Experimental Assessment of Probabilistic Fingerprinting Codes over AWGN Channel

Minoru Kuribayashi[1]

Graduate School of Engineering
1-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo, 657-8501 Japan.
kminoru@kobe-u.ac.jp

Abstract. The estimation of the false-positive probability has been an important concern for fingerprinting codes, and the formula of the probability has been derived under a restricted assumption and statistic model. In this paper, we first analyze the statistic behavior of the value of score derived from the correlation between a pirated codeword and codewords of all users when some bits are flipped. Then, the derivation of the score is adaptively designed to consider the attack model such that a pirated codeword is distorted by additive white Gaussian noise. The traceability and probability of false-positive are estimated by Monte-Carlo simulation, and the validity of the Gaussian approximation for the distribution of score is evaluated for probabilistic fingerprinting codes.

1 Introduction

Due to the progress in information technology, digital contents such as music, images, and movies are distributed from providers to multiple users connected with a network. Although it offers convenient means for users to obtain digital content, it also causes the threats of illegal distribution from malicious parties. In order to prevent users from distributing the pirated version of digital content, digital fingerprinting technique has been studied including the procedure of embedding and detecting fingerprints, secure protocol between buyer and seller, and the way of distribution and identification of illegal action. One of the critical threats for the fingerprinting system is the collusion of users who purchase a same content. Since their fingerprinted copies slightly differ with each other, a coalition of users can combine their fingerprinted copies of the same content for the purpose of removing/changing the original fingerprint. Such an attack is called a collusion attack.

An early work on designing collusion-resistant binary fingerprinting codes was presented by Boneh and Shaw [1] underlying the principle referred to as the *marking assumption*. In this case, a fingerprint is a set of redundant digits which are distributed in some random positions of an original content. When a coalition of users attempts to discover some of the fingerprint positions by comparing their copies for differences, the coalition may modify only those positions where they find a difference in their fingerprinted copies. A c-secure code guarantees the tolerance for the collusion attack with c pirates or less. Tardos

[12] has proposed a probabilistic c-secure code with negligible error probability which has a length of theoretically minimal order with respect to the number of colluders. One of the interesting reports about the characteristic of Tardos's code is presented by Škorić et al. [2] about the symmetric version of the tracing algorithm. In the algorithm, correlation scores are used to detect the pirates. In the report [3], Gaussian approximation of the value of score derived from the correlation between a pirated codeword and codewords of all users. Based on the report, the code length is further shortened under a given false-positive probability. The results are supported and further analyzed by Furon et al. [4]. Nuida et al. [9] studied the parameters to generate the codewords of Tardos's code which are expressed by continuous distribution, and presented a discrete version in attempts to reduce the code length and the required memory amount without degrading the traceability. Moreover, they gave a security proof under an assumption weaker than the marking assumption. However, the goal is to reduce the code length under the binary symmetric channel with a certain error rate. In addition, their estimation is based on the assumption that the number of colluders is less than c which is fixed in advance.

In this paper, we study the statistic behavior of the value of the score when some bits of pirated codeword are flipped, and estimate the attenuation of average value of the score for Nuida's code. In our attack model, a coalition of users produces a pirated copy under the marking assumption, and then, the pirated copy is distorted by attacks intended to remove/modify the watermarked signal embedded in digital content. We assume that the noise injected by the attacks is additive white Gaussian noise (AWGN). So, in our attack model, a pirated codeword produced by collusion attack is further distorted by transmitting over AWGN channel. In such a case, the symbols of received(extracted) codeword are represented by analog value. Considering the case of error correcting code, the soft decision detection can reduce more errors than the hard one which rounds the analog values into digital ones. In [6], the traceability and false positive probability of Tardos's code were analyzed by experiments introducing the soft and hard decision methods into the tracing algorithm, and revealed that the false positive probability is increased with the amount of noise for both methods. In this study, the reason is analyzed by the statistic behavior of the value of the score, and the analysis is further applied for Nuida's code. Moreover, for Nuida's code, the dependency of the number of colluders and the type of collusion attack is measured by the behavior of the value of the score.

It is remarkable that each symbol of a pirated codeword is rounded into binary digit which may be flipped by an additive noise if the hard decision method is used. Thus, the performance of the hard decision method is strongly related to the analysis of the statistic behavior of the value of the score. On the other hand, the soft decision method will be able to utilize the analogue signal to detect more colluders. The performance of the hard and soft decision methods are compared by Monte-Carlo simulation, and it is revealed that the soft decision method is suitable for the case that the amount of noise added to a pirated copy is very large. It is noted that the experimental results of Tardos's

code in [6] are derived under the only restricted environment such that SNR is more than 3 [dB]. In this paper, we evaluate the performance of Nuida's code as well as Tardos's code by varying the SNR from -4 to 10 [dB]. We further evaluate the probabilities of false-positive for various kinds of code length, and compare the performance of Tardos's code with that of Nuida's code from the probability of false-positive point of view. The experimental results reveal an interesting characteristic such that the false positive probability of Nuida's code is almost independent of the amount of noise, but is dependent heavily on it for Tardos's code.

2 Fingerprinting Code

In the fingerprinting system, a distributor provides digital contents to users which contain fingerprint information. The number of users is N. If at most c users are colluded to produce a pirated copy using their fingerprinted copies, a fingerprinting code ensures that at least one of them can be caught from the copy under the well-known assumption called the marking assumption [1].

At the collusion attack, a set of malicious users called colluders try to find the positions of the embedded codeword from differences of their copies, and then to modify bits of the codeword in these positions. Suppose that a codeword of fingerprint codes is binary and each bit is embedded into one of the segments of digital content without overlapping using a robust watermarking scheme. It is possible for colluders to compare their fingerprinted copies of the content with each other to find the differences. In the situation, the positions that the bit of their codewords is different are detectable. The marking assumption states that any bit within a detectable position can be selected or even erased, while any bit without the position will be left unchanged in the pirated codeword. A fingerprint code is called totally c-secure if at least one of the colluders is traceable under the marking assumption with the condition that the number of colluders is at most c. Boneh and Shaw, however, proved that when $c > 1$, totally c-secure code does not exist if the marking assumption is satisfied [1]. Under the weaker condition that one of innocent users will be captured with a tiny probability ϵ, a c-secure code with ϵ-error was constructed.

