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 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 point out some differences between our model and others in the literature.

Index Terms—Collusion-secure codes, copyright protection, digital fingerprinting (FP), traitor tracing.

I. INTRODUCTION

UNAUTHORIZED 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.

Estimated 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 increases sales. Still, it is clear that big money is 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 (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 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 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 II, we will define the FP model, which we refine a little compared to past literature. In Section III, we give the main result, which is an improved error analysis of the BS FP 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. FINGERPRINTING PROBLEM

A. Preliminaries From Coding Theory

An $(n, M)_q$ code is a set of M words of n symbols from an alphabet Q of q elements. 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 normalized minimum distance is $\delta = d/n$.

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

$$D(\sigma||p) = \sigma \log_2 \frac{\sigma}{p} + (1 - \sigma) \log_2 \frac{1 - \sigma}{1 - p}. \quad (1)$$

Manuscript received April 26, 2005; revised February 20, 2006. This work was supported in part by the Commission of the European Communities through the IST program under Contract IST-2002-507932, and in part by the Norwegian Research Council under Grant 146874/20. Research was conducted in the Department of Informatics, University of Bergen, Bergen N-5020, Norway. The information in this document is provided as is, and no warranty is given or implied that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Thomas Johansson.

The author is with the Department of Computing, School of Electronics and Physical Sciences, University of Surrey, Guildford GU2 7XH, U.K. (e-mail: h.schaathun@surrey.ac.uk).

Digital Object Identifier 10.1109/TIFS.2006.873596

Theorem 1 (Chernoff): Let $X_1, \dots, X_t$ 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

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

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 $X \sim \mathcal{B}(n, p)$. All logarithms will be to base 2 unless otherwise stated.

B. BS Model and Marking Assumption

Our model follows that of Boneh and Shaw (BS) [2]. Let $\mathbf{u} = (u_1, \dots, u_n)$ be a digital file divided into n segments $u_i \in \mathcal{U}$. We call $\mathcal{U}$ the file alphabet. Let $\mathcal{M}$ be the set of users. We can view $i \in \mathcal{M}$ as a customer account number. Write $M = \#\mathcal{M}$.

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 $w_i \in \mathcal{U}$. The *extractor* inverts the embedding; given w_i , it outputs x .

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

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

If user i naively copies and redistributes $\mathbf{w}$, then the watermarks can be extracted to obtain the fingerprint $\mathbf{x} = (x_1, \dots, x_n)$, 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 fingerprints, but the full sequence $(x_1, \dots, x_n)$ 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 (The Marking Assumption): Let $P \subseteq C_K$ 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 $\mathbf{x} \in F(P)$, where

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

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 3 (Traitor Tracing): Collusion-secure codes are used for traitor tracing [3], where Definition 2 is satisfied as follows. 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 $(a_1, \dots, a_n)$ 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 [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].

C. FP Scheme

An $(n, M)_q$ 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 $(n, M)_q$ code. The tracing algorithm A_K takes a hybrid fingerprint $\mathbf{x}$ as an input and outputs a set $L \subseteq \mathcal{M}$. 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 $P \subseteq \mathcal{M}$ of t (potential) pirates is thus assigned fingerprints $e_K(P) \subset C_K$, and receives a set $Y = \{\mathbf{w}_1, \dots, \mathbf{w}_t\}$ of fingerprinted copies.
- 4) The coalition of potential pirates gets together and compares the copies Y .
- 5) If $\Pr(\text{Error} | Y = y)$ is sufficiently low, the pirates seeing y will opt out, without committing any crime.
- 6) If the pirates choose 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 $\mathbf{x}$.
- 7) If and when an illegal copy is discovered, the vendor extracts the hybrid FP $\mathbf{x}$, computes $A_K(\mathbf{x})$, 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.

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

Definition 4 (Weak Security): An FP scheme S is (weak) (t, ϵ) -secure if, for any $v \leq t$, $P_K(\text{Error}) \leq \epsilon$ when a set $P \subset M$ of v 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., $P_K(\text{Error}|Y = y)$). This probability can be higher or lower than the unconditional probability $P_K(\text{Error})$. We expect that there is a threshold e_p such that the pirates choose to play if and only if $P_K(\text{Error}|Y = y) > e_p$. If $P_K(\text{Error}|Y = y) < e_p$, 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 5 (Strong Security): An FP scheme S is strongly (t, ϵ) -secure if, for any y seen by at most t pirates, we have $P_K(\text{Error}|Y = y) \leq \epsilon$.

