Des. Codes. Cryptogr. (2018) 86:785–802
Author's accepted manuscript. The final publication is available at
http://link.springer.com/article/10.1007/s10623-017-0359-z

Constructions of almost secure frameproof codes with applications to fingerprinting schemes

José Moreira · Marcel Fernández · Grigory Kabatiansky

Received: date / Accepted: date

Abstract This paper presents explicit constructions of fingerprinting codes. The proposed constructions use a class of codes called almost secure frameproof codes. An almost secure frameproof code is a relaxed version of a secure frameproof code, which in turn is the same as a separating code. This relaxed version is the object of our interest because it gives rise to fingerprinting codes of higher rate than fingerprinting codes derived from separating codes. The construction of almost secure frameproof codes discussed here is based on weakly biased arrays, a class of combinatorial objects tightly related to weakly dependent random variables.

Keywords Separating code · Secure frameproof code · Fingerprinting · Traitor tracing

Mathematics Subject Classification (2000) 94B60 · 94B65

1 Introduction

Copyright protection schemes are used to deter illegal distribution of digital objects. In order to offer copyright protection to content distributors, a different set of marks is embedded in each copy of a digital object. This makes each copy unique, and clearly rules out plain redistribution. Weakness of such schemes comes

The material in this paper was presented in part at the 2013 International Workshop on Security (IWSEC), November 2013, Okinawa, Japan [17].

José Moreira · Marcel Fernández
Department of Network Engineering, Universitat Politècnica de Catalunya (UPC), C. Jordi Girona 1-3, Building C3, 08034 Barcelona, Spain
E-mail: {jose.moreira, marcel}@entel.upc.edu

Grigory Kabatiansky
Skolkovo Institute of Science and Technology (Skoltech), 3 Nobel St., Skolkovo, Moscow Region, Russia 143025
and
National Research University Higher School of Economics (HSE), Myasnitskaya ul. 20, Moscow, Russia 101000
E-mail: g.kabatyansky@skoltech.ru

in the form of *collusion attacks*. In this case, by comparing their copies, attackers detect the positions where their copies differ. With this knowledge at hand, they can create a pirated copy, where they might modify the symbols in the detected positions. This pirated copy, which masks the attackers' identities, is the one being redistributed. The attackers are usually referred to as *traitors*, a term coined in [7, 8].

The goal in constructing codes for copyright protection is to come up with a set of codewords (embedded marks) that are resilient against collusion attacks. For instance, in a *c-secure frameproof* code two disjoint coalitions of traitors of size at most c, cannot create the same pirated copy. Stronger protection is given by *c-secure* families of fingerprinting codes. In this case, given a pirated copy created by a coalition of size at most c, it is possible to trace, with high probability, at least one of the traitors in the coalition.

A way to approach the fingerprinting problem is to try to "separate" all disjoint sets codewords of at most a given size. The notion of separation in coding theory has been studied for decades; see [22] and the references therein. In [3] it was shown that fingerprinting codes can be obtained by "amplifying" such separation.

While the separating property and the secure frameproof property are exactly the same thing, relaxing both definitions in the sense of not requiring neither strict separation nor strict frameproofness, leads to two different notions: *almost separating codes* and *almost secure frameproof codes* [13,18]. Since almost secure frameproof codes offer better rates [13,18], we will focus on obtaining explicit constructions of such codes over almost separating codes. This will allow us to obtain fingerprinting codes of higher rate.

1.1 Our contribution

As we have just said, in this paper we discuss an explicit construction of a family of (binary) fingerprinting codes. In order to do this, we depart from an earlier work on weakly biased arrays [19], and develop a new concept called *almost universal sets*, which might be of independent interest. We then link weakly biased arrays and almost universal sets with almost secure frameproof codes, and present explicit constructions of such codes. With these constructions at hand, we use standard code concatenation techniques [14,3,5,6] to obtain the family of fingerprinting codes.

One thing that is worth noting from the beginning is that our goal in this paper is to give explicit constructions of codes. While there are several elegant approaches to fingerprinting codes in the literature, they are probabilistic in nature [5,6,25], and therefore not comparable to our constructions here. Closer to our work are the results in [3], which we improve in two directions. One is to note that for their inner codes, absolute separation can be replaced by almost secure frameproofness, and this leads to codes with higher rate. The other is that we give explicit constructions for those inner codes.

We are now in position to underline the structure of the paper. In Section 2 we give some useful definitions and an overview of previous results. Our work begins in Section 3, where by adjusting the bias in weakly biased arrays we present explicit constructions for sets of vectors that we call almost universal. The main contribution is discussed in Section 4, where explicit constructions of almost secure

frameproof codes are given. Although these constructions might be of independent interest, we make of use them in Section 5 to obtain a secure family of fingerprinting codes with exponentially small error probability and an efficient identification algorithm.

2 Background and definitions

We begin by introducing some coding theory concepts.

Let $q \geq 2$ be an integer. A q-ary *alphabet* Q is a nonempty set of size q. If Q is the finite field of q elements, we denote it by $\mathbb{F}_q$. For any integer $n \geq 1$, let Q^n denote the set of all possible n-tuples over Q. We denote the elements of Q^n in boldface, e.g., $\mathbf{u} = (u_1, \ldots, u_n) \in Q^n$.

An (n, M)-*code* C over Q is a subset of Q^n of size M. The parameter n is called the *length* of the code. A code C is a *linear* $[n, k]$-*code* over $\mathbb{F}_q$ if $C \subseteq \mathbb{F}_q^n$ is a vector subspace of dimension k. The elements of a code are called *codewords*, and the *minimum distance* of a code is the smallest Hamming distance between any two of its codewords. Also, we say that an (n, M)-code over a q-ary alphabet has *rate* R, where

$$R = n^{-1} \log_q M.$$

2.1 Secure frameproof and almost secure frameproof codes

Let $U = \{\mathbf{u}_1, \ldots, \mathbf{u}_c\}$ be a subset of size $|U| = c$ of a code C. We denote by $P_i(U)$ the *projection* of U on the ith position, that is, the set of elements of the code alphabet at the ith position,

$$P_i(U) \stackrel{\text{def}}{=} \{u_{1i}, \ldots, u_{ci}\}.$$

Definition 1 A code C is (c, c')-*separating* if for every pair of disjoint subsets $U, V \subseteq C$, where $|U| = c$ and $|V| = c'$, there is a position $1 \leq i \leq n$ such that their projections on this position have empty intersection, i.e.,

$$P_i(U) \cap P_i(V) = \emptyset.$$

We call such a position a *separating position*.

Clearly, a code that is (c, c')-separating is also (t, t')-separating, for $t \leq c$ and $t' \leq c'$. The separating property was first discussed in [15], and has been subsequently investigated by many authors [22,20,21,16,10,11,12].

Recently, more attention has been paid to separating codes, in connection with digital fingerprinting schemes. In this context, we define a *c-coalition* as a subset $U \subseteq C$ of size $|U| \leq c$, where C is a code over an alphabet Q. A position i is *undetectable* for U if all the codewords in U match at this position, that is, $|P_i(U)| = 1$. A position that fails to satisfy this property is called *detectable*.

