Anti-Collusion Forensics of Multimedia Fingerprinting Using Orthogonal Modulation

Z. Jane Wang, *Member, IEEE*, Min Wu, *Member, IEEE*, Hong Vicky Zhao, *Member, IEEE*,
Wade Trappe, *Member, IEEE*, and K. J. Ray Liu, *Fellow, IEEE*

***Abstract*—Digital fingerprinting is a method for protecting digital data in which fingerprints that are embedded in multimedia are capable of identifying unauthorized use of digital content. A powerful attack that can be employed to reduce this tracing capability is collusion, where several users combine their copies of the same content to attenuate/remove the original fingerprints. In this paper, we study the collusion resistance of a fingerprinting system employing Gaussian distributed fingerprints and orthogonal modulation. We introduce the maximum detector and the thresholding detector for colluder identification. We then analyze the collusion resistance of a system to the averaging collusion attack for the performance criteria represented by the probability of a false negative and the probability of a false positive. Lower and upper bounds for the maximum number of colluders $K_{\max}$ are derived. We then show that the detectors are robust to different collusion attacks. We further study different sets of performance criteria, and our results indicate that attacks based on a few dozen independent copies can confound such a fingerprinting system. We also propose a likelihood-based approach to estimate the number of colluders. Finally, we demonstrate the performance for detecting colluders through experiments using real images.**

***Index Terms*—Colluder detection, collusion attacks, collusion resistance, digital fingerprinting, spread spectrum embedding.**

I. Introduction

WITH the rapid deployment of multimedia technologies and the substantial growth in the use of the Internet, digital representations of multimedia data have become increasingly popular. Due to the ease with which digital content can be accessed, retrieved and manipulated, there is a demand for methods to protect digital media and facilitate digital rights management. Several methods have emerged in the literature to offer this protection, and may be broadly classified into the categories of cryptographic solutions and steganographic solutions. The protection provided by cryptography technology disappears once data is decrypted, while digital watermarking and steganography offer a supplemental form of protection that extends into cleartext [15] and can be used to identify pirates and discourage the unauthorized use of digital content, such as redistribution [6], [8], [26].

Digital fingerprinting is one possible application of data embedding techniques, whereby some unique information, such as a serial number or a label assigned by the vendor to a given user/purchaser, is embedded in the multimedia content using watermarking techniques. A wide range of requirements have been proposed for watermarking [15], [18]. One requirement of watermarking is that the marked copy should differ imperceptibly from the original. An equally important requirement is that the watermark is able to perform its function in the presence of attacks mounted by adversaries. One powerful class of attacks that adversaries may employ against watermarks and the corresponding fingerprints is *collusion*, whereby a coalition of users combine their different marked copies of the same multimedia content in an attempt to attenuate/remove the trace of any original fingerprint. The fingerprint must, therefore, survive both standard distortions (such as compression, filtering, data conversion, and channel noise) and collusion attacks by users intending to destroy it.

Several methods have been proposed in the literature to embed and hide fingerprints (watermarks) into different media and, depending on the function they are intended to serve, these watermarks can be invisible or visible [2], [3], [8], [9], [13], [29]. Though most watermarking methods are easy to defeat by collusion attacks, the spread spectrum watermarking method proposed in [9], where the watermarks have a component-wise Gaussian distribution and are statistically independent, was argued to be highly resistant to collusion attacks [9], [15]. The basic intuition of this natural strategy is that the randomness inherent in such watermarks makes the probability of accusing an innocent user very unlikely. It was shown that randomness is needed to obtain collusion resistance [31]. Since collusion resistance is the main focus of this paper, we focus our study in this paper on Gaussian watermarks.

The research on collusion-resistant fingerprinting systems can be broadly divided into two main directions. One direction focuses on designing collusion-resistant fingerprint codes. One of the first such fingerprinting schemes was presented by Boneh and Shaw [4], [5] for generic data. They proposed a coding scheme requiring code length as $O(k^4 \log k)$ to capture at least one out of at most k colluders with high probability. This fingerprinting scheme is further improved in [30] by combining a Direct Sequence Spread Spectrum embedding layer with the Boneh-Shaw layer. Recently, a two-layer fingerprinting system

Manuscript received February 6, 2004; revised June 11, 2004. This work was supported in part by the Air Force Research Laboratory under DDET Grant F30602-03-2-0045 and in part by the National Science Foundation under CAREER Award CCR-0133704. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jelena Kovačević.

Z. J. Wang is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4 Canada (e-mail: wangzhen@glue.umd.edu).

M. Wu, H. V. Zhao, and K. J. R. Liu are with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: minwu@eng.umd.edu; hzhao@glue.umd.edu; kjrliu@eng.umd.edu).

W. Trappe is with WINLAB and the Electrical and Computer Engineering Department, Rutgers University, Piscataway, NJ 08854-8060 USA (e-mail: trappe@winlab.rutgers.edu).

Digital Object Identifier 10.1109/TIP.2005.847284

was presented in [7], where the inner code of the spread spectrum was combined with the outer Boneh–Shaw's fingerprint code. It was noted that, for a 2-h video distributed to 10 000 users, with no more than three colluders, that their system could detect at least one colluder correctly with probability 0.9. The Boneh-Shaw scheme has also been used to build more complex schemes to yield better collusion properties [20], [21]. A collusion-secure fingerprint built on top of a robust watermarking algorithm was presented in [11], where the security against two colluders was achieved by using dual Hamming codes. To reduce the computation time and memory requirements of the decoding algorithm, a two-level coding scheme in a c-secure fashion was presented in [31]. In this approach, an inner Cox watermarking code is combined with an outer error-correcting code with large minimum distance to maintain collusion resistance. By taking advantage of the overlap between different codevectors to identify up to k colluding users, the authors in [10] proposed a scheme based on finite geometries, and the authors in [26] developed a fingerprinting scheme based upon anti-collusion codes (ACC) that used the theory of combinatorial designs. It is worth mentioning that most of the above schemes implicitly use the Cox watermarking method in some manner.

The other direction of research is characterized by examining the resistance performance of specific watermarking schemes when considering different types of attacks. The main purpose is to study the relationships between the length of the data to be marked (N); the number of users accommodated in a fingerprinting system (n); and the number of colluding users (K). We are aware of only a few works that focus on analyzing the collusion resistance of digital watermarks [12], [15], [24], [25]. Focusing on a simple linear collusion attack that consists of adding noise to the average of K independent copies, the authors concluded in [15] that $O(\sqrt{N/\log n})$ independently marked copies are sufficient for an attack to defeat the underlying system with nonnegligible probability when the watermarks are Gaussian. It is further shown in [15] to be optimal: No other watermarking scheme can offer better collusion resistance. These results are also supported by [12]. By studying several types of attacks, including some nonlinear types, Stone suggested that the most powerful attack may succeed in defeating uniformly distributed watermarks if as few as one to two dozen independent copies are available [24]. Although these works study collusion, they do not provide a precise analysis of the collusion resistance of watermarks when employed with different possible detection schemes. This paper addresses this issue.

This paper focuses on issues related to collusion, and presents results quantifying the collusion resistance of a fingerprinting system by evaluating how many colluders are allowed before the collusion undermines the tracing capability of the system. We employ a few basic assumptions in this paper.

- We apply the spread spectrum watermarking method and consider independent Gaussian watermarks. That is, *iid* normally distributed random values are used as fingerprints, since this watermarking scheme has been shown to be highly robust to a variety of attacks [9] and it allows for theoretical statistical analysis. Further, we assume that the fingerprints use orthogonal modulation, or at least the correlation between different fingerprints can be ignored. This feature helps to decrease the probability of a false positive and leads to simple detection schemes employing correlation.
- A nonblind detection scenario is assumed. There are two common detection scenarios for data embedding, namely, *blind* and *nonblind* detection. They are characterized by the absence or presence of the host signal at the detector. In the blind scenario, the host signal is not available to the detector and serves as an additional noise source that hinders the detectability;[1] therefore, the blind scenario can be regarded as a nonblind scenario with very low watermark-to-noise ratio (WNR). Analysis provided later in this paper shows that the resistance capability of a system is proportional to the square root of the WNR value, and, thus, two or three independent copies may defeat the spread spectrum watermark under the blind scenario.
- The additive distortion is modeled as *iid* Gaussian noise.