Clearly, a strongly t -secure FP scheme will deter any pirate coalition 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 6: Let S be an FP scheme. The *a priori* error bound ϵ_1 is the smallest ϵ such that S is (weak) (t, ϵ) -secure. The *a posteriori* error bound ϵ_2 is the smallest ϵ such that S is strongly (t, ϵ) -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 $P_K(\text{Error}|P = p)$ for any p , which clearly bounds ϵ_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:

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

Of course, when the vendor finds an illegal copy, he or she knows that 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 lemmata gives some information about this.

Lemma 7: 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 that 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 8: Unless $P_K(\text{Error}|Y = y)$ is constant over all y , there is a pirate strategy such that $\epsilon_1 < \epsilon_A$.

Proof: Write $\epsilon_K(y) = P_K(\text{Error}|Y = y)$, and write

$$\epsilon_1 = \sum_y P_K(Y = y) \epsilon_K(y).$$

Suppose the pirates choose to play whenever $\epsilon_K(y) > \epsilon_1$. Write $\bar{Y} = \{y : \epsilon_K(y) > \epsilon_1\}$. Then, we get that

$$\epsilon_A = \sum_{y \in \bar{Y}} \frac{P_K(Y = y)}{\sum_{y \in \bar{Y}} P_K(Y = y)} \cdot \epsilon_K(y).$$

We clearly get that $\epsilon_A > \epsilon_1$ 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 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.

Definition 9 (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 I* (or a failure) is the event that $L = \emptyset$. An error of *Type II* (or false accusation) is the event that $L \not\subseteq P$.

In the context of criminal law, we know that Type II errors are a serious matter. Frequent Type I 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 II errors are frequent, the output cannot be trusted even when we have output. For the rest of this paper, ϵ_I will denote the *a posteriori* probability of Type I error, and ϵ_{II} the *a posteriori* probability of Type II error. A scheme with such error probabilities is said to be (strongly) $(t, \epsilon_I, \epsilon_{II})$ -secure.

III. 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 10 (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 $S = S_O \circ S_I = (e_K, A_K)$ 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 for S_I . The encoding is

$$e_K(u) = e_{K_1}^I(c_1) \| e_{K_2}^I(c_2) \| \dots \| e_{K_{n_O}}^I(c_{n_O})$$

where $(c_1, \dots, c_{n_O}) = e_{K_O}^O(u)$ and $u \in \mathcal{M}$. (2)

Each segment $e_{K_i}^I(c_i)$ of the word is called a block. The algorithm A_K first decodes each block using $A_{K_i}^I$, and then decodes the resulting word over Q using $A_{K_O}^I$.

Note that the FP code of S is the concatenation of the FP codes of S_I and S_O . Let R_I and R_O denote the rates of S_I and S_O , respectively. We demand that S_I is strongly $(t, \epsilon_{\text{in}})$ 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 itself.

We analyze 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 probability of Type II errors is bounded for each choice of outer codes. In Section III-C, we study parameters of the BS codes (i.e., using the inner code of [2] and random outer codes).

The original concatenated BS scheme will be named BS-RC. The acronym before the dash indicates the inner code, the BS 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 algebraic-geometry (AG) outer codes.

A. Decoding and Type I Errors

For the outer tracing algorithm A_{K_O} , we use list decoding, defined as follows.

Definition 11: 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

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

We have chosen list decoding because it gives us simple proofs. It has the additional advantage that we often get to trace several pirates.

Remark 12: The closest neighbor of $\mathbf{y}$ will be in the list L unless $L = \emptyset$. Hence, if list decoding is successful, then the closest neighbor decoding is successful too. Thus, we can use the closest neighbor decoding instead without increasing the total error probability. However, when $L = \emptyset$, list decoding gives Type I error and closest neighbor gives Type II.

Let F' be the number of blocks i where inner decoding $A_{K_i}^I$ is incorrect. The pirates match $\mathbf{y}$ in at least $(n_O - F')/t$ blocks on average, which means that if $F' \leq t\Delta n_O - (t-1)n_O$, then at least one guilty pirate is caught.

Observe that F' is the sum of n independent Bernoulli trials with probability at most ϵ_{in} . It follows that:

$$\epsilon_I \leq P(F' \geq (1-t+t\Delta)n_O) \leq P(F \geq (1-t+t\Delta)n_O)$$

where $F \sim \mathcal{B}(n_O, \epsilon_{\text{in}})$, and we get the following theorem by the Chernoff bound.

