raised in [13]:

> Given a t-i.p.p. code, traceability can be done in time $O(\binom{M}{t})$. Can this be improved, perhaps, for certain subclasses of t-i.p.p. codes?

Before stating our results, let us isolate the range of parameters where this problem is nontrivial. For that note that there is a subclass of i.p.p. codes, called *t-traceability* codes, for which identification can be performed more efficiently than in the general case. For a code $\mathcal{C}$ and a vector $y \in \mathcal{Q}^n$ let

$$\mathrm{d}(y, \mathcal{C}) = \min_{x \in \mathcal{C}} \mathrm{d}(y, x).$$

For a vector $y \in \mathcal{Q}^n$ let $S_{\mathcal{C}}(y) \subset \mathcal{C}, |S_{\mathcal{C}}(y)| \geq 1$ be the set of the nearest neighbors of y, *i.e.,* , the codewords $x \in C$ such that for every vector $x \in S_{\mathcal{C}}(y)$,

$$\mathrm{d}(y, x) = \mathrm{d}(y, \mathcal{C}).$$

Definition. A t-i.p.p. code is said to have the *t-traceability property* if for every $y \in E_t(\mathcal{C})$, there holds

$$S_{\mathcal{C}}(y) \subset P_{\mathcal{C}}(y), \tag{2}$$

i.e., if x is a nearest codevector to y then it is also one of its parents. Hence in a traceability code a parent of a given word $y \in E_t(\mathcal{C})$ can be identified by finding a closest codeword to y. This requires examining at most M codewords; thus, identification for traceability codes can be performed faster than for i.p.p. codes in general.

One simple sufficient condition for a code to possess the t-traceability property was given in [7].

Proposition 3 [7] *Let $\mathcal{C}$ be a q-ary code with distance d, where $d > n(1 - 1/t^2)$ for some $t > 0$. Then $\mathcal{C}$ is a t-traceability code.*

Proof : Let $\mathcal{X} = \{x^1, \dots, x^t\} \subset \mathcal{C}$ be a t-coalition and let $y \in e(\mathcal{X})$. By the definition of the envelope, for some $x \in \mathcal{X}$,

$$s(x, y) \geq n/t.$$

On the other hand, for any code vector w outside $\mathcal{X}$,

$$s(y, w) \leq \sum_{j=1}^{t} s(w, x^j) = \sum_{j=1}^{t} (n - \mathrm{d}(w, x^j)) < n/t.$$

∎

This proposition together with some known results on error-correcting codes leads to a more accurate formulation of the above problem. Namely, it is clear that a solution to it can be obtained if we can construct a sequence of t-traceability codes equipped with an efficient algorithm of finding a nearest codeword to a vector $y \in \mathcal{Q}^n$ (hence, in particular, to a vector $y \in E_t(\mathcal{C})$). By Prop. 3 such a sequence can be obtained from any family of error-correcting codes with large minimum distance. A natural candidate is q-ary $[N, RN]$ linear algebraic-geometric codes whose relative distance $\delta = \mathrm{d}(\mathcal{C})/N$ can be made arbitrarily close to one for sufficiently large q [15]. Moreover, a polynomial-time list decoding algorithm of [9] enables us to find a nearest codeword x to a given point y provided that $s(y, x) \geq \sqrt{N(N-D)}$. Choosing D sufficiently large, we obtain a sequence of efficient t-i.p.p. codes with the traceability property (After the draft version of this paper [5] was completed, we became aware of the paper [12] that works out the details of this idea.).

However, to construct codes of large size following this approach, we have to employ codes over an alphabet of a fairly large size. Namely, recall that by the Plotkin upper bound [15] for any q-ary (n, M) code with distance d,

$$M \leq \frac{d}{d - \frac{q-1}{q} n}.$$

Substituting $d > n(1 - 1/t^2)$, we see that for $q < t^2$ the denominator is positive, and therefore, the size of the code $M \leq n O_{t,q}(1)$. Thus, for small q it is not possible to construct codes whose rate R remains bounded away from zero as n grows. Therefore, the above problem can be formulated as follows:

> For $t^2 \geq q > t + 1$ construct a sequence of q-ary t-i.p.p. codes with rate $R > 0$ and an identification algorithm of complexity polynomial in the code length n.

A solution to this problem is given in this paper. Our main result is given by the following theorem.