In a collusion attack, the *marking assumption* [5,6] states that for a c-coalition U all the undetectable positions i must remain unchanged in the pirated word that the traitors generate $\mathbf{z} = (z_1, \ldots, z_n)$. Moreover, the *narrow-sense envelope*

model [3], states that $z_i \in P_i(U)$, for every position i. Hence, the set of all pirated words that the c-coalition U can generate under this model, denoted $\mathrm{desc}(U)$, is

$$\mathrm{desc}(U) \stackrel{\mathrm{def}}{=} \{\mathbf{z} = (z_1, \ldots, z_n) \in Q^n : z_i \in P_i(U),\ 1 \leq i \leq n\}.$$

Often, the codewords in U are called *parents* and the words in $\mathrm{desc}(U)$ are called *descendants*. Also, the *c-descendant code* of C, denoted $\mathrm{desc}_c(C)$, is defined as

$$\mathrm{desc}_c(C) \stackrel{\mathrm{def}}{=} \bigcup_{U \subseteq C, |U| \leq c} \mathrm{desc}(U).$$

A descendant $\mathbf{z} \in \mathrm{desc}_c(C)$ is called *c-uniquely decodable* if $\mathbf{z} \in \mathrm{desc}(U)$ for some c-coalition $U \subseteq C$ and $\mathbf{z} \notin \mathrm{desc}(V)$ for any c-coalition $V \subseteq C$ such that $U \cap V = \emptyset$.

Definition 2 A code C is *c-secure frameproof* if for any pair of c-coalitions $U, V \subseteq C$ such that $U \cap V = \emptyset$, then we have $\mathrm{desc}(U) \cap \mathrm{desc}(V) = \emptyset$. Equivalently, if all $\mathbf{z} \in \mathrm{desc}_c(C)$ are c-uniquely decodable.

The concept of secure frameproof code was introduced in [5,6,24,23]. It is not difficult to see that the definition of a c-secure frameproof code coincides with that of a (c, c)-separating code, as it was noticed, e.g., in [3]. Moreover, in the crypto literature, $(c, 1)$-separating codes are also called *c-frameproof codes*.

Lower bounds on the asymptotic rate of $(2, 2)$-separating codes were first studied in [15,20]. For the case of binary alphabets, it is known that a lower bound is $\geq 1 - \log_2(7/8) \approx 0.0642$ for arbitrary $(2, 2)$-separating codes [20,22], and this bound also holds for linear $(2, 2)$-separating codes [20]. Moreover, for arbitrary (c, c)-separating codes, it was shown in [3] that the lower bound is at least

$$-\frac{1}{2c-1} \log_2(1 - 2^{-2c+1}).$$

On the other hand, an upper bound on the asymptotic rate is < 0.2835 for arbitrary $(2, 2)$-separating codes [22,16], and it was shown in [22] that it is < 0.108 for linear $(2, 2)$-separating codes. We also refer the reader to [12], where further results are presented.

Note that the existence bounds for separating codes shown above give codes of low rate. In order to obtain codes with better rates, two relaxed versions of separating codes were presented in [13,18]: *almost separating codes* and *almost secure frameproof codes*. In this paper our focus is on almost secure frameproof codes, since it was shown in [13,18] that they have better rates than almost separating codes.

Definition 3 A code C is *ε-almost c-secure frameproof* if the ratio of c-uniquely decodable descendants in $\mathrm{desc}_c(C)$ is at least $1 - \varepsilon$.

The fact that the secure frameproof property is required with high probability, rather that in all the cases, is the main reason that allows us to obtain an improvement on the code rate, compared to ordinary secure frameproof (i.e., separating) codes. It was shown in [13,18] that the lower bound on the asymptotic rate of binary almost secure frameproof codes is at least

$$-\frac{1}{c} \log_2(1 - 2^{-c}).$$

2.2 Weakly biased arrays, weakly dependent arrays, and universal sets

Now we turn our focus to weakly biased and weakly dependent arrays, which are the combinatorial objects that will serve us to construct almost secure frameproof codes. We will focus on the binary case, since our final goal is to use these constructions to obtain families of binary fingerprinting codes. Also, we recall that weakly biased and weakly dependent arrays are strongly related to small-bias probability spaces. For a more detailed exposition, we refer the reader to [1,19,4].

A *binary* (n, M)*-array* is an n-by-M matrix whose entries are elements from $\mathbb{F}_2$. For a binary vector $\mathbf{u}$ of length n, $\mathbf{u} \in \mathbb{F}_2^n$, let us denote $\mu(0; \mathbf{u})$ and $\mu(1; \mathbf{u})$ the number of zeros and ones in $\mathbf{u}$, respectively. The *bias* of $\mathbf{u}$ is defined as

$$\frac{1}{n} |\mu(0; \mathbf{u}) - \mu(1; \mathbf{u})|.$$

That is, the vector $\mathbf{u}$ has low bias if it has roughly the same number of zeros and ones.

Definition 4 Let $0 \leq \varepsilon < 1$. A binary (n, M)-array A is *ε-biased* if every nontrivial linear combination of its columns has bias $\leq \varepsilon$.

In other words, the bias of a binary array A is the bias of the vector subspace spanned by the columns of A. By definition, the bias of A is low if and only if the bias of every nonzero vector of this subspace is low. Also, the above definition can be restricted by allowing a maximum number of columns in the linear combination.

Definition 5 Let $0 \leq \varepsilon < 1$. A binary (n, M)-array A is *t-wise ε-biased* if every nontrivial linear combination of at most t columns has bias $\leq \varepsilon$.

Throughout our discussion we refer to (t-wise) ε-biased arrays simply as *weakly biased* arrays when there is no need to make explicit its parameters.

Next, we will also draw our attention to binary arrays that exhibit "almost uniformity" in the following sense. Let A be a binary (n, M)-array, and let $S \subseteq \{1, \ldots, M\}$ be a subset of column indices of size s. Now, take a vector $\mathbf{a} \in \mathbb{F}_2^s$. The number of rows of A whose projection onto the indices of S equal $\mathbf{a}$ is denoted $\nu_S(\mathbf{a}; A)$. We are interested in the fact that every vector $\mathbf{a} \in \mathbb{F}_2^s$ appears "almost evenly" in the projection, for every possible choice of S. In other words, we want $\nu_S(\mathbf{a}; A)/n \approx 2^{-s}$. These kinds of binary arrays are referred to as *weakly dependent* arrays.

We consider two variants of weakly dependent arrays, by using the L_∞- and L_1-norms, as follows in the definitions below.

Definition 6 Let $0 \leq \varepsilon < 1$. A binary (n, M)-array A is *t-wise ε-dependent* (in L_∞-norm) if, for every subset $S \subseteq \{1, \ldots, M\}$ of $s \leq t$ columns, satisfies

$$\max_{\mathbf{a} \in \mathbb{F}_2^s} \left| \frac{\nu_S(\mathbf{a}; A)}{n} - 2^{-s} \right| \leq \varepsilon.$$

Definition 7 Let $0 \leq \varepsilon < 1$. A binary (n, M)-array A is *ε-away from t-wise independence* (in L_1-norm) if, for every subset $S \subseteq \{1, \ldots, M\}$ of $s \leq t$ columns, satisfies

$$\sum_{\mathbf{a} \in \mathbb{F}_2^s} \left| \frac{\nu_S(\mathbf{a}; A)}{n} - 2^{-s} \right| \leq \varepsilon.$$

As commented above, weakly dependent arrays have an interpretation as a small-bias probability space [19,1]. If M random variables $X_1, \ldots, X_M$ take uniformly at random the corresponding values of a row of a binary (n, M)-array A which is ε-away from t-wise independence, then any $s \leq t$ of such random variables behave like "almost independent" random variables, provided that ε is small. Hence, one would like to obtain such an array A with n (the size of the sample space) as small as possible.

For our purposes, we will also be interested in a certain kind of combinatorial objects known as universal sets. We will extend this notion to what we call almost universal sets in Section 3 below.

Definition 8 An (M, t)-*universal set* B is a subset of $\mathbb{F}_2^M$ such that for every subset $S \subseteq \{1, \ldots, M\}$ of t positions the set of projections of the elements of B onto the indices of S contains every vector $\mathbf{a} \in \mathbb{F}_2^t$.