2.1 Probabilistic Fingerprinting Codes

Tardos [12] has proposed a probabilistic c-secure code with error probability ϵ_1 which has a length of theoretically minimal order with respect to the number of colluders. On the binary digits of the codeword, the frequency of "0" and "1" is ruled by a specific probability distribution referred to as the *bias distribution*. The codewords are arranged as an $N \times L$ matrix $\boldsymbol{X}$, where the j-th row corresponds to the fingerprint given to the j-th user. The generation of the matrix $\boldsymbol{X}$ is composed of two steps.

1. A distributor is supposed to choose the random variables $0 < p_i < 1$ independently for every $1 \leq i \leq L$, according to a given bias distribution.

Table 1. Examples of discrete version of bias distribution.

c	p	q	c	p	q
1,2	0.50000	1.00000	7,8	0.06943	0.24833
3,4	0.21132	0.50000		0.33001	0.25167
	0.78868	0.50000		0.66999	0.25167
5,6	0.11270	0.33201		0.93057	0.24833
	0.50000	0.33598			
	0.88730	0.33201			

2. Each entry $X_{j,i}$ of the matrix $\boldsymbol{X}$ is selected independently from the binary alphabet $\{0,1\}$ with $\Pr(X_{j,i}=1)=p_i$ and $\Pr(X_{j,i}=0)=1-p_i$ for every $1 \le j \le N$.

In the case of Tardos's codes, a certain continuous distribution is used as the bias distribution. The values of p_i is selected from the range $[t, 1-t]$. Here, $t = 1/(300c)$ and $p_i = \sin^2 r_i$ is selected by picking uniformly at random the value $r_i \in [t', \pi/2 - t']$ with $0 < t' < \pi/4$, $\sin^2 t' = t$. Nuida et al. [9] proposed the specific discrete distribution introduced by a discrete variant [10] of Tardos's codes that can be tuned for a given number c of colluders. The bias distribution is called "Gauss-Legendre distribution" due to the deep relation to Gauss-Legendre quadrature in numerical approximation theory (see [9] for detail). The numerical examples of the discrete distribution are shown in Table 1, where q denotes the emerging probability of p.

Let $\mathcal{C}$ be a set of colluders and $\tilde{c}$ be the number of colluders. Then we denote by $\boldsymbol{X_C}$ the $\tilde{c} \times L$ matrix of codewords assigned to the colluders. Depending on the attack strategy ρ, the fingerprint $\boldsymbol{y} = (y_1, \ldots, y_L), y_i \in \{0,1\}$ contained in a pirated copy is denoted by $\boldsymbol{y} = \rho(\boldsymbol{X_C})$. For a given pirated codeword $\boldsymbol{y}$, the tracing algorithm first calculates a score $S_i^{(j)}$ for i-th bit $X_{j,i}$ of j-th user by a certain real-valued function, and then sums them up as the total score $S^{(j)} = \sum_{i=1}^{L} S_i^{(j)}$ of j-th user. For Tardos's code, if the score $S^{(j)}$ exceeds a threshold Z, the user is determined as guilty. The design of appropriate parameters has been studied in [12], [3], [10]. For Nuida's code [9], the tracing algorithm outputs only one guilty user whose score becomes maximum. Although no explicit description about the use of a threshold have been presented, it is supposed to be applicable for Nuida's code. In this paper, we calculate the threshold of Nuida's code in the same manner as that of Tardos's one, and evaluate the validity of the design of the threshold and measure the performance.

By introducing an auxiliary function $\sigma(p) = \sqrt{(1-p)/p}$, the scoring function $S_i^{(j)}$ in [12] is given as follows.

$$S_i^{(j)} = \begin{cases} \sigma(p_i) & \text{if } y_i = 1 \text{ and } X_{j,i} = 1 \\ -\sigma(1-p_i) & \text{if } y_i = 1 \text{ and } X_{j,i} = 0 \\ 0 & \text{if } y_i \in \{0, ?\}\ , \end{cases} \tag{1}$$

where "?" stands for erasure of element. The above scoring function ignores all position with $y_i \in \{0, ?\}$. For such positions, Škorić et al. [2] proposed a symmetric version of accusation sum which scoring function is given as follows.

$$S_i^{(j)} = \begin{cases} \sigma(p_i) & \text{if } y_i = 1 \text{ and } X_{j,i} = 1 \\ -\sigma(1-p_i) & \text{if } y_i = 1 \text{ and } X_{j,i} = 0 \\ \sigma(1-p_i) & \text{if } y_i = 0 \text{ and } X_{j,i} = 0 \\ -\sigma(p_i) & \text{if } y_i = 0 \text{ and } X_{j,i} = 1 \\ 0 & \text{if } y_i =? \end{cases} \tag{2}$$

Note that an erasure symbol "?" is regarded as $y_i = 0$ in Nuida's code.

The traceability is usually evaluated in terms of the probability ϵ_1 of accusing an innocent user and the probability ϵ_2 of missing all colluders. In order to guarantee that the probability of accusing an innocent user is below ϵ_1, Tardos's original code has length $L = 100c^2 \log(N/\epsilon_1)$ [12]. In [3], the constant "100" was reduced to $4\pi^2$ without changing the scheme. For the above symmetric conversion [2], the lower bound of the code length was given by $L > \pi^2 c^2 \log(N/\epsilon_1)$. In the same paper, it was shown that the code length was further reduced by converting the construction of the code from binary to q-ary alphabets. For simplicity, we consider only binary fingerprinting code in this paper.