Theorem 4 *For any $q \geq t + 1$ and any rate*

$$0 \leq R < \frac{1}{t^2} R_q(t),$$

there exists a family of q-ary (n, q^{Rn}) t-i.p.p. codes. Construction, fingerprint generation, and identification for codes $\mathcal{C}$ can be accomplished by algorithms of complexity that grows polynomially with the code length n.

2 Proof: A family of t-i.p.p. codes

Recall the notation $\mathcal{W}[N, K, D]$ for a linear code $\mathcal{W}$ of length N, dimension K and distance D. We will construct a family of q-ary codes $\mathcal{C}(n, M)$ by concatenating a q-ary t-i.p.p. code $\mathcal{V}(m, q^{mR(\mathcal{V})})$ and a Q-ary linear $[N, K, D]$ code $\mathcal{W}$, $Q = q^{mR(\mathcal{V})}$. The parameters of the code $\mathcal{C}$ are: length mN, rate $R(\mathcal{C}) = R(\mathcal{V})R(\mathcal{W})$. By carefully choosing the parameters of the constituent codes we will ensure that $R(\mathcal{C}) \geq \text{const} > 0$ for an infinite sequence of values of the code length. The value of q is fixed, independent of the code length. Finally, t can be any number such that $t + 1 \leq q$.

First we recall the concatenated construction. Let us fix an arbitrary one-to-one mapping ϕ between $\mathbb{F}_Q$ and $\mathcal{V}$. The code $\mathcal{C}$ can be viewed as a composite mapping $\mathcal{C} : (\mathbb{F}_Q)^K \to (\mathbb{F}_q)^{nN}$ formed of the following two mappings. First, a K-vector from $\mathbb{F}_Q$ is encoded by the code $\mathcal{W}$ into a vector w in $(\mathbb{F}_Q)^N$. Next, every coordinate of w is "encoded" with the code $\mathcal{V}$, *i.e.,* mapped by ϕ to a vector in $\mathbb{F}_q$. This results into a codeword in $(\mathbb{F}_q)^{Nm}$, which is a word of the concatenated code $\mathcal{C}$. It is convenient to have in mind a representation of this codeword as a matrix with N columns of length m. Clearly, $\mathcal{C}$ is an (n, M) code with $n = mN$, $M = Q^K = q^{mNR(\mathcal{V})R(\mathcal{W})}$, and hence of rate $R = R(\mathcal{V})R(\mathcal{W})$.

Let $t \le q - 1$ be fixed. As the code $\mathcal{V}$ we take a t-i.p.p $(m, q^{mR(\mathcal{V})})$ code, whose existence is proved in Theorems 1, 2. The rate $R(\mathcal{V})$ is chosen in such a way that $Q = q^{mR(\mathcal{V})}$ is an even power of a prime. We also take $\mathcal{W}$ to be a Q-ary one-point algebraic geometry (AG) code with the parameters $[N, K = R(\mathcal{W})N, D = \delta N]$. To obtain as high rate $R(\mathcal{C})$ as possible, we take algebraic-geometric (AG) codes from maximal curves so that their parameters for large N attain the bound [15]

$$\delta + \varepsilon \ge 1 - R(\mathcal{W}) - (\sqrt{Q} - 1)^{-1} \quad (\varepsilon > 0). \tag{3}$$

Let $y \in E_t(\mathcal{C})$ be an n-word, which is represented as an $(m \times N)$ matrix over $\mathbb{F}_q$. Let $y^1, \dots, y^N$ denote the columns of y. Our decoding algorithm is a two-stage procedure, which in the initial stage employs N parallel, independent decodings of the columns in y with the code $\mathcal{V}$, and uses the Guruswami-Sudan (GS) algorithm [9] in the second stage to find a codeword from $P_{\mathcal{C}}(y)$.

Guruswami-Sudan decoding [9]. Let $\mathcal{W}$ be a Q-ary Reed-Solomon or an (one-point) AG code with parameters $[N, RN, D]$. Then for any vector $z \in \mathbb{F}_Q^N$ there exists an algorithm that outputs a subset $Y \subset \mathcal{W}$ formed of all the vectors $w \in \mathcal{W}$ such that the distance $\mathrm{d}(w, z) \le N - \sqrt{N(N - D)}$. The size $|Y|$ of this list is bounded above by a polynomial function of N. The implementation complexity of the algorithm is also polynomial in N.

