# On the Assumption of Equal Contributions in Fingerprinting

Hans Georg Schaathun

***Abstract*—With a digital fingerprinting scheme a vendor of digital copies of copyrighted material 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. A coalition of pirates may combine their copies to produce an unauthorised copy with a false, hybrid fingerprint.**

It is often assumed in the literature that the members of the collusion will make equal contributions to the hybrid fingerprint, because nobody will accept an increased risk of being caught. We argue that no such assumption is valid *a priori*, and we show that a published solution by Sebé and Domingo-Ferrer can be broken by breaking the assumption.

***Index Terms*—Digital fingerprinting, collusion-secure code, watermarking, collusion-attack, scattering codes**

EDICS Category: WAT-FING

I. Background

The problem of digital fingerprinting was introduced in [1], and has received quite some attention following [2]. A vendor of digital copies of copyrighted material wants to prevent unauthorised copying. Digital fingerprinting makes it possible to trace the guilty user (pirate) when an illegal copy is found. This is done by embedding a secret identification mark (fingerprint) in each copy, making every copy unique.

Typically a robust watermarking (WM) scheme is used to hide the fingerprint in the file. WM schemes are designed to hide any message in a file, in such a way that they can be recovered, even after being subject to noise, signal processing operations, or even malicious attacks.

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. However, a collusion of users can compare their copies, and thereby find regions which differ and hence must be part of the fingerprint. A simple attack is to cut and paste segments from their individual copies, to produce a hybrid copy where the fingerprint does not match any of the colluders'.

Many authors [3], [4] assume that a collusion will always make a hybrid by combining *equal* shares from each of their fingerprints. This is based on the idea that the more the user fingerprint resembles the hybrid, the more likely the user is to be accused. Obviously nobody would accept a higher risk of being accused.

The assumption may be correct when closest neighbour or correlation decoding is used [3], but in general it is not valid. Obviously, if we prove that the system is secure assuming a certain user behaviour, then we are sure that a malicious (and intelligent) user will find some other behaviour. This is illustrated by the scattering codes [5], and we shall prove that they are indeed not secure when the users are not restricted to equal contributions.

The purpose of this paper is to highlight how important it is in information security, not to jump to conclusions about user behaviour. Any statement about user behaviour must be demonstrated based on the actual system.

The author is with the University of Surrey, Dept. of Computing, GU2 7XH Guildford, England. Email: ⟨H.Schaathun@surrey.ac.uk⟩

Part of the presented research was conducted under employment with the University of Bergen, Norway, with funding from the Norwegian Research Council under grant no. 146874/420.

II. Collusion-Secure Codes

A common model for fingerprinting combines a WM scheme with a collusion-secure code (CSC) [3]. An $(n, M)_q$ code C is a set of M words $(c_1, \ldots, c_n)$ over a q-ary alphabet Q. Each user is associated with a fingerprint (word) $\mathbf{c} \in C$. The file is divided into n segments, and each symbol c_i is embedded independently in a corresponding segment.

The code C is often viewed as an $n \times M$ matrix called the codebook, where the rows are codewords. Each column corresponds to a segment of the file.

A collusion of t pirates will have a set $\mathcal{P} \subseteq \mathcal{C}$ of fingerprints. We will also think of $\mathcal{P}$ as an $n \times t$ matrix, and refer to columns of $\mathcal{P}$. A column i, $1 \leq i \leq n$, is *detectable* if more than one element of Q occurs in column i of $\mathcal{P}$. We will assume that the correspondence between file segments and code columns is chosen pseudo-randomly by the vendor and kept secret [6]. Hence the colluders have no means of knowing which columns they detect.

By comparing their copies, the pirates are able to produce an unauthorised copy with a hybrid fingerprint $\mathbf{x} \in Q^n$. The pirates choose an *attack function* $A : Q^{n \times t} \to Q^n$, possibly stochastic, taking the pirate fingerprints $\mathcal{P}$ as input and returning the hybrid fingerprint $\mathbf{x}$.