The number of traceable colluders depends on the design of threshold Z. There are many statistical analyses of proper threshold Z for original and symmetric version of Tardos's fingerprinting code. By modeling the accusation sums as normally distributed stochastic variables, Škorić et al. presented simple approximate expressions for the false-positive and false-negative rates [3]. Moreover, due to the Central Limit Theorem, it is reported that the accusation sums is approximated to follow Gaussian distribution. Under the assumption that the score $S^{(j)}$ follows Gaussian distribution, the threshold Z is expressed by the complementary error function erfc() for a given ϵ_1 [7]:

$$Z = \sqrt{2L} \cdot \mathrm{erfc}^{-1}(2\epsilon_1/N) \ . \tag{3}$$

Furon et al. studied the statistics of the score $S^{(j)}$ in [4]. Without loss of generality, the probability density function (PDF) of $S^{(j)}$ are approximated by the normal distribution $N(0, L)$ when j-th user is innocent, and $N(2L/\tilde{c}\pi, L(1 - 4/\tilde{c}^2\pi^2))$ when he is involved in $\mathcal{C}$. In this study, they insisted that the use of the Central Limit Theorem was absolutely not recommended when estimating the code length because it amounts to integrate the distribution function on its tail where the Gaussianity assumption does not hold. The Berry-Esséen bound shows that the gap between the Gaussian law and the real distribution of the scores depends on their third moment. On the other hand, based on the above distributions of $S^{(j)}$, the probability of true-positive per each colluder and the expected number of detectable colluders are theoretically estimated in [7] when the threshold Z is calculated by Eq.(3) for a given false-positive probability ϵ_1, and the validity is evaluated through computer simulation. The simulation results also show that the probability of false positive is slightly less than the given ϵ_1.

Although the above threshold given by Eq.(3) is specified for the symmetric version of tracing algorithm of Tardos's code, it could be applicable for the Nuida's code. Since the theoretical analysis of the validity of such a threshold is difficult because of its complexity, experimental assessment is performed in this paper.

2.2 Relaxation of Marking Assumption

Although the marking assumption is reasonable to evaluate the performance of fingerprint codes, there is a big gap from practical cases. Even if a watermarking scheme offers a considerable level of robustness, it is still possible to erase/modify the embedded bits with a non-negligible probability due to the addition of noise to a pirated copy. Because of the noise, the extracted signal from such a pirated copy must be distorted from the original signal $y_i \in \{0, 1\}$. Therefore, the bits without the detectable position may be erased/modified by the attacks for the watermarked signal. In our assumption, the effects caused by attacks are modeled by additive white Gaussian noise, and the noise is added after collusion attack. The degraded codeword is represented by

$$\boldsymbol{y'} = \boldsymbol{y} + \boldsymbol{e} \ , \tag{4}$$

where $\boldsymbol{e}$ is the additive white Gaussian noise.

In order to cover more practical cases, various relaxation of the marking assumption have been introduced and several c-secure codes under those assumptions, called robust c-secure codes, have been proposed in [9], [5], [11], [8]. Among those assumptions, there are two common conditions: At least one of the colluders is traceable and the number of colluders is at most c. Their goal is mainly to estimate a proper code length L to satisfy that the probability of accusing an innocent user is below ϵ_1, which is dependent on the number of flipped/erased bits at the undetectable position.

Suppose that a fingerprint code is equipped in a fingerprinting system. Then, the code length must be determined under the considerations of system policy and attack strategies such as the number of colluders and the amount of noise. Here, our interest is how to design the good tracing algorithm that can detect more colluders and less innocent users no matter how many colluders get involved in to generate a pirated copy and no matter how much amount of noise is added to the copy. In this regard, it is meaningful to design a proper threshold Z for a given false probability ϵ_1 and a fixed code length. The threshold Z given by Eq.(3) could adjust well for the relaxed version of the marking assumption. In [6], the number of detectable colluders and false-positive probability for Tardos's code was presented under the relaxed version of the marking assumption. However, it merely showed the results obtained by experiments. Our contribution of this paper is to present the effect of bit flip caused by the additive noise from the viewpoint of the correlation score. Moreover, the performance between Tardos's code and Nuida's code is compared with each other.

In the following sections, we forget about the limitation of c-secure code such that the number of colluders is at most c. The performance of conventional

tracing algorithm based on a threshold Z and its variant is evaluated for arbitrary number of colluders $\tilde{c}$.

3 Distribution of Accusation Sum

3.1 Effect of Bit Flip

In this section, we consider the changes of accusation sum $S^{(j)}$ when arbitrary x bits of pirated codeword are flipped by attack under the assumption that each element of pirated codeword is rounded into a bit, namely, $y_i \in \{0, 1\}$.

Remember that the PDF of $S^{(j)}$ is approximated to be $N(2L/\tilde{c}\pi, L(1 - 4/\tilde{c}^2\pi^2))$ when j-th user is involved in $\mathcal{C}$, and the elements $S_i^{(j)}$ are independent with each other. Since the length of codeword is L, the PDF of $S_i^{(j)}$ is given by $N(2/\tilde{c}\pi, 1 - 4/\tilde{c}^2\pi^2)$. Suppose that i-th bit y_i of pirated codeword is flipped. Then, the corresponding score $S_i^{(j)}$ is changed to $-S_i^{(j)}$ from Eq.(2). It means that the variance of accusation sum $S^{(j)}$ is unchanged by the bit flip, while the average is changed from $2/\tilde{c}\pi$ to $-2/\tilde{c}\pi$. When arbitrary x bits of pirated codeword are flipped, the sum of x elements $S_i^{(j)}$ is expected to be $-2x/\tilde{c}\pi$, and that of the other unflipped $(L - x)$ elements is to be $2(L - x)/\tilde{c}\pi$. Therefore, without loss of generality, when x bits of pirated codeword are flipped, the PDF of $S^{(j)}$ is expected to be $N(2(L - 2x)/\tilde{c}\pi, L(1 - 4/\tilde{c}^2\pi^2))$.

On the other hand, the PDF of $S^{(j)}$ is approximated to be $N(0, L)$ when j-th user is innocent. Then, it is expected that the PDF is unchanged even if any number of bits of pirated codeword are flipped.