Though we focus on collusion attacks in this paper, it is worth mentioning that there are other noncollusion attacks, such as geometric distortions, which may be effective. Recent research showed that even very small geometric distortions, such as rotation, scale, shift, and cropping, can prevent the detection of a watermark [17]. One may argue that combined attacks, such as a collusion attack combined with a noncollusion attack (e.g., geometric distortion), can defeat a fingerprinting system more effectively. However, since we focus in this paper on the spread spectrum additive embedding technique, we benefit from the resilience of spread spectrum embedding to noncollusion attacks. For example, let us consider geometric distortion as a noncollusion attack. With appropriately chosen features and additional alignment procedures, it has been observed that a small set of salient points of the host signal available to the detector will suffice for the embedded watermark to survive moderate geometric distortions [14]. Further, since we assume the nonblind scenario, meaning the host signal is available in detection, the watermarked copy can often be registered to the original and the geometric distortion thereby inverted. Undoing geometrical distortions may inevitably leave some unrecovered residue. However, the alignment noise/error has been shown to be very small in general [19] and, thus, can be approximated by additive noise. Therefore, we concentrate on collusion attacks, though a real system should also include other decision components to combat with other types of distortions.

The paper is organized as follows. We begin with the description of the collusion problem of interest in Section II. We introduce two detection schemes, namely, the maximum detector and the thresholding detector, and also examine the theoretical collusion resistance of orthogonal fingerprinting when considering the average collusion attack. We represent the system performance by the probability of a false positive and the proba-

[1] Note that we can apply some preprocessing technique to reduce the noise effect of the host signal; however, it will also distort the fingerprint to some extent and make the detection of fingerprint vulnerable. We also note that there are other types of watermarking techniques that do not require the original unmarked copy at the detector. We are going to investigate their appropriateness for fingerprinting and their collusion resistance performance in our future work.

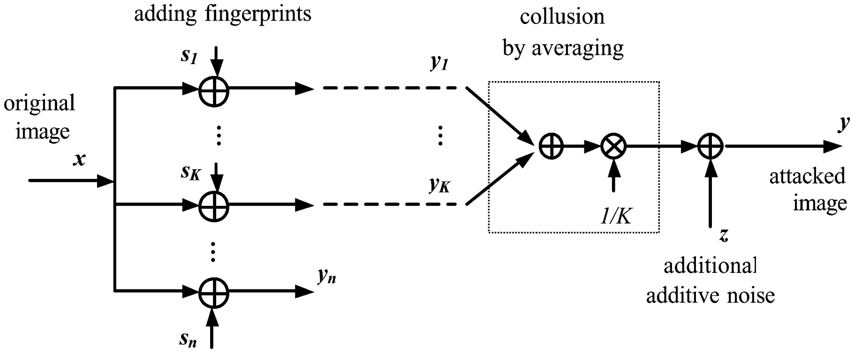


Fig. 1. Model for collusion by averaging.

bility of a false negative. Since different detection goals arise under different application scenarios, two more sets of performance criteria are examined in Section III. In Section IV, we further study other types of collusion. Since the knowledge of the number of colluders is normally not available in practice, we propose in Section V a maximum-likelihood (ML) approach to estimate the number of colluders K, and carry out simulations. Experiments using real images are demonstrated in Section VI. Finally, we present conclusions in Section VII.

II. Detection Approaches

In this paper, we use independent normally distributed random values as fingerprints. We first introduce the average collusion attack for these fingerprints. There are different types of collusion attacks in the literature [24]. We start with average collusion due to its popularity, its simple form and its feasibility for analysis. We will extend our study to other attacks later in Section IV.

Additive embedding is a widely used watermarking scheme, where a watermark signal $\mathbf{s}_j$ is added to a host signal $\mathbf{x}$. As shown in Fig. 1, the content owner has a family of watermarks, denoted by $\{\mathbf{s}_j\}$, which are used to mark copies of the content and facilitate colluder tracing. For the jth user, the owner computes the marked version of the content $\mathbf{y}_j$ by adding the watermark $\mathbf{s}_j$ to the host signal, $\mathbf{y}_j = \mathbf{x} + \mathbf{s}_j$. In addition to attacks operating on a single copy, collusion attacks are possible when several buyers/users having different marked copies of the same host signal come together and combine several copies to generate a new composite copy $\mathbf{y}$ such that the traces of each "original" fingerprint in the new version is removed or attenuated. We illustrate the average collusion attack in Fig. 1, a similar model was used in [12], [25], [26]. Based on this average attack model, the observed content $\mathbf{y}$ after collusion is

$$\mathbf{y} = \frac{1}{K}\sum_{j\in S_c}\mathbf{y}_j + \mathbf{d} = \frac{1}{K}\sum_{j\in S_c}\mathbf{s}_j + \mathbf{x} + \mathbf{d} \tag{1}$$

where all vectors have dimension N, K is the number of colluders, where $K \geq 1$ since each single copy is marked, and S_c indicates the colluder subset of size K, where $S_c \subseteq [1,\ldots,n]$ and n is the total number of users. The fingerprints $\mathbf{s}_j$ are assumed to be orthogonal to each other, have equal energy, and normally distributed. Due to the orthogonality of $\mathbf{s}_j$, we have $n \leq N$. We also assume the distortion $\mathbf{d}$ is an N-dimensional vector following an *iid* $\mathcal{N}\left(0,\sigma_d^2\right)$ distribution, and define the WNR as $\mathrm{WNR} = 10\log_{10}(\|\ \mathbf{s}\ \|^2\ /\ \|\ \mathbf{d}\ \|^2)$. In this paper, we will be concerned with detecting colluders, and will study the collusion resistance performance of the fingerprinting system. Our detection scheme seeks to identify the colluders based on the observations $\mathbf{y}$. Since we assume a nonblind detection scenario in this paper, the host signal $\mathbf{x}$ is always subtracted from $\mathbf{y}$. Because of the orthogonality of the basis $\{\mathbf{s}_j\}$, when performing detection it suffices to consider the correlator vector $\mathbf{T}_N$, where the jth component is given by

$$T_N(j) = \frac{(\mathbf{y}-\mathbf{x})^T\mathbf{s}_j}{\sqrt{\|\ \mathbf{s}_j\ \|^2}} \tag{2}$$

for $j = 1,\ldots,n$. It is straightforward to show that

$$p(T_N(j)|H_K,S_c) = \begin{cases} \mathcal{N}\left(\frac{\|\mathbf{s}\|}{K},\sigma_d^2\right), & \text{if } j\in S_c \\ \mathcal{N}\left(0,\sigma_d^2\right), & \text{otherwise} \end{cases} \tag{3}$$

where H_K represents the hypothesis that there are K colluders, $\|\mathbf{s}\| = \|\mathbf{s}_j\|$ for all j due to the equal energy assumption, and each component $T_N(j)$ is independent of each other due to the orthogonality of $\mathbf{s}_j$.

In this section, we are interested in the theoretical collusion resistance of such fingerprinting systems. When studying the efficiency of a detection algorithm in collusion applications, appropriate criterion should be used to address the need of each specific application. The probability of a false negative and the probability of a false positive are popular criteria explored by researchers [12], [15]. From the detector's (owner's) point of view, a detection approach fails if either the detector fails to identify any of the colluders (a false negative) or the detector falsely indicates that an innocent user is a colluder (a false positive). Therefore, it is desirable to find an efficient detector that minimizes the probability of a false negative (P_{fn}), with a given probability of a false positive (P_{fp}). In general, P_{fp} should be exceptionally low, since a false positive may have severe consequences, such as serving as false testimony in a court of law. Though we consider the criteria P_{fp} and P_{fn} in this section, it is worth mentioning that other performance criteria also deserve consideration. We will present the study of two additional sets of criteria in Section III.

Next, we will introduce two other detection approaches and study their collusion resistance under the average attack.

A. Maximum Detector

We have observed in collusion detection that the more colluders a detector aims to catch, the higher probability a false positive occurs. A detector designed to catch only one colluder should be capable of providing a smaller P_{fp}. A maximum detector is

$$T_{\max} = \max_{j=1}^{n} T_N(j) \tag{4}$$

where $T_N(j)$ is defined as in (2), can be applied to catch one colluder with high confidence. This maximum detector should be compared to a threshold h chosen to yield the desired P_{fp}. Thus, we have the following test [see (5), shown at the bottom of the page] where $\hat{j}$ indicates the index of the accused user, and $\hat{j} = \emptyset$ means that no accusation is made. In practice, it is possible that more than one j maximizes $T_N(j)$ simultaneously. In this case, the test randomly accuses one of these users. The following analysis reveals that the threshold h is determined by parameters including the length of the host signal N, the total number of users n, the number of colluders K and the WNR.

