ANALYSIS OF NONLINEAR COLLUSION ATTACKS ON FINGERPRINTING SYSTEMS FOR COMPRESSED MULTIMEDIA

Avinash L. Varna, Shan He, Ashwin Swaminathan and Min Wu

ECE Department, University of Maryland, College Park, USA

ABSTRACT

In this paper, we analyze the effect of various collusion attacks on fingerprinting systems for compressed multimedia. We evaluate the effectiveness of the collusion attacks in terms of the probability of detection and accuracy in estimating the host signal. Our analysis shows that applying averaging collusion on copies of moderately compressed content gives a highly accurate estimation of the host, and can effectively remove the embedded fingerprints. Averaging is thus the best choice for an attacker as the probability of detection and the distortion introduced are the lowest.

Index Terms— Digital Fingerprinting, Collusion Resistance, Compressed Signals, Anti-Collusion Dither.

1. INTRODUCTION

With the proliferation of the internet and consequent ease of redistribution of multimedia, intellectual property protection has become a challenging problem. Digital fingerprinting is an important tool for traitor tracing and copyright enforcement. A unique fingerprint signal is embedded in every legally distributed copy of the multimedia that can be used to identify the recipient. Upon obtaining an illegal copy, this fingerprint is extracted and used to identify the person(s) responsible for the leak. However, a group of users can mount powerful collusion attacks whereby the attackers try to create a copy of the multimedia that does not contain traces of any of their individual fingerprints. Techniques for systematic construction of fingerprints with collusion resistance have been proposed in [1, 2, 3]. Independent Gaussian based spread spectrum sequences are often used for modulation as they have been shown to have good collusion resistance when fingerprinting uncompressed signals.

In most cases, however, multimedia content is stored and transmitted in compressed form to conserve bandwidth. Consider, for example, a cable TV distribution service which has millions of subscribers. The service provider transmits video in compressed form to conserve precious bandwidth. To deter and identify pirates in the case of illegal redistribution, fingerprints are embedded in the video by the set-top box. However, a group of users may capture and store the output of the set-top box using devices such as Digital Video Recorders (DVR) and then collude to remove the embedded fingerprints. Online music and video stores may also require fingerprints to be embedded in compressed multimedia signals.

To the best of our knowledge, the problem of collusion resistant fingerprinting for compressed multimedia signals has not been addressed in prior work. Recently, we have shown that traditional independent Gaussian fingerprinting does not perform well for compressed signals, and is easily defeated by averaging and median attacks at moderate levels of compression [4]. Our results indicate that by applying *Anti-Collusion Dither*, the collusion resistance of the fingerprinting system can be approximately quadrupled [4].

In this paper, we present a theoretical framework to analyze the effect of various collusion attacks on compressed multimedia fingerprinting systems. We first compute the probability of successfully catching a guilty colluder for various collusion attacks. We then examine collusion in an estimation framework and evaluate the effectiveness of the attack in terms of the Mean Squared Error of the estimate.

2. SYSTEM MODEL

The system model for fingerprinting compressed signals is shown in Fig. 1. The host signal can be represented by the vector $\mathbf{S}$ consisting of M components $[S_1, S_2, \ldots, S_M]$. For simplicity, we consider $\mathbf{S}$ to consist of elements from one frequency band, such as one frequency location in the 8×8 block DCT of images or video. We model the compression of the host image/video as quantization of the DCT coefficients, so that $S_j = m\Delta$, where $m = 0, \pm 1, \pm 2, \ldots$ and Δ is the quantization step size for the particular frequency band. The fingerprint is then embedded into the quantized host signal $\mathbf{S}$.

After the embedding process, since the fingerprinted signal for user α, $\mathbf{X}^{(\alpha)}$, is also stored in compressed form, it is quantized, *i.e.,* $X_j^{(\alpha)} = m\Delta_e$. The quantization step size Δ_e models the compression of the fingerprinted signal and is chosen by the embedder to achieve a tradeoff between the distortion introduced and communication bandwidth. Choice of $\Delta_e > \Delta$ will result in larger distortion and choosing $\Delta_e < \Delta$ will result in greater bandwidth requirements. Hence, a reasonable choice for the embedder is to set $\Delta_e = \Delta$.

