The Boneh-Shaw fingerprinting scheme is better than we thought

Hans Georg Schaathun *Member, IEEE,*

***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 adresses binary collusion-secure codes in the setting of Boneh and Shaw (1995/98). We prove that the Boneh-Shaw scheme is much more efficient than originally proved, and we propose adaptions further improving the scheme. We also point out some differences between our model and others in the literature.

***Index Terms*—collusion-secure codes, digital fingerprinting, copyright protection, traitor tracing**

I. Introduction

Unauthorised copying and distribution of copyrighted material has received increasing attention over many years, both in research communities and in the daily press. Authors and artists depend on their income from legal sales, and unauthorised copying is often seen as a threat to these sales. For the movie or music industry, this is a question of big money.

Estimates of the losses due to illegal copying are generally disputable. There is no generally accepted method to estimate the sales that would have been achieved without illegal copying. For example, it is sometimes claimed that illegally distributed copies have a promotional effect which actually increase sales. Still, it is clear that big money are at stake and the issue receives interest from many different angles. Several countries these days change their legislation to deal more effectively with illegal distribution in new media.

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

The fingerprint must be embedded in such a way that it does not disturb the information in the data file too much. 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 lossy compression. This embedding problem is essentially the same as the problem of watermarking.

If a single pirate distributes unauthorised copies, they will carry his fingerprint. If the vendor discovers the illegal copies he can trace them back to the pirate and prosecute him. 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 to trace at least one pirates where a coalition of pirates have colluded.

Collusion-secure coding is also employed in traitor tracing [3]. Whereas fingerprinting protects the digital data in themselves, traitor tracing protects broadcast encryption keys. Many other related problems have been studied, but space does not permit us to mention them.

A collusion-secure code can be probabilistic or combinatorial. Combinatorially collusion-secure codes are able to successfully trace at least one pirate with probability 1. Using probabilistic schemes, we are satisfied with successful tracing with probability at least $1-\epsilon$ for some small error rate ϵ.

In this paper, we study binary, concatenated, fingerprinting schemes generalising and improving the approach of [2]. In Section II, we will define the fingerprinting model, which we refine a little compared to the past literature. In Section III, we give the main result, which is an improved error analysis of the Boneh-Shaw fingerprinting scheme and new variants of it. Section IV gives further improvements in the two pirate case, and we finish with a conclusion and comparison with other schemes in Section V.

II. The fingerprinting problem

A. Preliminaries from coding theory

We use notation and terminology from coding theory. The set of fingerprints is an $(n, M)_q$ code, which provides for up to M buyers, uses an alphabet of q symbols, and requires n such symbols embedded in the digital file. The Hamming distance between two words $\mathbf{x}$ and $\mathbf{y}$ is denoted $d(\mathbf{x}, \mathbf{y})$, and the minimum distance of a code C is denoted d. The normalised minimum distance is $\delta = d/n$. The rate of the code is $R = (\log M)/n$.

Closest neighbour decoding is any algorithm which takes a word $\mathbf{x}$ and returns a word $\mathbf{c} \in C$ such that $d(\mathbf{c}, \mathbf{x})$

The author is with The Selmer Centre, Department of Informatics, University of Bergen, PB 7800, N-5020 Bergen, Norway (email: georg@ii.uib.no)

This work has in part been supported by the Commission of the European Communities through the IST program under contract IST-2002-507932, and by the Norwegian Research Council under Grant 146874/20.

is minimised. 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 e.g. [4] for a proof. The relative entropy function is defined as

$$D(\sigma||p) = \sigma \log \frac{\sigma}{p} + (1-\sigma) \log \frac{1-\sigma}{1-p}. \quad (1)$$

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

$$P\left(\sum_{i=1}^{t} X_i \leq t\delta\right) \leq 2^{-tD(\delta||x)}.$$

We write $\mathcal{B}(n,p)$ for the binomial distribution with n trials with probability p. If X is distributed as $\mathcal{B}(n,p)$, we write $X \sim \mathcal{B}(n,p)$.

B. The fingerprinting scheme

Let U denote the set of users. An $(n,M)_q$ fingerprinting (FP) scheme is an ensemble $S = \{(C_K, e_K, A_K) : K \in \mathcal{K}\}$, where

1) $\mathcal{K}$ is called the key space, and a a secret key $K \in \mathcal{K}$ is randomly chosen when the scheme is initialised.
2) C_K is an (n,M) code.
3) e_K is an encoding function $e_K : U \hookrightarrow C_K$.
4) A_K is called the tracing algorithm, taking as input any q-ary vector $\mathbf{x}$ and outputting a subset $L \subseteq U$.

When a certain user buys a digital copy, the corresponding fingerprint $\mathbf{u} \in C_K$ is embedded in it. If several users collude to make illegal copies, they can make copies with some hybrid fingerprint $\mathbf{x}$ which combines information from their respective fingerprints. The algorithm A_K takes $\mathbf{x}$ as the input, and outputs a set $L \subseteq U$ of users to be accused. If successful, the output is a non-empty subset of the pirates.

The fingerprinting model is made according to Kerchoff's principles, in that the key is random and everything but the key is assumed to be public knowledge. If the entire system is compromised, 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 S to use for the product he 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 fingerprinting scheme and the key, and distributed to the users. A coallition $P \subset U$ of t (potential) pirates is thus assigned fingerprints $e_K(P) \subset C_K$.
4) The coalition of potential pirates get together and compare their copies. We let Y be a random variable comprising all the information the pirates obtain at this point. Generally $0 < I(Y;K) < H(K)$, which means that the pirates get some, but imperfect information about the unknown variable K.
5) If $P(\text{Error}|Y=y)$ is sufficiently low, the pirates seeing y will opt out, without comitting any crime.
6) If the pirates choose to play, they choose a strategy for garbling the fingerprint, make the copies, and sell the copies with the false fingerprint $\mathbf{x}$.
7) If and when an illegal copy is discovered, the vendor runs the tracing algorithm $A_K(\mathbf{x})$ 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 $\neg$Error where the pirates lose. Event 0 is not random; it is an unrestricted choice of the pirates. Therefore, all probabilities will be taken under the assumption that the pirates do play.