Let A be a binary (n, M)-array. If for every subset $S \subseteq \{1, \ldots M\}$ of t columns and every vector $\mathbf{a} \in \mathbb{F}_2^t$ we have $\nu_S(\mathbf{a}; A) > 0$, then the rows of A yield an (M, t)-universal set. Observe trivially, that an (M, t)-universal set is also an (M, t')-universal set for any $t' \leq t$.

2.3 Relationships

The combinatorial objects presented in the previous sections can be related to each other through several important results.

It is clear, as remarked in [1], that if a binary (n, M)-array A is ε-away from t-wise independence, then it is t-wise ε-dependent, and if A is t-wise ε-dependent, then it is $2^t\varepsilon$-away from t-wise independence. The following results relate weakly biased arrays, weakly dependent arrays, universal sets and secure frameproof codes.

Lemma 1 ([26]) *Let A be a binary (n, M)-array. If A is t-wise ε-biased, then A is $2^{t/2}\varepsilon$-away from t-wise independence.*

The above result is attributed to Vazirani [26], but we also refer the reader to [27,9,1,19]. An obvious consequence of this lemma is that an ε-biased array is also $2^{t/2}\varepsilon$-away from t-wise independence.

Proposition 1 ([19]) *Let A be a binary (n, M)-array. If A is 2^{-t}-away from t-wise independence, then the rows of A yield an (M, t)-universal set of size n.*

It is not difficult to see how universal sets can be used to construct separating (i.e., secure frameproof) codes. This result was clearly shown in [2]. We provide here a proof sketch for completeness.

Corollary 1 ([2]) *Let $c \geq 2$ be an integer. Then, an $(M, 2c)$-universal set B of size n yields a binary (c, c)-separating (n, M)-code.*

Proof (Sketch) Let A be a binary (n, M)-array whose rows are the n vectors of B. We claim that the columns of A form a (c, c)-separating (n, M)-code C. To see this, consider any two disjoint subsets of codewords $U, V \subseteq C$ of size c, and call $S_U, S_V \subseteq \{1, \ldots, M\}$ the corresponding disjoint subsets of column indices from A.

The $(M, 2c)$-universal property of B implies that, in particular, there is a row i in A for which all the c columns in S_U contain a 0 and all the c columns in S_V contain a 1, at this row index i. In terms of the code C, this means that at position i we have $P_i(U) \cap P_i(V) = \emptyset$, and the proof follows. □

By virtue of Lemma 1, an explicit construction of weakly biased arrays will provide an explicit construction of weakly dependent arrays immediately. This construction, will in turn provide an explicit construction of universal sets by Proposition 1, which automatically leads to secure frameproof codes, by Corollary 1.

2.4 Existing explicit constructions

As it has been shown, in order to construct secure frameproof codes of good rate the problem can be reduced to the construction of weakly biased arrays with the smallest number of rows possible. We will deal with such explicit constructions in this section. We will follow a convention similar to [4] to present the results.

An important, well-known explicit construction of ε-biased arrays is presented in [19].

Theorem 1 ([19]) *There is an explicit construction of a binary (n, M)-array that is ε-biased, with*

$$n \leq 2^{2(\log_2 M + \log_2 \frac{1}{\varepsilon})}. \tag{1}$$

The arrays from the previous theorem can be used to construct ε-away from t-wise independence arrays with $n = 2^{O(t+\log M+\log \frac{1}{\varepsilon})}$, which, in turn, lead to (M, t)-universal sets of size $n = 2^{O(t+\log M)}$. That is, following this procedure we obtain c-secure frameproof codes of length

$$n = M \cdot 2^{O(c)}. \tag{2}$$

In order to obtain codes with shorter lengths, it is also noted in [19] that it suffices to consider t-wise ε-biased arrays (instead of ε-biased arrays). This is a less restrictive condition. The next result enables us to obtain t-wise ε-biased arrays from ε-biased arrays.

Theorem 2 ([19]) *Let A be an ε-biased binary (n, M')-array, and let H be the parity-check matrix of a binary $[M, M - M']$-code with minimum distance $t + 1$. Then, the matrix product $A \times H$ is a t-wise ε-biased binary (n, M)-array.*

Usually, the matrix H employed in Theorem 2 above is the parity-check matrix of a binary $[M, M-M']$-BCH code with minimum distance $t+1$. In this case, H has M columns and $M' = t \log_2 M$ rows. Therefore, by using Theorem 2 together with Lemma 1, the number of rows of an ε-away from t-wise independence (n, M)-array can be reduced from $n = 2^{O(t+\log M+\log \frac{1}{\varepsilon})}$ to $n = 2^{O(t+\log\log M+\log \frac{1}{\varepsilon})}$. Finally, the explicit construction of (M, t)-universal sets from [19] has size $n = 2^{O(t+\log\log M)}$, which for $t = 2c$, generate c-secure frameproof codes of length

$$n = \log_2 M \cdot 2^{O(c)}, \tag{3}$$

and rate $R = 2^{-O(c)}$ [19], a clear improvement compared to (2).

We conclude this section by noting that better explicit constructions of ε-biased arrays are presented in [4], compared to Theorem 2, when the parameters satisfy some conditions. The best construction shown there is based on Suzuki codes.

Theorem 3 ([4]) *If* $\log_2 M > 3\log_2 \frac{1}{\varepsilon}$, *then there is an explicit construction of a binary* (n, M)*-array that is* ε*-biased, with*

$$n \leq 2^{3/2\,(\log_2 M + \log_2 \frac{1}{\varepsilon}) + 2}.$$

However, we will show below that the condition imposed by Theorem 3 will prevent us from using this improved construction in practical scenarios.

3 Almost universal sets

In this section we relax the concept of universal set presented in Section 2, and obtain what we call almost universal sets. We note that the results obtained in this section are of independent interest, and they will also be useful to analyze the properties of the codes that we construct in Section 4.

Lemma 2 *Let* A *be a binary* (n, M)*-array. Suppose that there is a subset* $S \subseteq \{1, \ldots, M\}$ *of* t *columns for which there are* m *vectors* $\mathbf{a}_1, \ldots, \mathbf{a}_m \subseteq \mathbb{F}_2^t$ *such that*

$$\frac{\nu_S(\mathbf{a}_1; A)}{n} = \cdots = \frac{\nu_S(\mathbf{a}_m; A)}{n} = p,$$

with $0 \leq p \leq 1/m$. *Then,* A *is not* ε*-away from* t*-wise independence for any*

$$\varepsilon < 2m|p - 2^{-t}|.$$

Proof Consider the subset S of t columns and the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ stated above. For this particular subset of columns, we have

$$\begin{aligned}
&\sum_{\mathbf{a} \in \mathbb{F}_2^t} \left| \frac{\nu_S(\mathbf{a}; A)}{n} - 2^{-t} \right| \\
&\quad = \sum_{\mathbf{a} \in \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}} \left| \frac{\nu_S(\mathbf{a}; A)}{n} - 2^{-t} \right| + \sum_{\mathbf{a} \notin \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}} \left| \frac{\nu_S(\mathbf{a}; A)}{n} - 2^{-t} \right| \\
&\quad \geq m|p - 2^{-t}| + (2^t - m) \left| \frac{1 - mp}{2^t - m} - 2^{-t} \right| \\
&\quad = 2m|p - 2^{-t}|.
\end{aligned}$$

The above inequality is derived as follows. On one hand, the sum over the vectors $\mathbf{a} \in \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ is readily seen to equal $m|p - 2^{-t}|$. On the other hand, the sum over the vectors $\mathbf{a} \notin \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ can be viewed as the objective of a resource allocation problem, and we are interested in finding the minimum of this objective. A solution to this minimization problem is achieved when each such $\mathbf{a}$ appears the same number of times. These $\mathbf{a}$ account for $n - nmp$ rows of A, which implies that each $\nu_S(\mathbf{a}; A)$ should equal $n(1 - mp)/(2^t - m)$, and the inequality follows.

