On Watermarking/Fingerprinting for Copyright Protection

Hans Georg Schaathun
Dept. of Computing, SEPS
University of Surrey
Guildford, GU2 7XH
England
georg@ii.uib.no

Abstract

Digital fingerprinting has been suggested for copyright protection. Using a watermarking scheme, a fingerprint identifying the buyer is embedded in every copy sold. If an illegal copy appears, it can be traced back to the guilty user. By using collusion-secure codes, the fingerprinting system is made secure against cut-and-paste attacks.

In this paper we study the interface between the collusion-secure fingerprinting codes and the underlying watermarking scheme, and we construct several codes which are both error-correcting and collusion-secure. Error-correction makes the system robust against successful attacks on the watermarking layer.

Keywords

collusion-secure fingerprinting, copyright protection, error-correcting codes, watermarking, soft-decision decoding

1. Introduction

Protecting copyright is one of the hottest topics in information and media technology at the moment. Digital technology enables perfect copying on amateur equipment, and artists and companies worry about lost sales. Digital fingerprinting is one of many proposed countermeasures. This was first proposed in [6] and has received increasing interest following [1].

The basic idea in fingerprinting is to make every sold copy unique, by embedding a fingerprint identifying the buyer. Thereby any emerging illegitimate copy can be traced back to the guilty party. For instance the Beazley Archive in Oxford uses fingerprinting technology from IBM and DataMark on images on their web site.

A major challenge is to make this system secure against coalitions of pirates. By comparing their different copies, colluding pirates can identify, and consequently remove or damage, part of the fingerprint. A typical attack would be the cut-and-paste attack, where a collusion of pirates cut segments from their individual copies and paste them together to form a hybrid copy with a hybrid fingerprint.

Fingerprinting (FP) is often divided into two modules. A collusion-secure code is used in order to select fingerprints which are resistant against collusive attacks. An underlying watermarking (WM) scheme (see e.g. [2]) is used to embed the fingerprint in the digital file.

Most of the literature studies the two modules separately. The fingerprinting literature has defined its requirements for the WM scheme as a Marking Assumption. Unfortunately, most of the research has been based on very unrealistically strong Marking Assumptions, for instance that of [1]. In this paper, following Guth and Pfitzmann [3], we use a weaker Marking Assumption, which allows for some successful attacks in the watermarking layer as well as the cut-and-paste attack. The solution is codes that are both collusion-secure and error-correcting.

2. The layered FP/WM model

Figure 1 shows the layered structure of a fingeprinting system with an FP and a WM module. The buyer's identity is input to an encoder to produce a fingerprint consisting of n symbols. The file is divided into n segments, and each symbol is embedded independently embedded in one segment by the WM embedder.

Watermarking, briefly defined, is a technique to embed a message in a digital file in such a way that an adversary is unable to remove or change the message. Neither the existence nor the contents of the message is assumed to be secret, contrary to the scenario in steganography.

The purpose of the FP layer is to make the system resistant against coalitions of pirates. If several users collude and compare their copies, they will observe some differences which must be part of the watermark/fingerprint. This allows them to damage the fingerprint.

The independent study of the FP and WM layers raises one key issue: Is there an agreed interface between the layers? We shall address this interface shortly.

The pirates can essentially mount attacks on each layer. A watermarking attack would operate on the individual segments, whereas a fingerprinting attack works on the sequence as a whole.

Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06)
0-7695-2616-0/06 $20.00 © 2006 IEEE

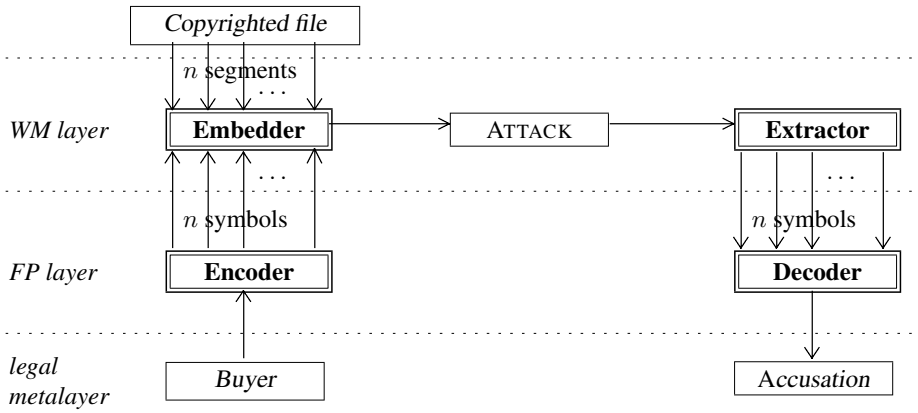

Figure 1. Watermarking/Fingerprinting model.