When the pirates choose to play, they will behave adversarially, choosing the strategy that maximises their chance of escape. Therefore we assume that the probabilities $P(\text{Error})$ and $P(\text{Error}|Y=y)$ are maximal over all pirate strategies. The traditional definition of collusion-secure codes is based on the unconditional probability of error, given no information about K.

Definition 1 (Weak security) *An FP scheme S is (weakly) (t,ϵ)-secure, if $P_K(\text{Error}) \leq \epsilon$ for any t-set $P \subset U$.*

We will sometimes say that C_K is (t,ϵ)-secure if it is part of a (t,ϵ)-secure scheme, when misunderstanding is improbable.

Remark 1 *If, for each distinct (C_K, A_K), e_K is uniformly random over all possible encodings, then $P_K(\text{Error}) \leq \epsilon$ for some t-set $P \subset U$ implies that $P_K(\text{Error}) \leq \epsilon$ for any t-set $P \subset U$. This is because the set of fingerprints $e_K(P) \subset C_K$ becomes a random t-set C_K, and it explains why some authors assume that the pirates do not know their identities.*

Unauthorised 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 sort of a stalemate where nobody wins and nobody loses is perfectly satisfactory.

We expect that there is a threshold e_p such that if the pirates think the error probability is greater than e_p, then they will play. If it is less than or equal e_p, then they will not. If this is the case, we get Event 0 if $P(\text{Error}|Y=y) < e_p$. Note that a (weakly) (t,e_p)-secure code is not sufficient to deter all pirate coallitions of size t or less. This is why we introduce a new and stronger definition.

Definition 2 (Strong security) *An FP scheme S is strongly (t,ϵ)-secure, if for any y the pirates could see, we have $P(\text{Error}|Y=y) \leq \epsilon$ for any t-set $P \subseteq U$.*

Clearly a strongly t-secure FP scheme will deter any pirate coallition of size at most t if $\epsilon \leq e_p$. By abuse of language, we

shall sometimes say that C_K is (t, ϵ)-secure when the scheme is.

Definition 3 *If a scheme is (weakly) (t, ϵ_1)-secure, then ϵ_1 is an* a priori *error bound. If it is strongly (t, ϵ_2)-secure, then ϵ_2 is an* a posteriori *error bound.*

Note that $\epsilon_1 \leq \epsilon_2$. Even though the explicit definition of strongly t-secure codes is new, many previous scheme Boneh-Shaw do meet the definition. Some authors bound $P(\text{Error}|P = p)$ for any p, which gives an error bound no weaker than ϵ_2.

Since the pirate can choose whether they want to play or not, it is reasonable to assume that the coallitions playing have a higher average chance of escaping than has the average coallition. The economists call this *adverse selection*, and it affects the following interesting probability

$$\epsilon_A = P(\text{Error}|\text{The pirates did play}).$$

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

Lemma 1 *When the vendor obtains an illegal copy having only the knowledge of the key K and the false fingerprint $\mathbf{x}$, the probability ϵ_A of getting an incorrect output from $A_K(\mathbf{x})$ is at most the a posteriori error bound ϵ_2.*

Proof: Since ϵ_2 bounds the conditional error probability for any information the pirate could have, in particular it bounds this probability for any information which would induce the pirates to play. Hence ϵ_2 also bounds the probability of error under the condition that the pirates play. ∎

Lemma 2 *If $P(\text{Error}|Y = y)$ not constant over all y, then there is an a priori error bound ϵ_1 such that $\epsilon_1 < \epsilon_A$ for some pirate strategy.*

Proof: Suppose ϵ_1 is tight, then it is given as the average

$$\epsilon_1 = \sum_y P(Y = y)P(\text{Error}|Y = y).$$

Suppose the pirates choose to play whenever $P(\text{Error}|Y = y) > \epsilon_1$. Then we get that

$$\epsilon_A = \sum_{\substack{y \\ P(\text{Error}|Y=y)>\epsilon_1}} P(\text{Error}|Y = y).$$

We clearly get that $\epsilon_A > \epsilon_1$, since we have removed only small terms from the average. ∎

To summarise, if the pirates decide to do illegal copying before they see their copies, their chance of escape is at most ϵ_1. For any pirate collusion of size at most t having compared their copies, the chance of escape is at most ϵ_2. Which error bound is the most important, will depend on the application.

C. The marking assumption

The fingerprinting system must include some method to embed the fingerprints in the digital data, in addition to the fingerprinting code and tracing algorithm briefly described above. Some security assumptions must be made about this embedding in order to devise FP schemes. We will stick to the Boneh-Shaw model throughout the paper; and this is defined by the following Marking Assumption. It is hard to construct general embeddings satisfying the assumption, but some theoretical examples are suggested in [2], and one is given as the example below.

Definition 4 (The Marking Assumption) *Let $P \subseteq C$ be the set of fingerprints held by a coalition of pirates. The pirates can produce a copy with a false fingerprint $\mathbf{x}$ for any $\mathbf{x} \in F_C(P)$, where*

$$F_C(P) = \{(c_1, \ldots, c_n) : \forall i, \exists (x_1, \ldots, x_n) \in P, x_i = c_i\}.$$

We call $F_C(P)$ the feasible set of P with respect to C.

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

Example 1 (Traitor Tracing) *Collusion-secure codes are used for traitor tracing [3], where the Marking Assumption is satisfied as following. The system uses a $q \times n$ matrix of permanent keys $K_{j,i}$. Each row corresponds to an 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 Assumptions. 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 positions. See [5], [6] for details. Muratani [7] use a stronger assumption where the pirates output each word in the feasible set, allowing much shorter codewords.