Finally, we conclude that A cannot be ε-away from t-wise independence for any $\varepsilon < 2m|p - 2^{-t}|$. □

Recall from Section 2.3 that if A is ε-away from t-wise independence, then it is t-wise ε-dependent. However, this result can be somewhat improved as follows.

Corollary 2 *Let $0 \leq \varepsilon < 1$, and let A be a binary (n, M)-array. If A is ε-away from t-wise independence, then A is t-wise $\varepsilon/2$-dependent.*

Proof Let A be ε-away from t-wise independence, and assume by contradiction that A is not t-wise $\varepsilon/2$-dependent. In other words, there is a subset $S \subseteq \{1, \ldots, M\}$ of size t for which there is a vector $\mathbf{a}' \in \mathbb{F}_2^t$, with $p = \nu_S(\mathbf{a}'; A)/n$, such that

$$|p - 2^{-t}| > \frac{\varepsilon}{2}.$$

Substituting $m = 1$ and $|p - 2^{-t}|$ in Lemma 2 shows that A is not ε-away from t-wise independence, which contradicts the starting premise. □

In an (M, t)-universal set $B \subseteq \mathbb{F}_2^M$ we have that every possible vector $\mathbf{a} \in \mathbb{F}_2^t$ appears in the set of projections of the elements of B onto any subset of t positions. Nevertheless, for our purposes in the construction of a secure family of fingerprinting codes, strict universality will not be completely necessary. That is, it will not be required that every possible vector $\mathbf{a} \in \mathbb{F}_2^t$ appears in such projections. More concretely, we will be interested in the fact that, at least, one of the following two vectors from $\mathbb{F}_2^t$ appears:

$$(\underbrace{0, \ldots, 0}_{c}, \underbrace{1, \ldots, 1}_{c}) \text{ or } (\underbrace{1, \ldots, 1}_{c}, \underbrace{0, \ldots, 0}_{c}), \tag{4}$$

being $t = 2c$ for some integer $c \geq 2$. This suggests replacing the original definition of universal set, and consider the case where it suffices to obtain a set of vectors $B \subseteq \mathbb{F}_2^M$ where the universal property is guaranteed in a sufficiently high number of cases. Namely, the main idea is to require that a high number of vectors from $\mathbb{F}_2^t$ appear in each set of projections. To this end we consider the following definition.

Definition 9 An *ε-almost (M, t)-universal set* B is a subset of $\mathbb{F}_2^M$ such that for every subset $S \subseteq \{1, \ldots, M\}$ of t positions the set of projections of the elements of B onto the indices of S contains, at least, a fraction of $1 - \varepsilon$ vectors $\mathbf{a} \in \mathbb{F}_2^t$.

Of course, the idea of relaxing the constraint imposed by the universal property is to obtain a smaller set of vectors, which will enable us to improve the rate of the codes that we will construct below. However, substituting an universal set by an almost universal set implies that there could be some subset of t positions S such that the set of projections of the elements of B onto S does not contain any of the desired vectors from $\mathbb{F}_2^t$ mentioned above. In our analysis below, we will have to handle this situation, ensuring that this undesired event occurs with sufficiently small probability.

Also, note again that if A is a binary (n, M)-array, the rows of A generate an ε-almost (M, t)-universal set provided that there is a fraction of, at least, $1 - \varepsilon$ vectors $\mathbf{a} \in \mathbb{F}_2^t$ such that $\nu_S(\mathbf{a}; A) > 0$, for every subset $S \subseteq \{1, \ldots, M\}$ of t columns.

In the following proposition, we show the relationship between almost universal sets and weakly dependent arrays.

Proposition 2 *Let A be a binary (n, M)-array. If A is $(2^{-t}+\varepsilon)$-away from t-wise independence, then the rows of A yield an ε-almost (M, t)-universal set of size n.*

Proof Let A be $(2^{-t}+\varepsilon)$-away from t-wise independence, and assume by contradiction that the rows of A do not yield an ε-almost (M, t)-universal set. In other words, there is a subset $S \subseteq \{1, \ldots, M\}$ of t columns for which there are strictly more than $2^t\varepsilon$ vectors $\mathbf{a} \in \mathbb{F}_2^t$ with $\nu_S(\mathbf{a}; A) = 0$. Substituting $m = \lfloor 2^t\varepsilon \rfloor + 1$ and $p = 0$ in Lemma 2, we obtain

$$\sum_{\mathbf{a}\in\mathbb{F}_2^t} \left| \frac{\nu_S(\mathbf{a}; A)}{n} - 2^{-t} \right| \geq (\lfloor 2^t\varepsilon \rfloor + 1)2^{-t+1}.$$

It is routine to check that $(\lfloor 2^t\varepsilon \rfloor + 1)2^{-t+1} > 2^{-t} + \varepsilon$. Hence, it follows that A is not $(2^{-t} + \varepsilon)$-away from t-wise independence, a contradiction. □

Therefore, Proposition 2 shows that the construction of almost universal sets is reduced, again, to the construction of weakly dependent arrays, and by virtue of Lemma 1 it is reduced to the construction of weakly biased arrays.

We conclude this section by showing the connection between universal and almost universal sets in the following lemma.

Lemma 3 *Let B be an ε-almost (M, t)-universal set. Then, B is an (M, t')-universal set for any $t' \leq \min\{t, \lceil \log_2 \frac{1}{\varepsilon} \rceil - 1\}$.*

Proof Let A be a binary (n, M)-array whose rows are the elements of B. Using our notation, this means that for any given subset $S \subseteq \{1, \ldots, M\}$ of t columns there are, at most, $2^t\varepsilon$ vectors $\mathbf{a} \in \mathbb{F}_2^t$ such that $\nu_S(\mathbf{a}; A) = 0$.

On one hand, if $\varepsilon < 2^{-t}$, then there is no vector $\mathbf{a} \in \mathbb{F}_2^t$ with $\nu_S(\mathbf{a}; A) = 0$, and B is (M, t)-universal. On the other hand, assume that $\varepsilon \geq 2^{-t}$ and consider any subset $S' \subseteq S$ of $t' \leq t$ columns. Note that for each vector $\mathbf{a}' \in \mathbb{F}_2^{t'}$ there are exactly $2^{t-t'}$ vectors from $\mathbb{F}_2^t$ that coincide with $\mathbf{a}'$ in the columns with indices in S'. This implies that if there is a vector $\mathbf{a}' \in \mathbb{F}_2^{t'}$ such that $\nu_{S'}(\mathbf{a}'; A) = 0$, then it must be the case that $2^t\varepsilon \geq 2^{t-t'}$. Equivalently, if $\varepsilon < 2^{-t'}$, that is, $t' \leq \lceil \log_2 \frac{1}{\varepsilon} \rceil - 1$, then no vector $\mathbf{a}' \in \mathbb{F}_2^{t'}$ has $\nu_{S'}(\mathbf{a}'; A) = 0$. It follows that B is (M, t')-universal. □

4 Construction of almost secure frameproof codes

Now, we aim to explicitly construct almost secure frameproof codes from weakly biased arrays and almost universal sets. Before dwelling into explicit details, we give an intuitive reasoning of our motivation to use these combinatorial objects.