Relying on t-i.p.p. codes and GS decoding, we will show that under certain conditions, the concatenated code $\mathcal{C}$ admits an efficient identification algorithm described as follows.

Identification Algorithm for the code $\mathcal{C}$.

- For every $i, 1 \le i \le N$ find a vector $v^i \in P_{\mathcal{V}}(y^i)$. Form a vector $z = (z_1, z_2, \dots, z_N) \in \mathbb{F}_Q^N$, where $z_i = \phi^{-1}(v^i)$.
- Decode z with the code $\mathcal{W}$ using the GS algorithm with decoding radius $N(1 - 1/t)$. This step results in a list $Y \subset \mathcal{W}$.
- Take an arbitrary vector w from the list Y at distance at most $N(1 - 1/t)$ from z (we will show that there always exists such a codeword w). Encode w with the map φ as described in the beginning of this section to obtain a vector $x \in \mathcal{C}$. Output x.

Proposition 5 *Consider a code $\mathcal{C}(n, q^{Rn})$ constructed by concatenating a q-ary t-i.p.p code $\mathcal{V}$ of length m and a Q-ary $[N, K, D = N\delta]$ code $\mathcal{W}$, $\delta > 1 - 1/t^2$. The code $\mathcal{C}$ is t-i.p.p., and for every $y \in E_t(\mathcal{C})$ the identification algorithm will output a vector $x \in P_{\mathcal{C}}(y)$.*

Proof : Consider the vector z constructed by the algorithm. We claim that

(A) the sphere of radius $N(1 - 1/t)$ with center at z contains at least one codeword $x \in \mathcal{W}$, and that

(B) all codewords of $\mathcal{W}$ contained in this sphere correspond to parents of y.

Let $\mathcal{X} = \{x^1, \dots x^t\} \subset \mathcal{C}$ be the coalition. Consider an (Nm)-vector $y = (y^1, y^2, \dots, y^N) \in e(\mathcal{X})$, where $y^i = (y^i_1, y^i_2, \dots, y^i_m)$ for $i = 1, 2, \dots, N$. Denote by $w^1, \dots, w^t$ the codewords of $\mathcal{W}$ that correspond to $x^1, \dots, x^t$, respectively. Denote by $v^{i,j} \in \mathcal{V}$ the ith column of the codeword $x^j, 1 \le j \le t$. Then

$$y^i \in e(v^{i,1}, \dots v^{i,t}), \quad 1 \le i \le N.$$

Furthermore, since $\mathcal{V}$ is a t-i.p.p. code, for each $1 \le i \le N$ it is possible to find at least one parent codeword of the column y^i. Let $j(i)$ be a number such that $v^{i,j(i)} \in P_{\mathcal{V}}(y^i)$. Let

$$z = (\phi^{-1}(v^{1,j(1)}), \dots, \phi^{-1}(v^{N,j(N)})) = (w_i^{j(1)}, \dots, w_N^{j(N)}).$$

The next step is to decode z with the code $\mathcal{W}$. Since y is a descendant of the coalition $\mathcal{X}$, for every i there exists a number ℓ such that $z_i = w_i^\ell$. Therefore there exists at least one $\widehat{w} \in \{w^1, \dots, w^t\}$, such that

$$s(z, \widehat{w}) \ge N/t. \tag{4}$$

On the other hand, let us show that for any vector $w \in \mathcal{W} \backslash \{w^1, \dots, w^t\}$, the number of agreements with z will be smaller than N/t. This is done by repeating the proof of Prop. 3. Thereby part (B) of our claim is established.

Finally, let us show that only codewords from the set $\{w^1, \dots, w^t\}$ can fall in the list Y constructed by the GS algorithm. Any member u of this list is at distance $N - \sqrt{N(N-D)}$ or less from z. Since $D > N - N/t^2$,

$$s(u, z) = \sqrt{N(N-D)} > N/t,$$

so by (4) the vector u is one of the vectors $w^1, \dots, w^t$. This proves part (A) and the fact that the Algorithm always produces a vector in $P_{\mathcal{C}}(y)$, and therefore also the fact that $\mathcal{C}$ is t-i.p.p. ∎