Due to the complexity of the parameters introduced in the discrete version of bias distribution in Nuida's code, we skip the theoretical analysis of the distribution of accusation sum under the Gaussian assumption in this paper. Instead, we derive a conjecture of the distribution of accusation sum from the experimental results.

3.2 Numerical Evaluation

The above analysis is evaluated by implementing Tardos's code with the following parameters. The number of users is $N = 10^4$ and the code length is $L = 10000$. The range of bias distribution p_i is fixed by setting $t = 0.000167$ ($c = 20$). Under a constant number of colluders $\tilde{c} = 10$, the PDF of accusation sum $S^{(j)}$ is calculated using Monte-Carlo simulation with 10^6 trials. Table 2 shows the mean and variance of accusation sum when x symbols of pirated codeword are flipped, where the values in parenthesis are theoretical ones. In this experiment, the performed collusion attack is "majority attack": If the sum of i-th bit exceeds $\tilde{c}/2$, then $y_i = 1$, otherwise, $y_i = 0$. The PDF of the distribution is also described in Fig.1, where solid and dashed lines are the experimental and theoretical values, respectively. These results confirm that the PDF of $S^{(j)}$ actually follows $N(2(L - 2x)/\tilde{c}\pi, L(1 - 4/\tilde{c}^2\pi^2))$ in this experiment.

Table 2. The mean and variance of accusation sum $S^{(j)}$ of Tardos's code when $\tilde{c} = 10$, where the values in parenthesis are theoretical ones.

	innocent		colluders	
x	mean	variance	mean	variance
0	-2.6 (0.0)	10499.5 (10000)	644.9 (636.6)	9955.6 (9959.5)
1000	-0.8 (0.0)	10253.5 (10000)	511.6 (509.3)	10042.0 (9959.5)
2000	-8.9 (0.0)	10501.0 (10000)	382.0 (382.0)	10318.3 (9959.5)

The mean and variance of accusation sum for Nuida's code is calculated using the following parameters. The discrete version of bias distribution is selected by the case $c = 7, 8$ in Table 1. The number of colluders is $\tilde{c} = 10$, the code length is $L = 10^4$, and the trials for Monte-Carlo simulation is 10^6, which are the same parameters to Tardos's code. Table 3 shows the mean and variance when x symbols of pirated codeword are flipped. From this table, we make a conjecture of the distribution of accusation sum. At first, it seems difficult to extract useful information from the values of variance. Because the values are almost equal to L and are very similar to that of Tardos's code which variance of colluders' $S^{(j)}$ are expected to be $L(1 - 4/\tilde{c}^2\pi^2)$ from the theoretical analysis. Then, we focus on the mean values of colluders' $S^{(j)}$. Referring to the mean value $2(L - 2x)/\tilde{c}\pi$ of Tardos's code, that of Nuida's code can be experimentally estimated by $2(L - 2x)/2.826\tilde{c}$ from the three mean values in Table 3. In other word, the parameter "π" in the mean value of Tardos's code is replaced by "$A = 2.826$" in that of Nuida's one under the above condition. So, we make the following conjecture for the distribution of accusation sum of Nuida's code; $N(2(L - 2x)/A\tilde{c}, L(1 - 1/2\tilde{c}^2))$, where $A = 2.826$ under "majority attack" and $L = 10000$. In order to confirm the validity of the conjecture, the PDF of $S^{(j)}$ are depicted in Fig.2, where solid and dash lines are the experimental and conjectured values, respectively. From the figure, we can see that the conjectured values are very close the experimental values. These results are derived by using the discrete version of Nuida's bias distribution for $c = 7, 8$ in Table 1. However, the number of colluders is fixed by $\tilde{c} = 10$ in the experiment and only "majority attack" is tested. Considering the design of the bias distribution, the parameter A may be sensitive for the change of $\tilde{c}$. Moreover, the value of A should be measured for different types of collusion attack. The changes of the

Table 3. The mean and variance of accusation sum $S^{(j)}$ of Nuida's code when $\tilde{c} = 10$.

	innocent		colluders	
x	mean	variance	mean	variance
0	-9.5	10456.7	708.3	10316
1000	-6.9	10833.8	562.9	10039
2000	0.1	10119.4	421.5	10332

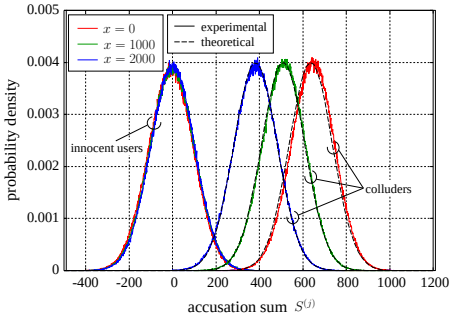


Fig. 1. The PDF of accusation sum $S^{(j)}$ of Tardos's code when $\tilde{c} = 10$.

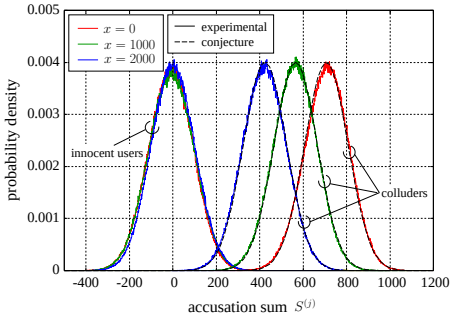


Fig. 2. The PDF of accusation sum $S^{(j)}$ of Nuida's code when $\tilde{c} = 10$.

value of A are depicted in Fig.3 by changing the number $\tilde{c}$ for 5 types of collusion attack; "majority", "minority", "random", "all-0", and "all-1". Under the marking assumption, if i-th bit of c colluders' codewords is different, that of pirated codeword y_i is selected by the following manner.

- majority: If the sum of i-th bit exceeds $c/2$, $y_i = 1$, otherwise, $y_i = 0$.
- minority: If the sum of i-th bit exceeds $c/2$, $y_i = 0$, otherwise, $y_i = 0$.
- random: $y_i \in_R \{0, 1\}$.
- all-0: $y_i = 0$.
- all-1: $y_i = 1$.

