Journal of Computations & Modelling, vol.3, no.3, 2013, 57-73
ISSN: 1792-7625 (print), 1792-8850 (online)
Scienpress Ltd, 2013

Analysis total and better the Boneh–Shaw Fingerprinting codes

Solmaz Sharifnia[1]

Abstract

Digital fingerprinting is a forensic method against illegal copying. The distributor marks each individual copy with a unique fingerprint. If an illegal copy appears, it can be traced back to one or more guilty pirates due to this fingerprint. To work against a coalition of several pirates, the fingerprinting scheme must be based on a collusion– secure code. This paper addresses binary collusion–secure codes in the setting of Boneh and Shaw (1995/1998). We prove that the Boneh-Shaw scheme is more efficient than originally proven, and we propose adaptations to further improve the scheme.

We also show how to use these codes to construct binary fingerprinting codes with length L = o (C^6 log c log$n$).

Mathematics Subject Classification: 94A60; 68R05; 68P30

Keywords: Fingerprinting; Collusion; Traitor tracing; Watermarking

[1] Department of Mathematics and Computer Science, Shahed University, Tehran, Iran.

Article Info: *Received* : June 8, 2013. *Revised* : July 28, 2013
Published online : September 15, 2013

1 Introduction

Digital fingerprinting (FP) was introduced in [1], and given increasing attention following [2]. A vendor selling digital copies of copyrighted material wants to prevent illegal copying.

Digital FP is supposed to make it possible to trace the guilty user (pirate) when an illegal copy is found. This is done by embedding a secret identification mark, called a fingerprint, in each copy, making every copy unique. FP can also be used to protect broadcast encryption keys (i. e, cable TV decoders), in which context it is usually referred to as traitor tracing [3].

The fingerprint must be embedded in such a way that it does not disturb the information in the data file too mach. It must also be impossible for the user to remove or damage the fingerprint, without damaging the information contents beyond any practical use. In particular, the fingerprint must survive any change of file format (e.g. gif to tiff) and any reasonable loss compression. This embedding problem is essentially the same as the problem of watermarking.

If a single pirate distributes unauthorized copies, they will carry his or her fingerprint. If the vendor discovers the illegal copies he or she can trace them back to the pirate and prosecute him or her. If several pirates collude, they can to some extent tamper with the fingerprint. When they compare their copies, they see some bits (or symbols) which differ and, thus, must be part of the fingerprint. Identified bits may be changed and, thus, the pirates create a hybrid copy with a false fingerprint. Collusion – secure coding is required to enable tracing of at least one pirate where a coalition of pirates have colluded.

In this paper, we study binary, concatenated, FP schemes generalizing and improving the approach of [2]. In section 2, we will define the FP model, which we refine a little compared to past literature. In section 3, we give the main result, which is an improved error analysis of the BS FP scheme and new variants of it.
Section 4 gives deals with the construction of binary fingerprinting codes by concatenating algebraic – geometric codes with the Boneh – Shaw code.

2 Fingerprinting Problem

2.1 Preliminaries from Coding Theory

An $(n, M)_{\mathrm{q}}$ code is a set of M words of n symbols from an alphabet Q of q elements. The hamming distance between two words x and y is denoted d(x, y), and the minimum distance of a code C is denoted d. The normalized minimum distance is $\delta = d/n$.

Closet neighbor is any algorithm which takes a word x and returns a word c ∈ C such that d(c, x) is minimized. This can always be performed in O(nM) operations and, for some codes, it may be faster. For the error analysis, we will use the well-known Chernoff bound as given in the following theorem. See, for example, [5] for a proof. The relative entropy function is defined as

$$\mathrm{D}\,(\sigma||p) = \sigma \log_2 \frac{\sigma}{p} + (1 - \sigma) \log_2 \frac{1-\sigma}{1-p} \tag{1}$$

Theorem 2.1(Chernoff) Let X_1 ,…, X_t be bounded, independent, and identically distributed stochastic variables in the range [0,1]. Let x be their (common) expected value. Then, for any $\delta > x$, we have

$$\mathrm{P}\ (\textstyle\sum_{i=1}^{t} Xi \geq t\delta) \leq 2^{-tD(\delta\|x)}.$$

We write $B(n, p)$ for the binomial distribution with n trials with probability$p$. If X is distributed as $B(n, p)$, we write $X{\sim}B(n, p)$. All logarithms will be to base 2 unless otherwise stated.