The rate of the concatenated code $\mathcal{C}$ equals $R(\mathcal{C}) = R(\mathcal{V})R(\mathcal{W})$. If the inner and outer codes are chosen as described, then it is possible to construct codes whose rate $R(\mathcal{C})$ is

$$\begin{aligned} R(\mathcal{C}) &\ge R_q(t)\Big(1 - \delta - \frac{1}{\sqrt{Q} - 1}\Big) - \alpha \\ &= R_q(t)\Big(\frac{1}{t^2} - \frac{1}{\sqrt{Q} - 1}\Big) - \alpha', \end{aligned}$$

where α, α' are arbitrarily small positive numbers. Next we choose the length m of the inner code $\mathcal{V}$ sufficiently large (but fixed) so that, for a given ε, the term $(\sqrt{Q} - 1)^{-1} R_q(t) + \alpha' \le \varepsilon$. This proves the estimate on the rate $R(\mathcal{C})$ from Theorem 4.

It remains to prove the claims about the complexity of the codes $\mathcal{C}$. The construction, encoding, and decoding of the code $\mathcal{V}$ whose length does not depend on n, are of constant complexity. Therefore, the construction and the encoding complexity of the codes $\mathcal{C}$ up to a constant factor are equal to the construction (resp., encoding) complexity of the sequence of AG codes, which is known to be polynomial (in fact, a very recent result of [11] provides for them a construction algorithm of low complexity $(N \log_Q N)^3$). The complexity of identification is governed by the decoding complexity of the GS algorithm which is known to be polynomial [9]. This concludes the proof of Theorem 4.

3 Tracing traitors

This section is devoted to an application of the above result to the "tracing traitors problem" [7], [8]. Following these papers, we give a short a description of how i.p.p. codes fit into a broader context of data distribution schemes that have to deal with pirate decoders.

A traceability scheme consists of a user initialization scheme, an encryption-decryption scheme, and a traitor tracing, or identification algorithm. The user initialization scheme is a procedure of assigning to each of the M users of the system his personal key u, which is an n-string over an alphabet of a certain size q. The actual data is transmitted over a public channel in an encrypted form. The encrypted message is formed of two parts, the enabling block of length $r = nq$, and the actual data. To access the data, a registered user needs to obtain a secret key s that is encrypted in the enabling block. To encrypt s the distributor chooses a random string $(s_1, s_2, \dots, s_n)$ which is encrypted in the enabling block in such a way that every authorized user can access it using his personal key. Upon retrieving the values $(s_1, s_2, \dots, s_n)$ the secret key is found as $s = \sum s_i \pmod 2$.

The enabling block is an $(n \times m)$ matrix such that the part s_i is encrypted under every element of the ith row. User u accesses one element in each row of the enabling block in the column whose number is given by the ith entry in the user's personal key.

A threat to the integrity of the system is posed by users manufacturing and reselling pirate decoders which can decrypt the secret key. Such a decoder produced by a single user i will contain the keys s_i and therefore can be traced back to the user. A more serious problem arises if we assume that a coalition of t or fewer registered users can cooperate in order to produce a pirate decoder. In this case the decrypting key installed in it may contain parts of the keys of the users from the coalition. The task of the distributor is to find at least one of the members of the coalition. Since every element of the secret key in the pirate decoder belongs to one of the users of the coalition, this key is in the envelope of the secret keys of the coalition members.

A formal definition of traceability schemes can be given in terms of incidence systems of finite sets; see [14] where a generalization of this definition is also suggested. This description leads to a link of traceability schemes with incidence structures such as combinatorial designs.

Paper [8] presents two "open fully t-resilient" tracing traitor schemes, *i.e.,* schemes that will enable the distributor to locate with certainty at least one traitor (and a few more schemes that allow for a nonzero probability of misidentification). The approach to traitor tracing described corresponds to the first of these two schemes (the second scheme is a bit more complicated in the sense that every element of the enabling matrix is replaced with a string of elements).

Traitor tracing schemes are characterized by the key length n and the length r of the enabling block (and, generally, the number of decryptions performed by a user, although in our context it always equals n). Such schemes are optimized to achieve small values of n and r as functions of the number of users M. The parameters of the traitor tracing scheme in [8] are obtained by an application of the probabilistic method. In our notation they are summarized in the following theorem.