The results indicate that the value of A is almost constant when the number $\tilde{c}$ of colluders is below c, and that the value of A is widely varied with the type of collusion attack if $\tilde{c}$ exceeds c. Interestingly, we can see from Fig.3 that the behavior of the values for $\tilde{c} > c$ is completely different with the selection of discrete version of bias distribution in Table 1. The reason will come from the generation of the bias distribution. The detail analysis is left for the future work.

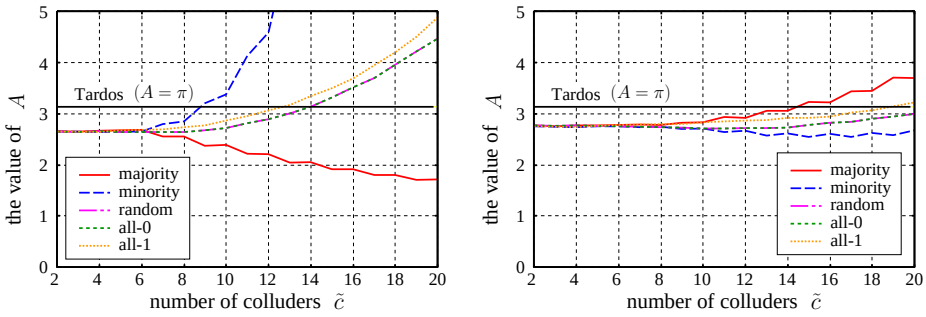
3.3 Estimation of True-Positive and False-Positive

Based on the statistical behavior of the colluders' accusation sum derived by the above experiments, the number of detectable colluders from a pirated copy can be estimated by referring to the analysis in [7]. For Tardos's code, the probability $\Pr[TP]$ of true-positive per each colluder is given by

$$\Pr[TP] = \frac{1}{2}\mathrm{erfc}\left(\frac{1}{\sqrt{2\sigma^2}}\left(\hat{\mathrm{Z}} - \frac{2\mathrm{L}}{\tilde{c}\pi}\right)\right), \tag{5}$$

where

$$\sigma^2 = L\left(1 - \frac{4}{\tilde{c}^2\pi^2}\right). \tag{6}$$



(a) $c = 5, 6$ (b) $c = 7, 8$

Fig. 3. The value of parameter A for 5 types of collusion attack when $L = 10000$.

Using the probability $\Pr[TP]$, the expected number of detectable colluders is given by

$$N_{TP} = \tilde{c}\Pr[TP] = \frac{\tilde{c}}{2}\mathrm{erfc}\left(\frac{1}{\sqrt{2\sigma^2}}\left(\hat{\mathrm{Z}} - \frac{2\mathrm{L}}{\tilde{c}\pi}\right)\right). \tag{7}$$

These analyses are based on the Gaussianity assumption for the distribution of accusation sum. The numerical results of the distribution confirm the validity of the assumption for both Tardos's code and Nuida's code. Therefore, it is expected for Nuida's code that $\Pr[TP]$ and N_{TP} can be represented by Eq.(5) and Eq.(7) where the parameter "π" is replaced by "A".

On the other hand, even if the accusation sum of innocent users can be approximated by Gaussian distribution $N(0, L)$ from the experimental results, the probability of false-positive cannot be simply expressed by Gauss error function as reported in [4]. Thus, the experimental evaluation is required for the probability of false-positive, which is discussed in Sect.5.

4 Soft Decision Method

The signal extracted from a pirated copy is represented by analog value $\boldsymbol{y'}$ because of the addition of noise $\boldsymbol{e}$ in our assumption. Considering the scoring function given by Eq.(2), each symbol of the pirated codeword $\boldsymbol{y'}$ must be rounded into a bit $\{0, 1\}$ or erasure symbol "?". Hence, an extracted signal from a pirated copy is first rounded into digital value, and then the tracing algorithm is performed to identify the colluders. This method is analogous to the hard decision (HD) method in error correcting code. Here, there is an interesting question whether a soft decision (SD) method is applicable to the tracing algorithm by adaptively designing a proper threshold or not. In general, the performance of SD method is much better than the HD method in error correcting code.

Suppose that in the HD method each symbol of the pirated codeword $\boldsymbol{y'}$ is rounded into a bit, which is denoted by $y_i^\star \in \{0, 1\}$ for $1 \le i \le L$. If an erasure

error is occurred, such a symbol is regarded as $y_i^\star = 0$ similar to the tracing algorithm in Nuida's code. Based on Eq.(2), a score $\hat{S}_i^{(j)}$ for i-th bit $X_{j,i}$ of j-th user is represented by

$$\hat{S}_i^{(j)} = \begin{cases} (2y_i^\star - 1)\sigma(p_i) & \text{if } X_{j,i} = 1 \\ -(2y_i^\star - 1)\sigma(1-p_i) & \text{if } X_{j,i} = 0 \ . \end{cases} \tag{8}$$

The design of threshold in Eq.(3) is based on the Gaussian approximation of the score $\hat{S}_i^{(j)}$. From the discussion in Sect.3.1, the PDF of $\hat{S}^{(j)} = \sum_{i=1}^{L} \hat{S}_i^{(j)}$ is $N(0, L)$ when j-th user is innocent even if any symbols in $\boldsymbol{y'}$ are flipped from that in $\boldsymbol{y}$, and hence, the proper threshold Z_{HD} is calculated by Eq.(3). In the SD method, $y_i^\star$ in Eq.(8) is replaced by y_i' to calculate the score directly from the extracted analog signal $\boldsymbol{y'}$. Since $\boldsymbol{y'}$ is distorted by AWGN channel, the effect on the score is also approximated to follow Gaussian distribution. Hence, if the variance σ_{SD}^2 of the accusation sum is obtained, the proper threshold Z_{SD} can be designed using the same equation as the case of HD method:

$$Z_{SD} = \sqrt{2\sigma_{SD}^2}\mathrm{erfc}^{-1}(2\epsilon_1/N) \ . \tag{9}$$

Because of the randomness in the generation of codeword, the variance σ_{SD}^2 can be calculated as follows.