Our analysis considers the scenario of additive embedding under the setting $\Delta_e = \Delta$. Denoting the fingerprint for user α as $\mathbf{W_c}^{(\alpha)}$, the fingerprinted signal is obtained as $\mathbf{X}^{(\alpha)} = \mathbf{S} + \mathbf{W}^{(\alpha)}$, where $\mathbf{W}^{(\alpha)} = \text{round}\left(\frac{\mathbf{W_c}^{(\alpha)}}{\Delta}\right) \times \Delta$.

Email contact: {varna, shanhe, ashwins, minwu}@eng.umd.edu.

The energy of the fingerprint is chosen such that embedding does not introduce perceptual distortion:

$$\mathrm{E}[\|\mathbf{S}-\mathbf{X}^{(\alpha)}\|^2]=\mathrm{E}[\|\mathbf{W}^{(\alpha)}\|^2]\leq M\cdot D(\Delta), \quad (1)$$

where $D(\Delta)$ is the maximum allowed squared distortion given the quantization step size Δ.

A group of K users S_c may collude and attempt to create an unauthorized copy $\mathbf{V}$ that does not contain traces of their fingerprints. The colluded signal may be compressed for easy storage and transmission. Let Δ_c be the quantization step size corresponding to the compression of the colluded signal so that $V_j = m\Delta_c$. The attackers' choice of Δ_c is affected by the value of Δ. Choosing $\Delta_c < \Delta$ would not improve the quality of the colluded signal as the host has already been quantized with step Δ. Also, smaller quantization step size may not be effective in removing traces of the fingerprint leading to a higher probability of a colluder being caught. On the other hand, choice of $\Delta_c > \Delta$ would introduce further distortion. Hence, in this paper, we consider the case $\Delta_c = \Delta$ as a reasonable tradeoff for the colluders. The colluded signal $\mathbf{V}$ is obtained from the fingerprinted versions as $\mathbf{V} = g(\{\mathbf{X}^{(k)}\}_{k\in S_c})$, where $g(\cdot)$ is the collusion function.

Colluders may use linear or nonlinear collusion functions $g(\cdot)$ such as those studied in [5] for uncompressed signals. These collusion functions can be extended to compressed systems by adding quantization and are defined as follows:

$$\begin{aligned}
\text{Average}: V_j^{\text{avg}} &= \text{round}\left(\frac{\sum_{k\in S_c} X_j^{(k)}}{K\Delta}\right)\times\Delta,\\
\text{Median}: V_j^{\text{med}} &= \text{round}\left(\frac{\text{median}(\{X_j^{(k)}\}_{k\in S_c})}{\Delta}\right)\times\Delta,\\
\text{Minimum}: V_j^{\text{min}} &= \min(\{X_j^{(k)}\}_{k\in S_c}),\\
\text{Maximum}: V_j^{\text{max}} &= \max(\{X_j^{(k)}\}_{k\in S_c}),\\
\text{Minmax}: V_j^{\text{minmax}} &= \text{round}\left(\frac{V_j^{\text{max}}+V_j^{\text{min}}}{2\Delta}\right)\times\Delta,\\
\text{Modneg}: V_j^{\text{modneg}} &= V_j^{\text{max}}+V_j^{\text{min}}-V_j^{\text{med}}, \quad (2)
\end{aligned}$$

where Modneg represents the modified negative attack. Colluders may apply further processing such as adding noise or filtering which we model as additive white Gaussian noise, $\mathbf{n}$, to obtain the attacked signal $\mathbf{Z} = \mathbf{V} + \mathbf{n}$ as shown in Fig. 1.

Upon obtaining the attacked signal, a correlation based detector is used to identify *at least one* of the guilty users. Interference from the host signal is first removed by subtracting the host $\mathbf{S}$, which is usually available to the detector in fingerprinting applications, from the attacked signal. The detector then subtracts the mean of the extracted fingerprint to obtain the test signal. The user q whose fingerprint has the maximum correlation $T^{(q)}$ with the test signal is declared guilty, *i.e.,* $q = \arg\max_{\alpha=1,2,\ldots,N} T^{(\alpha)}$, where

$$T^{(\alpha)}=\frac{1}{M}\langle h(\mathbf{Z}-\mathbf{S}),\mathbf{W_c}^{(\alpha)}\rangle, \text{ with } h(\mathbf{Y})=\mathbf{Y}-\text{mean}(\mathbf{Y}).$$