The set of hybrid fingerprints producible by P is called the *feasible set* $F_{\mathcal{C}}(\mathcal{P})$. This restricts the attack function, so that $A(\mathcal{P}) \in F_{\mathcal{C}}(\mathcal{P})$. The most common model, due to [2], assumes that

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

In other words, each symbol x_i in the hybrid fingerprint $\mathbf{x}$ must occur in the i-th column of $\mathcal{P}$. This is known as the Marking Assumption.

A *tracing algorithm* for the code $\mathcal{C}$ is any algorithm $T : Q^n \to \{L : L \subseteq C\}$. The input is the hybrid fingerprint $\mathbf{x}$ from an unauthorised copy, and the output L is a list of users who are accused of the copyright violation. If $\mathcal{P}$ is a set of pirate fingerprints and A is an attack function producing $\mathbf{x} = A(\mathcal{P})$, then T is successful if $L \subseteq \mathcal{P}$ and $L \neq \emptyset$. If T is not successful, we say that there is an error. A (probabilistically) collusion-secure code is one with a tracing algorithm with bounded error probability.

III. Scattering codes (SC)

Scattering codes were introduced in [5], [7] and used in conjunction with a simplex code to give a probabilistically 3-secure code. An alleged attack [8] was rebutted in [6].

The scattering code $\mathrm{SC}(r, t)$ is a probabilistic encoding of a single bit. Each bit value is encoded as one out of t possible words, chosen uniformly at random. The code has $2t + 1$

Encodes	Zone A	Zone B	Zone C
1	1111	1111 0000 0000	0000 0000 0000
	1111	0000 1111 0000	0000 0000 0000
	1111	0000 0000 1111	0000 0000 0000
0	0000	0000 0000 0000	1111 0000 0000
	0000	0000 0000 0000	0000 1111 0000
	0000	0000 0000 0000	0000 0000 1111

Table I
THE SCATTERING CODE $\mathrm{SC}(4,3)$.

A	B_1	B_2	B_3	C_1	C_2	C_3
r bits	r bits	r bits	r bits	r bits	r bits	r bits
1111	1111	0000	0000	0000	0000	0000
1111	0000	1111	0000	0000	0000	0000
0000	0000	0000	0000	1111	0000	0000

Table II
THREE PIRATE CODEWORDS.

distinct columns replicated r times. We divide the columns into three zones. Zone A has r identical columns where a word has one if and only if it encodes one. Zone B has t distinct columns of weight one replicated r times, and all words encoding zero are zero. Zone C is similar, with t distinct columns of weight 1, and words encoding one are zero. Table I gives an example.

As part of the embedding, the fingerprint is xor-ed with a random, secret bit string $\mathbf{k}$. Similarly, the extracted hybrid fingerprint is xor-ed with $\mathbf{k}$ before descattering. The effect of this is that the colluders cannot tell whether a segment hides a 0 or a 1; they can only tell whether or not two segments are different. (This randomisation prevents the attack from [8].)

Assuming that we detect a hybrid fingerprint produced by three colluders, the following decoding algorithm aims to recover a symbol seen by at least two of the colluders.

Algorithm 1 (Descattering [5]) *The decoding algorithm for scattering codes (*descattering*) uses the first applicable rule in the following list. One block is one set of r identical columns.*

1) *If there are at least two blocks of Zone B with at least one one-bit, then decode as 1.*
2) *If there are at least two blocks of Zone C with at least one one-bit, then decode as 0.*
3) *If there are more ones than zeroes in Zone A, then decode as 1.*
4) *If there are more zeroes than ones in Zone A, then decode as 0.*
5) *With the same number of zeroes and ones in Zone A, decode as erasure.*

It is easy to validate that the algorithm is always correct if the rows of $\mathcal{P}$ encode the same bit. Table II shows a typical example where the collusion see two different bits. We can see that if the pirates use a minority choice strategy with high probability, they will probably output at least one 1-bit in each of B_1 and B_2, and decoding rule 1 will cause decoding to 1. If they use a majority choice strategy with high probability, they are likely to produce a majority of ones in block A, and cause correct decoding of 1 by Rule 3. In the lemma below, we will establish the exact probability of correct decoding.