1. Generate $\tilde{N}$ fingerprint codewords $\boldsymbol{X}_{\tilde{j}}$ for $\tilde{j} \notin \{1, \ldots, N\}$.
2. Calculate the correlation scores $\hat{S}^{(\tilde{j})}$.
3. Compute the variance of $\hat{S}^{(\tilde{j})}$, and output it as σ_{SD}^2.

The generated $\tilde{N}$ codewords $\boldsymbol{X}_{\tilde{j}}$ are statistically uncorrelated with the pirated codeword. If $\tilde{N}$ is sufficiently large, a proper variance σ_{SD}^2 can be obtained by the above procedure, and finally, a proper threshold Z_{SD} is derived.

It is noticed that the model of noisy channel is regarded as the binary symmetric channel (BSC) when the HD method is used, which is the same model as the report in [4]. Since an erasure symbol "?" is regarded as "0" in [8], the erasure channel assumed in the analysis is also equal to BSC. Even if AWGN channel is assumed in our paper, the HD method replaces the channel into BSC. On the other hand, the introduction of SD method enables us to utilize the characteristic of AWGN channel. In the next section, we experimentally evaluate the performance of these methods.

5 Experimental Results

The HD and SD methods are applicable for both Tardos's code and Nuida's code when a pirated codeword is distorted by AWGN channel. The performance of such methods are evaluated by experiments under the following conditions. The number of users is $N = 10^4$ and the number of trials for Monte-Carlo simulation is 10^5. The range of bias distribution p_i for Tardos's code is fixed by setting

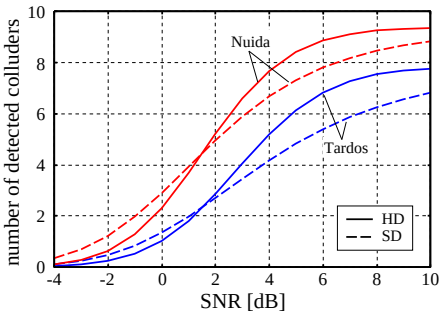


Fig. 4. The number of detected colluders when $\tilde{c} = 10$ and $L = 10000$.

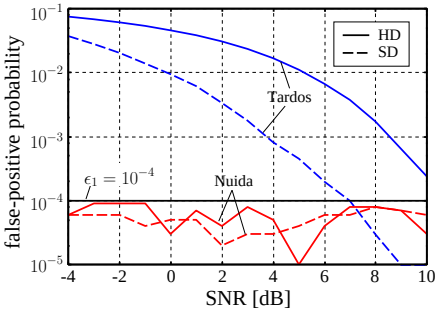


Fig. 5. The false-positive probability when $\tilde{c} = 10$ and $L = 10000$.

$t = 0.000167$ ($c = 20$), and the discrete version of bias distribution of Nuida's code is selected by the case of $c = 7, 8$ shown in Table 1. In the SD method, the number of codewords to calculate σ^2_{SD} is $\tilde{N} = 1000$. The designed false-positive probability is $\epsilon_1 = 10^{-4}$. It is reported in [9] that the performance of Nuida's code is better than that of Tardos's code. So, we mainly compare the HD and SD methods from the behavior of the traceability point of view, and assess the validity of Gaussian assumption of accusation sum for innocent users.

As shown in Fig.3, the attenuation of accusation sum for Nuida's code, which are measured by the parameter A, becomes maximum when the majority attack is performed by colluders for the case that the discrete version of bias distribution is the case of $c = 7, 8$. So, a pirated copy is produced by the majority attack, and it is distorted by transmitting through AWGN channel. By fixing the number of colluders $\tilde{c} = 10$ and the code length $L = 10000$, the number of detected colluders and false-positive probability for HD and SD methods are measured, which results are plotted in Fig.4 and Fig.5, respectively. For both codes, the HD method detects more colluders than the SD method when SNR is more than 2 [dB], and the SD method is suitable only when SNR is less than 2 [dB]. On the other hand, the characteristics of two codes are apparently appeared in the false-positive probability. For Tardos's code, the probability of HD method is higher than that of SD method, and both of the probabilities are drastically increased with the amount of additive noise. Meanwhile for Nuida's code, the probability is almost constant and is below ϵ_1. The results mean that the Gaussian assumption of the distribution of accusation sum is invalid for Tardos's code, while it is valid for Nuida's code under the above conditions. By changing the number $\tilde{c}$, the number of detected colluders and the false-positive probability are measured for two cases that SNR is 1 [dB] and 2 [dB], which results are shown in Fig.6 and Fig.7. Figure 6 indicates that the traceability of HD method is better than that of SD method when SNR is 2 [dB], while the performance of these methods is exchanged when SNR is 1 [dB]. It is remarkable that the false-positive probability is almost constant even if $\tilde{c}$ is changed from 2 to 20. Hence, we can say that the probability is independent on the number $\tilde{c}$ of colluders.

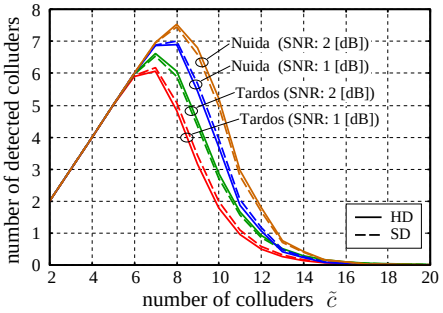


Fig. 6. The number of detected colluders for various number of colluders.

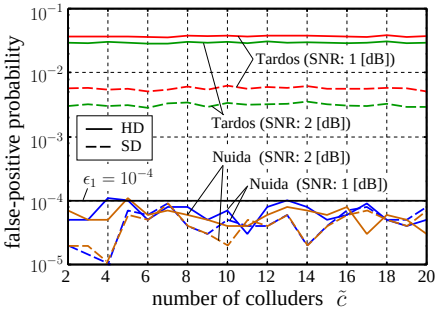


Fig. 7. The false-positive probability for various number of colluders.