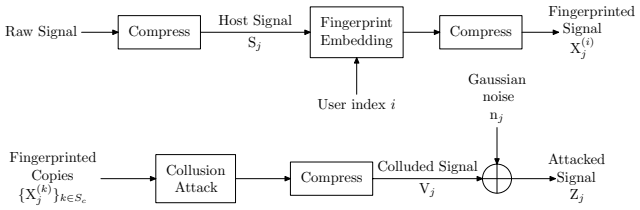


Fig. 1. System Model

3. STATISTICAL ANALYSIS OF COLLUSION

In this section, we characterize the statistical behavior of the detection statistics $T^{(\alpha)}$ under collusion attacks and compute the probability of catching a guilty user.

Theoretical Analysis Framework: The test signal $h(\mathbf{Z}-\mathbf{S})$ can be represented as

$$h(\mathbf{Z}-\mathbf{S})=h(\mathbf{V}+\mathbf{n}-\mathbf{S})=h(g(\{\mathbf{W}^{(k)}\}_{k\in S_c}))+\mathbf{n},$$

since $g(\{X_j^{(k)}\}_{k\in S_c}) = S_j + g(\{W_j^{(k)}\}_{k\in S_c})$ for the attacks considered and $\mathbf{n}$ is zero mean. Denoting $g'(\cdot) = h(g(\cdot))$, we have $T^{(\alpha)} = \frac{1}{M}\sum_{j=1}^{M}(g'(\{W_j^{(k)}\}_{k\in S_c}) + n_j)\times W_{c_j}^{(\alpha)}$. As the $W_{c_j}^{(\alpha)}$ are assumed independent and identically distributed (*i.i.d.*), $T^{(\alpha)}$ follows a Gaussian distribution from the Central Limit Theorem. Further, the mean and variance of the Gaussian distribution are independent of j due to the *i.i.d.* property, and depend only on whether α belongs to the set of colluders S_c or not. Dropping the subscript j, the mean and variance of $T^{(\alpha)}$ for $\alpha \notin S_c$ can be shown to be

$$\begin{aligned}
\text{mean: } \mu_0 &= E[g'(\{W^{(k)}\}_{k\in S_c})+n]E[{W_c}^{(\alpha)}]=0,\\
\text{variance: } \sigma_0^2 &= \frac{1}{M}E[((g'(\{W^{(k)}\}_{k\in S_c})+n){W_c}^{(\alpha)})^2]\\
&= \frac{1}{M}E[(g'(\{W^{(k)}\}_{k\in S_c}))^2+n^2]E[({W_c}^{(\alpha)})^2].
\end{aligned}$$

Here, the equalities follow due to the independence assumption and since the ${W_c}^{(\alpha)}$ are zero mean. Similarly, for $\alpha \in S_c$, we can derive the mean and variances to be

$$\begin{aligned}
\text{mean: } \mu_1 &= E[g'(\{W^{(k)}\}_{k\in S_c}){W_c}^{(\alpha)}],\\
\text{variance: } \sigma_1^2 &= \frac{1}{M}\Big(E[(g'(\{W^{(k)}\}_{k\in S_c})W_c^{(\alpha)})^2]\\
&\quad + E[n^2]E[(W_c^{(\alpha)})^2]\Big)-\mu_1^2.
\end{aligned}$$

The quantities μ_1, σ_0^2, and σ_1^2 can be computed from the joint probability density $f(g(\{W^{(k)}\}_{k\in S_c}), W_c^{(\alpha)})$, $\alpha \in S_c$ and the distribution of $g(\{W^{(k)}\}_{k\in S_c})$. The probability of successfully catching one colluder is then given by the probability that the detection statistic for one of the colluders is larger than the detection statistics of all the innocent users:

$$P_d = \Pr(\max_{k\in S_c} T^{(k)} > \max_{\alpha\notin S_c} T^{(\alpha)}).$$

Analysis of Averaging Collusion: Due to space constraints, we illustrate using the averaging attack as an example in this paper, and our approach can be extended to other attacks in (2). Let $W' = \frac{1}{K}\sum_{k\in S_c} W^{(k)}$ and W^{avg}=round $\left(\frac{W'}{\Delta}\right)\times\Delta$.