Consider a random binary (n, M)-code C such that each codeword is generated according to a vector of probabilities $\mathbf{p} = (p_1, \ldots, p_n)$, where $\mathbf{p}$ is a random vector with pmf $f_{\mathbf{p}}$. That is, we first generate a vector of probabilities $\mathbf{p}$ of length n, distributed according to $f_{\mathbf{p}}$, and then we randomly generate M binary codewords $\mathbf{u} = (u_1, \ldots, u_n)$ such that $\Pr\{u_i = 1\} = p_i$. Now, let p_{sep} be the probability that two randomly chosen subsets $U, V \subseteq C$ of size c have a separating position. That

is, p_{sep} is a random variable dependent on $\mathbf{p}$. We would like to know which pmf $f_{\mathbf{p}}$ maximizes the expected value of p_{sep},

$$E_{f_{\mathbf{p}}}[p_{\text{sep}}] = E_{f_{\mathbf{p}}}\Big[1 - \prod_{i=1}^{n}(1 - 2p_i^c(1-p_i)^c)\Big].$$

Observe that this expectation is maximized simply by considering a pmf that takes 1 on the maximum of the argument of the expectation and 0 otherwise. The product $\prod_{i=1}^{n}(1 - 2p_i^c(1-p_i)^c)$ attains its minimum value when each one of its n terms attains individually its minimum value. This occurs when $p_i = 1/2$, for $1 \leq i \leq n$, that is, for $\mathbf{p} = (1/2, \ldots, 1/2)$. Therefore,

$$\max\{E_{f_{\mathbf{p}}}[p_{\text{sep}}]\} = 1 - 2^{1-2c}.$$

If each pair of subsets $U, V \subseteq C$ of size c have a separating position, then the code is (c, c)-separating, or what is the same, c-secure frameproof.

The above argument shows that randomly generated codes using a vector of probabilities $\mathbf{p} = (1/2, \ldots, 1/2)$ seem good candidates to have good separating properties. Also, from Corollary 1 we see that universal sets enable us to construct secure frameproof codes and, in fact, almost universal sets derived from Proposition 2 exhibit a behavior close to the uniform distribution.

Since we are interested in almost secure frameproof codes, we do not require absolute separation, and we can tolerate a small bias on the elements of the probability vector $\mathbf{p}$. This is what motivates our choice of weakly dependent arrays in our constructions below.

4.1 Using weakly dependent arrays to obtain almost secure frameproof codes

As it has been shown in Corollary 1, an $(M, 2c)$-universal set of size n generates a c-secure frameproof (n, M)-code. Now, consider an ε-almost (M, t)-universal set B, and rearrange its elements as the rows of a binary (n, M)-array A. Also, regard the columns of A as the codewords of a code C, similarly as in the proof of Corollary 1. Hence, C is an (n, M)-code of rate $R = \log_2 M/n$.

Let us focus on the frameproof properties of C. If $t < 2c$, then C need not be c-secure frameproof. However, if $t \geq 2c$ and $\varepsilon < 2^{-2c+1}$, then C is c-secure frameproof. In fact, to see that C is c-secure frameproof, we can use an argument similar to that from the proof of Lemma 3. The code C is c-secure frameproof if in each projection of B onto every possible subset S' of $2c$ positions it contains at least one separating vector like the ones in (4). Assume that there is one such S' for which both separating vectors from $\mathbb{F}_2^{2c}$ are missing in the projection. Then, there must be one subset S of t positions, $S' \subseteq S$, such that there are at least 2^{t-2c+1} missing vectors from $\mathbb{F}_2^t$ in the projection of B onto the indices of S. This event would contradict the fact that B is an ε-almost (M, t)-universal set with $\varepsilon < 2^{-2c+1}$.

From Lemma 3 it follows that for $t \geq 2c$ an $(M, 2c)$-universal set can be derived from an ε-almost (M, t)-universal set, provided that $\varepsilon < 2^{-2c}$. As we have argued above, the condition for the case of a c-secure frameproof code is less strict on ε, and only requires an ε-almost (M, t)-universal set with $\varepsilon < 2^{-2c+1}$. Also, we remark that for $\varepsilon < 2^{-2c+1}$, an ε-almost (M, t)-universal set is an $(M, 2c-1)$-universal

set, again from Lemma 3, in addition to generate a c-secure frameproof code. But in general, an $(M, 2c-1)$-universal set need not yield a c-secure frameproof code.

Again, if $t < 2c$, or if $\varepsilon \geq 2^{-2c+1}$, then C need not be c-secure frameproof, and we have to regard it as an almost secure frameproof code. To this end, we will extend the use of weakly dependent arrays, leading to almost universal sets, to the construction of such codes. This is formalized in the following result.

Proposition 3 *Let $c \geq 2$, t, M be integers such that $M \geq 2c$, and let one of the following conditions be satisfied*

$$\begin{array}{lllll} \text{Condition 1:} & c \leq t < 2c, & \text{and} & 0 \leq \varepsilon' < 2^{-c+1}, & \text{or} \\ \text{Condition 2:} & t \geq 2c, & \text{and} & 2^{-2c+1} + 2^{-t} < \varepsilon' < 2^{-c+1}. & \end{array}$$

Then, an ε'-away from t-wise independence (n, M)-array generates an ε-almost c-secure frameproof (n, M)-code, for any

$$\varepsilon \geq M^c (1 - 2^{-c} + \varepsilon'/2)^n. \tag{5}$$

Proof Let A be a binary (n, M)-array which is ε'-away from t-wise independence, and regard its columns as the codewords of a code C.

First, observe that under the conditions given in the statement of the proposition, the array A need not yield a c-secure frameproof code. To see this, observe that for Condition 1 we have $t < 2c$, and for Condition 2 the code C would be c-secure frameproof code if A yielded a (2^{-2c+1})-almost universal set. According to Proposition 2 this occurs for $\varepsilon' \leq 2^{-2c+1} + 2^{-t}$, which contradicts the statement of the proposition.

Now, for a randomly chosen c-coalition $V \subseteq C$, let us denote p_0 (resp. p_1) the probability that all the codewords of V equal 0 (resp. 1) at a given position $1 \leq i \leq n$. Since the codewords from V correspond to, at most, $c \leq t$ columns of A, we have

$$p_0, p_1 \geq 2^{-|V|} - \varepsilon'/2 \geq 2^{-c} - \varepsilon'/2,$$

because A is also t-wise $\varepsilon'/2$-dependent, by virtue of Corollary 2.

Now, let $\mathbf{z}$ be a descendant generated by some c-coalition U of the code, i.e., $\mathbf{z} \subseteq \mathrm{desc}(U)$. Then, for a randomly chosen c-coalition V,

$$\Pr\{\mathbf{z} \in \mathrm{desc}(V)\} = \prod_{i=1}^{n} \Pr\{z_i \in P_i(V)\} \leq (1 - 2^{-c} + \varepsilon'/2)^n.$$

Indeed, the probability that the ith symbol of $\mathbf{z}$ is in $P_i(V)$ satisfies

$$\Pr\{z_i \in P_i(V)\} \leq \Pr\{z_i = 0\}(1 - p_1) + \Pr\{z_i = 1\}(1 - p_0) \leq 1 - 2^{-c} + \varepsilon'/2.$$