The archetypical fingerprinting attack is cut-and-paste; a collusion of pirates cut segments from their various copies and paste them together. The result is a hybrid copy where every segment contains a symbol matching some pirate, but where the full string of symbols is no fingerprint of the pirates'.

Many traditional works on fingerprinting considered only the cut-and-paste fingerprint attack, and assumed that the WM extractor was infallible. In other words, x_i from the watermarking layer would always match the i-th symbol of at least one of the pirates. The classic phrasing of this assumption is as follows [1].

Definition 1 (Boneh-Shaw)
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 $\vec{x}$ for any $\vec{x} \in F(P)$, where

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

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

A code C is said to be (t, ϵ)-secure under the Marking Assumption, if, when there are at most t pirates, the output L of the fingerprinting decoder is a non-empty subsets of the pirates with probability at least $1 - \epsilon$. The most well-known such code is due to Boneh and Shaw [1].

A real watermarking scheme cannot be expected to be infallible. We say that the extraction algorithm fail in position i if the output x_i does not match the i-th symbol of any of the pirate fingerprints. Such failure can be either accidental or due to pirate attacks.

The pirates can attack the watermarking layer by modifying the segment(s). Possible attacks depend on the WM scheme in use, but typically they include random noise and geometrical distortion. If the attack is too strong, the file would be rendered useless, but we expect that the extractor will fail with a certain probability p_e. This leads to a weaker Marking Assumption [3] as follows.

Definition 2 (Guth-Pfitzmann)
Let $P \subseteq C$ be the set of fingerprints held by a coalition of pirates, and let x_i be the output x_i from the watermarking layer in position i. The probability that for all $(c_1 \ldots c_n) \in P$, $c_i \neq x_i$, is at most p_e, independently of the output x_j for all other columns $j \neq i$.

Note that when $p_e = 0$, this coïncides with the Boneh-Shaw Marking Assumption. Guth and Pfitzman [3] presented an adapted version of the Boneh-Shaw scheme for this new Marking Assumption. Our solution in this paper will be more efficient than that, by incorporating the latest improvements on Boneh-Shaw [4, 5].

The assumption of independent segments is crucial in order to use simple statistical models and formulæ. In real applications it may not be true. It is likely that some segments are independent whereas others are more or less correlated. Now it is important to remember that the pirates do not know to which code column a given segment corresponds. Thus, they will have no means to predict the correlation between two code columns, and it seems reasonable to assume independence as a fair approximation, though we have to assert it for potential watermarking schemes.

3. The Boneh-Shaw code

We present a novel, error-correcting version of Boneh-Shaw; it is simpler and more efficient than the one from [3]. We view the Boneh-Shaw FP system as a three-layer system, where we split the FP layer into a collusion-secure (CS) and an error-correcting (EC) layer. Even though this layering was not explicit in [1], the original scheme fits well.

We introduce a couple of defintions from coding theory. An $(n, M)_q$ code is a set of M words of length n over an alphabet of q symbols. If q is suppressed, it is generally assumed to be 2. The minimum (Hamming) distance between two distinct codewords is denoted by d. An $(n, M, d)_q$ code is an $(n, M)_q$ code with minimum distance d.

The inner code in [1], fitting in the CS layer, was a q-secure, binary $(r(q-1), q)$ code called Γ. In the EC layer, an outer random code was used.

Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06)
0-7695-2616-0/06 $20.00 © 2006 IEEE

Each buyer is represented in the EC layer by a q-ary word of length n_{O} drawn uniformly at random. By concatenation of codes, this is converted in the CS layer to a binary word by mapping each q-ary symbol to a codeword of Γ. The bits arising from the same symbol in the outer code constitute a *block*.

The vendor applies a random, secret permutation of the code bits between the CS and WM layer. This ensures that the correspondence between file segments and codeword bits is known only to the vendor.

The outer code is inherently error-correcting, even though this fact was not originally exploited. Once this was realised, it was evident that that we could improve the rate by using a weaker FP code and let the EC code compensate for that [4].

The key to the present work is to recognise that the error-correcting capability can also be used to correct errors from any underlying layer, in particular the WM layer.

The original scheme used hard decision decoding, such that the decoder in the FP layer would output for each block, one symbol from the q-ary alphabet. Following [5], we use soft decision decoding, which means that the decoder output a heuristic for each possible symbol reflecting its likelihood. Since this gives more fine-grained information, it allows the EC decoder to make a more reliable accusation.