Theorem 6 [8] *There exists a t-traceability scheme with $n = 4t^2 \ln M$ and $q = 2t^2$, i.e., $r = 8t^4 \ln M$.*

Proposition 3 together with a relation to i.p.p. codes outlined above implies that any code with a large minimum Hamming distance defines a traceability scheme. Therefore, it makes sense to compare the existence result of Theorem 6 to the parameters of the scheme obtained by choosing a code with minimum distance at least $n(1-1/t^2)$ whose rate attains the Gilbert-Varshamov bound. In this way we obtain a t-traceability scheme with $q = 2t^2$ and $n = 2t^2 \ln M/(2\ln 2 - 1)$, which is only by a factor $(4\ln 2 - 2)^{-1} \approx 1.29$ inferior to Theorem 6.

We will use t-i.p.p. codes to present constructive alternatives to the existence results just cited.

In coding terms the problem can be formulated as follows: *given t and M construct a t-i.p.p. code of size M so that the length n and the value $r = qn$ are as small as possible.* One obvious avenue to follow is again to take a code with a stronger property, a t-traceability code. For instance, let us take algebraic geometry codes with relative distance $\delta > 1 - 1/t^2$ assuming that for large length N the code parameters attain the bound (3). Let $Q = (2t^2)^2$. We obtain:

$$1 - \frac{1}{t^2} < \delta = 1 - R - \frac{1}{2t^2 - 1} - \varepsilon.$$

Omitting small terms, we obtain the following rate: $R = 1/(2t^2)$, or

$$\ln M = \ln(Q^{nR}) = \frac{n}{t^2} \ln 2t^2.$$

Therefore, $n = t^2(\ln M)/(\ln 2t^2)$, and the length of the enabling block equals $r = 4t^6(\ln M)/(\ln 2t^2)$.

Proposition 7 *There is a constructive t-traceability scheme with $n = t^2(\ln M)/(\ln 2t^2)$ and $r = 4t^6(\ln M)/(\ln 2t^2)$ and with an identification algorithm of complexity polynomial in n.*

The comparison of the parameters of this scheme with Theorem 6 shows that from algebraic-geometric traceability codes we obtain a shorter key length but a longer enabling block. Recall that Theorem 6 is only an existence result. The parameters of constructive schemes cited in [8] are $n = O(t^6 \ln M), r = O(t^8 \ln M)$, so Prop. 7 gives a substantial improvement of that result.

Now let us use t-i.p.p. codes constructed in this paper to construct traceability schemes. Let

$$v = \frac{(q-t)!q^u}{(q-t)!q^u - q!(q-t)^{u-t}}.$$

Let us take a code of length n and rate R such that $nR = (\log_q v)/t^2(u-1)$ obtained by combining Theorems 1 and 4. Assuming that $q \gg t$, after a series of simplifications, we obtain

$$n = t^2(u-1)(e^{tu/q} - 1) \ln M.$$

Choosing $q = t^3$, we get for the code length the inequality $n \le 2t^2(u-1)\ln M$ and consequently, $r \le 2t^5(u-1)\ln M$. Again the parameters of this scheme are better than constructive schemes previously known.

4 Concluding remarks

1. *Traceability codes.* As discussed above, traceability codes form a subclass of i.p.p. codes. Similarly to the problem addressed in this paper, one can ask if there exist t-traceability codes with $t^2 \ge q \ge t+1$ and code rate bounded away from zero for growing length n. This problem so far remains open.

2. *Digital fingerprinting.* The technique of this paper can be employed to construct codes for a more difficult problem, that of digital fingerprinting. Codes for digital fingerprinting satisfy the same condition (1) as i.p.p. codes except that the envelope $e(\mathcal{X})$ is formed under a more relaxed rule: if $x_i^1 = x_i^2 = \cdots = x_i^t$, then also y_i takes on the same value; however, if $|\{x_i^1, \dots, x_i^t\}| \geq 2$, then y_i can be any letter of the alphabet $\mathcal{Q}$. This problem was suggested in [6], where it was also observed that unless we allow some probability p_e of identification error, there are no binary fingerprinting codes of rate $R > 0$. Binary fingerprinting codes with positive rate R and $p_e = \exp(-\Omega(n))$ are constructed in [2]. These codes afford a $\mathsf{poly}(n)$-time identification of a vector from the set of parents of a given vector y. The decoding (identification) algorithm of [2] also relies upon list decoding algorithms of one-point AG codes, although some additional arguments are involved to estimate the probability p_e of identification error.