1) Performance Analysis: To analyze the detection performance of the maximum detector, we assume that the number of colluders K is known, and without loss of generality, we set the subset $S_c = [1, 2, \ldots, K]$, indicating that the first K users are colluders. We now have

$$\begin{aligned} P_{fp} &= P_r\{T_{\max} > h, \hat{j} \notin S_c\} = P_r\{T_1 < T_2, T_2 \geq h\} \\ &= P_r\{T_2 \geq h\} P_r\{T_1 < h\} \\ &\quad + \int_h^{\infty} P_r\{T_2 \geq T_1\} p(T_1) dT_1 \end{aligned} \tag{6}$$

with the statistics $T_1 = \max_{j=1}^{K} T_N(j)$ and $T_2 = \max_{j=K+1}^{n} T_N(j)$. Here, n is the total number of users, and $p(T_1)$ is the *pdf* of the random variable T_1. Clearly, T_1 is independent of T_2 due to the independency of $T_N(j)$. We also define the detection probability P_d as

$$\begin{aligned} P_d &= 1 - P_{fn} = P_r\{T_{\max} > h, \hat{j} \in S_c\} \\ &= P_r\{T_1 > T_2, T_1 \geq h\} \\ &= P_r\{T_1 \geq h\} P_r\{T_2 < h\} \\ &\quad + \int_h^{\infty} P_r\{T_1 \geq T_2\} p(T_2) dT_2. \end{aligned} \tag{7}$$

Since $p(T_N(j)|H_K, S_c)$ is given as in (3), we have

$$\begin{aligned} P_r(T_1 \leq t) &= \left(1 - Q\left(\frac{t - \frac{\|\mathbf{s}\|}{K}}{\sigma_d}\right)\right)^K \\ P_r(T_2 \leq t) &= \left(1 - Q\left(\frac{t}{\sigma_d}\right)\right)^{n-K} \end{aligned} \tag{8}$$

where the Q-function is defined as $Q(t) = \int_t^{\infty} (1/\sqrt{2\pi}) \exp(-x^2/2) dx$. The pdf $p(T_1)$ and $p(T_2)$ can be derived correspondingly from the above cdf. Therefore, for a given small value of ϵ, we can numerically solve for h to yield $P_{fp} = \epsilon$ for different K, n and WNR, and then numerically compute the corresponding P_d.

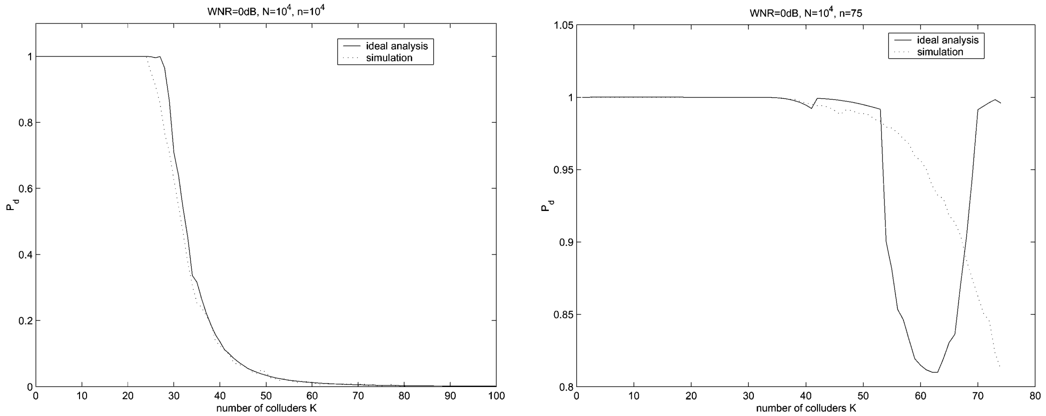
One important efficiency measure of a fingerprint detector is the maximum number of colluders that can be tolerated by a fingerprinting system with a total of n different N-point fingerprints. Specifically, with a given P_{fp}, we explore how many differently marked copies of the host signal are required for an averaging attack to generate a colluded copy from which no colluder's fingerprint can be detected with a high probability. A reasonably high P_d and a reasonably low P_{fp} are necessary to maintain the system's resistance to collusion.

We illustrate the resistance performance using an example, where WNR $= 0$ dB and the vector length is $N = 10^4$. Since 0-dB WNR corresponds to a *nonblind* scenario, the distortion $\mathbf{d}$ only consists of the additional additive noise. The variance σ_d^2 is assumed known and set to 1 for simplicity. In this example, the system requirements are expressed as $P_d \geq 0.8$ and $P_{fp} \leq 10^{-3}$. The symbol $K_{\max}$ represents the maximum number of colluders the fingerprinting system can successfully resist. In the examples shown in Fig. 2(a) and (b), when the number of users n is as high as 10^4, the fingerprinting system can resist up to 29 colluders; while, when n is set as a small number 75, the fingerprinting system can resist up to 75 colluders. It is also noted in Fig. 2(a) that, if an attacker can collect 50 independent copies, the chance that the system can trace any original copy is only 4%. We note in Fig. 2(b) that, as K increases, P_d first decreases slowly, then decreases quickly over the range $50 < K < 65$, and then increases. This behavior is determined by the expressions of P_{fp} and P_d in (6) and (7). We will give a similar explanation in Section II-B, where a similar behavior is observed for the thresholding detector and the reason is more obvious. To have an overall understanding of the collusion resistance of this scheme, in Fig. 3, we also plot the maximum resistible number of colluders $K_{\max}$ as a function of the total number of users n, under $N = 10^4$ and WNR $= 0$ dB. It is noted that the system can resist up to n colluders when the total number of users (fingerprints) n is less than 75. However, as a system accommodates more than 75 users, the collusion resistance of the system starts to decrease. For a system accommodating more than one thousand users, the maximum number of colluders that the system can handle is 30.

B. Thresholding Detector

Although the goal of this section is to identify at least one of the colluders, from the content owner's point of view, it is beneficial to catch as many colluders as possible as long as we sat-

$$T_{\max} = \max_{j=1}^{n} T_N(j), \text{ and } \begin{cases} \hat{j} = \arg\max_{j=1}^{n} T_N(j), & \text{if } T_{\max} \geq h \\ \hat{j} = \emptyset, & \text{if } T_{\max} < h \end{cases} \tag{5}$$



(a) (b)

Fig. 2. Probability of detection P_d as a function of the number of colluders K when apply the maximum detector, with $\mathrm{WNR} = 0$ dB, $N = 10^4$, and $P_{fp} \leq 10^{-3}$. In (a), the number of users n is 10^4. In (b), $n = 75$.

isfy the P_{fp} requirement. We employ the traditional correlator $T_N(j)$ and compare it to a threshold h, and finally report that the jth fingerprint is present if $T_N(j)$ exceeds h. This simple approach is described as

$$\hat{\mathbf{j}} = \arg_{j=1,\ldots,n} \{T_N(j) \geq h\} \tag{9}$$

where the set $\hat{\mathbf{j}}$ indicates the indices of colluders, and an empty set means that no user is accused. Similar to the case of the maximum detector, the threshold h here is determined by such parameters as the document length N, the total number of users n, the number of colluders K, and the WNR.

1) Performance Analysis: The threshold h in test (9) is chosen to yield $P_{fp} = \epsilon$, where ϵ is a desired small value. Same as in Section II-A.I, to analyze the theoretical performance, we assume that the number of colluders K is known. And without loss of generality, we set the subset $S_c = [1, 2, \ldots, K]$. We now have

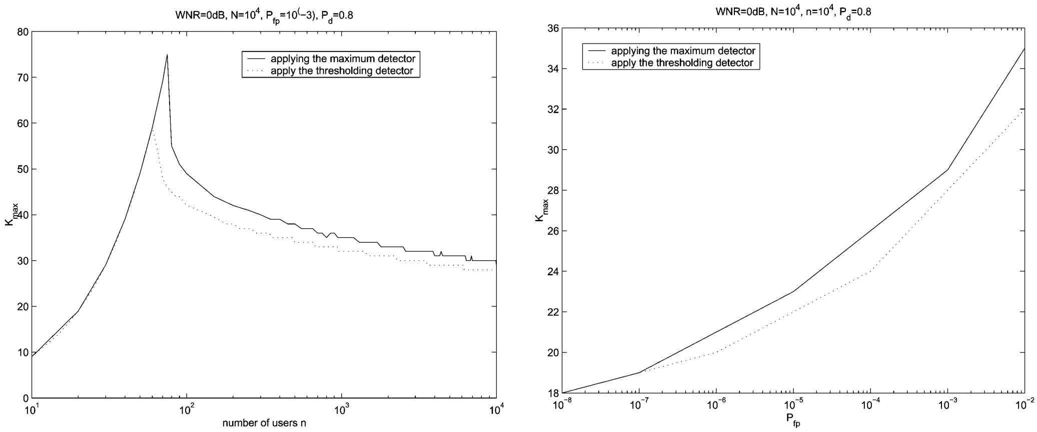
$$\begin{aligned} P_{fp} &= P_r\{\hat{\mathbf{j}} \cap \bar{S}_c \neq \emptyset\} = P_r\{T_2 \geq h\} \\ &= 1 - \left(1 - Q\left(\frac{h}{\sigma_d}\right)\right)^{n-K} \\ P_d &= 1 - P_{fn} = P_r\{\hat{\mathbf{j}} \cap S_c \neq \emptyset\} \\ &= P_r\{T_1 \geq h\} = 1 - \left(1 - Q\left(\frac{h - \frac{\|\mathbf{s}\|}{K}}{\sigma_d}\right)\right)^{K} \end{aligned} \tag{10}$$

where $\bar{S}_c$ is the complement set of S_c, $T_1 = \max_{j \in S_c} T_N(j)$, $T_2 = \max_{j \in \bar{S}_c} T_N(j)$, and n the total number of users. Due to the independency among $T_N(j)$, T_1 is independent of T_2. The cdfs of the order statistics T_1 and T_2 are given as in (8). Therefore, according to (10), we can numerically calculate h to yield $P_{fp} = \epsilon$ with given K, n, and WNR, and then compute the corresponding P_d. Similar to the analysis in Section II-A, our goal is to study the resistance of the fingerprinting system to averaging collusion when employing the thresholding detector (9). A sufficiently high P_d and a sufficiently low P_{fp} are required to make a fingerprinting system resistant to collusion attacks.