Theorem 13: Let S_I be a strongly $(t, \epsilon_{\text{in}})$ -secure FP scheme, and $S_O = (e_{K_O}^O, A_{K_O}^O)$ a scheme where $A_{K_O}^O$ is list decoding with threshold Δ . Then, using $S = S_I \circ S_O$, the probability of identifying no guilty user is

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

Corollary 14: If $\epsilon_{\text{in}} < 1-t+t\Delta$ and $D(1-t+t\Delta\|\epsilon_{\text{in}}\|) > 0$, 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 probability 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)

BS used random codes for S_O . Let K_O be an $M \times n$ matrix over Q where every entry is chosen independently and uniformly at random. Suppose some arbitrary ordering on $\mathcal{M}$. The encoding e_{K_O} maps the i th user to the i th row of K_O .

Theorem 15: Let $S = S_I \circ S_O$ 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

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

and the total Type II error probability is bounded as

$$\epsilon_{\text{II}} \leq 2^{n_O (R_O \log q - D(1-\Delta\|1/q\|))}.$$

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

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

and the theorem follows by Chernoff's bound. $\blacksquare$

Corollary 16: The Type II error probability tends to zero with an increasing length if $R_O < D(1-\Delta\|1/q\|)/\log q$ and $1/q < 1 - \Delta$.

One great advantage of random codes is that they can be made for any number of users quite trivially. Observe that for increasing M , the bound on ϵ_I is unaltered, and the bound on ϵ_{II} degrades gracefully.

C. BS Concatenated Code

The following (q, ϵ) -secure scheme S_I was used in [2]. Let C_t^I 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 $(1 \dots 10 \dots 0)^T$, such that the i th ($1 \leq i \leq q$) user has zeroes in the first $i-1$ types and a one in the rest.

Example 17: The BS inner code for $r = 3$ and $q = 5$ is the set of the following five codewords:

$$\begin{aligned} \mathbf{c}_1 &= (111\ 111\ 111\ 111) \\ \mathbf{c}_2 &= (000\ 111\ 111\ 111) \\ \mathbf{c}_3 &= (000\ 000\ 111\ 111) \\ \mathbf{c}_4 &= (000\ 000\ 000\ 111) \\ \mathbf{c}_5 &= (000\ 000\ 000\ 000). \end{aligned}$$

The key K maps the code C_t^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 i th 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

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

Theorem 18 (BS): The BS inner code with replication factor r is strongly (q, ϵ) secure whenever $r \geq 2q^2 \log(2q/\epsilon)$.

Let BS-RC be the scheme $S = S_O \circ S_I$ 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 19: If we use

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

then BS-RC is a strongly (t, ϵ) -secure $(n, M)_2$ FP scheme, where

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

$$n_O = \frac{\max\{-\log \epsilon_I, \log M - \log \epsilon_{II}\}}{D\left(\frac{1}{t+1} \parallel \frac{1}{2t}\right)}. \quad (4)$$

Asymptotically, the length is

$$n = \Theta(t^4(\log t)(\log M - \log \epsilon)). \quad (5)$$

In order to use Theorem 13, we require $\Delta > (t-1)/t$ so that the first argument to D is positive. In this theorem, Δ is only slightly greater than this minimum. In order to use Corollary 14, we require $\epsilon_{in} < 1/(t+1)$, but to make n_O linear in t , ϵ_{in}

TABLE I
SOME LENGTHS WHEN $t = \log M$

$t = \log M$	Bonch 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	$8.68 \cdot 10^{10}$	$4.09 \cdot 10^{10}$

must, in fact, be much smaller than $1/(t+1)$. The choice of r gives $\epsilon_{in} = 1/(2t)$.

Proof: We have $n = (q-1)rn_O$. We get (3) by expanding the parameters r and q in the definition. Equation (4) follows from the two bounds on n_O from Theorems 13 and 15. Using the fact that $D(1/(t+1) \parallel 1/(2t)) = \Theta(t^{-1})$, we get

$$n_O = \Theta(t(\log M - \log \epsilon)).$$

Equation (5) follows easily by combining this with (3). ■

Sample rates are presented in Table I, both for our analysis and for that of BS. Considering asymptotic classes of codes, Δ can be made smaller. The following theorem gives the better rates.