The Marking Assumption may be too strong for many applications. Some authors relax it by assuming that the pirates have a certain probability p_e of outputting a random symbol from the alphabet in each column. Thus they call for codes which are error-secure in addition to collusion-secure [8].

D. Two types of error

Let $L \subseteq U$ be the output of the tracing algorithm and $P \subseteq U$ the pirate collusion. We have defined an error to be the event that $L = \emptyset$ or $L \not\subseteq P$. Some fingerprinting schemes distinguish between two types of errors.

A Type I error (or *failure*) is the event that $L \cap P = \emptyset$, i.e. that the vendor fails to identify any guilty pirate. If the tracing algorithm outputs one or more innocent users, i.e. if $L \backslash P \neq \emptyset$, then we say that we have an error of Type II.

In the context of criminal law, we know that Type II errors is a serious matter. Frequent Type I errors means that we often do not get useful output, but they do not affect the reliability of the output which is obtained. If Type II errors are frequent, the output cannot be trusted even when we have output. We will sometimes say that a code is strongly (or weakly) $(t, \epsilon_{\mathrm{I}}, \epsilon_{\mathrm{II}})$-secure if the a posteriori (or a priori) probability of error is at most ϵ_{I} for Type I and at most ϵ_{II} for Type II.

Our analysis will give separate bounds for Types I and II. This is mainly because it is the simplest way to do the proofs. Since many existing schemes have been analysed only with one error bound, we will usually state results with one error bound only, to simplify the presentation and allow for comparison. In this case, the error bound is the sum of the Type I and Type II bounds.

III. Concatenated schemes

In this chapter we develop a general analysis of concatenated fingerprinting schemes. Such concatenation was applied in [2], but our error analysis will prove that those constructions have better error rate than originally proved. We make the following formal definition of concatenated schemes. Note that the involved code C_K is a concatenated code in coding terms.

Definition 5 (Concatenated Fingerprinting Scheme) *Let $S_{\mathrm{I}} = (C_k^{\mathrm{I}}, e_k^{\mathrm{I}}, A_k^{\mathrm{I}})$ and $S_{\mathrm{O}} = (C_k^{\mathrm{O}}, e_k^{\mathrm{O}}, A_k^{\mathrm{O}})$ be FP schemes, where C_k^{I} is $(n_{\mathrm{I}}, q)_2$ and C_k^{O} is (n_{O}, M). Let Q denote the alphabet of C^{O}. A concatenated FP scheme $S = S_{\mathrm{O}} \circ S_{\mathrm{I}} = (C_K, e_K, A_K)$ consists of the following elements. The key is a tuple $K = (k_{\mathrm{O}}, k_1, \ldots, k_{n_{\mathrm{O}}})$, where k_{O} is a key for S_{O} and k_i are keys for S_{I}. The encoding is*

$$\begin{aligned} e_K(u) &= e_{k_1}^{\mathrm{I}}(c_1) || e_{k_2}^{\mathrm{I}}(c_2) || \ldots || e_{k_{n_{\mathrm{O}}}}^{\mathrm{I}}(c_{n_{\mathrm{O}}}), \\ &\text{where } (c_1, \ldots, c_{n_{\mathrm{O}}}) = e_{k_{\mathrm{O}}}^{\mathrm{O}}(u) \text{ and } u \in U. \end{aligned} \tag{2}$$

Each segment $e_{k_i}^{\mathrm{I}}(c_i)$ of the word is called a block. The code is $C_K = \{e_K(u) : u \in U\}$ which is $(n_{\mathrm{I}} n_{\mathrm{O}}, M)_2$. The algorithm A_K first decodes each block using A_{k_i}, and then decodes the resulting word over Q using $A_{k_{\mathrm{O}}}$.

Let R_{I} and R_{O} denote the rates of C^{I} and C^{O} respectively.

We demand that S_{I} is strongly (t, ϵ)-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 itself.

We analyse two different kinds of outer codes, namely random codes as suggested in [2] in Section III-B, and codes with large minimum distance in Section III-D. The outer code is list decoded with respect to the Hamming distance. The probability of Type II errors is bounded for each choice of outer codes. In Section III-C, we study parameters of the Boneh-Shaw codes, i.e. using the inner code of [2] and random outer codes.

The original concatenated Boneh-Shaw scheme will be named BS-RC. The acronym before the dash indicates the inner code, the Boneh-Shaw inner code, and the last letters indicate the outer code, a random code. Other schemes discussed in this section are BS-RS with Reed-Solomon outer codes, and BS-AG with asymptotic AG outer codes.

A. Decoding and Type I errors

We will use list decoding for the outer tracing algorithm $A_{k_{\mathrm{O}}}$.

Definition 6 *A list decoding algorithm A for a code C takes as input an n-word $\mathbf{y}$ and a threshold Δ and returns the set*

$$L = \{\mathbf{c} \in C | d(\mathbf{c}, \mathbf{y}) \leq \Delta n\}. \tag{3}$$

We have chosen list decoding because it gives us simple proofs. It has the additional advantage that we often get to trace several pirates. Observe that the closest neighbour of $\mathbf{y}$ will be in the list L unless $L = \emptyset$. Hence if list decoding is successful, then closest neighbour decoding is successful too. Hence an application can use closest neighbour decoding instead, without increasing the error rate.

Let F be the number of positions i where inner decoding $A_{k_i}^{\mathrm{I}}$ is incorrect. Clearly, $F \sim \mathcal{B}(n, \epsilon_{\mathrm{in}})$. The pirates match $\mathbf{y}$ in at least $(n-F)/t$ positions on average, which means that if $F \leq t\Delta n - (t-1)n_{\mathrm{O}}$, then at least one guilty pirate is caught. The following theorem follows by the Chernoff bound.