In [7], attacks were considered where the colluders make independent, random choices for each segment. We describe the strategy as a tuple (p_1, p_2, p_3) where p_i is the probability that the attack outputs the bit seen by two colluders (majority choice) in a column where colluder no. i differs from the other two. Due to the assumption of equal contributions, [7] assumed $p_1 = p_2 = p_3$. The following lemma is a generalisation of a result from [7], [5]. The proof is a trivial extension of the original, but is included for completeness.

Lemma 1 *Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ be three codewords held by the collusion, where $\mathbf{a}_i$ encodes the opposite value of the other two codewords. Suppose the pirates pick the majority bit with probability p_j in any column where user j is the minority. Then the probability of correct descattering r_i is given as*

$$r_i = 1 - \frac{1 + (t-1)(\sum_{j\neq i} p_j^r - \prod_{j\neq i} p_j^r)}{t} \cdot \sum_{j=0}^{\lfloor r/2 \rfloor} \binom{r}{j} p_i^j (1-p_i)^{r-j}.$$

Proof: We prove the lemma for $i = 3$, assuming that $\mathbf{a}_3$ encodes a 0. The general case follows by symmetry.

We consider first the case where $\mathbf{a}_1 \neq \mathbf{a}_2$. Suppose the pirate codewords are as depicted in Table II. In order to get a decoding error, both Rule 1 and 3 have to fail. The first rule fails if least one of Block B_1 and B_2 is all zero, and this happens with probability

$$P_1 = p_1^r + p_2^r - p_1^r p_2^r.$$

The other rule fails if Zone A has a majority of zeros, which happens with probability

$$P_3 = \sum_{j=0}^{\lfloor r/2 \rfloor} \binom{r}{j} p_3^i (1-p_3)^{r-j}.$$

The two events are independent, so the error probability is $P_1 \cdot P_3$.

If $\mathbf{a}_1 = \mathbf{a}_2$, there is only one block Zone B, say B_1, where the pirates see two different bits. Hence decoding Rule 1 always fails, and we have a decoding error with probability P_3.

For each bit, one of the t codewords is chosen uniformly at random. Hence $P(\mathbf{a}_1 = \mathbf{a}_2) = 1/t$, and we get the following total error probability,

$$P_{\mathrm{E}} = P_3 \cdot \frac{(t-1)P_1 + 1}{t},$$

which is equivalent to the formula in the theorem. Note that the error probability increases in p_1 and p_2 and decreases in p_3. ∎

If $p_1 = p_2 = p_3 =: p$, then clearly $r_1 = r_2 = r_3 =: \rho(p)$ by symmetry. The worst-case probability of successful descattering, $p^*(r,t) := \min_p \rho(p)$, is calculated in [5].

In [5], a scattering code S was concatenated with a simplex code C as outer code. In other words, each user was represented by a binary codeword $\mathbf{c} \in C$, and each bit of $\mathbf{c}$ was encoded using S. The decoder would first descatter block by block, and then decode the resulting vector with respect

	Strategy			
Innocent user	(111)	(001)	(010)	(100)
$\mathbf{c}_1 = (1000)$	(0110)	(0000)	(1100)	(1010)
$\mathbf{c}_2 = (0100)$	(1010)	(0000)	(1100)	(0110)
$\mathbf{c}_3 = (0010)$	(1100)	(0000)	(1010)	(0110)
$\mathbf{c}_4 = (1110)$	(0000)	(1100)	(1010)	(0110)

Table III
EXAMPLE OF THE STRATEGIES IN THE PROOF OF THEOREM 1.

to C using closest neighbour decoding. It was proved that if the descattering is successful with sufficiently high probability, then the error probability of the outer decoding can be made arbitrarily close to 0. As we shall see, this result is only valid for strategies (p, p, p).

Remark 1 *Closest neighbour decoding invariably returns one and only one codeword which is accused. There is no distinction between false negatives and false positives. An error means that the returned user is innocent, and both a false positive and a false negative are implied. If the returned user is guilty, we say that decoding is correct, but there are still two unidentified members of the collusion (false negatives).*

We can view the outer code as an fingerprinting code in itself. For a column of C where colluder i has a fingerprint different from the other two, we can define a probability r_i that the resulting hybrid fingerprint *after* descattering matches the other two colluders. This is the probability r_i given in Lemma 1. In effect, we get a colluder strategy (r_1, r_2, r_3) with respect to the outer code.