We illustrate the resistance performance using an example, where $\mathrm{WNR} = 0$ dB, and $N = 10^4$. The variance σ_d^2 is set to 1 like before. The system requirements are defined as $P_d \geq 0.8$ and $P_{fp} \leq 10^{-3}$. As shown in Fig. 9(a) and (b), when the number of users n is on the order of 10^4, the fingerprinting system can resist to up to 28 colluders; when n is set as a small number 75, the system can resist to up to 46 colluders. Similar to Section II-A, Fig. 9 shows that P_d first decreases slowly, then decreases quickly, and then increases, as K increases. This behavior can be intuitively explained by the expressions of P_{fp} and P_d in (10). The sudden quick decrease is due to the exponential nature of the Q function; when K is reasonably small, the term $\| \mathbf{s} \| /K$ in $Q(\cdot)$ function is the dominating factor in deciding P_d, this term decreases as K increases and, therefore, results in a decreasing P_d. On the other hand, when K is sufficiently large, the exponent term K is the dominating factor in deciding P_d, and, thus, P_d increases as K increases. To have an overall understanding of the collusion resistance of the orthogonal fingerprinting scheme, we plot the maximum resistible number of colluders $K_{\max}$ as a function of the total number of users n in Fig. 3, where $N = 10^4$ and $\mathrm{WNR} = 0$ dB. It is noted that the system can resist to up to n colluders when the total number of users n is less than 60. However, for a system accommodating more than 60 users, its collusion resistance starts to decrease. For a system accommodating more than one thousand users, the number $K_{\max}$ is 28, meaning that the system requirements for the fingerprinting system is no longer met if the number of colluders is larger than 28.

We also compare the collusion resistance of the orthogonal fingerprinting scheme when applying both test (5) and test (9). Fig. 3(a) shows $K_{\max}$ as a function of the total number of users n, with $N = 10^4$ and $\mathrm{WNR} = 0$ dB. In Fig. 3(b), we present $K_{\max}$ as a function of P_{fp} for a specific system with 10^4 users. We note that the maximum detector provides better performance than the thresholding detector. The intuitive explanation for this observation is that the maximum detector is designed to catch only one colluder. The overall difference is small, however, especially when the total number of users is large.

2) Lower and Upper Bounds of $K_{\max}$: Next, we provide analytic bounds on the maximum number of colluders $K_{\max}$ for



(a) (b)

Fig. 3. Collusion resistance of the orthogonal fingerprinting system to the averaging attack. Here, $\mathrm{WNR} = 0$ dB, $N = 10^4$ and $\epsilon = 10^{-3}$, and $\beta = 0.8$.

an orthogonal fingerprinting system employing the thresholding detector. Since the above analysis is based on numerical computation, it does not provide an explicit understanding of the relationships between $K_{\max}$ and other system parameters, such as the sample length N, the WNR, the total number of users n, and the performance requirements of P_{fp} and P_d. To get more insight into the collusion resistance of the thresholding detector, it is useful to study the analytic lower and upper bounds of $K_{\max}$. We begin by introducing two important lemmas.

Lemma 1: Define the Gaussian tail integral as $Q(t) = \int_t^\infty (1/\sqrt{2\pi}) \exp(-x^2/2)dx$. $Q(t)$ is nonnegative for all t and monotonously decreases as t increases for $t > 0$. We have $Q(t) = 1 - Q(-t)$ by definition. This tail integral $Q(t)$ can be lower and upper bounded by

$$\begin{aligned} Q_a(t) = &\frac{1}{\sqrt{2\pi}t}\left(1 - \frac{1}{t^2}\right)\exp\left(-\frac{t^2}{2}\right) \\ &< Q(t) < \frac{1}{\sqrt{2\pi}t}\exp\left(-\frac{t^2}{2}\right) = Q_b(t) \end{aligned} \tag{11}$$

for $t > 0$, respectively. Please refer to [23] for a detailed proof.

Lemma 2: Let n be a positive integer. For $0 < x < 1/n$, $(1-x)^n$ can be bounded by

$$1 - nx < (1-x)^n < 1 - nx + \frac{n(n-1)}{2}x^2. \tag{12}$$

Proof: We first expand $(1-x)^n$ as

$$(1-x)^n = \sum_{i=0}^{n} \binom{n}{i} (-x)^{n-i} \tag{13}$$

and utilizing the fact that $\binom{n}{i} x^i > \binom{n}{i+1} x^{i+1}$ for $0 < x < 1/n$, we derive the above inequality.

Setting $\sigma_d^2 = 1$ for convenience, note that now $||s|| = \sqrt{\eta N}$ with the WNR $\eta = ||s||^2/||d||^2$. Recalling the expressions for P_{fp} and P_d in (10), we restate the system requirements as

$$\begin{aligned} P_{fp} &= 1 - (1 - Q(h))^{n-K} \le \epsilon, \\ P_d &= 1 - \left(1 - Q\left(h - \frac{\sqrt{\eta N}}{K}\right)\right)^K \ge \beta \end{aligned} \tag{14}$$

where ϵ is a small number and β is close to 1. For instance, a typical setting is $\epsilon = 10^{-3}$ and $\beta = 0.8$. A key step in determining $K_{\max}$ is to figure out the appropriate threshold h in (14). Though the explicit solution of h is hard to obtain, we can take advantage of the lower and upper bound of the threshold h by linking it to the lower and upper bounds of $K_{\max}$. The following inequalities are observed by applying Lemma 1:

$$\begin{aligned} P_{fp} &< (n-K)Q_b(h), \\ P_{fp} &> (n-K)Q_a(h) - \frac{(n-K)(n-K-1)}{2}Q_b^2(h). \end{aligned} \tag{15}$$

The assumption that ϵ is small implies that the choice of h can meet the condition $Q(h) \ll 1/n$. Based on this observation and inequality (15), we obtain a lower and upper bound of h as

$$\begin{aligned} h &< h_H = \sqrt{\log\left(\frac{n^2}{2\pi\epsilon^2 \log\left(\frac{0.5n^2}{\pi}\right)}\right)} \\ h &> h_L = \max\{h_{L1}, h_{L2}\} \end{aligned} \tag{16}$$

where the bounds are defined as $h_{L1} = \sqrt{\log(0.5n^2/\pi)}$ and

$$h_{L2} = \sqrt{2\log\left(\frac{2h_{L1}^2 - h_{L1} - 2}{2\sqrt{2\pi}\epsilon h_H h_{L1}^2}\right)}.$$

The detailed derivation of (15) and (16) is given in Appendix A.

So far, we have obtained a lower and upper bound for the threshold h with a few reasonable assumptions. We now proceed to show that a lower and upper bound of the maximum number of colluders $K_{\max}$ can be obtained by using the bounds of h in (16) to evaluate the probability of accurate detection, P_d, in (14). The basic idea is to find a lower bound K_L of $K_{\max}$ such that the resulting pair (K_L, h_H) simultaneously satisfies the conditions that the corresponding P_d is larger than but close to the requirement β, and P_{fp} is smaller than but close to the requirement ϵ. Similarly, an upper bound K_H is chosen such that the pair (K_H, h_L) results in a P_d, which is smaller than but close to the requirement β, and a P_{fp}, which is larger than but close to the requirement ϵ. The smaller the difference between the two sets of results, the tighter the bounds represented by K_L and K_H. A detailed derivation, given in Appendix B, leads to the following collusion resistance:

$$\begin{aligned} &K_{\max} \geq \min\{n, K_L\}, \text{with} \\ &K_L = \frac{\sqrt{\eta N}}{h_H} = \sqrt{\frac{\eta N}{\log\left(\frac{n^2}{2\pi\epsilon^2 \log\left(\frac{0.5n^2}{\pi}\right)}\right)}} \\ &K_{\max} \leq \min\{n, K_H\}, \text{with} \\ &K_H = \frac{\sqrt{\eta N}}{h_L - Q^{-1}(1 - \sqrt[\tilde{k}]{1-\beta})} \end{aligned} \tag{17}$$

where $Q^{-1}(\cdot)$ represents the inverse Q function, and $\tilde{K}$ serves as an upper bound of K_H

$$\tilde{K} = \frac{\sqrt{\eta N}}{h_L - Q^{-1}(1 - \sqrt[n]{1-\beta})}. \tag{18}$$

It is worth mentioning that a tighter lower and upper bound of $K_{\max}$ can be obtained by solving the one-dimensional problem $P_d = \beta$ when h_H and h_L are considered, respectively. However, this would require more computation and no explicit expressions of K_H and K_L as in (17) would be available due to the complex nature of P_d. In addition, though the bounds (17) are derived for the thresholding detector, they are also applicable to the maximum detector since, as shown in Fig. 3, the overall performance difference between these two schemes is small and can be neglected.