Theorem 2 *Let S_{I} be a strongly $(t, \epsilon_{\mathrm{in}})$-secure scheme, and $S_{\mathrm{O}} = (C_{k_{\mathrm{O}}}^{\mathrm{O}}, e_{k_{\mathrm{O}}}^{\mathrm{O}}, A_{k_{\mathrm{O}}}^{\mathrm{O}})$ a scheme where $A_{k_{\mathrm{O}}}^{\mathrm{O}}$ is list decoding with threshold Δ. Then using $S = S_{\mathrm{I}} \circ S_{\mathrm{O}}$, the probability of identifying no guilty user is*

$$\epsilon_{\mathrm{I}} \leq P(F \geq (1 - t + t\Delta)n_{\mathrm{O}}), \quad F \sim \mathcal{B}(n_{\mathrm{O}}, \epsilon_{\mathrm{in}}),$$

and

$$\epsilon_{\mathrm{I}} \leq 2^{-n_{\mathrm{O}} D(1-t+t\Delta || \epsilon_{\mathrm{in}})}, \quad \text{if} \quad \epsilon_{\mathrm{in}} < 1 - t + t\Delta.$$

Corollary 1 *If $D(1 - t + t\Delta || \epsilon_{\mathrm{in}}) > 0$, then the probability of Type I error tends to zero with increasing code length n_{O}.*

Note that the bound on ϵ_{I} depends only on n_{O}, Δ, t, and ϵ_{in}. It is oblivious to S_{O}. The Type II error rate ϵ_{II} will depend on the design of S_{O}. The inner code keys k_i have to be independent so that errors in two distinct blocks are independent events. Otherwise the Chernoff bound would not be applicable.

B. Random codes (RC)

Boneh and Shaw used random codes for S_{O}. Let k_{O} be an $M \times n$ matrix over Q were every entry is chosen independently and uniformly at random. Suppose some arbitrary ordering on U. The encoding $e_{k_{\mathrm{O}}}$ maps the i-th user to the i-th row of k_{O}, so that $C_{k_{\mathrm{O}}}^{\mathrm{O}}$ is the set of rows from k_{O}.

Theorem 3 *Let $S = S_{\mathrm{O}} \circ S_{\mathrm{I}}$ be a scheme using random codes for S_{O}. If $1/q < 1 - \Delta$, the probability of including a given innocent user $\mathbf{c}$ in the output list is bounded as*

$$P(\mathbf{c} \in L) \leq \hat{\epsilon} \leq 2^{-n_{\mathrm{O}} D(1-\Delta||1/q)},$$

and the total Type II error rate is bounded as

$$\epsilon_{\mathrm{II}} \leq 2^{n_{\mathrm{O}}(R_{\mathrm{O}} \log q - D(1-\Delta||1/q))}.$$

Proof: Consider the output $\mathbf{y}$ from inner decoding and an innocent user $\mathbf{c} \notin P$. Let $X = n_{\mathrm{O}} - d(\mathbf{c}, \mathbf{y})$. Clearly X is a stochastic variable with distribution $B(n_{\mathrm{O}}, 1/q)$, and $P(\mathbf{c} \in L) = P(X \geq n_{\mathrm{O}} - D)$. The error probability is bounded as

$$\epsilon_{\mathrm{II}} \leq \sum_{\mathbf{c} \in C \setminus P} P(\mathbf{c} \in L) \leq M \cdot P(X \geq n_{\mathrm{O}}(1 - \Delta)),$$

and the theorem follows by Chernoff's bound. ∎

Corollary 2 *The Type II error rate tends to zero with increasing length if $R_{\mathrm{O}} < D(1 - \Delta||1/q)/\log q$ and $1/q < 1 - \Delta$.*

One great advantage of random codes is that they can be made for any number of users quite trivially. Observing the error bounds, we note that ϵ_{I} is unaltered, and ϵ_{II} degrades gracefully when M increases.

C. The Boneh-Shaw concatenated code

The following (M, ϵ)-secure scheme S_{I} was used in [2]. Let C_ι^{I} be a $(r(M-1), M)_2$ code with a codebook consisting of $M - 1$ distinct columns, each replicated r times. A set of identical columns will be called a type. Every column has the form $(1 \ldots 10 \ldots 0)$, such that the i-th $(1 \leq i \leq M)$ user has zeroes in the first $i - 1$ types and a one in the rest.

The key k maps the code C_ι^{I} onto an equivalent code C_k 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 i-th type. Hence they have to use the same probability of choosing a 1 in both these types. The tracing algorithm $A^{\mathrm{I}}_{k_{\mathrm{I}}}$ 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

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

The probability of accusing a given innocent user is bounded as

$$\hat{\epsilon} \leq 2^{1 - \frac{r}{2M^2}}.$$

Theorem 4 (Boneh and Shaw) *The BS inner code with replication factor r is strongly (M, ϵ)-secure whenever $r \geq 2M^2 \log(2M/\epsilon)$.*

Let BS-RC be the scheme $S = S_{\mathrm{O}} \circ S_{\mathrm{I}}$ 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 Boneh and Shaw and set $q = 2t$. Obviously n_{O} and r control a trade-off between code length and error rate. Finally, we have Δ to control the trade-off between the two error types.

$t = \log M$	Boneh and Shaw	New analysis
10	$6.64 \cdot 10^8$	$3.06 \cdot 10^8$
15	$3.91 \cdot 10^9$	$1.79 \cdot 10^9$
20	$1.40 \cdot 10^{10}$	$6.44 \cdot 10^9$
25	$3.80 \cdot 10^{10}$	$1.77 \cdot 10^{10}$
30	$8.68 \cdot 10^{10}$	$4.09 \cdot 10^{10}$

Table I
SOME LENGTHS WHEN $t = \log M$.