2.2 BS Model and Marking Assumption

Our model follows that of Boneh and Shaw (BS) [2]. Let u = $(u_1, \dots, u_n)$ be a digital file divided into n segments $u_i \epsilon U$. We call U the file alphabet. Let M be the set of users. We can view $i \in M$ as a customer account number. Write M = $\#M$. The model assumes a watermarking scheme with an embedder allowing us to hide a symbol x from some alphabet Q in a file segment $u_{i,}$ producing a watermarked file segment $w_i \epsilon U$. The extractor inverts the embedding; given $w_{i\,,}$ it outputs x.

The fingerprint encoder is an injective map e_K: $M \hookrightarrow Q^n$, identifying each user $i \epsilon M$ by a fingerprint $c = (c_1 \ , ... \ , \ c_n)$, where $c_i \epsilon$ Q. The image under e_K of a subset $P \subseteq M$ is denoted e_K(p). The secret key K is drawn uniformly at random from akey space $\mathcal{K}$ when the system is initialized, and kept secret by the vendor. The set $C_{K=}e_K$(M) is an $(n, M)_{\mathrm{q}}$ code called the FP code (corresponding to K).

When user i busy a copy of file u, the vendor obtains the fingerprint $(c_1, ..., c_n)$: = e_K (i), and embeds each c_i in u_i to obtain the fingerprinted file w = $(w_1 \, , ... \ , \ w_n)$, which is handed to the user.

If user i naively copies and redistributes w, then the watermarks can be extracted to obtain the fingerprint $X = (x_1, ..., x_n)$, which identifies the pirate i who can be prosecuted. If several pirates collude, they can cut and paste segments from their different copies, thus, the output x_i from the extractor will match one of the pirate fingerprint, but the full sequence $(x_1, ..., x_n)$ is a hybrid fingerprint which matches none of the pirates.

The BS model presumes that this cut – and – paste attack is the only one available to the pirates. This is expressed formally by the marking assumption.

Definition 2.1 (The Marking Assumption) Let $P \subseteq \ C_K$ be the set of fingerprint held by a coalition of pirates. The pirates can produce a copy with a false fingerprint x for any $x \ \in F(P)$, where

$$F(P) = \{(c_1, ..., c_n) : \forall_{i\,.} \exists (x_1, ..., x_n) \in p, x_{i=} \, c_i \, \}.$$

We call F(P) the feasible set of P.

A position where the pirates see at least two symbols and, thus, have a choice is called a detectable position.

Example 2.1 (Traitor Tracing) Collusion-Secure codes are used for traitor tracing [3], where Definition 2.1 is satisfied as follows. The system uses a $q \times n$

matrix of permanent keys $K_{j,i}$. Each row corresponds to an $(c_1, \ldots, c_n)$ alphabet symbol and each column to a coordinate position. The user with fingerprint $(a_{1,}, \ldots, a_n)$ receives the key$K_{a_i,i}$for every i. The session key is the exclusive or of n elements s_1 to s_n. An enabling block is transmitted at the start of each session consisting of $e_{K_{j,i}}(s_i)$ for each i and j, where e_K is the encryption function for key K. To get the session key, one key from each column of the matrix is required, and that is exactly what each user has. When the pirates make a pirate decoder box, they must supply it with a key for each coordinate position from one of their true fingerprints and, thus, the marking assumption is satisfied.

Some authors use alternative marking assumption. Some models assume that the pirates can output any symbol in a detectable position, or they may be allowed to output an erasure (no valid symbol) in detectable position. See [6] and [7] for details.