Theorem 20: There exists an asymptotic class of BS-RC FP schemes with exponentially declining error probability for any rate R satisfying

$$R < \frac{D\left(\frac{1-2q2^{-r/(2q^2)}}{t} \parallel 1/q\right)}{r(q-1)} \quad (6)$$

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 $1 - t + t\Delta$, and by Theorem 18, we get that

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

By Theorem 15, the outer rate can be chosen arbitrarily close to $D(1 - \Delta \parallel 1/q)/\log q$. We get the following component code rates:

$$R_O \approx \frac{D\left(\frac{1-2q2^{-r/2q^2}}{t} \parallel 1/q\right)}{\log q}, \quad R_I = \frac{\log q}{r(q-1)}$$

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 [6] Barg Blakley, and Kabatiansky (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 maximizing 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

TABLE II
ASYMPTOTIC RATES AND MAXIMIZING VALUES OF q AND r FOR THE
BS-RC VODES FOR SOME NUMBERS OF PIRATES

t	BS-RC			Theoretical BBK
	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	$8.04 \cdot 10^{-7}$	$2.76 \cdot 10^{-7}$

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, ϵ) -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 FP 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 do not require randomization for the outer scheme when it is based on a large minimum distance (i.e., the key space is a Singleton set), and the ensemble S_O contains a unique pair (e^O, A^O) . 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 $P = \{\mathbf{a}_1, \dots, \mathbf{a}_t\} \subseteq C^O$ be 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 error free, $\mathbf{x}$ 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 at most $\hat{\epsilon}_{\text{in}}$ and 0 otherwise. The probability $\pi(\mathbf{c}) = P(s(\mathbf{c}, \mathbf{x}) \geq (1 - \Delta)n)$ is bounded as

$$\pi(\mathbf{c}) \leq P\left(X \geq \frac{(1 - \Delta) - t(1 - \delta)}{1 - t(1 - \delta)} N\right).$$

Using the Chernoff bound, we get the following theorem.

Theorem 21: The probability of accusing a given innocent user in $S = S_O \circ S_I$ as described in this section is

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

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

The total probability of Type II error is

$$\epsilon_{\text{II}} \leq 2^{nE}, \quad E = R_O \log q - (1 - t(1 - \delta)) D(\sigma||\hat{\epsilon}_{\text{in}}).$$

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.

Combining Theorems 13 and 21, we get that

$$1 - \delta < \frac{1 - \epsilon_{\text{in}} - t\hat{\epsilon}_{\text{in}}}{t^2(1 - \hat{\epsilon}_{\text{in}})}$$

which implies $1/(1 - \delta) > t^2$. Any asymptotic code requires $q > 1/(1 - \delta) > t^2$ by the Plotkin bound. A good candidate as an outer code with large minimum distance is the $[n_O, k_O, n_O - k_O + 1]_q$ RS codes, which can be decoded with the Guruswami–Sudan (GS) algorithm [10], with polynomial complexity in n_O .

Example 22: 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 $[690, 2]_q$ generalized RS 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 are still good enough to be interesting for applications where decoding complexity is important.

Concatenations of BS inner codes and RS outer codes will be denoted as BS-RS. We have $n_I = \Theta(q^3 \log q)$ 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.

To make asymptotic classes of codes, we can use the AG codes guaranteed by the following result from [11].

Theorem 23: 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 $q > t^4$. The problem is that large q require very long codewords in the BS inner code. Using AG and RS codes would be much more effective if the inner code could be improved.

Proposition 24: If there is an (n_I, q) strongly $(t, \epsilon_{\text{in}})$ -secure code where the probability of accusing a given innocent user is at most $\hat{\epsilon}_{\text{in}}$, then there is an asymptotic family of strongly

(t, ϵ) -secure codes with any rate of less than $R_O(\log q)/n_I$, where R_O solves

$$R_O \log q = [1 - t(1 - \delta)] D \left(\frac{\frac{1 - \epsilon_{\text{in}}}{t} - t \left(R_O + \frac{1}{\sqrt{q} - 1} \right)}{1 - t \left(R_O + \frac{1}{\sqrt{q} - 1} \right)} \parallel \hat{\epsilon}_{\text{in}} \right)$$

and where ϵ vanishes exponentially.