The construction of [2] makes use of a more powerful version of the GS algorithm associated with soft decision, or weighted decoding of the outer codes. Weighted decoding is proposed in the same paper [9] as the version of list decoding (the so-called hard decoding) employed in this paper. The difference between weighted decoding and hard decoding in our context can be briefly explained as follows. In the first step of our identification algorithm, for a given column i, we find an element z_i of the field $\mathbb{F}_Q$ that corresponds to a member of the coalition $\mathcal{X}$. The vector z thus formed is then submitted to the decoding algorithm of the outer code. Note however that to find z_i we first find a t-subset of vectors of the inner code (or of elements of the field) that supposedly forms the coalition. This t-subset may contain more information about the parents of y than just one element. The purpose of weighted decoding is to make use of just this information. In other words, instead of decoding from one vector z we attempt to decode from a subset of vectors of $(\mathbb{F}_Q)^N$ obtained by taking a product of the N t-subsets that correspond to the coordinates $i = 1, \dots, N$.

While in the problem of digital fingerprinting the use of weighted decoding yields an improvement in the rate and error bounds of the codes constructed, it does not seem to improve over hard decision decoding in the problem of identification with i.p.p. codes.

References

[1] N. Alon, G. Cohen, M. Krivelevich, and S. Litsyn, "Generalized hashing and applications to digital fingerprinting," Proc. 2002 IEEE Int. Symposium on Information Theory, Lausanne, Switzerland, p. 436.

[2] A. Barg, G. R. Blakley and G. Kabatiansky, "Digital fingerprinting codes: Problem statements, constructions, identification of traitors," DIMACS Techn. Rep. 2001-52 (http://dimacs.rutgers.edu), submitted for publication.

[3] ———, "On digital fingerprinting codes," Proc. 2001 IEEE Int'l Symposium Inform. Theory, Washington, DC, p. 161.

[4] A. Barg, G. Cohen, S. Encheva, G. Kabatiansky, and G. Zémor, "A hypergraph approach to the identifying parent property: the case of multiple parents," *SIAM J. Discrete Math.*, **14** (2001), pp. 423–431.

[5] A. Barg and G. Kabatiansky, "A class of i.p.p. codes with efficient identification," Proc. 39th Annual Allerton Conf. Commun. Control Comput., Monticello, IL, Oct. 2001, pp. 885–890.

[6] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inform. Theory*, **44** (1998), pp. 1987–1905.

[7] B. Chor, A. Fiat and M. Naor, "Tracing traitors," *Crypto'94* LNCS 839 (1994), pp. 257–270.

[8] B. Chor, A. Fiat, M. Naor, and B. Pinkas, "Tracing traitors," *IEEE Trans. Inform. Theory*, **46** (2000), pp. 893–910.

[9] V. Guruswami and M. Sudan, "Improved decoding of Reed-Solomon and algebraic-geometry codes," *IEEE Trans. Inform. Theory*, **45**, no.6 (1999), pp. 1757–1767.

[10] 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*, Series A, **82** (1998) pp. 121–133.

[11] K. W. Shum, I. Aleshnikov, P. V. Kumar, H. Stichtenoth, and V. Deolalikar, "A low-complexity algorithm for the construction of algebraic-geometric codes better than the Gilbert-Varshamov bound," *IEEE Trans. Inform. Theory*, **47**, no 6 (2001), pp. 2225–2241.

[12] A. Silverberg, J. Staddon, and J. L. Walker, "Efficient traitor tracing algorithm using list decoding," *ASIACRYPT 2001*, Lect. Notes Comput. Sci. 2248 (2001), pp. 175–192.

[13] J. N. Staddon, D. R. Stinson, and R. Wei, "Combinatorial properties of frameproof and traceability codes," *IEEE Trans. Inform. Theory*, **47** (2001), pp. 1042–1049.

[14] D. R. Stinson and R. Wei, "Combinatorial properties and constructions of traceability schemes and frameproof codes," *SIAM J. Discrete Math.*, **11** (1998), pp. 41–53.

[15] M. Tsfasman and S. Vlăduţ, *Algebraic-geometric codes*, Kluwer, Dordrecht, 1991.