Theorem 5 *If we use*

$$q = 2t, \quad \Delta = \frac{t}{t+1}, \quad \epsilon_{\mathrm{in}} = \frac{1}{2t},$$

then BS-RC is a strongly (t, ϵ)-secure FP scheme accommodating M users requiring length

$$n = (2t - 1)\left\lceil 8t^2(3 + 2\log t)\right\rceil n_{\mathrm{O}},$$

where

$$n_{\mathrm{O}} = \frac{\max\{-\log \epsilon_{\mathrm{I}}, \log M - \log \epsilon_{\mathrm{II}}\}}{D(\frac{1}{t+1}||\frac{1}{2t})}.$$

Asymptotically, the length is

$$n = \Theta\big(t^4 (\log t)(\log M - \log \epsilon)\big).$$

In this theorem, Δ is made only slightly greater than the minimum value of $(t-1)/t$. By Corollary 1 we require $\epsilon_{\mathrm{in}} < 1/(t+1)$, but to make n_{O} linear in t, ϵ_{in} must in fact be much smaller than $1/(t+1)$.

Proof: Theorems 2 and 3 give two bounds on n_{O}, so we get

$$n_{\mathrm{O}} = \max\left\{\frac{-\log \epsilon_{\mathrm{I}}}{D(\frac{1}{t+1}||\frac{1}{2t})}, \frac{\log M - \log \epsilon_{\mathrm{II}}}{D(\frac{1}{t+1}||\frac{1}{2t})}\right\}.$$

It can be shown that $D(1/(t+1)||1/(2t)) = \Theta(t^{-1})$, and hence

$$n_{\mathrm{O}} = \Theta(t(\log M - \log \epsilon)).$$

For the inner code, we have

$$\begin{aligned} n_{\mathrm{I}} &= (q-1)2q^2(\log(2q) - \log \epsilon_{\mathrm{in}}) \\ &= (2t-1)8t^2(3 + 2\log t) = \Theta(t^3 \log t). \end{aligned}$$

The theorem follows since $n = n_{\mathrm{I}} n_{\mathrm{O}}$. ∎

Considering asymptotic classes of codes, Δ can be made smaller. The following theorem gives the better rates.

Theorem 6 *There exists an asymptotic class of BS-RC FP schemes with exponentially declining error rate for any rate R satisfying*

$$R < \frac{D(\frac{1 - 2q2^{-r/(2q^2)}}{t}||1/q)}{r(q-1)}, \tag{4}$$

	BS-RC			Theoretical BBK
t	q	r	Rate	Rate
2	4	238	$2.42 \cdot 10^{-4}$	0.016
3	5	410	$3.62 \cdot 10^{-5}$	0.00102
4	7	847	$9.62 \cdot 10^{-6}$	$1.01 \cdot 10^{-4}$
5	9	1457	$3.53 \cdot 10^{-6}$	$1.25 \cdot 10^{-5}$
7	13	3223	$8.04 \cdot 10^{-7}$	$2.76 \cdot 10^{-7}$

Table II
ASYMPTOTIC RATES AND MAXIMISING VALUES OF q AND r FOR THE BS-RC CODES FOR SOME NUMBERS OF PIRATES.

if q and r are natural numbers such that $(1-2q2^{-r/(2q^2)})/t > 1/q$.

Proof: Asymptotically, ϵ_{in} can be taken arbitrarily close to $1 - t + t\Delta$, or in other words

$$\Delta \approx \frac{t - 1 + \epsilon_{\text{in}}}{t} = \frac{t - 1 + 2q2^{-r/2q^2}}{t}.$$

By Theorem 3, the outer rate can be chosen arbitrarily close to $D(1-\Delta||1/q)/\log q$. We get the following component code rates

$$R_{\text{O}} \approx \frac{D(\frac{1-2q2^{-r/2q^2}}{t}||1/q)}{\log q}, \quad R_{\text{I}} = \frac{\log q}{r(q-1)},$$

which gives the total rate as stated in the theorem. ∎

In Table II, we can see some asymptotic rates for our codes, as well as those of [5] (BBK). We note that BBK has the better rate for few pirates, whereas ours is better against seven pirates. It is also interesting to note that $2t$ is not the maximising value of q asymptotically, except for $t = 2$.

D. Outer code with large distance

We recall that codes with sufficiently large distance give combinatorially secure codes. The BBK scheme introduced outer codes where the minimum distance is large enough not only to successfully trace, but also to correct for some decoding errors from the inner decoding. We shall see how this can be combined with strongly $(t, \epsilon_{\text{in}})$-secure inner codes following the lines of BS-RC.

Let S_{I} be a strongly $(t, \epsilon_{\text{in}})$-secure scheme as before. Let $\hat{\epsilon}_{\text{in}}$ be an upper bound on the probability of accusing a given innocent user $\mathbf{c}$. Even though this is a parameter traditionally not explicitly stated for constructed fingerprinting schemes, it is often known by a bound at least as good as that for ϵ_{in}, which is often bounded as $\epsilon_{\text{in}} \leq M\hat{\epsilon}_{\text{in}}$.

We use an outer scheme $S_{\text{O}} = \{(C^{\text{O}}, e^{\text{O}}, A^{\text{O}})\}$ with a constant key. The encoding e^{O} is arbitrary and the tracing A^{O} is list decoding with threshold Δ as before.

Let δn_{O} be the minimum distance of C^{O}, and $P = \{\mathbf{a}_1, \ldots, \mathbf{a}_t\} \subseteq C^{\text{O}}$ a pirate coalition. Consider a false fingerprint $\mathbf{x}$ after inner decoding and an arbitrary innocent user $\mathbf{c} \notin P$. For each i, $\mathbf{c}$ matches $\mathbf{a}_i$ in at most $n(1-\delta)$ positions. If inner decoding were perfect, $\mathbf{x}$ would match $\mathbf{c}$ in at most $nt(1-\delta)$ positions.