Proof: We see from Theorem 13 that exponentially declining ϵ_I is obtained if $\Delta > 1 - 1/t + \epsilon_{\text{in}}/t$, but Δ can be taken arbitrarily close to this bound. From Theorem 21, we get that ϵ_{II} will decline exponentially if $R_O \log q < [1 - t(1 - \delta)] D(\sigma \parallel \hat{\epsilon}_{\text{in}})$, where

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

Again, R_O can be taken arbitrarily close to this bound. By Theorem 23, we get

$$1 - \delta \approx R_O + \frac{1}{\sqrt{q} - 1}$$

giving

$$\sigma \approx \frac{\frac{1 - \epsilon_{\text{in}}}{t} - t \left(R_O + \frac{1}{\sqrt{q} - 1} \right)}{1 - t \left(R_O + \frac{1}{\sqrt{q} - 1} \right)}.$$

Now, R_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 proven in [12] that so-called separating systems (SS) give $(2, \epsilon, 0)$ -secure schemes (e_K, A_K) . The FP code C_K is equivalent to a $(2, 2)$ -separating base code C_L , e_K is arbitrary, and A_K is an exhaustive search through all possible two sets $\hat{P} \subseteq C_K$. The $[31, 5, 15]$ 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 codes 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 [6] was 0.015, using the $[126, 14]$ BCH dual as inner code and an AG outer code. On the other hand, [12] offered a rate of 0.026 for an asymptotic class of $(2, 2)$ -SS.

Theorem 25: The SS-RC scheme with the $[126, 14]$ punctured dual of the two-error-correcting BCH code as an inner code, forms an infinite class of strongly $(2, \epsilon)$ -secure schemes

with rate R , for any $R < 0.0476$ and exponentially declining error probability given as

$$\begin{aligned} \epsilon_I &\leq 2^{-n \frac{D(2\Delta - 1 \parallel 0.3 \cdot 10^{-10})}{126}} \\ \epsilon_{\text{II}} &\leq 2^{n(R - D(1 - \Delta \parallel 2^{-14})/126)} \end{aligned}$$

for any Δ such that $1/2 + 1.5 \cdot 10^{-11} < \Delta < 1 - 2^{-14}$.

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

For any innocent user $\mathbf{c}$ and a hybrid fingerprint $\mathbf{x}$, we have $s(\mathbf{c}, \mathbf{x}) \leq 2(1 - \delta)n_O$. 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 ϵ is found from Theorem 13

$$\epsilon_I \leq 2^{-n_O D(-1 + 2\Delta \parallel \epsilon_{\text{in}})}.$$

It is necessary that $\Delta > (1 + \epsilon)/2$, which gives us

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

By Theorem 23, we obtain outer code rate arbitrarily close to

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

The $[126, 14]$ inner code, gives $R_O \approx 0.242$ and an overall rate of 0.0269. This is not as good as using random codes, but it is better than the BBK scheme [6], and like BBK, it can be GS decoded. An alternative inner code is the $[15, 4, 8]$ code. This is too small to use AG codes (cf. Thm. 23), but it works well with random codes.

Theorem 26: The SS-RC scheme with a $[15, 4, 8]$ inner code forms an infinite class of strongly $(2, \epsilon)$ -secure codes with rate R , for any $R < 0.0688$, and exponentially declining error probabilities given as

$$\epsilon_I \leq 2^{-n \frac{D(2\Delta - 1 \parallel 1/140)}{15}} \quad \text{and} \quad \epsilon_{\text{II}} \leq 2^{n(R - D(1 - \Delta \parallel 1/16)/15)}$$

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

Corollary 27: The SS-RC codes with $[15, 4, 8]$ inner codes are strongly $(2, \epsilon)$ secure with length

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

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

B. Practical Codes

In Table III, we present code lengths for 1000 to a billion users with the schemes we know, including those of [13] and [14]. The RS-RC codes are computed with $q = 4$, $\epsilon_{\text{in}} = 0.002$. The error probabilities were set such that 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.

TABLE III
CODE LENGTHS AGAINST TWO PIRATES FOR 1000 TO A BILLION
USERS AND ERROR PROBABILITY $\epsilon \leq 10^{-10}$

$\log M$	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
SOME SS-RS CODES

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]	$0.1 \cdot 10^{-10}$

The constructions of SS-RS are shown in Table IV and the general result is given in the theorem below. Also, the [126, 14] and [254, 24] BCH duals are $(2, 10^{-10})$ secure.