It was found in [5] that a replication factor $r = 1$ was optimal, such that Γ is an upper triangular 0-1 matrix, i.e. with ones on and above the main diagonal.

Let $(X_1, \ldots, X_{q-1})$ be a block of a hybrid fingerprint. Let $X_0 = 0$ and $X_q = 1$ by convention. Note that unless user i is seen by a pirate, the pirates cannot distinguish between the $(i-1)$-th and the i-th column. Hence the probability of outputting a 1 is equal for the two columns, i.e. $X_i \sim X_{i-1}$.

The original scheme used hard decision decoding, which means that the CS decoder commits to a certain symbol for each block. In [5] improvements where made by using soft decision decoding instead; i.e. a heuristic is output for every possible symbol in each block, where a high heuristic indicates a symbol which is likely to be correct. It was suggested that the CS decoder output the vector $(V_j : j \in \Gamma)$ where $V_j = X_j - X_{j-1}$. Observe that all the V_j sum to 1 and $V_j \in [-1, 1]$ for all j. Furthermore, if the pirates cannot see symbol j and $j \notin \{1, q\}$, then $E(V_j) = 0$.

We now turn to the decoder for the EC layer. As input, it will receive, for each of the n_{O} blocks, a vector $(V_j; j \in \Gamma)$.

After inner decoding of each block, we form the $q \times n$ reliability matrix $R = [r_{i,j}]$ where the j-th column is the vector $(V_1, \ldots, V_q)$ from inner decoding of the j-th block. The output of the soft decision list decoder is a list $L \subseteq C$ of codewords

$$L = \{\vec{c} : W(\vec{c}) \geq \Delta n\}, \tag{1}$$

$$W((c_1, \ldots, c_n)) = \sum_{i=1}^{n} r_{i,c_i}. \tag{2}$$

For random codes, the list decoding has to be implemented as an exhaustive search with complexity $O(M)$. It is possible to use algebraic codes with more efficient decoding, and we expect to present this in a more comprehensive version of this paper.

We employ the common assumption that the pirates make independent decisions in each column (segment), such that all the X_i are independent and distributed as $B(1, p_i)$ for some probability p_i. This assumption is reasonable by the laws of large numbers, if there is at least a moderately large number of columns indistinguishable for the pirates. Most importantly, this assumption implies that the r_{i,c_i} for different i are stochastically independent, allowing us to use the well-known Chernoff bound, defined as follows.

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 $\delta \in [0, 1]$, we have

$$P\left(\sum_{i=1}^{t} X_i \leq t\delta\right) \leq 2^{-tD(\delta||x)}, \quad \textit{when } \delta < x,$$

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

where

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

We distinguish between two types of errors: Type I is the case where the output list L contains no guilty pirate, and Type II the case where L contains innocent users. Let ϵ_{I} and ϵ_{II} be the probabilities of Type I and Type II errors. The following bounds are proved in the full version of this paper. The total length of the scheme is the total number of segments used, and it is given as $n = (q-1)n_{\mathrm{O}}$.

Theorem 2 (Error probabilities)
Suppose there are at most t pirates, and that they have probability at most $p_e < 1/2$ of causing an error in an undetectable position. Using the concatenated code with a BS inner code and a random code with soft input list decoding, with threshold Δ such that $1/q < \Delta < (1-2p_e)/t$, the error probabilites are as follows

$$\epsilon_{\mathrm{I}} \leq \exp -n_{\mathrm{O}} D\left(\frac{1+\Delta}{2} || \frac{t+1}{2t} - \frac{p_e}{t}\right),$$

$$\epsilon_{\mathrm{II}} \leq \exp\left(R_{\mathrm{O}} \log q - D\left(\frac{1+\Delta}{2} || \frac{q+1}{2q}\right)\right) n_{\mathrm{O}}$$

The bound on ϵ_{I} is independent of the choice of outer code, whereas the bound on ϵ_{II} must be derived specifically for each choice of outer code. We observe that for $p_e = 0$, the above theorem reduces to the original result of [5].