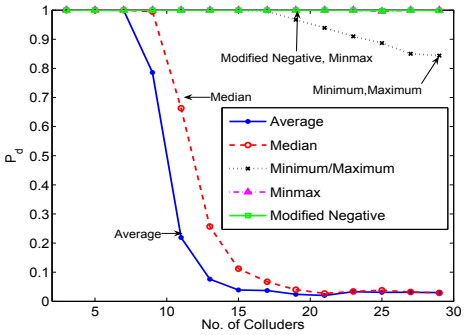


Fig. 2. Probability of catching one colluder using Gaussian based fingerprints at WNR = 0dB, 1024 users, $M = 10^4$, $D(\Delta) = 15$.

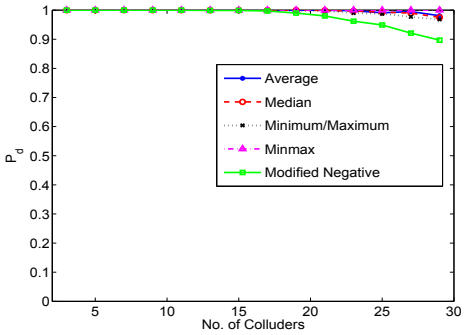


Fig. 3. Probability of catching one colluder for fingerprinting with ACD at WNR = 0dB, 1024 users, $M = 10^4$, $D(\Delta) = 15$.

Then, $\Pr(W^{\text{avg}} = m\Delta) = \Pr(W' \in I_m)$, where $I_m = \left[m\Delta - \frac{\Delta}{2}, m\Delta + \frac{\Delta}{2}\right)$. The characteristic function of W', $M'(t) = E[\exp(itW')]$ is related to the characteristic function of $W^{(\alpha)}$, $M(t)$, as $M'(t) = [M(\frac{t}{K})]^K$, where K is the number of colluders. The probability mass function (pmf) of W' is then given as

$$\Pr\left(W'=m\frac{\Delta}{K}\right)=\frac{1}{2\pi K}\int_{-\pi K}^{\pi K}\exp\left(-\frac{itm\Delta}{K}\right)\left[M\left(\frac{t}{K}\right)\right]^K \mathrm{d}t.$$

The joint distribution $f(W^{\text{avg}} = m\Delta, W_c^{(\alpha)} = w), \alpha \in S_c$ can be written as the product of the conditional distribution $\Pr(W^{\text{avg}} = m\Delta | W^{(\alpha)} = n\Delta)$ and the marginal distribution $f(W_c^{(\alpha)} = w)$. The conditional distribution can then be computed as

$$\Pr(W^{\text{avg}} = m\Delta | W^{(\alpha)} = n\Delta) = \Pr\left(W' \in I_m | W^{(\alpha)} = n\Delta\right)$$
$$= \Pr\left(\frac{1}{K}\sum_{k\in S_c\setminus\{\alpha\}} W^{(k)} \in I_{m,n}\right)$$

where $I_{m,n} = \left[m\Delta - \frac{\Delta}{2} - \frac{n\Delta}{K}, m\Delta + \frac{\Delta}{2} - \frac{n\Delta}{K}\right)$. The conditional distribution can now be computed from the pmf

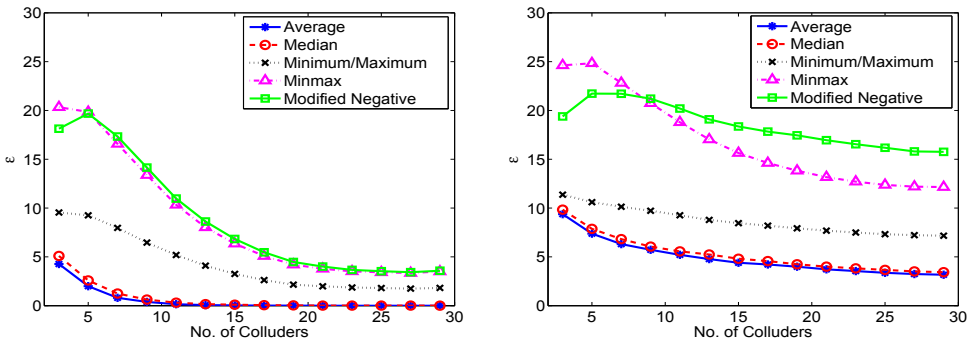
$$\Pr\left(\frac{1}{K}\sum_{k\in S_c\setminus\{\alpha\}} W^{(k)}=\frac{m\Delta}{K}\right)=\frac{1}{2\pi K}\int_{-\pi K}^{\pi K}\exp\left(-\frac{itm\Delta}{K}\right)\left[M\left(\frac{t}{K}\right)\right]^{K-1} \mathrm{d}t.$$