We can bound the probability that $\mathbf{z}$ is generated by some other c-coalition disjoint from U. In order to apply the union bound, we argue that it suffices to take into account only the disjoint coalitions of maximum size. That is, only consider the subsets $V \subseteq C$ of size exactly c (instead of all c-coalitions, of size $\leq c$). To see this, observe that for a coalition V' which is a subset of another coalition V,

$$\Pr\{\mathbf{z} \in \mathrm{desc}(V') \text{ or } \mathbf{z} \in \mathrm{desc}(V)\} = \Pr\{\mathbf{z} \in \mathrm{desc}(V)\},$$

because $\mathrm{desc}(V') \subseteq \mathrm{desc}(V)$. There are at most $\binom{M}{c}$ different coalitions of size c disjoint from U. Therefore, the probability that $\mathbf{z}$ is generated by some other disjoint c-coalition, or equivalently, the ratio of descendants $\mathbf{z} \in \mathrm{desc}_c(C)$ that are not c-uniquely decodable is at most $M^c(1-2^{-c}+\varepsilon'/2)^n$. We conclude that C can be regarded as an ε-almost c-secure frameproof for any $\varepsilon \geq M^c(1-2^{-c}+\varepsilon'/2)^n$, because the ratio of non-uniquely decodable descendants will not exceed ε. □

In order to ease the analysis, one could assume that for every subset of at most c indices, each possible vector from $\mathbb{F}_2^c$ appears with uniform probability in the (M,c)-universal sets in the proof above, obtaining ε-almost c-secure frameproof codes for $\varepsilon \geq M^c(1-2^{-c})^n$. This is a reasonable assumption, since universal sets generated from weakly biased arrays are indeed "almost uniform" probability sample spaces. However, the error probability in (5) is already negligible, and this assumption would not handle the case $t = c$ properly.

4.2 Explicit constructions of almost secure frameproof codes

In this section, we show how to derive an explicit construction of almost secure frameproof codes. Armed with the machinery we have developed, it follows that a construction for almost secure frameproof codes can be reduced to the construction of weakly biased arrays. We have the following construction.

Construction 1 *Let M, c, t and ε' be values satisfying the conditions of Proposition 3.*

1. *Construct a binary (n, M')-array A' that is $2^{-t/2}\varepsilon'$-biased, where we take $M' = t\log_2 M$.*
2. *Take the parity-check matrix H of a binary $[M, M-M']$-BCH code, of length M, codimension $M' = t\log_2 M$ and minimum distance $t+1$.*
3. *The matrix product $A = A' \times H$ generates a t-wise $2^{-t/2}\varepsilon'$-biased (n, M)-array.*
4. *A is also ε'-away from t-wise independence.*
5. *Hence, the rows of A generate an $(\varepsilon' - 2^{-t})$-almost (M,t)-universal set of size n.*
6. *Moreover, the rows of A also generate an ε-almost c-secure frameproof code, for any*
$$\varepsilon \geq M^c(1-2^{-c}+\varepsilon'/2)^n.$$

In the above construction, Step 3 follows from Theorem 2, Step 4 from Lemma 1, Step 5 from Proposition 2, and Step 6 from Proposition 3.

Therefore, it remains to choose an appropriate explicit construction of a weakly biased array in Step 1, either from Theorem 1 or from Theorem 3. Observe that the conditions of Theorem 3 apply in Construction 1 when $\varepsilon' > 0$, and $\log_2 M' > -3\log_2(2^{-t/2}\varepsilon')$, i.e., when the size of the desired code satisfies

$$\log_2\log_2 M > 3t/2 - 3\log_2\varepsilon' - \log_2 t. \tag{6}$$

The resulting ε-almost c-secure frameproof code has length

$$n \leq 2^{3/2(t/2+\log_2 t+\log_2\log_2 M-\log_2\varepsilon')+2}.$$

We note that condition (6), even though analytically meaningful, it is only satisfied for impractically large values of M. That is, it will lead to codes with an excessively large number of codewords. For example, consider the simple case of $c = 2, t = 4$ and $\varepsilon' \lesssim 2^{-c+1}$. Then, the condition states that the construction is only valid for codes of size $M > 2^{128}$, in the most optimistic case.

On the other hand, Theorem 1 does not impose any restriction in the design parameters M, t, ε'. That is, for realistic scenarios, we have to consider this result to construct weakly biased arrays in Step 1 of Construction 1 above. Hence, plugging the parameters from Construction 1 into (1), the resulting code has length

$$n \leq 2^{2(t/2+\log_2 t+\log_2\log_2 M-\log_2 \varepsilon')} = \left(\frac{t}{\varepsilon'}\right)^2 2^t \log_2^2 M,$$

and rate

$$R = \left(\frac{\varepsilon'}{t}\right)^2 \cdot \frac{1}{2^t \log_2 M}. \tag{7}$$

In both cases, following the same approach as in [19] to derive (3) for $t = O(c)$, the length of the construction is

$$n = \log_2 M \cdot 2^{O(c-\log \varepsilon')},$$

which is an improvement compared to (3), for the case of ordinary secure frameproof (i.e., separating) codes.

Given the design parameters M, c, ε of an ε-almost c-secure frameproof code, it remains to optimize the value of ε' in (7), for the corresponding weakly dependent array, in order to obtain codes with the highest possible rate. Note that, as expected, the rate in (7) is an increasing function of ε'. Therefore, we must pick the maximum value of ε' such that (5) holds. For practical values of the design parameters, such value of ε' occurs at $\varepsilon' \lesssim 2^{-c+1}$. Hence, taking $t = c$ (the minimum allowable value according to Proposition 3), a good approximation on the rate R of the best codes derived from Construction 1 leads us to conclude the following result.

Corollary 3 *There is an explicit construction of ε-almost c-secure frameproof code of rate*

$$R \lesssim \left(c^2 2^{3c-2} \log_2 M\right)^{-1}. \tag{8}$$

In contrast, to obtain secure frameproof codes using the above construction and Corollary 1 we have to consider $t = 2c$ and $\varepsilon' = 2^{-t}$, and the resulting code rate drops to $R = \left(4c^2 2^{6c} \log_2 M\right)^{-1}$.

In Table 1 we show some code rates for ε-secure frameproof codes from Proposition 3, for the case of coalitions of size $c = 2$ to $c = 6$, considering the case $t = c$, and for code sizes ranging from $M = 10^2$ to $M = 10^8$. We have taken a design parameter $\varepsilon = 10^{-10}$, and the code rates are obtained from (7), through obtaining numerically the maximum allowable parameter ε' of the associated ε'-away from t-wise independence array in Construction 1. For comparison, we also present the rates of ordinary c-secure frameproof codes using the same construction, which in this particular case are derived from universal sets, as noted in Corollary 1. Observe that the rates obtained for almost secure frameproof codes are several orders of magnitude higher than those for secure frameproof codes.

ε-almost c-secure frameproof codes, with $\varepsilon = 10^{-10}$, from Construction 1

c	$M = 10^2$	$M = 10^3$	$M = 10^4$	$M = 10^5$	$M = 10^6$	$M = 10^7$	$M = 10^8$
2	2.1598E-3	1.4993E-3	1.1427E-3	9.2161E-4	7.7169E-4	6.6350E-4	5.8180E-4
3	1.2916E-4	8.6575E-5	6.5069E-5	5.2111E-5	4.3454E-5	3.7262E-5	3.2614E-5
4	9.1699E-6	6.1184E-6	4.5903E-6	3.6729E-6	3.0611E-6	2.6239E-6	2.2961E-6
5	7.3470E-7	4.8987E-7	3.6742E-7	2.9395E-7	2.4496E-7	2.0997E-7	1.8372E-7
6	6.3793E-8	4.2530E-8	3.1898E-8	2.5518E-8	2.1265E-8	1.8227E-8	1.5949E-8

c-secure frameproof codes from Corollary 1

c	$M = 10^2$	$M = 10^3$	$M = 10^4$	$M = 10^5$	$M = 10^6$	$M = 10^7$	$M = 10^8$
2	2.2967E-6	1.5311E-6	1.1483E-6	9.1867E-7	7.6556E-7	6.5619E-7	5.7417E-7
3	1.5949E-8	1.0633E-8	7.9746E-9	6.3797E-9	5.3164E-9	4.5569E-9	3.9873E-9
4	1.4018E-10	9.3452E-11	7.0089E-11	5.6071E-11	4.6726E-11	4.0051E-11	3.5045E-11
5	1.4018E-12	9.3452E-13	7.0089E-13	5.6071E-13	4.6726E-13	4.0051E-13	3.5045E-13
6	1.5210E-14	1.0140E-14	7.6051E-15	6.0841E-15	5.0701E-15	4.3458E-15	3.8026E-15

Table 1 Some attainable code rates for explicit constructions of ε-almost c-secure frameproof codes and c-secure frameproof codes. The rates of the ε-almost c-secure frameproof codes have been computed numerically, rather than using the approximation (8).