Muratani [8] uses a stronger assumption where the pirates output each word in the feasible set with equal probability, allowing much shorter codewords. Some authors assume that the pirates have a certain probability $p_e$of outputting a random symbol from the alphabet in every column. These lead to the study of error- and collusion- secure codes [9].

2.3 FP Scheme

An $(n, M)_{\mathrm{q}}$ FP scheme is an ensemble S = $\{(e_{K},A_K): K \in \mathcal{K}\}$, where e_K is the encoding as defined in the previous section, and the FP code C_K is an $(n, M)_{\mathrm{q}}$ code. The tracing algorithm A_K takes a hybrid fingerprint X as an input and outputs a set $L \subseteq M$. It is successful if L is a nonempty subset of the pirates. The rate of the FP scheme is the rate of the code C_K, namely $R = (log_q M)/n$.

We adhere to Kerchoff’s principles, so the key is random and secret, and everything but the key is public information. That is, the pirates know the definition of e_k and A_k. If the entire system is leaked, a new random key can be chosen for the same scheme, and it will be secure for future applications (until the key is again compromised).

$$The\ game\ proceeds\ in\ the\ following\ steps.$$

1. The vendor chooses the FP scheme σ to use for the product he or she is selling; this is the vendor strategy. We assume that this is known to the pirates.
2. The key K is chosen at random, and kept secret by the vendor.
3. The copies of the digital data are generated using the FP scheme and the key, and distributed to the users. A coalition $P \subseteq M$ of t(potential) pirates is thus assigned fingerprints$e_K(P)\ \subset C_K$, and receives a set $Y = \{\, w_1, \dots, w_t\, \}$ of fingerprinted copies.
4. The coalition of potential pirates gets together and compares the copies Y.
5. If P_r(Error|Y=y) is sufficiently low, the pirates seeing y will opt out, without committing any crime.
6. If the pirates chose to play, they choose a strategy for selecting segments from the different copies, paste together a hybrid copy, and sell the copies with a hybrid fingerprint x.
7. If and when an illegal copy is discovered, the vendor extracts the hybrid FP x, computes $A_K(x)$, and prosecutes any users traced.

Note that we have three outcomes of the game. The pirates can choose not to play (Event 0). If they do play, we get a random outcome, either error where the pirates win, or ¬error where the pirates lose.

For an event E, Let $P_{K,\psi}(E)$ denote the probability of E under the distribution induced by the uniform distribution of K under the assumption that the pirates always play regardless of y and choose Strategy ψ in Step 6. The pirates want to escape and will therefore choose ψ to maximize $P_{K,\psi}$(E), so let P_K (E) =

ma$x_{\psi}P_{K,\psi}(E)$. The traditional definition of collusion-secure codes is based on the unconditional probability P_K(Error), given no information about K or Y.

Definition 2.2 (Weak Security) An FP scheme S is (weak) (t,∈)- secure if, for any υ ≤ t,P_K (Error) ≤ ∈ when a set P ⊂ M of υ pirates is drawn uniformly at random. Unauthorized copying is a criminal act (in most countries), and pirates that are caught will therefore be subject to punishment. The primary reason for assigning punishment is to deter potential pirates. The vendor's goal is not necessarily to win the game (make the pirates lose). Deterring the pirates (Event 0) obtaining a stalemate where nobody wins and nobody loses is perfectly satisfactory.

The pirates choose whether to play or not in Step 5, according to their perceived probability of escaping ($i.e., P_K$ (Error |Y =y)). This probability can be higher or lower than the unconditional probability p_K(Error). We expect that there is a threshold e_p such that the pirates choose to play if and only if P_K(Error |Y =y)>e_p. If P_K(Error |Y =y) $< e_p$, we get Event 0. Note that a (weak) (t, e_p)-secure code is not sufficient to deter all pirate coalitions of size t or less. This is why we introduce a new and stronger definition.