First we study the probability $\pi(\mathbf{c})$ that an innocent user $\mathbf{c}$ be accused. Let V be the set of coordinates where $\mathbf{c}$ is different from any pirate, and let V^C be the complement, i.e. the set of positions where $\mathbf{c}$ match at least one pirate. Let X_i be a stochastic variable which is one if and only if $c_i = x_i$, and $s(\mathbf{x}, \mathbf{c})$ the total number of such matches. We get that

$$s(\mathbf{c}, \mathbf{x}) = \sum_{i \in V} X_i + \sum_{i \in V^C} X_i \leq \sum_{i \in V} X_i + \#V^C.$$

We have $\#V^C \leq nt(1-\delta)$. If we let $V' \subseteq V$ be any subset of size $N = n(1 - t(1-\delta))$, we get

$$s(\mathbf{c}, \mathbf{x}) \leq X + nt(1-\delta), \quad \text{where } X = \sum_{i \in V'} X_i.$$

We have that X_i is 1 with probability $\hat{\epsilon}_{\text{in}}$ and 0 otherwise. We get

$$\hat{\epsilon} \leq P(s(\mathbf{c}, \mathbf{x}) \geq (1-\Delta)n) \leq P\left(X \geq \frac{(1-\Delta) - t(1-\delta)}{1 - t(1-\delta)} N\right).$$

Using Chernoff, we get the following theorem.

Theorem 7 *The probability of accusing a given innocent user for $S = S_{\text{O}} \circ S_{\text{I}}$ as described in this section is*

$$\hat{\epsilon} \leq 2^{-(1-t(1-\delta))nD(\sigma||\hat{\epsilon}_{\text{in}})},$$

if

$$\sigma = \frac{(1-\Delta) - t(1-\delta)}{1 - t(1-\delta)} > \hat{\epsilon}_{\text{in}}.$$

This error bound is rather pessimistic. We assume that the hybrid fingerprint matches the innocent user in every block where one pirate matches the innocent. This pessimistic approach also means that we do not need any further secret key, except the inner code key for each block.

Combining Theorems 2 and 7, we get that

$$\delta > 1 - \frac{1 - \epsilon_{\text{in}} - t\hat{\epsilon}_{\text{in}}}{t^2(1 - \hat{\epsilon}_{\text{in}})}. \tag{5}$$

It follows immediately that $q = \Omega(t^2)$. A good candidate as an outer code with large minimum distance is the $[n_{\text{O}}, k_{\text{O}}, n_{\text{O}} - k_{\text{O}} + 1]_q$ Reed-Solomon (RS) codes, which can be decoded with the Guruswami-Sudan algorithm, with complexity $O(n_{\text{O}})$.

Example 2 *An RS outer code can be combined with a BS inner code. Take for instance, $t = 20$ and $M = 2^t$. Let $q = 2^{10}$ and $r = 3.1 \cdot 10^7$, and use a $(r(q-1), q)$ BS as inner code. As an outer code, we use a $[690, 2]_q$ generalised Reed-Solomon code. With a decoding threshold of $\Delta = 0.958$, we get a total error rate of $\epsilon \leq 0.356 \cdot 10^{-10}$. The total length is $2.139 \cdot 10^{10}$. These parameters are inferior to BS-RC, but still good enough to be interesting for application where decoding complexity is important.*

Concatenations of BS inner codes and RS outer codes will be denoted BS-RS. We have $n_{\text{I}} = \Theta(q^3 \log q)$ from the inner code, and $q = \Omega(t^2)$ due to the distance requirement. This gives us $n = \Omega(t^6 \log t)$, which is inferior to BS-RC. Furthermore, it is rather difficult to find the optimal choices for the various parameters.

Asymptotic classes of codes are possible using asymptotic AG codes as by the following theorem. The problem with this approach is that a large inner code is needed, and the codewords of the Boneh-Shaw code get very long. Using AG and RS codes would be much more effective if the inner code can be improved.

Proposition 1 *If there is an $(n_{\rm I}, q)$ strongly $(t, \epsilon_{\rm in})$-secure code where the probability of accusing a given innocent user is at most $\hat{\epsilon}_{\rm in}$, then there is an asymptotic family of strongly (t, ϵ)-secure codes with any rate less than $R_{\rm O}(\log q)/n_{\rm I}$, where $R_{\rm O}$ solves*

$$R_{\rm O} \log q = D\Bigg(\frac{\frac{1-\epsilon_{\rm in}}{t} - t\Big(R_{\rm O} + \frac{1}{\sqrt{q}-1}\Big)}{1 - t\Big(R_{\rm O} + \frac{1}{\sqrt{q}-1}\Big)} \,\Big|\Big|\, \hat{\epsilon}_{\rm in} \Bigg),$$

and where ϵ vanishes exponentially.

Proof: We see from Theorem 2, that exponentially declining $\epsilon_{\rm I}$ is obtained if $\Delta > 1 - 1/t + \epsilon_{\rm in}/t$, but Δ can be taken arbitrarily close to this bound. From Theorem 7, we get that $\epsilon_{\rm II}$ will decline exponentially if $R_{\rm O} \log q < D(\sigma||\hat{\epsilon}_{\rm in})$, where

$$\sigma = \frac{(1-\Delta) - t(1-\delta)}{1 - t(1-\delta)} \approx \frac{\frac{1-\epsilon_{\rm in}}{t} - t(1-\delta)}{1 - t(1-\delta)}.$$

Again $R_{\rm O}$ can be taken arbitrarily close to this bound. Using AG outer codes, we get

$$\delta \approx 1 - R_{\rm O} - \frac{1}{\sqrt{q}-1},$$

giving

$$\sigma \approx \frac{\frac{1-\epsilon_{\rm in}}{t} - t\Big(R_{\rm O} + \frac{1}{\sqrt{q}-1}\Big)}{1 - t\Big(R_{\rm O} + \frac{1}{\sqrt{q}-1}\Big)},$$

Now, $R_{\rm O}$ can be taken arbitrarily close to the solution of the equation stated in the proposition. ∎

IV. Fighting two pirates

We mentioned that the BS replication codes may not be the ideal choice for inner codes. For two pirates we have good alternatives, which we consider in this section.