Table 4. The comparison of number of detected colluders.

(a) $L = 1000, \tilde{c} = 3$

SNR	HD		SD	
[dB]	Tardos	Nuida	Tardos	Nuida
1	0.76	1.55	0.85	1.65
2	1.16	2.03	1.13	1.97
5	2.20	2.79	1.85	2.58
10	2.61	2.94	2.40	2.87

(b) $L = 2000, \tilde{c} = 5$

SNR	HD		SD	
[dB]	Tardos	Nuida	Tardos	Nuida
1	0.37	1.02	0.42	1.11
2	0.65	1.59	0.62	1.49
5	1.84	3.28	1.33	2.60
10	2.70	4.05	2.20	3.59

(c) $L = 5000, \tilde{c} = 8$

SNR	HD		SD	
[dB]	Tardos	Nuida	Tardos	Nuida
1	0.56	1.53	0.64	1.71
2	0.98	2.40	0.94	2.31
5	2.76	5.03	2.00	4.05
10	4.09	6.31	3.36	5.66

(c) $L = 10000, \tilde{c} = 10$

SNR	HD		SD	
[dB]	Tardos	Nuida	Tardos	Nuida
1	1.80	3.70	2.00	3.94
2	2.86	5.22	2.72	4.97
5	6.13	8.42	4.84	7.33
10	7.75	9.36	6.83	8.84

The comparison of the number of detected colluders for various kinds of code length is shown in Table 4. From the table, it is confirmed that the HD method is better than the SD method to detect as many colluders as possible if SNR is more than 2 [dB], and vice versa. The probabilities of false-positive are also evaluated by changing the parameters $\tilde{c}$ and L, which results are shown in Table 5. The probabilities for Tardos's code are much higher than the given $\epsilon_1 = 10^{-4}$ though the values are decreased with the code length L. Such characteristics are also appeared when the number $\tilde{c}$ of colluders is much higher than c. On the other hand, the probabilities for Nuida's code are almost constant and slightly less than ϵ_1 no matter how many users are colluded to produce a pirated copy and no matter how much noise is added to the codeword.

The traceability and the probability of false-positive are further measured for some typical collusion attacks when $\tilde{c} = 10$ and $L = 10000$. The results are shown in Table 6 and Table 7. As shown in Fig. 3, the attenuation of accusation

Table 5. The comparison of probability of false-positive.

(a) $L = 1000$

SNR [dB]	$\tilde{c}$	HD [$\times 10^{-4}$]		SD [$\times 10^{-4}$]	
		Tardos	Nuida	Tardos	Nuida
1	3	1108.5	0.1	167.3	0.0
	20	1105.7	0.3	169.2	0.0
	100	1089.3	0.4	153.4	0.0
2	3	871.5	0.1	94.2	0.0
	20	881.4	0.4	93.2	0.0
	100	858.6	0.3	85.8	0.0
5	3	300.1	0.1	9.2	0.2
	20	313.3	0.3	10.8	0.2
	100	297.4	0.1	5.6	0.0
10	3	5.5	0.1	0.1	0.2
	20	6.8	0.1	0.0	0.1
	100	4.3	0.1	0.1	0.1

(b) $L = 2000$

SNR [dB]	$\tilde{c}$	HD [$\times 10^{-4}$]		SD [$\times 10^{-4}$]	
		Tardos	Nuida	Tardos	Nuida
1	5	851.3	0.1	133.7	0.5
	20	843.1	0.1	128.8	0.1
	100	817.1	0.3	123.6	0.0
2	5	673.6	0.2	76.9	0.5
	20	674.0	0.7	73.6	0.2
	100	655.1	0.1	70.1	0.0
5	5	242.2	0.6	8.7	0.6
	20	246.9	0.1	7.7	0.2
	100	232.8	0.1	7.0	0.1
10	5	5.1	0.1	0.0	0.2
	20	5.8	0.3	0.0	0.2
	100	5.5	0.0	0.0	0.1

(c) $L = 5000$

SNR [dB]	$\tilde{c}$	HD [$\times 10^{-4}$]		SD [$\times 10^{-4}$]	
		Tardos	Nuida	Tardos	Nuida
1	8	523.1	0.6	82.6	0.3
	20	528.1	0.5	78.5	0.6
	100	532.6	0.4	76.2	0.5
2	8	417.6	0.4	46.6	0.6
	20	424.9	0.4	45.8	0.7
	100	419.0	0.8	45.9	0.3
5	8	151.4	0.6	7.1	0.5
	20	146.8	0.6	4.5	0.8
	100	151.3	0.3	4.9	0.5
10	8	4.0	0.5	0.1	0.6
	20	2.4	0.5	0.3	0.8
	100	2.6	0.6	0.4	0.7

(d) $L = 10000$

SNR [dB]	$\tilde{c}$	HD [$\times 10^{-4}$]		SD [$\times 10^{-4}$]	
		Tardos	Nuida	Tardos	Nuida
1	10	383.8	0.7	62.7	0.5
	20	380.1	0.5	51.1	0.8
	100	365.1	0.6	54.7	0.5
2	10	307.4	0.4	34.3	0.2
	20	222.0	0.3	22.2	0.7
	100	287.9	0.5	28.7	0.6
5	10	111.6	0.8	4.6	0.4
	20	102.8	0.5	3.3	1.0
	100	98.3	0.6	3.6	0.7
10	10	2.4	0.3	0.1	0.6
	20	2.5	0.6	0.3	0.6
	100	2.1	0.4	0.3	0.6

sum for colluders is varied for five types of collusion attack. The number of detected colluders is varied in a similar fashion. Moreover, the HD method is better than the SD method when SNR is more than 2 [dB] for every types of collusion attack. There is a remarkable tendency for Nuida's code in the probability of false-positive against the type of collusion attack. The less the attenuation of accusation sum is, the more the probability of false-positive becomes in this experiment. For example, the parameter "A" of minority attack in Fig.3 becomes minimum among five types of collusion attack, and then the probability of false-positive shwon in Table 7 becomes maximum in most cases. The detailed theoretical analysis for such a characteristic is left for the future work.