Definition 2.3 (Strong Security) An FP scheme S is strongly $(t, \in)$- secure if, for any y seen by at most t pirates, we have P_K(Error |Y =y)≤ ∈.

Clearly, a strongly t-secure FP scheme will deter any pirate coalition of size at most t if $\in \leq\ e_p$. By abuse of language, we shall sometimes say that C_K is $(t, \in)$-secure when the scheme is.

Definition 2.4 Let S be an FP scheme. A priori error bound $\in_1$is the smallest ∈ such that S is (weak) $(t, \in)$-secure. A posteriori error bound $\in_2$ is the smallest ∈ such that S is strongly $(t, \in)$- secure.

Even though the explicit definition of strong t-secure codes is new, many previous schemes do meet the definition, including the BS scheme. Some authors bound P_K (Error |P = p) for any p, which clearly bounds $\in_2$.

The phenomenon that pirates with a higher escape probability are more likely to play is called adverse selection in economics and game theory. This affects the following interesting probability:

$$\in_A = \mathrm{pr}(\mathrm{Error|the\ pirates\ did\ play}).$$

Of course, when the vendor finds an illegal copy, he or she knows the pirates have played, and it is interesting for him or her (and for the court if the FP scheme is used as evidence). What the error probability is under this condition. The following lemma gives some information about this.

Lemma 2.2 When the vendor obtains an illegal copy having only the knowledge of the key K and the false fingerprint x, the probability $\in_A$ of getting an incorrect output from A_K(x) is, at most, the posteriori error bound $\in_2$.

Proof. Since $\in_2$ bounds the conditional error probability for any information that the pirate could have; in particular, it bounds this probability for any information which would induce the pirates to play. Hence, $\in_2$ also bounds the probability of error under the condition that the pirates play. □

Lemma 2.3 Unless P_K(Error |Y = y) is constant over all y, there is a pirate strategy suchthat

$$\in_1 < \in_A.$$

Proof. Write $\varepsilon_K(y) = P_K(Error\ |Y = y)$, and write

$$\in_1 = \sum_y P_K\,(Y = y)\varepsilon_K(y).$$

Suppose the pirates choose to play whenever $\varepsilon_K(y) > \in_1$.

Write $\overline{Y} = \{y\colon \varepsilon_K(y) > \in_1\}$. Then, we get that

$$\in_A = \sum_{y\,\in \bar{Y}} \frac{P_K\,(\,Y = y)}{\sum_{y\,\in\,\bar{Y}} P_{K\,(Y=y)}} . \varepsilon_K(y).$$

We clearly get that $\in_A > \in_1$ since we have removed only small terms from the average. □

To summarize, if the pirates decide to do illegal copying before they see their copies, their chance of escape is at most $\in_1$. For any pirate collusion of size, at most, t having compared their copies, the chance of escape is, at most, $\in_2$. Which error bound is the most important will depend on the application.

Definition 2.5 (Errors and Failures) Let L $\subseteq M$ be the output of the tracing algorithm and P $\subseteq M$ be the pirate collusion. An error of type 1 (or a failure) is the event that L=$\emptyset$. An error of type 2 (or false accusation) is the event that L $\nsubseteq$ P. In the context of criminal law, we know that type 2 errors are a serious matter. Frequent type 1 errors mean that we often do not get useful output, but they do not affect the reliability of the output, which is obtained. If type 2 errors are frequent, the output cannot be trusted even when we have output. For the rest of this paper, $\in_1$will denote a posteriori probability of type 1 error, and $\in_2$ a posteriori probability of type 2 error.

A scheme with such error probabilities is said to be (strongly) $(t, \in_{1,} \in_2)$- secure.

3 Concatenated Schemes

In this chapter, we develop a general analysis of concatenated FP schemes. Such concatenation was applied in [2], but our error analysis will prove that those constructions have a better error rate than originally proven. We make the following formal definition of concatenated schemes.