It was proved in [9] that socalled separating codes give $(2, \epsilon, 0)$-secure schemes (C_K, e_K, A_K). The code C_K is equivalent to a $(2,2)$-separating base code C_ι, e_K is arbitrary, and A_K is an exhaustive search through all possible two-sets $\hat{P} \subseteq C_K$. The $[31, 5, 16]$ Simplex code is $(2, 2^{-11}, 0)$-secure, and the $[126, 14]$ dual BCH code is $(2, 0.292 \cdot 10^{-10}, 0)$-secure.

We define three new concatenated schemes, all using separating inner codes as described above. The SS-RC scheme uses random outer code as described in Section III-B. The SS-RS and SS-AG codes use respectively Reed-Solomon and AG codes as described in Section III-D.

A. Asymptotic constructions

The best asymptotic rate offered for $t = 2$ in [5] was 0.015, using the $[126, 14]$ BCH-dual as inner code and an AG outer code. On the other hand, [9] offered a rate of 0.026 for an asymptotic class $(2,2)$-SS.

Theorem 8 *The SS-RC scheme with the $[126, 14]$ punctured dual of the two-error-correcting BCH code as inner code, forms an infinite class of $(2, \epsilon)$-secure schemes with rate R, for any $R < 0.0476$ and exponentially declining error rates given as*

$$\epsilon_{\rm I} \le 2^{-n\frac{D(2\Delta-1||0.3\cdot 10^{-10})}{126}} \quad \textit{and} \quad \epsilon_{\rm II} \le 2^{n(R - D(1-\Delta||2^{-14})/126)},$$

where Δ may be chosen freely in the interval $1/2 + 1.5 \cdot 10^{-11} < \Delta < 1 - 2^{-14}$.

The algebraic structure of SS-AG makes it possible to take advantage of the fact that the inner codes have $\epsilon_{\rm II} = 0$ and make concatenated schemes which also have $\epsilon_{\rm II} = 0$.

For any innocent user $\mathbf{c}$, we have $s(\mathbf{c}, \mathbf{x}) \le 2(1-\delta)n_{\rm O}$. Hence $\mathbf{c}$ will never be accused if $\Delta > 1 - 2(1-\delta)$. Asymptotically, δ can be taken arbitrarily close to $(1+\Delta)/2$. The bound on $\epsilon_{\rm I}$ is found from Theorem 2,

$$\epsilon_{\rm I} \le 2^{-n_{\rm O} D(-1+2\Delta||\epsilon_{\rm in})}.$$

It is necessary that $\Delta > (1 + \epsilon_{\rm in})/2$, which gives us

$$\delta \approx 1/2 + (1 + \epsilon_{\rm in})/4.$$

The outer code rate can be brought arbitrarily close to

$$R_O \approx 1 - \delta - \frac{1}{\sqrt{q}-1} \approx \frac{1}{2} - \frac{1+\epsilon_{\rm in}}{4} - \frac{1}{\sqrt{q}-1}.$$

The $[126, 14]$ inner code, gives $R_O \approx 0.242$ and overall rate 0.0269. This is not as good as using random codes, but it is better than the BBK scheme [5], and like BBK, it can be GS decoded in time $O(\log M)$.

An alternative inner code is the $[15, 4, 8]$ code. This is too small for AG codes, but it works well with random codes.

Theorem 9 *The SS-RC scheme with a $[15, 4, 8]$ inner code forms an infinite class of $(2, \epsilon)$-secure codes with rate R, for any $R < 0.0688$, and exponentially declining error rates given as*

$$\epsilon_{\rm I} \le 2^{-n\frac{D(2\Delta-1||1/140)}{15}} \quad \textit{and} \quad \epsilon_{\rm II} \le 2^{n(R - D(1-\Delta||1/16)/15)},$$

where Δ may be chosen freely in the interval $1/2 + 1/280 < \Delta < 15/16$.

Corollary 3 *The SS-RC codes with $[15, 4, 8]$ inner codes are probabilistically $(2, \epsilon)$-secure with length*

$$n = 15 \left\lceil \max\left\{ \frac{\log \epsilon_{\rm I}}{D(2\Delta - 1||1/140)}, \frac{\log \epsilon_{\rm II} - \log M}{D(1-\Delta||1/16} \right\} \right\rceil,$$

for any Δ such that $1/2 + 1/280 < \Delta < 15/16$.

Inner code	Outer code	Concatenated code	Error rate
$[31, 5]$	$[4, 2]_{32}$	$[124, 10]$	$0.2 \cdot 10^{-12}$
$[31, 5]$	$[6, 3]_{32}$	$[186, 15]$	$0.2 \cdot 10^{-11}$
$[31, 5]$	$[8, 4]_{32}$	$[248, 20]$	$0.3 \cdot 10^{-11}$
$[31, 5]$	$[10, 5]_{32}$	$[310, 25]$	$0.7 \cdot 10^{-11}$
$[31, 5]$	$[12, 6]_{32}$	$[372, 30]$	$0.1 \cdot 10^{-10}$

Table III
SOME SS-RS CODES.

$\log M$	RS-RC	Simplex	SS-RC	SS-RS	Tardos	LBH
10	299 889	1 023	1 305	124	12 000	25 884
15	334 359	32 767	1 455	186	13 600	28 878
20	367 359	1 048 575	1 545	248	14 800	31 872
25	401 001	$2^{25} - 1$	1 605	310	16 400	34 867
30	435 471	$2^{30} - 1$	1 695	372	17 600	37 861

Table IV
CODE LENGTHS AGAINST TWO PIRATES FOR 1000 TO A BILLION USERS AND ERROR RATE $\epsilon \leq 10^{-10}$.

B. Practical codes

In Table IV, we present code lengths for 1000 to a billion users with the schemes we know. The RS-RC codes are computed with $q = 4$, $\epsilon_{\text{in}} = 0.002$. The error rates were set such that both ϵ_{I} and ϵ_{II} both are less than $10^{-10}/2$. We used $\Delta = 0.655$ for 2^{10} users, $\Delta = 210/320$ for 2^{15} users, $\Delta = 52/80$ for 2^{20} users, $\Delta = 41/64$ for 2^{25} users, and $\Delta = 203/320$ for 2^{30} users. Also the $[126, 14]$ and $[254, 24]$ BCH duals are $(2, 10^{-10})$-secure, but these do not fit well in the table.