We illustrate the resistance analysis in Fig. 4, where $\sigma_d^2 = 1$, WNR = 0 dB, and $N = 10^4$. Setting the requirements $P_{fp} \leq 10^{-3}$ and $P_d \geq 0.8$, we plot the lower and upper bound of $K_{\max}$ versus the number of users n, along with the numerical result $K_{\max}$. It is noted that the lower and upper bounds are within a factor of 2 of the true value of $K_{\max}$. Given the lower and upper bounds, some interesting observations are noted from this example. From the attacker point of view, if an attacker can only collect up to 20 copies, he/she can never succeed in removing all trace of the fingerprints; however, an attacker is guaranteed success if 80 independent copies are available. From the owner (detector) point of view, if the owner has a means to ensure that a potential attacker has no way to obtain 20 or more independent copies, the fingerprinting system is essentially collusion resistant. Further, in order to maximize the worst case of P_d, the owner should limit the number of independent distributions. For instance, if the number of independent copies is less than 60, the system is also collusion resistant.

III. Extensions to Other Performance Criteria

In Section II, we were concerned with capturing one true colluder with high confidence. The motivating application was to provide digital evidence in the court of law. However, different goals arise under different situations, and there are other possible performance measures for colluder identification. These measures place a varying amount of emphasis on capturing colluders and placing innocents under suspicion. In fact, colluder identification might only be one component of the evidence gathering process. Since the final decision will depend upon many types of evidence, there might be different roles that collusion detection will play in protecting content value. For example, it might be desirable to use colluder identification to identify a set of suspects and then perform other types of surveillance on these suspects to gather the remaining evidence. This suggests that researchers should consider a wider spectrum of performance measures.

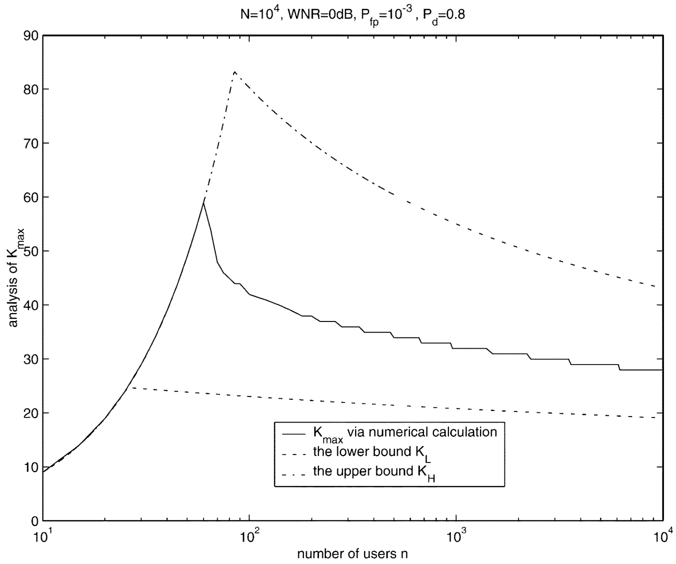


Fig. 4. Lower and upper bound of $K_{\max}$ as a function of the number of users n when apply the thresholding detector in (9). Here, $\mathrm{WNR} = 0$ dB, $N = 10^4$, $\epsilon = 10^{-3}$, and $\beta = 0.8$.

We consider two additional sets of performance criteria in this section and study the thresholding detector under the average attack. The analysis of the thresholding detector is easier than the maximum detector. However, the results are similar, and for that reason we will omit the analysis of the maximum detector.

Case 1: Capture More : This set of performance criteria consists of the expected fraction of colluders that are successfully captured, denoted by r_c, and the expected fraction of innocent users that are falsely placed under suspicion, denoted by r_i. Here, the major concern is to catch as many colluders as possible, though potentially at a cost of accusing more innocents. The balance between capturing colluders and placing innocents under suspicion is represented by these two expected fractions. We define

$$\gamma_j = \begin{cases} 1, & \text{if } j\text{th user is accused} \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Considering the thresholding detector and the average attack, we have

$$\begin{aligned} r_c &= \frac{E\left(\sum_{j\in S_c} \gamma_j\right)}{K} = \frac{\sum_{j\in S_c} P_r\{\gamma_j = 1\}}{K} \\ &= \frac{KQ\left(\frac{h - \frac{\|s\|}{K}}{\sigma_d}\right)}{K} = Q\left(\frac{h - \frac{\|s\|}{K}}{\sigma_d}\right) \\ r_i &= \frac{E\left(\sum_{j\notin S_c} \gamma_j\right)}{n-K} = Q\left(\frac{h}{\sigma_d}\right). \end{aligned} \tag{20}$$

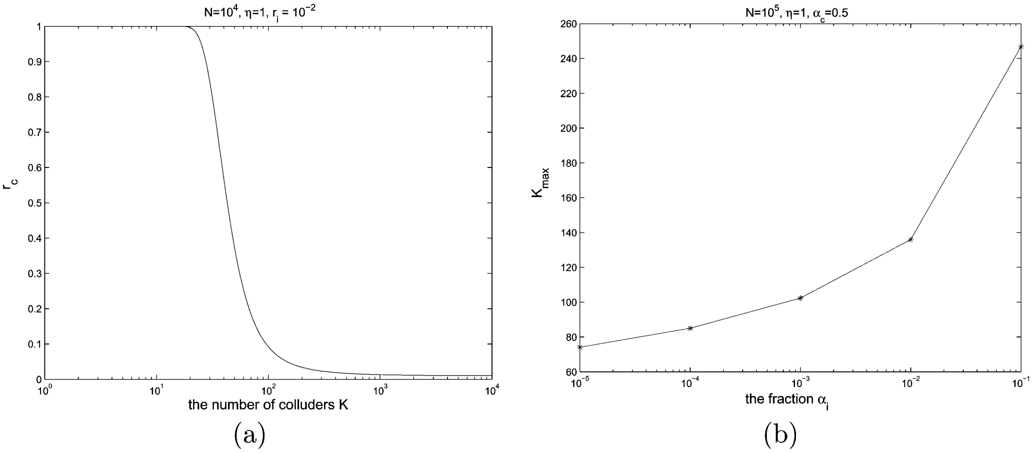


Fig. 5. Resistance performance under the criteria r_c and r_i, when applying the thresholding detector in (9). In (a), we plot the expected fraction r_c versus the number of colluders K, with $N = 10^4$, the WNR $\eta = 1$, and the expected fraction $r_i = 0.01$. $K_{\max}$ under different requirements of α_i is illustrated in (b).

The above observation indicates that studying the behavior of the fractions r_c and r_i is equivalent to studying the probability of correctly detecting a specific colluder and the probability of falsely accusing a specific innocent user. Based on this pair $\{r_c, r_i\}$, now the system requirements are

$$r_i = Q\left(\frac{h}{\sigma_d}\right) \leq \alpha_i \text{ and } r_c = Q\left(\frac{h - \frac{\|s\|}{K}}{\sigma_d}\right) \geq \alpha_c \qquad (21)$$

meaning a reasonably high r_c and a reasonably low r_i are required to keep the fingerprinting system safe to attacks.