Theorem 1 *A fingerprinting scheme with scattering inner codes and a linear outer code has error rate at least* $1/4$ *if the pirates use an optimal strategy, regardless of the outer decoding algorithm used.*

In particular, the simplex code C is linear. In the case of list decoding, an error rate of 25% means that, on average, 25% of all accused codewords are false positives.

Proof: We propose a mixed strategy where the colluders choose a pure strategy (p_1, p_2, p_3) uniformly at random from $\{(1,1,1),(1,0,0),(0,1,0),(0,0,1)\}$. Observe that each of these four strategies gives $(r_1, r_2, r_3) = (p_1, p_2, p_3)$.

Consider three linearly independent codewords $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$, and $\mathbf{c}_4 = \mathbf{c}_1 + \mathbf{c}_2 + \mathbf{c}_3$. By linearity $\mathbf{c}_4 \in C$. Any collusion of three out of these four codewords using our proposed strategy will produce the same four false fingerprints with equal probabilities. Hence when one of these false fingerprints is detected, there are four users who are equally likely to be guilty, and one of them is innocent. ∎

Example 1 *Table III shows an example of the strategies used in the proof. Since the attack works independently on each column, we have truncated the codewords to display each column type (up to equivalence) once. The first column shows the four codewords, and each row then shows the four hybrid words generated when the corresponding user is innocent (by the other three codewords). Note that each one of the codewords the decoder can observe appears once in each row. Consequently any one of the four users may be innocent.*

Remark 2 *The problem with the original construction is clearly in the outer code. Our attack only works because the outer code is linear. It was proved in [9], [10], that a secure construction can be made by using a non-linear outer code (so-called* $(2,2)$*- and* $(3,1)$*-separating codes). Unfortunately, the rate of such a code is inferior to other codes in the literature, and therefore we have omitted the details.*

IV. CLOSING WORDS

We conclude that the fingerprinting code of [7] is broken if the colluders refuse to accept the assumption of equal contributions, and an optimal attack gives an error rate of at least 25%. This proves that the assumption of equal contributions is not valid *in general*, and it is worrying that this assumption so often is accepted in the literature (e.g. [4]) without argument.

It is an open question to check the assumption for other proposed solutions where it has been applied. In many cases it can almost certainly be proved that there exists an optimal attack where equal contributions are used, but then this would be a property of the particular fingerprinting scheme and not of the fingerprinting model.

REFERENCES

[1] N. R. Wagner, "Fingerprinting," in *Proceedings of the 1983 Symposium on Security and Privacy*, 1983.

[2] D. Boneh and J. 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] M. Wu, W. Trappe, Z. J. Wang, and K. J. R. Liu, "Collusion resistant fingerprinting for multimedia," *IEEE Signal Processing Magazine*, 2004.

[4] S. He and M. Wu, "Joint coding and embedding techniques for multimedia fingerprinting," *IEEE Trans. Information Forensics and Security*, vol. 1, pp. 231–248, Jun. 2006.

[5] F. Sebé and J. Domingo-Ferrer, "Short 3-secure fingerprinting codes for copyright protection," in *ACISP 2002*, ser. Springer Lecture Notes in Computer Science, vol. 2384. Springer-Verlag, 2002, pp. 316–327.

[6] ——, "Critique to Burmester and Le attack on sebé and dommingo-ferrer fingerprinting scheme," *Electronics Letters*, vol. 40, Sep. 2004.

[7] ——, "Scattering codes to implement short 3-secure fingerprinting for copyright protection," *Electronics Letters*, vol. 38, pp. 958–959, Aug. 2002.

[8] M. Burmester and T. Le, "Attack on Sebé, Dommingo-Ferrer and Herrera-Joancomarti fingerprinting schemes," *Electronics Letters*, vol. 40, Feb. 2004.

[9] H. G. Schaathun, "Fighting three pirates with scattering codes," in *Proc. IEEE Intern. Symp. Inform. Theory*, Jun. 2004.

[10] ——, "Fighting three pirates with scattering codes," Dept. of Informatics, University of Bergen, Tech. Rep. 263, 2004, also available at http://www.ii.uib.no/~georg/sci/inf/coding/public/.