Constructions are shown in Table III, and the general result is given in the following theorem.

Theorem 10 *A concatenated code of an $(2, \epsilon_{\text{in}}, 0)$-secure inner code and a $[n_{\text{O}}, k_{\text{O}}]_q$ RS code is 2-secure with $\epsilon_{\text{II}} = 0$ and*

$$\epsilon_{\text{I}} \leq P(Y \leq 2k_O - 3), \quad \textit{where } Y \sim \mathcal{B}(n_{\text{O}}, 1 - \epsilon_{\text{in}}). \quad (6)$$

Tracing is done with the Guruswami-Sudan algorithm.

Proof: We have $\epsilon_{\text{I}} \leq P(X > n_{\text{O}}(1 - 2(1 - \delta)) + 1)$ where $X \sim \mathcal{B}(n_{\text{O}}, \epsilon_{\text{in}})$, from the proof of Theorem 2. Setting $Y = n_{\text{O}} - X$ gives the theorem. ∎

There is a variant from [10] using Simplex inner codes, Reed-Solomon outer codes, and a more complicated inner decoding algorithm.

V. OTHER KNOWN SCHEMES AND CONCLUSION

We have studied concatenated collusion-secure codes. As inner codes we suggest separating codes in the two pirate case, and the Boneh-Shaw inner code in the general case. As outer codes, we propose random codes, Reed-Solomon codes, or asymptotic AG codes. One of the schemes, BS-RC, is the classic of [2], but our analysis show length can be less than previously assumed. Samples for a thousand to a billion user show a reduction by a factor of about 2.1.

We know of one other strongly (t, ϵ)-secure scheme for our model, namely the BBK scheme of [5]. The BBK scheme is very good against a few pirates. Against sufficiently many pirates the Boneh-Shaw based schemes are better. Asymptotically, BS-RC has the best rate for seven pirates and more. Against two pirates, our construction of SS-RS appears to give the best codes for a thousand to a billion users, whereas SS-RC has the best known asymptotic rate.

There are many other fingerprinting schemes in the literature, but most of them use different Marking Assumption, and thus fall outside the scope of this paper. There are two schemes which are weakly (t, ϵ)-secure under the Boneh-Shaw Marking Assumption, due to Tardos [11] and LBH [12]. Tardos has a code length of $n = 100t^2 \ln(M/\epsilon)$, giving it the best known rate for many parameters. LBH is very good against three pirates but $n = \Theta(2^t)$.

The best decoding complexity in M is achieved when we can use the Sudan-Guruswami algorithm, that is when using RS or AG outer codes. This includes the BBK scheme, and gives $O(\log M)$. Using random codes, i.e. for BS-RC and SS-RC as well as Tardos, a linear search through the code is needed for decoding. Decoding of BBK however, is exponential in t, and this problem is avoided by BS-RS. Thus, against many pirates, BS-RS has the most efficient decoding algorithm known, even for weak (t, ϵ)-security.

An interesting open problem is lower bounds on the code length in terms of t and M. The few known bounds are independent of M. Another open issue is optimising the construction parameters of our schemes.

REFERENCES

[1] Neal R. Wagner, "Fingerprinting," in *Proceedings of the 1983 Symposium on Security and Privacy*, 1983, pp. 18–22.

[2] Dan Boneh and James Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inform. Theory*, vol. 44, no. 5, pp. 1897–1905, 1998, Presented in part at CRYPTO'95.

[3] B. Chor, A. Fiat, M. Naor, and B. Pinkas, "Tracing traitors," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 893–910, May 2000, Presented in part at CRYPTO'94.

[4] Torben Hagerup and Christine Rüb, "A guided tour of Chernoff bounds," *Information Processing Letters*, vol. 33, pp. 305–308, 1990.

[5] A. Barg, G. R. Blakley, and G. A. Kabatiansky, "Digital fingerprinting codes: Problem statements, constructions, identification of traitors," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 852–865, Apr. 2003.

[6] Katsunari Yoshioka, Junji Shikata, and Tsutomu Matsumoto, "Collusion secure codes: Systematic security definitions and their relations," *IEICE Trans. Fundamentals*, vol. E87-A, no. 5, May 2004.

[7] Hirofumi Muratani, "Optimization and evaluation of randomized c-secure CRT code defined on polynomial ring," in *Information Hiding 2004*, J. Fridrich (Ed.), Ed., vol. 3200 of *Springer Lecture Notes in Computer Science*, pp. 282–292. Springer-Verlag, 2004.

[8] Hans-Jürgen Guth and Birgit Pfitzmann, "Error- and collusion-secure fingerprinting for digital data," in *Information Hiding '99, Proceedings*. 2000, vol. 1768 of *Springer Lecture Notes in Computer Science*, pp. 134–145, Springer-Verlag.

[9] Hans Georg Schaathun, "Fighting two pirates," in *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*. May 2003, vol. 2643 of *Springer Lecture Notes in Computer Science*, pp. 71–78, Springer-Verlag.

[10] Marcel Fernandez and Miguel Soriano, "Protecting intellectual property by guessing secrets," 2003, EC-Web Sept. 2-5 2003 Prague, Czech Rep.

[11] Gábor Tardos, "Optimal probabilistic fingerprint codes," *Journal of the ACM*, 2005, http://www.renyi.hu/~tardos/fingerprint.ps. To appear. In part at STOC'03.

[12] Tri Van Le, Mike Burmester, and Jiangyi Hu, "Short c-secure fingerprinting codes," in *Proceedings of the 6th Information Security Conference*, Oct. 2003, Available at http://websrv.cs.fsu.edu/~burmeste/.