Results for Gaussian Fingerprints: We study the performance of *compressed multimedia* fingerprinting systems under the traditional independent Gaussian based fingerprints. For our experiments, we focus on one frequency band in the 8×8 block DCT domain and the results can be extended to the multi-channel case. Since the host signal, the fingerprinted signal and the colluded signal are all quantized with the same step size Δ, the results obtained are independent of the host signal. To construct the fingerprint sequences, zero mean Gaussian random variables are generated and quantized with step size Δ to obtain $W_j^{(\alpha)}$. The variance of the Gaussian random variables is chosen such that the distortion constraint in (1) is satisfied. We consider a system with $N = 1024$ users and set the fingerprint length $M = 10^4$ which represents the approximate number of embeddable DCT coefficients in a typical 256×256 image. The maximum allowed distortion $D(\Delta)$ is set to 15 such that if every embeddable DCT coefficient were used for embedding with the same allowed distortion, the PSNR would be approximately 36 dB. Here, we present results for $\Delta = 6$ that corresponds to the quantization step size of the AC_{11} band at a JPEG quality factor setting of 75 as it generally provides a good tradeoff between signal quality and bit rate.

Fig. 2 shows the probability of successfully catching one colluder P_d versus the number of users participating in the collusion for various attacks. The power of additive noise is set to be the same as the power of the watermark, *i.e.,* the Watermark-to-Noise Ratio (WNR) is set to 0 dB for each of the attacks. From the figure, we observe that the probability of catching a guilty user is the lowest for averaging attack and the system can resist only 7 colluders with $P_d \approx 1$. The median attack is also very effective at removing traces of the fingerprints. The minimum and maximum attacks are less effective, and the modified negative and the minmax attacks are the least effective attacks.

Performance Analysis under Anti-Collusion Dithering: To improve the collusion resistance of compressed domain fingerprinting systems, we have recently proposed a dithering technique to make the embedded fingerprint appear more continuous [4]. The fingerprinted signal for user α is obtained as $X_j^{(\alpha)} = \text{round}\left(\frac{S_j + d_j + W_{c_j}^{(\alpha)}}{\Delta}\right)$, where d_j is uniformly distributed on $[-\Delta/2, \Delta/2]$. The dither $\mathbf{d}$ is added to make the host appear more continuous and is referred to as Anti-Collusion Dither (ACD) since it has been shown to improve the collusion resistance [4]. The embedded fingerprint is then detected by computing the correlation $\frac{1}{M}\langle h(\mathbf{Z}-\mathbf{S}-\mathbf{d}), \mathbf{W_c}^{(\alpha)}\rangle$.

A similar theoretical analysis can be performed for the attacks in (2) under Anti-Collusion dithering. Fig. 3 shows the probability of catching one colluder versus the number of colluders for fingerprinting using ACD. We observe that the collusion resistance against averaging and median attacks has now approximately quadrupled and approximately 30 colluders can be resisted; and the collusion resistance for the mini-



(a) Without ACD (b) With ACD.

Fig. 4. MSE of various estimators for fingerprinting (a) without ACD and (b) with ACD for $\Delta = 6$.

mum and maximum attacks has also increased. For the modified negative attack, the probability of catching a colluder is the lowest making it the best choice for an attacker purely from the probability of detection point of view. In the next section, we examine collusion from an estimation perspective and evaluate the effectiveness of collusion attacks in terms of the accuracy of estimating the host signal.