Definition 3.1 (Concatenated Fingerprinting Scheme) Let$S_I = (e_K^I, A_K^I)$ be an$(n_I, q)_2$ FP scheme, and $S_o = (e_K^o, A_K^o)$ an $(n_o, M)_q$ FP scheme. Let Q denote the alphabet of S_o.

A concatenated FP scheme $S = S_O$o $S_I = (e_K, A_K)$ consists of the following elements. The key is a tuple K = $(K_o, K_1, \dots, K_{n_o})$, where K_o is a key for S_o and K_i are keys forS_I. The encoding is

$$e_K(u) = e^I_{K_1}(c_1)\|e^I_{K_2}(C_2)\|\ldots\|e^I_{K_{n_o}}(C_{n_o}) \qquad (2)$$

Where $(C_1, \dots, C_{n_o}) = e^O_{K_o}(u)$ and $u \in M$.

Each segment $e^I_{k_i}(c_i)$ of the word is called a block. The algorithm A_K first decodes each block using $A^I_{K_i}$, and then decodes the resulting word over Q using $A^I_{k_o}$.

Note that the FP code of S is the concatenation of the FP codes of S_I and-S_o. Let R_1 and R_o denote the rates of S_I and S_o, respectively. We demand that S_I is strongly $(t, \in_{in})$ secure, but our analysis is otherwise oblivious to its structure.
On the other hand, the error analysis must be made separately for each type of outer scheme S_O, but this scheme does not have to be collusion-secure in it.

3.1 BS Concatenated Code

The following $(q, \in)$-secure scheme S_1 was used in [2]. Let C^I_l be a (r(q – 1), q)$_2$ code with a codebook consisting of q – 1 distinct columns, each replicated r times. A set of identical columns will be called a type. Every column has the form (1…10…0)^T, such that the ith (1≤ $i \leq q$) user has zeroes in the first $i - 1$ types and a one in the rest.

Example 3.1 The BS inner code for r = 3 and q = 5 is the set of the following five codewords:

$$c_1 = (111\ 111\ 111\ 111)$$
$$c_2 = (000\ 111\ 111\ 111)$$
$$c_3 = (000\ 000\ 111\ 111)$$

$$c_4 = (000\ 000\ 000\ 111)$$

$$c_5 = (000\ 000\ 000\ 000)$$

The key K maps the code C_ι^I onto an equivalent code C_K^I by permuting the columns. View ι as the identity. We can see that unless user i is a pirate, the pirates cannot distinguish between the $(i-1)$th and the ith type. Hence, they have to use the same probability of choosing a 1 in both of these types. The tracing algorithm $A_{K_I}^I$ uses statistics to test the null hypothesis that user *i* be innocent. The output is some $user(s)\ for\ whom\ the\ null\ hypothesis\ may\ be\ rejected.$

The key size in bits is

$Log\#\mathcal{K} = log\frac{(r\,(q-1))!}{(r!)^{q-1}}.$

$$The\ probability\ of\ accusing\ a\ given\ innocent\ user\ is\ bounded\ as$$

$$\hat{\in} \le 2^{1-\frac{r}{2q^2}}.$$

Theorem 3.1 (BS) The BS inner code with replication factor r is strongly (q,∈) secure wheneverr $\geq 2q^2 \log(2q/\in)$.

Let RC (Random Codes) BS-RC be the scheme S= S_O o S_I with S_I as described above and a random code with list decoding for S_O. There are several control parameters which may be used to tune the performance of the system. The inner code cardinality q is the trickiest one. Most of the time, we will follow BS and set q = 2t.

Obviously, n_O and r control a tradeoff between the code length and error rate, and Δ controls the tradeoff between the two error types.