Table 6. The number of detected colluders for various kinds of collusion attack when $\tilde{c} = 10$ and $L = 10000$.

SNR		majority		minority		random		all-0		all-1	
[dB]	code	HD	SD	HD	SD	HD	SD	HD	SD	HD	SD
1	Tardos	1.80	2.00	1.79	1.96	1.78	1.92	1.78	1.92	1.79	1.94
	Nuida	3.70	3.94	4.24	4.51	3.97	4.16	3.96	4.14	3.97	4.16
2	Tardos	2.86	2.72	2.81	2.65	2.82	2.61	2.82	2.61	2.83	2.63
	Nuida	5.22	4.97	5.81	5.58	5.53	5.20	5.52	5.19	5.53	5.20
5	Tardos	6.13	4.84	6.04	4.72	6.10	4.70	6.10	4.69	6.09	4.72
	Nuida	8.42	7.33	8.82	7.87	8.66	7.55	8.65	7.54	8.64	7.54
10	Tardos	7.75	6.83	7.71	6.74	7.75	6.72	7.75	6.72	7.75	6.74
	Nuida	9.36	8.84	9.59	9.19	9.50	9.00	9.50	9.00	9.50	8.99

Table 7. The probability of false-positive [$\times 10^{-4}$] for various kinds of collusion attack when $\tilde{c} = 10$ and $L = 10000$.

SNR		majority		minority		random		all-0		all-1	
[dB]	code	HD	SD	HD	SD	HD	SD	HD	SD	HD	SD
1	Tardos	383.8	62.7	377.6	68.8	391.6	58.8	384.8	58.0	381.2	65.2
	Nuida	0.7	0.5	1.1	1.4	0.9	1.0	0.6	0.7	0.7	0.3
2	Tardos	307.4	34.3	315.7	33.8	297.7	29.4	310.4	34.4	302.8	32.4
	Nuida	0.4	0.2	1.8	1.6	1.0	1.0	0.8	0.7	0.6	0.3
5	Tardos	111.6	4.6	111.9	3.9	105.2	2.3	113.3	4.2	110.6	5.5
	Nuida	0.8	0.4	1.1	1.4	0.3	0.9	0.9	1.2	0.2	0.4
10	Tardos	2.4	0.1	3.1	0.5	1.5	0.0	2.8	0.4	2.8	0.3
	Nuida	0.3	0.6	1.0	1.2	0.7	0.7	0.6	0.7	0.4	0.6

6 Conclusion

In this paper, we statistically estimate the distribution of accusation sum under a relaxed marking assumption, and experimentally evaluate the validity of the estimation. In the attack model, a pirated codeword is distorted by additive white Gaussian noise after performing collusion attack. The experimental results confirm that the estimation of the distribution of colluders' accusation sum is valid for Tardos's code when some bits are flipped.

Assuming that each symbol of the pirated codeword is extracted from a pirated copy with analog value, hard and soft decision methods for calculating the accusation sum are proposed. The experimental results indicate that the hard decision method is better than the soft one if SNR is more than 2 [dB], and vice versa. It is also revealed that the probability of false-positive is almost constant for Nuida's code, while it is drastically increased for Tardos's code in proportion to the amount of noise.

Acknowledgment

This research was partially supported by the Ministry of Education, Culture, Sports Science and Technology, Grant-in-Aid for Young Scientists (B) (21760291), 2010.

References

1. Boneh, D., Shaw, J.: Collusion-secure fingerprinting for digital data. IEEE Trans. Inform. Theory 44(5), 1897–1905 (1998)
2. Škorić, B., Katzenbeisser, S., Celik, M.: Symmetric Tardos fingerprinting codes for arbitrary alphabet sizes. Designs, Codes and Cryptography 46(2), 137–166 (2008)
3. Škorić, B., Vladimirova, T.U., Celik, M., Talstra, J.C.: Tardos fingerprinting is better than we thought. IEEE Trans. Inform. Theory 54(8), 3663–3676 (2008)
4. Furon, T., Guyader, A., Cérou, F.: On the design and optimization of Tardos probabilistic fingerprinting codes. In: IH 2008. LNCS, vol. 5284, pp. 341–356. Springer, Heidelberg (2008)
5. Guth, H.J., Pfitzmann, B.: Error- and collusion-secure fingerprinting for digital data. In: IH 1999. LNCS, vol. 1768, pp. 134–145. Springer, Heidelberg (2000)
6. Kuribayashi, M.: Tardos's fingerprinting code over AWGN channel. In: IH 2010. LNCS, Springer, Heidelberg (2010) (in press)
7. Kuribayashi, M., Morii, M.: Systematic generation of Tardos's fingerprinting codes. IEICE Trans. Fundamentals E93-A(2), 508–515 (2009)
8. Nuida, K.: Making collusion-secure codes (more) robust against bit erasure. In: eprint. 2009-549 (2009)
9. Nuida, K., Fujitu, S., Hagiwara, M., Kitagawa, T., Watanabe, H., Ogawa, K., Imai, H.: An improvement of discrete Tardos fingerprinting codes. Designs, Codes and Cryptography 52(3), 339–362 (2010)
10. Nuida, K., Hagiwara, M., Watanabe, H., Imai, H.: Optimization of Tardos's fingerprinting codes in a viewpoint of memory amount. In: IH 2007. LNCS, vol. 4567, pp. 279–293. Springer, Heidelberg (2008)
11. Safavi-Naini, R., Wang, Y.: Collusion secure q-ary fingerprinting for perceptual content. In: DRM 2001. LNCS, vol. 2320, pp. 57–75. Springer, Heidelberg (2002)
12. Tardos, G.: Optimal probabilistic fingerprint codes. J. ACM 55(2), 1–24 (2008)