4. COLLUSION AS AN ESTIMATION PROBLEM

Collusion attacks to remove traces of the fingerprints can be considered as estimating the host signal, S_j, given the fingerprinted versions $X_j^{(k)} = S_j + W_j^{(k)}$, $k \in S_c$. Let the estimate of the host signal be represented as $\hat{S}_j = G'(\{X_j^{(k)}\}_{k \in S_c})$, where $G'(\cdot)$ is some suitable estimator. The accuracy of the estimate or, equivalently, the effectiveness of the collusion attack can be measured in terms of the Mean Squared Error (MSE), given by $\epsilon = E[(S_j - \hat{S}_j)^2]$. The collusion attacks considered in (2), can be considered as estimators if we set $G'(\cdot) = h(g(\cdot))$ for the collusion attack $g(\cdot)$. These estimators satisfy $G'(\{X_j^{(k)}\}_{k \in S_c}) = S_j + G'(\{W_j^{(k)}\}_{k \in S_c})$. Thus, the MSE simply becomes the variance of the colluded fingerprint. The MSE can thus be obtained from the distribution of the colluded signal as derived in Section 3. In [6], the authors adopt a similar approach to study uncompressed domain fingerprinting but do not provide explicit evaluation of the estimation accuracy for the various attacks.

Fig. 4 shows the MSE of the various estimators as a function of the number of colluders for the experimental setup described in Section 3. From Fig. 4(a), we notice that averaging collusion has the lowest MSE followed by median, minimum, minmax, and modneg attacks for fingerprinting using independent Gaussian based fingerprints and thus averaging gives the best estimate. Fig. 4(b) shows that the MSEs of all the estimators are significantly higher than without dithering. This suggests that these collusion attacks are not as effective in this case as in the case of fingerprinting without dither. In both cases, averaging is the most accurate estimator of the host signal.

The distortion introduced by the collusion attack, measured with respect to the host, is given by the second moment of the colluded fingerprint and is equal to the sum of the MSE and the square of the mean. For averaging, median, minmax and modified negative, the mean of the colluded fingerprint is zero and the distortion introduced is equal to the MSE. For the minimum and maximum attacks, the colluded fingerprint has non-zero mean and the distortion increases with the number of colluders. From Fig. 4(a) and (b), we observe that averaging introduces the lowest distortion. Thus, from the colluders' perspective, averaging is the best attack as it provides accurate estimate of the host signal and also introduces the lowest distortion. The modified negative attack introduces the highest distortion and is hence not preferable.

5. CONCLUSIONS

In this paper, we provide theoretical analysis of various nonlinear collusion attacks on fingerprinting systems for compressed multimedia signals. We evaluate the effectiveness of collusion attacks in terms of the probability of detection P_d and the accuracy of estimating the host signal. We show that averaging collusion gives a highly accurate estimate of the host signal and can effectively remove the embedded fingerprints. Averaging is thus the best choice for an attacker as the probability of detection and the distortion introduced is the lowest.

6. REFERENCES

[1] D. Boneh and J. Shaw, "Collusion-Secure Fingerprinting for Digital Data," *IEEE Trans. Info. Theory*, vol. 44, no. 5, pp. 1897–1905, 1998.

[2] W. Trappe, M. Wu, Z. J. Wang, and K. J. R. Liu, "Anti-collusion Fingerprinting for Multimedia," *IEEE Trans. Signal Proc.*, vol. 51, no. 4, pp. 1069–1087, Apr. 2003.

[3] S. He and M. Wu, "Joint Coding and Embedding Techniques for Multimedia Fingerprinting," *IEEE Trans. Information Forensics and Security*, vol. 1, no. 2, pp. 231–247, Jun. 2006.

[4] A. L. Varna, S. He, A. Swaminathan, M. Wu, H. Lu, and Z. Lu, "Collusion-Resistant Fingerprinting for Compressed Multimedia Signals," in *IEEE ICASSP*, Apr. 2007, vol. 2, pp. 165–168.

[5] H. V. Zhao, M. Wu, Z. J. Wang, and K. J. R. Liu, "Forensic Analysis of Nonlinear Collusion Attacks for Multimedia Fingerprinting," *IEEE Trans. Image Proc.*, vol. 14, no. 5, pp. 646 – 661, May 2005.

[6] N. Kiyavash and P. Moulin, "A Framework for Optimizing Nonlinear Collusion Attacks on Fingerprinting Systems," in *Proc. of the Conf. on Info. Sciences and Sys.*, Mar. 2006.