Theorem 3.2 If we use

$$q = 2t, \qquad \Delta = \frac{t}{t+1}, \qquad r = 2q^2 \log(4qt)$$

Then BS-RC is a strongly $(t, \in)$-secure $(n, M)_2$ FP scheme, where

$$n = (2t-1)\lceil 8t^2\ (3 + 2\log t)\rceil n_O, \tag{3}$$

$$n_O = \frac{maX\{-log\in_1, logM - log\in_2\}}{D\left(\frac{1}{t+1}\middle\|\frac{1}{2t}\right)}. \tag{4}$$

Asymptotically, the length is

$$n = \theta\left(t^4(logt)(logM - log\epsilon)\right). \quad (5)$$

Table 1: Some Lengths When t = log M, $\mathcal{E} < \frac{1}{t+1}$

t = log M	Boneh and Shaw	BS - RC
10	$6.64 .10^{18}$	$\mathbf{3.06 .10^{8}}$
15	$3.91 .10^{9}$	$\mathbf{1.76 .10^{9}}$
20	$1.40 .10^{10}$	$\mathbf{6.44 .10^{9}}$
25	$3.80 .10^{10}$	$\mathbf{1.77 .10^{10}}$
30	$8.68 .10^{10}$	$\mathbf{4.09 .10^{10}}$

4 Concatenated Algebraic Geometric (AG) Codes with Boneh-Shaw codes

Let W be a [*n*,*k*,*d*] code over F_q, where *n* is the code length, *k*the dimension, that is, the code has N = q^k codewords(users), and *d* the minimum Hamming distance of the code.

Let V be a BS (q, r) c-secure code. Then the concatenated code C = V o W is the code obtained by taking the words W= $(w_1 , \dots , w_n) \in W \subset F_q^n$ and mapping every symbol $w_i \in F_q$ on a word V $(w_i) \in V$. The code W is called the outer code and V the inner code.

Concatenated codes are often decoded by first decoding the inner code, thus obtaining a word of symbols from the outer code alphabet. Then, in a second step, this word is decoded with a decoding algorithm designed for the outer code. In what follows, we first determine the properties that the outer code needs to meet in order to determine a good fingerprinting code when concatenated with a

BS code. Then we show that such a code exists and finally we show how to efficiently identify a coalition member using this code [10].

Theorem 4.1 Let W be a [n, d] cod over F_q, with d> $n - n(1-\sigma)/c^2$, with 0 $< \sigma < 1$. Let V=BS(q, r) be a c-secure Boneh-Shaw code with error probability ϵ_B, where $\epsilon_B < \sigma$. Then the concatenated code C = V o W is a c-secure fingerprinting code with error probability

$$p_e = \exp\big(-\Omega(n)\big).$$

Proof. Let U = { $c_1, \dots, c_c$} be the codewords associated to a c-coalition, where $c_j = \left(V\big(c_1^j\big), \dots, V\big(c_n^j\big)\right) \in C$ and $c^j = \big(c_1^j, \dots, c_n^j\big) \in W$, for 1≤ $j \leq c$.

Given a false fingerprint

Z = $(z_1^1, \dots, z_{(q-1)_r}^1, \dots, z_1^n, \dots, z_{(q-1)r}^n)$

When decoding each block $Z^j = (z_1^j, \dots, z_{(q-1)r}^j)$ using the decoding algorithm for the BS inner code V, we obtain a symbol $Z_j \in F_q$ that, with probability 1- ϵ_B belongs to one of the codewords of some member of the c-coalition, in other words, $p\big(Z_j \in \{c_j^1, \dots, c_j^c\}\big) \geq 1 - \epsilon_B$.

Clearly, the error probability of each symbol is independent, thus we can model the number of errors produced in the inner decoding process by n Bernoulli random variables θ_i, equal to 1 with probability $\epsilon_B$and 0 with probability 1 - ϵ_B .

The probability of the tail can be bounded as

$$p(\textstyle\sum_{i=1}^{n} \theta_i \geq n\sigma) \leq 2^{-nD(\sigma||\epsilon_B)}.$$

Where $\sigma > \epsilon_B$and $D(\sigma||\epsilon_B) = \sigma log_2(\sigma/\epsilon_B) + (1-\sigma)log_2((1-\sigma)/(1-\epsilon_B))$.