We now study the resistance performance of orthogonal fingerprints under requirements (21). In our analysis, $\|s\| = \sqrt{\eta N}\sigma_d$ with η being the WNR and N being the vector length. Based on (20) and (21), we can obtain the threshold h and the maximum number of colluders $K_{\max}$ as

$$\begin{aligned} h &= Q^{-1}(\alpha_i)\sigma_d \\ K_{\max} &= \frac{\sqrt{\eta N}}{Q^{-1}(\alpha_i) - Q^{-1}(\alpha_c)}. \end{aligned} \qquad (22)$$

It is interesting to note that the threshold h is a constant value determined by α_i, and $K_{\max}$ is not affected by the total number of users n. The collusion resistance $K_{\max}$ is proportional to the square root of the vector length N and the WNR η. To illustrate this, in Fig. 5(a), we observe that a system with the requirements $r_i \leq 0.01$ and $r_c \geq 0.5$, which involves $N = 10^4$ fingerprints, can withstand 43 colluders. If we allow a larger fraction of innocents to be placed under suspicion, then the system can resist more colluders, as depicted in Fig. 5(b). Here, let us look at an example represented by the point with coordinate values $\{10^{-2}, 136\}$ in Fig. 5(b). In this example, since $N = 10^5$, $\alpha_i = 10^{-2}$ and $\alpha_c = 0.5$, the decision maker will have to identify 68 suspected colluders (calculated as $136 \times \alpha_c$) from a pool of people containing up to one thousand innocent users (calculated as $N \times \alpha_i$).

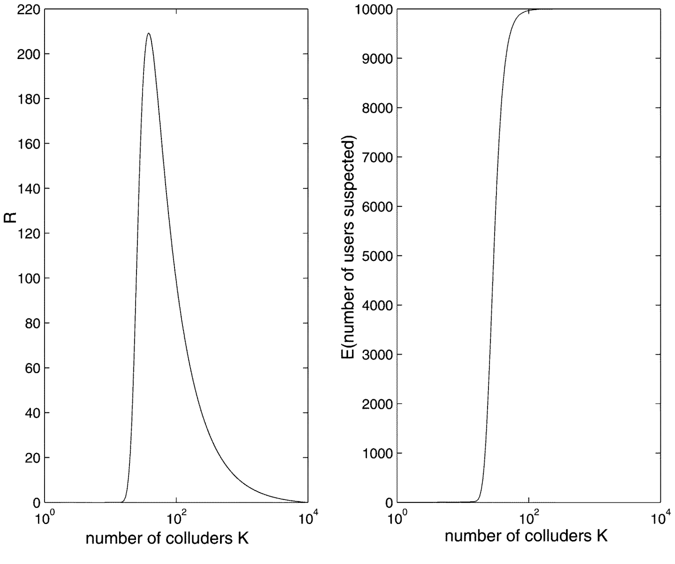
Case 2: Capture All: This set of performance criteria consists of the efficiency rate R, which describes the expected number of innocents accused per colluder, and the probability of capturing all K colluders, which we denote by P_d. Here, the goal is to capture all colluders with a high probability. The tradeoff between capturing colluders and placing innocents under suspicion is managed through the adjustment of the efficiency rate R. More specifically, when considering the thresholding detector and the average attack, we have

$$\begin{aligned} R &= \frac{E\left(\sum\limits_{j \notin S_c} \gamma_j\right)}{E\left(\sum\limits_{j \in S_c} \gamma_j\right)} = \frac{(n-K)Q\left(\frac{h}{\sigma_d}\right)}{KQ\left(\frac{h - \frac{\|s\|}{K}}{\sigma_d}\right)} \\ P_d &= P_r\{S_c \subseteq \hat{\mathbf{j}}\} = P_r\left\{\min_{j \in S_c} T_N(j) \geq h\right\} \\ &= Q\left(\frac{h - \frac{\|s\|}{K}}{\sigma_d}\right)^K. \end{aligned} \qquad (23)$$

Based on this pair $\{R, P_d\}$, the system requirements are expressed as

$$R \leq \alpha; \quad P_d \geq \beta. \qquad (24)$$

We first illustrate the resistance performance of the fingerprinting system under these requirements by examples, where $N = 10^4$ and $\eta = 1$. We set $\sigma_d^2 = 1$ for simplicity and recall that $\|s\| = \sqrt{\eta N}$. First, for a system accommodating as many as 10^4 users and requiring $P_d = 0.99$, we study the behavior of R when the number of colluders K increases as shown in Fig. 6. For each choice of K, the threshold h is chosen to yield $P_d = 0.99$ and then the corresponding R is calculated. It is clear that almost all users will be placed under suspicion if more than 100 users come together and perform the collusion. The decision of placing all users under suspicion certainly provides no useful clues to the identity of the colluders. If the rate R is set as 0.01, the system can resist to up to 13 colluders. To obtain an overall understanding of the collusion resistance of the system, we further study the performance of the system when different amounts of users are involved, as illustrated in Fig. 7 by requiring $R \leq 0.01$ and $P_d \geq 0.99$. It is clear that the system can afford up to n colluders if the number of total users n is smaller than 21. The resistance performance degrades when more than 21 users are accommodated. In situations where the system is required to distribute more than one thousand independently marked copies, an attacker having as few as



(a) (b)

Fig. 6. Behaviors of the efficiency rate R and the expected number of users suspected as K increases. Here, $N = 10^4$, $\eta = 1$, $n = 10^4$, and $P_d = 0.99$. We plot the rate R versus the number of colluders K in (a). The expected number of users suspected is plotted against K in (b).

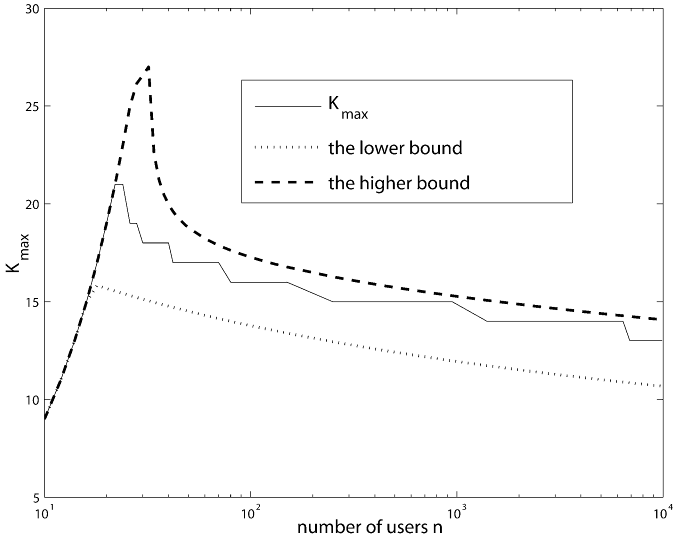


Fig. 7. Resistance performance of the orthogonal fingerprinting system under the criteria R and P_d. Here, $N = 10^4$, $\eta = 1$, $\alpha = 0.01$, and $P_d = 0.99$. The lower and upper bound is also plotted.

15 independent copies has the capability to break down the system. Similar to Section II-B, we provide a lower and upper bound of $K_{\max}$ under this set of criteria. Assume $\sigma_d^2 = 1$ for convenience. A derivation similar to that in Section II-B leads to the following bounds:

$$
\begin{aligned}
K_{\max} &\geq \min\{n, K_L\}, \text{with} \\
K_L &= \frac{\sqrt{\eta N}}{Q^{-1}\left(\frac{2\alpha}{n}\right) - Q^{-1}(\sqrt[n]{\beta})} \\
K_{\max} &\leq \min\{n, K_H\}, \text{with} \\
K_H &= \frac{\sqrt{\eta N}}{Q^{-1}\left(\frac{\alpha\tilde{K}}{(n-\tilde{K})}\right) - Q^{-1}(\sqrt[K_L]{\beta})}
\end{aligned} \tag{25}
$$

with $\tilde{K}$ being

$$\tilde{K} = \frac{\sqrt{\eta N}}{-Q^{-1}(\sqrt[K_L]{\beta})}.$$

The details of this derivation are omitted due to the limitation of space and due to its similarity to the derivations in Section II-B. An example is given in Fig. 7.

The analysis in this section reveals that the maximum number of colluders allowed by a Gaussian fingerprinting system is on the same order, under three different sets of criteria. Basically, a few dozen colluders could break down the orthogonal Gaussian fingerprinting system by generating a new composite copy such that the identification of the original fingerprints would unlikely be successful.

IV. Extensions to Other Types of Attacks

So far, we have studied the collusion resistance of the Gaussian fingerprinting system for the average attack. When an attacker has access to multiple independently watermarked copies of the same host signal, attacks other than the averaging attack are also possible. In this section, we consider several nonlinear attacks suggested by Stone in [24], and we evaluate the resistance of the maximum detector and the thresholding detector. We have further considered a few other collusion attacks (see [32]), such as randomly copying and pasting parts of content from individual copies, or randomly choosing any value between the minimum and the maximum values. Our study has shown that this additional set of attacks can be approximated as the collusion attacks discussed in this paper followed by additive noise. Thus, the attacks studied here represent a wide range of attacks.

1) Attacks based on the median operation.

   Under this attack, the attacker obtains K independently marked copies of the same host signal, and computes the composite observation $\mathbf{y}$ such that the ith component of y is

$$
\begin{aligned}
y(i) &= \operatorname*{median}_{j \in S_c}\{x(i) + s_j(i)\} + d(i) \\
&= \operatorname*{median}_{j \in S_c}\{s_j(i)\} + x(i) + d(i)
\end{aligned} \tag{26}
$$

   for $i = 1, 2, \ldots, N$, where the subset S_c indicates the colluder index and $\mathrm{median}(\cdot)$ represents the median operation. This attack is named the *median* attack, as indicated by its definition.