Example 3.1 *Suppose we require a Boneh-Shaw scheme with $t = 20$, $M = 2^{20}$, $p_e = 2\%$, $\epsilon_{\mathrm{II}} = 10^{-3}$, and $\epsilon_{\mathrm{I}} = 10^{-6}$. We use $q = 3t$. Setting equality in Theorem 2, we get*

$$3 \log 10 = D\left(\frac{1+\Delta}{2} || \frac{21}{40} - \frac{p_e}{20}\right) n_{\mathrm{O}},$$

$$6 \log 10 = D\left(\frac{1+\Delta}{2} || \frac{61}{120}\right) n_{\mathrm{O}} - 20.$$

We solve the equations to get $\Delta \approx 0.0376$, $n_{\mathrm{O}} \approx 126\,660$, and consequently $n = 7\,472\,940$.

Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06)
0-7695-2616-0/06 $20.00 © 2006 IEEE

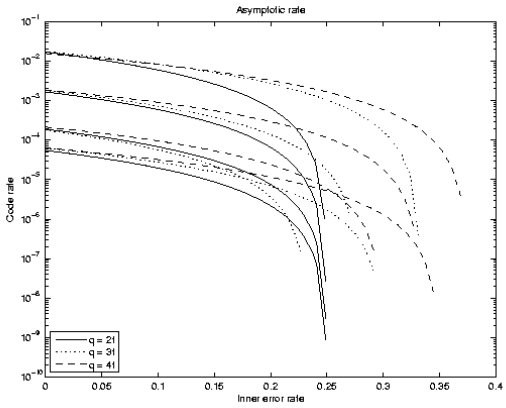

Figure 2. Code rates for concatenated codes with BS inner codes and random codes for varying underlying error rates and varying q for $t = 2, 4, 8, 12$.

Remark 3.1
Similar calculations to the example for fewer pirates give length $5\,655$ *for* $t = 2$, $21\,744$ *for* $t = 3$, $109\,074$ *for* $t = 5$, *and* $915\,385$ *for* $t = 10$, *still assuming a million users and* 2% *WM errors.*

The code rate is defined as $R = (\log_2 M)/n$. To get a general impression of the efficiency, it is useful to study the asymptotic rate, i.e. $\lim_{M\to\infty} R$. The following theorem gives the result, and Figure 2 compares some sample figures.

Theorem 3
There is an asymptotic class of (t, ϵ)*-secure codes with* $\epsilon \to \infty$ *and rate given by*

$$R_t \approx \frac{D\left(\frac{t+1-2p_e}{2t} || \frac{q+1}{2q}\right)}{q-1}, \quad \text{for any} \quad q > \frac{t}{1-2p_e}.$$

The theorem obviously demands $q = \Omega(t)$, but we cannot see any nice expression for the optimal value of q.

4. Conclusion and future research

The layered WM/FP models illustrate how fingerprinting and watermarking can be studied separately and combined as black boxes, *if* we have a clear and common understanding of the interface. Past works on fingerprinting for the Boneh-Shaw model have often suggested to use an underlying WM scheme, without comparing the assumptions about the interface. Similarly, watermarking works have referred to the Boneh-Shaw FP scheme without discussing the interface.

In this work, we adapt the Boneh-Shaw fingerprinting scheme to allow for random errors from the watermarking layer. Even allowing for errors, our codewords are shorter than those of Boneh-Shaw for no errors. This makes also makes it more efficient than that of [3], which had to increase the code length over Boneh-Shaw in order to allow for errors. The modular description also points out various ways to get further improvements, and versions using algebraic outer codes have been constructed. Their rate is inferior, but the decoding complexity is superior.

Acknowledgements

The author is grateful for many useful discussions with dr. Stefan Katzenbeisser of Munich, dr. Marcel Fernandez of Barcelona, and prof. Gérard Cohen of Paris.

References

[1] D. Boneh and J. Shaw. Collusion-secure fingerprinting for digital data. *IEEE Trans. Inform. Theory*, 44(5):1897–1905, 1998. Presented in part at CRYPTO'95.

[2] J. Cox, M. Miller, and J. Bloom. *Digital Watermarking*. Morgan Kaufmann, 2002.

[3] H.-J. Guth and B. Pfitzmann. Error- and collusion-secure fingerprinting for digital data. In *Information Hiding '99, Proceedings*, volume 1768 of *Springer Lecture Notes in Computer Science*, pages 134–145. Springer-Verlag, 2000.

[4] H. G. Schaathun. The boneh-shaw fingerprinting scheme is better than we thought. *IEEE Transaction on Information Forensics and Security*, June 2006.

[5] H. G. Schaathun and M. Fernandez-Muñoz. Boneh-Shaw fingerprinting and soft decision decoding. In *Information Theory Workshop*, 2005. Rotorua, NZ.

[6] N. R. Wagner. Fingerprinting. In *Proceedings of the 1983 Symposium on Security and Privacy*, 1983.