Thus, after decoding the inner code, we recover a false fingerprint Z= (Z$_1$,…, Z$_n$) $\in F_q^n$ where $p(|\{Z_j \,|Z_j \in \{c_j^1, \dots, c_j^c\}\}\,| \geq n - n\sigma) \geq 1 - 2^{-nD(\sigma||\epsilon_B)}$.

That is, with error probability less than $2^{-nD(\sigma||\epsilon_B)}$, there exists some coalition codeword $c_j \in U$ such that $d\big(Z, c^j\big) \leq n - n(1-\sigma)/c$.

From the hypothesis of the theorem, any two codewords u, w $\in W$, satisfy

$$d(u, w) > n - n\frac{1-\sigma}{c^2}.$$

Now, for any $v \in W$, not a coalition member codeword,

$$n - d(v, Z) \leq \sum_{j=1}^{c} (n - d(v, c^j)) \; < \; c(n\frac{1-\sigma}{c^2}) = n\frac{1-\sigma}{c}.$$

Thus, for any $v \; \in W$, not a coalition member codeword, $d(v, Z) > d(c^j$, Z), for some $c_j \; \in U$, with error probability less that $2^{-nD(\sigma||\epsilon_B)}$. As the code contains q^k codewords, we have that with error probability $p_e \; \leq \; q^k 2^{-nD(\sigma||\epsilon_B)}$, a codeword in W associated to a coalition member is close to the false fingerprint Z than any other code word in W, thus we can identify by minimum Hamming distance one of the members of the coalition, proving the theorem. □

Theorem 4.2 For any $\alpha \; > 0$, there are constructible, infinite families of codes with parameters $[N, NR, N\delta]_q$ for $N \geq \; N_0(\alpha)$ and

$$R + \; \delta \; \geq 1 - (\sqrt{q} - 1)^{-1} - \alpha.$$

Observe that these codes require $q > 1/(1-\delta)^2$, and recall that 1 - $\delta \; < t^{-2}$. Hence, the AG codes require$q > t^4$.

From the previous theorem, we know the properties of the desired codes, but it remains to prove the existence of such codes.

Next theorem proves the existence of the codes determined by theorem 4.1, and relates the number of users (code words) with the length and the error probability of the c-secure concatenated fingerprinting code.

Theorem 4.3 There exist c-secure fingerprinting codes with N codewords, length $L = O\;(c^6 \log c \;\; logN), and$ error probability $p_e = O(1/N)$.

Proof. By [11] we have families of Algebraic-Geometric codes (AG), with parameters [n, k, d], over a finite field F_q, whose parameters asymptotically approach the Tsfasman-Vlăduţ-Zink bound $k/n \geq 1 - 1/\left(\sqrt{q} - 1\right) - d/n$.

These codes satisfy $n = O(logN)$, where N is the number of codewords.

Let W be one of the AG codes that approach the Tsfasman-Vlăduţ-Zink bound, with $d \; > n - n(1-\sigma)/c^2$, where $0 \; < \; \sigma \; < 1$, then

$$n\left(1-\frac{1-\sigma}{c^2}\right)<d\ <n\left(1-\frac{1}{\sqrt{q}-1}\right)$$

that is, a sufficient condition for the existence of such a code is

$$1-\ \sigma\ >\frac{c^2}{\sqrt{q}-1}. \tag{6}$$

but as $0\ <\ \sigma\ <1,$ if $\sqrt{q}-1\ >c^2$ the code exists.

The length L_B of the inner code BS (q, r), by [10] we have:

$$L_B\ \geq\ 8q(c+\ \overline{\sqrt{c-1})}^2\ log\frac{4q}{\epsilon_B},$$

Where $\epsilon_B<\ \sigma$. By the inequality in (6) we have $q=O(c^4)$, thus

$$L_B=O(c^6\log c).$$

Therefore, the length of the concatenated code $C=BS(q,r)oW$ is $L=\ L_{B^n}=O(c^6\ logc\ logN)$.