2) Attacks based on the minimum operations.

   Under the *minimum* attack, the attacker creates a copy $\mathbf{y}$ whose ith component is the minimum of the ith components of the observed copies plus a noise term. Similarly, we can define the *maximum* attack and the so called *randomized negative* attack (also referred as Kilian's attack) [32]. Since our statistical analysis reveals that these three attacks share the same property in terms of collusion resistance, we study only the minimum attack here to save space.

3) Attacks based on the average of the minimum and maximum operations.

Under the *minmax* attack, the attacker creates a copy $\mathbf{y}$ whose ith component is

$$\begin{aligned} y(i) = & \frac{\left(\min_{j\in S_c}\{x(i)+s_j(i)\} + \max_{j\in S_c}\{x(i)+s_j(i)\}\right)}{2} \\ & + d(i) \\ = & \frac{\left(\min_{j\in S_c}\{s_j(i)\} + \max_{j\in S_c}\{s_j(i)\}\right)}{2} \\ & + x(i) + d(i) \end{aligned} \tag{27}$$

for $i = 1, 2, \ldots, N$, where $\min(\cdot)$ and $\max(\cdot)$ are the minimum and maximum operations, respectively.

4) Attacks based on the median, minimum, and maximum operations.

Since Kilian's attack produces unacceptable distortion, Stone suggested a modified version of Kilian's attack such that

$$\begin{aligned} y(i) = & \left(\min_{j\in S_c}\{s_j(i)\} + \max_{j\in S_c}\{s_j(i)\}\right. \\ & \left. - \operatorname*{median}_{j\in S_c}\{s_j(i)\}\right) + x(i) + d(i) \end{aligned} \tag{28}$$

for $i = 1, 2, \ldots, N$. It is noted that Stone's attack produces less distortion than Kilian's.

For a specific attack, we should examine the overall distortion introduced to the host signal, and the efficiency comparison of different attacks should be carried out under the assumption that the distortion level created by different attacks is approximately equal. The purpose of this section is to show that the nonlinear attacks described above can be regarded as attacks by averaging in the sense that they yield pretty similar performance when employing the maximum and the thresholding detectors, as long as the overall MSE (mean-square-error) introduced to the host signal by different attacks is the same. More specifically, our goal is to demonstrate that the attacks

$$\begin{aligned} \mathbf{y}_g &= g(\mathbf{y}_j, j \in S_c) + \mathbf{d}_g \\ \text{and } \mathbf{y}_{\text{mean}} &= \frac{1}{K}\sum_{j\in S_c}\mathbf{y}_j + \mathbf{d}_{\text{mean}} \end{aligned} \tag{29}$$

provide close collusion resistance performance as long as

$$E\{\parallel \mathbf{y}_g - \mathbf{x} \parallel^2\} = E\{\parallel \mathbf{y}_{\text{mean}} - \mathbf{x} \parallel^2\} \triangleq \xi_0 \tag{30}$$

where $g(\cdot)$ represents the attack operation, and the additive noise $\mathbf{d}_g$ are $\mathcal{N}\left(0, \sigma_{d,g}^2\right)$ distributed where the variance $\sigma_{d,g}^2$ is determined by the power ξ_0. Note that the power of the composite observation indicates the level of MSE introduced to the host signal. Therefore, given the MSE level allowed by the system, we want to show that the underlying attack model does not matter from the detector point of view. In other words, we want to demonstrate that the thresholding detector is robust to different attacks. A similar argument can be made for the maximum detector.

First, we illustrate an example based on 10^4 simulation runs in Fig. 8, where $N = 10^4$, $n = 100$, and thresholds are chosen to yield $P_{fp} = 10^{-2}$. Three types of attacks are studied: the average, minmax, and minimum attacks. The fingerprints $\mathbf{s}_j$ are taken as $\mathcal{N}\left(0, \sigma_s^2\right)$ distributed random values with $\sigma_s^2 = 1$, and the additive noise added to the minimum attack $\mathbf{d}_{\min}$ follows $\mathcal{N}(0, 1)$ distribution. Thus, the additive noises introduced by the average attack and the minmax attack are correspondingly generated to provide the same MSE level as by the minimum attack. From Fig. 8, it is noted that the performance curves are close to each other, with the minimum attack marginally superior to the other two attacks from the detector's point of view (i.e., worse from the attacker's point of view).

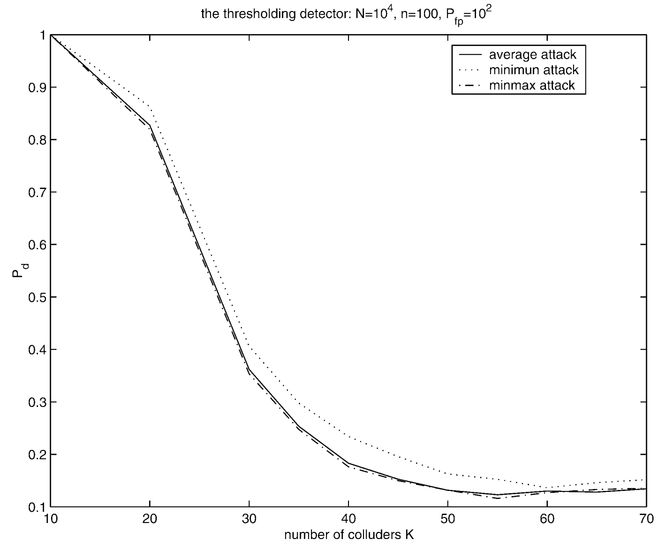


Fig. 8. Probability of detection as a number of colluders K under different attacks, when applying the thresholding detector and the same MSE level is introduced. Here, $N = 10^4$, $n = 100$ and $P_{fp} = 10^{-2}$.

The observation noted in the above example is encouraging. We intuitively explain the reasons by referring to the statistical analysis in [32]. We need to analyze the statistical behavior of the test $T_N(j)$ under different collusion attacks. Due to the *iid* Gaussian assumption of the fingerprint components and since N is generally in the order of 10^4 for 256 × 256 images, by applying the central limit theorem (CLT), we propose to approximate the distribution of $T_N(j)$ with a Gaussian distribution. Our results show that the correlator $T_N(j)$ still yields zero mean for $j \notin S_c$, and the mean of $T_N(j)$, for $j \in S_c$, is the same under different attacks. By calculating the corresponding mean and variance, we have that the correlator $T_N(j)$ is approximately distributed as (31), shown at the bottom of the next page, in which, for $\forall\, l \in S_c$

$$\begin{aligned} \sigma_{g0}^2 &= E\{g(s_j(i), j \in S_c)^2\}, \\ \sigma_{g1}^2 &= \frac{Var\{g(s_j(i), j \in S_c)s_l(i)\}}{\sigma_s^2}. \end{aligned}$$

Under each attack, $T_N(j)$, for $j \notin S_c$, is independent of each other. It is clear that for a given K, the behavior of $T_N(j)$, for $j \notin S_c$, is fully characterized by the overall power ξ_0; therefore, the threshold and P_{fp} are not affected by the type of attack. The derivation of the mean and the variance σ_{g1}^2 under the minimum attack is given in Appendix C. The analysis of other attacks can be similarly derived. We refer the interested readers to [32] for more details. It is worth mentioning that there is no closed form expression for the variance σ_{g1}^2 available under most attacks,

TABLE I
CORRESPONDING σ_{g0}^2, σ_{g1}^2, $\sigma_{d,g}^2$, AND $var\{T_N(j)\}$ UNDER DIFFERENT ATTACKS, WHERE $K = 15, \sigma_s^2 = 1$

variance \ attack	Average	minimum	median	minmax	Stone
σ_{g0}^2	0.0667	3.3144	0.1017	0.1581	0.5757
$\sigma_{d,g}^2$	3.2477	0	3.2127	3.1563	2.7387
σ_{g1}^2	0.0711	3.7519	0.1108	0.1747	0.6480
$\text{var}\{T_N(j)\}$, $j \in S_c$	3.3188	3.7519	3.3235	3.3310	3.3867

due to the existence of $Q(\cdot)$ terms in the distributions. Therefore, in our implementation, we numerically evaluate the integrals by employing the recursive adaptive Simpson quadrature method. As an example, suppose $\sigma_s^2 = 1$ and no noise is added to the minimum attack, we report the results in Table I, where we can see that the variance of $T_N(j)$, for $j \in S_c$, is comparable under different attacks and, thus, results in comparable P_d under different attacks. Our results also reveal that the difference of this variance among different attacks gets smaller as the number of colluders K increases.