5 Explicit constructions of fingerprinting codes

Finally, we show how the almost secure frameproof codes derived from Construction 1 can be used to explicitly construct a family of fingerprinting codes with small error and an efficient identification algorithm.

The contents of this section are based on the construction of fingerprinting codes presented in [13,18], which uses almost secure frameproof codes as their building block. While [13,18] show the existence of such constructions, here we make these constructions explicit.

Let M be the total number of users to whom the distributor wishes to deliver his content. For a fingerprinting scheme to achieve a small error probability a single code is not sufficient, and a *family of codes* $\mathcal{C} = \{C_j\}_{j\in T}$ is needed [5,6], where T is some finite set of keys, and each C_j is an (m, M)-code. The family $\mathcal{C}$ is publicly known, but the distributor chooses secretly a code $C_j \in \mathcal{C}$ with probability $\pi(j)$. Moreover, the family $\mathcal{C}$ also requires an *identification algorithm*, which is actually a set of functions $\mathcal{A} = \{A_j\}_{j\in T}$, where each A_j is a mapping from the set of descendants of C_j to the set of c-coalitions of C_j, i.e.,

$$A_j : \mathrm{desc}_c(C_j) \to \{U \subseteq C_j : |U| \leq c\}.$$

We usually require that the identification algorithm $\mathcal{A}$ is *efficient*, meaning that each A_j can be executed in $O(\mathrm{poly}(m))$ time.

We say that a family $\mathcal{C} = \{C_j\}_{j\in T}$ is *c-secure with ε-error* if for any descendant $\mathbf{z} \in \mathrm{desc}(U)$ of a c-coalition U the set $A_j(\mathbf{z})$ is not empty, and

$$\Pr\{A_j(\mathbf{z}) \subseteq U\} > 1 - \varepsilon,$$

where the probability is taken over the random choices made by the coalition when creating the descendant, and over the pmf π. Moreover, it was noted in [3] that in order to achieve an exponential decline on the error, i.e., $\varepsilon = O(-\exp(m))$, the size of the family must grow exponentially in the code length. Our focus is on families of binary fingerprinting codes based on code concatenation [14].

Construction 2 *Consider an outer (n, M)-code C_{out} over a q-ary alphabet Q, an inner binary (l, q)-code C_{in}, and a vector of n mappings $\Phi_j = (\phi_{j1}, \ldots, \phi_{jn})$, where each ϕ_{ji} is a bijection $\phi_{ji} \colon Q \to C_{\text{in}}$. Note that there are a total of $(q!)^n$ of such vector mappings Φ_j, and we can index them by $j \in T = \{1, \ldots, (q!)^n\}$ under an arbitrary order. The concatenated code C_j is defined as*

$$C_j \stackrel{\text{def}}{=} \{(\phi_{j1}(w_1), \ldots, \phi_{jn}(w_n)) : (w_1, \ldots, w_n) \in C_{\text{out}}\},$$

which is a binary (m, M)-code, with $m = n\,l$. A family of binary concatenated codes consists of the set of all the codes C_j,

$$\mathcal{C} = \{C_j\}_{j \in T}.$$

If R_{out} and R_{in} are the rates of the outer and inner codes, respectively, the rate of the family $\mathcal{C}$ is $R = R_{\text{out}} R_{\text{in}}$.

The family of fingerprinting codes proposed in [3] particularizes Construction 2 with C_{out} being a Reed-Solomon or an algebraic-geometric code, and C_{in} being a (c, c)-separating (i.e., c-secure frameproof) code. For instance, for the case of Reed-Solomon codes they prove the existence of c-secure with ε-error families of fingerprinting codes, with exponentially small error and an efficient identification algorithm, for any rate

$$R < \frac{1}{c(c+1)} \frac{-\log_2(1 - 2^{-2c+1})}{2c - 1}. \tag{9}$$

In [13,18] it was shown the existence of almost secure frameproof codes of rate significantly higher than that of ordinary secure frameproof codes. Therefore, by replacing the inner secure frameproof codes from [3] by almost secure frameproof codes (and by appropriately modifying the identification algorithm) it was shown the existence of c-secure with ε-error families of fingerprinting codes with the same factor of increase in the code rate. The following result is an excerpt of [18, Corollary 2].

Corollary 4 *Let C_{out} be an extended $[n, k]$-Reed-Solomon code over $\mathbb{F}_q$ of rate R_{out}, and let C_{in} be a binary ε_{in}-almost c-secure frameproof (l, q)-code of rate R_{in}. Let $\mathcal{C} = \{C_j\}_{j \in T}$ be the family of codes from Construction 2 with outer code C_{out}, inner code C_{in}, the mappings Φ_j, the set of keys T, and $\pi(j) = |T|^{-1}$. For $q > c^2$, and any R_{out}, σ satisfying*

$$R_{\text{out}} < \frac{1 - \sigma}{c(c+1)}, \qquad \varepsilon_{\text{in}} < \sigma < \frac{q - c^2}{q - c},$$

there exists a c-secure with ε-error family of binary codes $\mathcal{C} = \{C_j\}_{j \in T}$ with length $m = n\,l$, rate

$$R = R_{\text{out}}\, R_{\text{in}},$$

error probability ε decreasing exponentially as

$$\varepsilon \leq 2^{-m\left(\frac{1-\sigma}{c} R_{\text{in}} - (c+1)R + o(1)\right)} + 2^{-qD(\sigma \| \varepsilon_{\text{in}})},$$

and with a polynomial-time identification algorithm that runs in $O(\text{poly}(m))$.

Above, $D(\sigma\|\varepsilon_{\text{in}})$ denotes the Kullback-Leibler divergence between two Bernoulli random variables of parameters σ and ε_{in}, respectively, which satisfies $D(\sigma\|\varepsilon_{\text{in}}) > 0$ for $\sigma \neq \varepsilon_{\text{in}}$.

Recall that the rate of the family of codes (9) as well as the rate from Corollary 4 are only existential results, derived from probabilistic arguments in [3] and [13,18], respectively. Now, combining the novel Construction 1 together with Corollary 4 we can derive an explicit construction of a family of fingerprinting codes, with exponentially small error and an efficient identification algorithm.

Corollary 5 *Let q be a prime power, and c be an integer $c^2 < q$. Moreover, let ε_{in} and σ be such that*

$$\varepsilon_{\text{in}} < \sigma < \frac{q - c^2}{q - c}.$$

Then, for any fixed rate R satisfying

$$R < \frac{1-\sigma}{c(c+1)} \left(c^2 2^{3c-2} \log_2 q\right)^{-1},$$

there is an explicit construction of a c-secure with ε-error family of binary codes $\mathcal{C} = \{C_j\}_{j \in T}$ with length m, rate R, error probability ε decreasing exponentially as

$$\varepsilon \leq 2^{-m\left(\frac{1-\sigma}{c} R_{\text{in}} - (c+1)R + o(1)\right)} + 2^{-qD(\sigma\|\varepsilon_{\text{in}})},$$

and with a polynomial-time identification algorithm that runs in $O(\text{poly}(m))$.