Moreover, as the code satisfies the conditions in theorem 4.1 we have that $p_e\ \leq\ q^k2^{-nD(\sigma||\epsilon_B\,)}$, thus proving the theorem. □

Theorem 4.4 Let W be an *Algebraic-Geometric* code [n, k, d] $overF_q$ with N = q^kcode words, where $d\ >n-n(1-\sigma)/c^2and\ \sqrt{q}>c^2\ /(1-\sigma)^2+1$. Let V be a BS (q, r) c-secure Boneh-Shaw code with error probability $\epsilon_B<\ \sigma$. Then the concatenated code C = V o W is a c-secure fingerprinting code with error probability$p_e\ \leq\ q^k2^{-nD(\sigma||\epsilon_B\,)}$, length L= O($c^6\ logc\ logN$) and identification algorithm complexity poly(logN).

5 Conclusion

We have studied concatenated collusion-secure codes. As inner codes, we suggest separating codes in the two pirate case, and the BS inner code in the general case. As outer codes, we propose random codes.

One of the schemes, BS-RC, is the classic of [2] but our analysis shows length can be less than previously assumed. Samples with an error rate of 10^{-10} show a reduction by a factor of about 2.1.

We know of one other strongly (t,ϵ)-secure scheme and define new codes that use asymptotically good AG codes concatenated with Boneh-Shaw codes. The error probability of the concatenated construction is O(1/N)=$exp(-\Omega(logN))$, with length of order L=O($c^6\, logc\, logN$), and a decoding (tracing) algorithm of complexity poly (log N).

References

[1] N. R. Wagner,Fingerprinting, in *Proc. Symp. Security Privacy*, (1983), 18-22.

[2] D. Boneh and Shaw, Collusion-secure fingerprinting for digital data, *IEEE Trans. Inf. Theory*, **44**(5), (Sept., 1998), 1897- 1905.

[3] B. Chor, A. Fiat, M. Naor and B. Pinkas, Tracing traitors, *IEEE Trans. Inf. Theory*, **46**(3), (May, 2000), 893-910.

[4] A. Somekh-Baruch and N. Merhav, On the capacity game of private fingerprinting systems under collusion attacks, *IEEE Trans. Inf. Theory*, **51**(3), (Mar., 2005), 884-899.

[5] T. Hagerup and C. R$\ddot{u}$b, A guided tour of Chernoff bounds, *Inf. Process. Lett.*, **33**, (1990), 305-308.

[6] A. Barg, G. R. Blakley and G. A. Kabatiansky, Digital fingerprinting codes: Problem statements, constructions, identification of traitors, *IEEE Trans. Inf. Theory*, **49**(4), (Apr., 2003), 852-865.

[7] K. Yoshioka, J. Shikata and T. Matsumoto, Collusion secure codes: Systematic security definitions and their relations, I*EICE Trans. Fundamentals*, **E87-A**(5), (May, 2004).

[8] H. Muratani, Optimization and evaluation of randomized c-secure CRT code defined on polynomial ring, in Information Hiding 2004, J. Fridrich, Ed., 2004, *Lecture Notes Comput. Sci*., **3200**, 282-292.

[9] H.-J. Guth and B. Pfitzmann, Error- and collusion-secure fingerprinting for digital data, *in Proc. Information Hiding*, New York, 2000, *Lecture Notes Comput. Sci*., Springer-Verlag, **1768**, (2000), 134-145.

[10] H. Lipmaa, M. Yung and D. Lin, Obtaining Asymptotic Fingerprint Codes Through a New Analysis of the Boneh-Shaw codes, *Inscrypt 2006*, *LNCS*, Springer-Verlag, Berlin, Heidelberg, **4318**, (2006), 289-303.

[11] M. Tsfasman and S. Vl*ădut, Algebraic-geometric codes*, Dordrecht, The Netherlands, Kluwer, 1991.