Theorem 28: Let C be a concatenated code of an $(2, \epsilon_{\text{in}}, 0)$ -secure inner code and a $[n_O, k_O]_q$ RS code C_O . If we use GS decoding for C_O , then C is 2-secure with $\epsilon_{\text{II}} = 0$ and

$$\epsilon_{\text{I}} \leq P(Y \leq 2k_O - 3), \quad \text{where } Y \sim \mathcal{B}(n_O, 1 - \epsilon_{\text{in}}). \quad (7)$$

Proof: We have $\epsilon_{\text{I}} \leq P(X > n_O(1 - 2(1 - \delta)) + 1)$ where $X \sim \mathcal{B}(n_O, \epsilon_{\text{in}})$, from the proof of Theorem 13. Using the facts that $n_O = k_O + d_O - 1$ and $Y = n_O - X$, we get the theorem. ■

There is a variant from [15] 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 BS 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 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 for our model, namely the BBK scheme of [6]. The BBK scheme is very good against a few pirates. Against sufficiently many pirates the BS-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 FP schemes in the literature, but most of them use different marking assumptions and, thus, fall outside the scope of this paper. There are two schemes which are weakly (t, ϵ) -secure under the BS marking assumption, due to Tardos [13] and LBH [14]. 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 asymptotic decoding complexity is polynomial in $\log M$, obtained by the GS algorithm for AG outer codes, including the BBK scheme. 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 optimizing the construction parameters of our schemes.

ACKNOWLEDGMENT

The author would like to thank some of the anonymous referees for useful comments and many anonymous colleagues for encouragement and discussions.

REFERENCES

- [1] N. R. Wagner, "Fingerprinting," in *Proc. Symp. Security Privacy*, 1983, pp. 18–22.
- [2] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1897–1905, Sep. 1998.
- [3] B. Chor, A. Fiat, M. Naor, and B. Pinkas, "Tracing traitors," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 893–910, May 2000.
- [4] A. Somekh-Baruch and N. Merhav, "On the capacity game of private fingerprinting systems under collusion attacks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 884–899, Mar. 2005.
- [5] T. Hagerup and C. Rüb, "A guided tour of Chernoff bounds," *Inf. Process. Lett.*, vol. 33, pp. 305–308, 1990.
- [6] A. Barg, G. R. Blakley, and G. A. Kabatiansky, "Digital fingerprinting codes: Problem statements, constructions, identification of traitors," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 852–865, Apr. 2003.
- [7] K. Yoshioka, J. Shikata, and T. Matsumoto, "Collusion secure codes: Systematic security definitions and their relations," *IEICE Trans. Fundamentals*, vol. E87-A, no. 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, vol. 3200, Lecture Notes Comput. Sci., pp. 282–292.
- [9] H.-J. Guth and B. Pfitzmann, "Error- and collusion-secure fingerprinting for digital data," in *Proc. Information Hiding*, New York, 2000, vol. 1768, Lecture Notes Comput. Sci., pp. 134–145, Springer-Verlag.
- [10] V. Guruswami and M. Sudan, "Improved decoding of Reed-Solomon and algebraic-geometry codes," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 1757–1767, Sep. 1999.
- [11] M. A. Tsfasman, "Algebraic-geometric codes and asymptotic problems," *Discrete Appl. Math.*, vol. 33, no. 1–3, pp. 241–256, 1991.
- [12] H. G. Schaathun, "Fighting two pirates," in *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*. New York: Springer-Verlag, May 2003, vol. 2643, Lecture Notes Comput. Sci., pp. 71–78.
- [13] G. Tardos, "Optimal probabilistic fingerprint codes," *J. ACM*, 2005, To appear. In part at STOC'03.
- [14] T. Van Le, M. Burmester, and J. Hu, "Short c-secure fingerprinting codes," in *Proc. 6th Information Security Conf.*, Bristol, U.K., Oct. 2003.
- [15] M. Fernandez and M. Soriano, Protecting intellectual property by guessing secrets. Prague, Czech Republic, 2003, EC-Web, Sep. 2–5, 2003.

Hans Georg Schaathun (M'00) was born in Bergen, Norway, in 1975.

He was a Lecturer in coding and cryptography with the University of Bergen, Bergen, Norway, in 2002 and Postdoctoral from 2003 to 2006. He began as a Permanent Lecturer with the Department of Computing, School of Electronics and Physical Sciences (SEPS), University of Surrey, Surrey, U.K. in 2006. His research interests are multimedia security, algebraic coding theory, and cryptography.