As noted in [13,18], the use of almost secure frameproof codes instead of secure frameproof codes introduces an additional error term in the identification process, compared to the codes presented in [3]. Note again that this error term decreases exponentially with the outer code length.

6 Conclusion

In this paper, we have presented explicit constructions of almost secure frameproof codes, which are a relaxed version of secure frameproof (i.e., separating) codes. Our work has started with the study of the connection between weakly dependent arrays and universal sets, and the subsequent connection between universal sets and secure frameproof codes.

Starting with these ideas, we have first introduced a relaxation in the definition of a universal set, which helped us to transition from secure frameproof codes to almost secure frameproof codes. We show how almost universal sets and weakly biased arrays can be used to derive almost secure frameproof codes. This observation has lead us to explicit constructions of such codes.

As expected, these explicit constructions are somewhat far from the theoretical existence bounds shown in earlier works. For example, probabilistic arguments show the existence of asymptotically almost 2-secure frameproof families of codes of rate $R = 0.2075$, whereas the constructions that we have discussed provide codes of rate below this figure. Nevertheless, our work shows the existence of constructible almost secure frameproof codes of much higher rate than secure

frameproof codes based on weakly biased arrays. Also, the main point of our work is to present the first explicit and practical-use constructions for such codes.

Additionally, we have also shown how the proposed constructions can be used to explicitly construct a secure family of fingerprinting codes. The construction presented is based on the theoretical existence results of previous works, which assume the existence of almost secure frameproof codes. Hence, another of the main contributions of the present work has been to provide a real implementation of such a theoretical existence result for a fingerprinting scheme. Replacing separating codes by almost secure frameproof codes introduces an additional error term in the identification of guilty users that decreases exponentially with the outer code length.

Finally, we would like to note that even though a universal set is a separating code, the relationship between an almost universal set and an almost separating code is by no means evident and will we the subject of future research.

Acknowledgements We would like to thank the anonymous reviewers, whose valuable comments helped to improve the contents and presentation of this paper.

M. Fernández has been supported by the Spanish Government through projects Consolider Ingenio 2010 CSD2007-00004 "ARES", TEC2011-26491 "COPPI", and TEC2015-68734-R (MINECO/FEDER) "ANFORA".

References

1. Alon, N., Goldreich, O., Håstad, J., Peralta, R.: Simple constructions of almost k-wise independent random variables. Random Struct. Alg. **3**(3), 289–304 (1992)
2. Alon, N., Guruswami, V., Kaufman, T., Sudan, M.: Guessing secrets efficiently via list decoding. ACM Trans. Alg. **3**(4), 1–16 (2007)
3. Barg, A., Blakley, G.R., Kabatiansky, G.: Digital fingerprinting codes: Problem statements, constructions, identification of traitors. IEEE Trans. Inf. Theory **49**(4), 852–865 (2003)
4. Bierbrauer, J., Schellwat, H.: Almost independent and weakly biased arrays: Efficient constructions and cryptologic applications. In: Proc. Int. Cryptol. Conf. (CRYPTO), *Lecture Notes Comput. Sci. (LNCS)*, vol. 1880, pp. 533–544. Santa Barbara, CA (2000)
5. Boneh, D., Shaw, J.: Collusion-secure fingerprinting for digital data. In: Proc. Int. Cryptol. Conf. (CRYPTO), *Lecture Notes Comput. Sci. (LNCS)*, vol. 963, pp. 452–465. Santa Barbara, CA (1995)
6. Boneh, D., Shaw, J.: Collusion-secure fingerprinting for digital data. IEEE Trans. Inf. Theory **44**(5), 1897–1905 (1998)
7. Chor, B., Fiat, A., Naor, M.: Tracing traitors. In: Proc. Int. Cryptol. Conf. (CRYPTO), *Lecture Notes Comput. Sci. (LNCS)*, vol. 839, pp. 480–491. Santa Barbara, CA (1994)
8. Chor, B., Fiat, A., Naor, M., Pinkas, B.: Tracing traitors. IEEE Trans. Inf. Theory **46**(3), 893–910 (2000)
9. Chor, B., Goldreich, O., Hasted, J., Freidmann, J., Rudich, S., Smolensky, R.: The bit extraction problem or t-resilient functions. In: Proc. IEEE Symp. Found. Comput. Sci. (FOCS), pp. 396–407. Singer Island, FL (1984)
10. Cohen, G.D., Encheva, S., Schaathun, H.G.: On separating codes. Tech. Rep. 2001D003, ENST, Paris, France (2001)
11. Cohen, G.D., Schaathun, H.G.: Asymptotic overview on separating codes. Tech. Rep. 248, Deptartment of Informatics, University of Bergen, Norway (2003)
12. Cohen, G.D., Schaathun, H.G.: Upper bounds on separating codes. IEEE Trans. Inf. Theory **50**(6), 1291–1294 (2004)
13. Fernández, M., Kabatiansky, G., Moreira, J.: Almost separating and almost secure frameproof codes. In: Proc. IEEE Int. Symp. Inform. Theory (ISIT), pp. 2696–2700. Saint Petersburg, Russia (2011)
14. Forney, G.D.: Concatenated codes. Tech. Rep. 440, Res. Lab. Electron., MIT, Cambridge, MA (1966)

15. Friedman, A.D., Graham, R.L., Ullman, J.D.: Universal single transition time asynchronous state assignments. IEEE Trans. Comput. **C-18**(6), 541–547 (1969)
16. Körner, J., Simonyi, G.: Separating partition systems and locally different sequences. SIAM J. Discr. Math. (SIDMA) **1**(3), 355–359 (1988)
17. Moreira, J., Fernández, M., Kabatiansky, G.: Constructions of almost secure frameproof codes based on small-bias probability spaces. In: Proc. Int. Workshop Security (IWSEC), *Lecture Notes Comput. Sci. (LNCS)*, vol. 8231, pp. 53–67. Okinawa, Japan (2013)
18. Moreira, J., Fernández, M., Kabatiansky, G.: Almost separating and almost secure frameproof codes over q-ary alphabets. Des. Codes Cryptogr. **80**(1), 11–28 (2016)
19. Naor, J., Naor, M.: Small-bias probability spaces: Efficient constructions and applications. SIAM J. Comput. (SICOMP) **22**(4), 838–856 (1993)
20. Pinsker, M.S., Sagalovich, Y.L.: Lower bound on the cardinality of code of automata's states. Probl. Inform. Transm. **8**(3), 59–66 (1972)
21. Sagalovich, Y.L.: Completely separating systems. Probl. Inform. Transm. **18**(2), 140–146 (1982)
22. Sagalovich, Y.L.: Separating systems. Probl. Inform. Transm. **30**(2), 105–123 (1994)
23. Staddon, J.N., Stinson, D.R., Wei, R.: Combinatorial properties of frameproof and traceability codes. IEEE Trans. Inf. Theory **47**(3), 1042–1049 (2001)
24. Stinson, D.R., van Trung, T., Wei, R.: Secure frameproof codes, key distribution patterns, group testing algorithms and related structures. J. Stat. Plan. Infer. **86**(2), 595–617 (2000)
25. Tardos, G.: Optimal probabilistic fingerprint codes. J. ACM **55**(2), 1–24 (2008)
26. Vazirani, U.V.: Randomness, adversaries and computation. Ph.D. thesis, Dept. Elect. Eng. Comp. Sci., Univ. California, Berkeley (1986)
27. Vazirani, U.V., Vazirani, V.V.: Efficient and secure pseudo-random number generation. In: Proc. IEEE Symp. Found. Comput. Sci. (FOCS), pp. 458–463. Singer Island, FL (1984)