The above fact that different attacks provide comparable performance from the detector's point of view suggests, for the same MSE distortion, the average attack is the most efficient from the attacker point of view. This is because, from the detector point of view, there exists better detection schemes than detectors based on the correlators $T_N(j)$ for attacks other than the average attack. For this reason, we have concentrated only on the average attack in this paper, and we only address the collusion resistance of a fingerprinting system under the average attack.

In addition, to maintain an acceptable quality of the image, a basic requirement is that the collusion attack is unlikely to generate noticeable distortion. Therefore, three types of attacks, namely the minimum attack, the maximum attack, and Kilian's attack, should be excluded from consideration, since our analysis indicates that the energy of the composite watermark generated by these attacks is greater than that of the original watermark (e.g., large σ_{g0}^2 and σ_{g1}^2), and grows with the number of colluders K. This unfortunate feature of these attacks suggests that these attacks are likely to produce noticeable distortion which increases with K.

V. Practical Estimator for K

In the above analysis, we have assumed the number of colluders K is known. However, knowledge of K is normally not available in a practical collusion scenario. Therefore, in real colluder-identification situations, we need to estimate the number of colluders K. To start, we present the problem in a multiple-hypotheses-testing framework, where the different hypotheses lead to different $\mathbf{y}$ as

$$H_K : \mathbf{y} = \frac{1}{K}\sum_{j \in S_c} \mathbf{s}_j + \mathbf{x} + \mathbf{d} \tag{32}$$

for $1 \le K \le n$. An optimal way to estimate K can be based on the Bayesian classifier

$$\hat{K} = \arg\max_K p(H_K|\mathbf{y}) = \arg\max_K \sum_{S_c} p(H_K, S_c|\mathbf{y}) \tag{33}$$

where $p(\cdot)$ represents likelihood functions. However, it is immediately noted that the probability summation over all subsets S_c with size K is infeasible in practice. We address this issue by obtain the maximum-likelihood (ML) estimates of K and S_c jointly based on the observations $\mathbf{y}$

$$(\hat{K}, \hat{S}_c) = \arg\max_{K,S_c} p(\mathbf{y}|H_K, S_c). \tag{34}$$

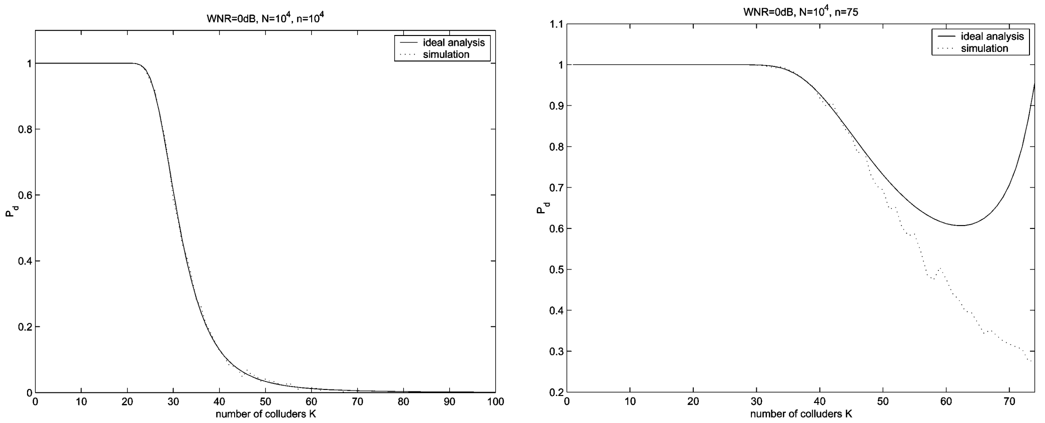
Because of the orthogonality of the basis $\{\mathbf{s}_j\}$, it suffices to consider the correlator vector $\mathbf{T}_N$, defined in (2). Now the estimator equals to

$$\begin{aligned}(\hat{K}, \hat{S}_c) &= \arg\max_{K,S_c} p(\mathbf{y}|H_K, S_c) = \arg\max_{K,S_c} p(\mathbf{T}_N|H_K, S_c)\\ \text{thus}\hat{K} &= \arg\max_K \left\{ \frac{2 \parallel \mathbf{s} \parallel}{K} \sum_{j=1}^{K} T_N^{(j)} - \frac{\parallel \mathbf{s} \parallel^2}{K} \right\}\\ \hat{S}_c &= \text{the set of indices of } \hat{K} \text{ largest } T_N(j)'s \end{aligned} \tag{35}$$

where $T_N^{(j)}$ are the order statistics of the sample $\mathbf{T}_N$ such that $T_N^{(1)} \ge T_N^{(2)} \ge \cdots \ge T_N^{(n)}$. We refer the interested reader to Appendix D for the detailed derivation of (35).

Based on $(\hat{K}, \hat{S}_c)$ obtained from the above approach, a fingerprinting system may accuse all users indicated by $\hat{S}_c$ as colluders. However, the above approach is aimed at jointly finding the ML estimates of K and the colluder set S_c. Although it might be interesting to study P_{fn} and P_{fp} of this approach in (35), this approach is not designed to allow one to adjust the

$$T_N(j)|K, S_c, g \sim \begin{cases} \mathcal{N}\left(0, \sigma_{g0}^2 + \sigma_{d,g}^2\right) = \mathcal{N}\left(0, \frac{E\{\|\mathbf{y}_g - \mathbf{x}\|^2\}}{N}\right), & \text{if } j \notin S_c \\ \mathcal{N}\left(\frac{\sqrt{N\sigma_s^2}}{K}, \sigma_{g1}^2 + \sigma_{d,g}^2\right), & \text{if } j \in S_c \end{cases} \tag{31}$$



(a) (b)

Fig. 9. Probability of detection P_d as a function of the number of colluders K when applying the thresholding detector, with $\mathrm{WNR} = 0$ dB, $N = 10^4$ and $P_{fp} \leq 10^{-3}$. In (a), the number of users n is 10^4. In (b), $n = 75$.

tradeoff between P_{fn} and P_{fp}, which is a desirable functionality for colluder tracing applications. Therefore, the above approach is not appropriate to meet our specific detection goal. Thus, we only use it to estimate the total number of colluders K, and examine the effects of the estimated K next.

A. Simulations for the Maximum Detector

Since K is unknown in a practical collusion scenario, we need to estimate K first before setting a suitable threshold h for the detection process. With given N, WNR and n, the colluder identification algorithm using the maximum detector becomes as follows.

1) Estimate the number of colluders K via (35).
2) Determine the threshold h correspondingly to yield a desired P_{fp}, according to (6). It is clear that the threshold h is only a function of $\hat{K}$ when N, WNR, and n are given.
3) Apply the maximum test statistic described in (5) and return the index $\hat{j}$.

In Fig. 2(a) and (b) for $N = 10^4$ and $\mathrm{WNR} = 0$ dB, the simulation results are compared to the ideal performance analysis shown in Section II where K is assumed known. Unlike the ideal case that K is assumed known, when K is estimated based on simulated observations, the resulting P_d always decreases with the increasing of K. Good match is observed over the nonincreasing part of the ideal case, i.e., when K is small. Mismatch is noted over the increasing part of the ideal case, i.e., when K is close to n, since K is underestimated in this situation due to the increasing overlap between the two Gaussian distributions $\mathcal{N}\left(0, \sigma_d^2\right)$ and $\mathcal{N}\left(\|\ \mathbf{s}\ \| / K, \sigma_d^2\right)$ as K increases. However, using an estimate of K will not alter $K_{\max}$ significantly from the results when we use the exact value of K since only the nonincreasing part (also the matched part) of the ideal case in the P_d versus K curve is evaluated to decide $K_{\max}$, the maximum number of colluders a system can afford.

B. Simulations for the Thresholding Detector

As in Section V-A, we need to first estimate K before setting a threshold h for the detection process. We introduce the following implementation

1) Estimate the number of colluders K via (35).
2) Determine the threshold h correspondingly to yield a desired P_{fp}, according to (10). It is clear that the threshold h is only a function of $\hat{K}$ when N, WNR, and n are given.
3) Apply the thresholding test statistic described in (9) and return the set $\hat{j}$.

We compare the simulation results with the ideal performance analysis in Fig. 9(a) and (b). We can see that, with the estimated number of colluders, the observation when employing the thresholding detection is similar to that of the maximum detection.

VI. Experiments With Images

In order to demonstrate the performance of a Gaussian fingerprinting system using orthogonal modulation on real images for identifying colluders, we apply an additive spread spectrum watermarking scheme similar to that in [22], where the original host image is divided into 8 × 8 blocks, and the watermark (fingerprint) is perceptually weighted and then embedded into the block DCT coefficients. The detection of the fingerprint is performed with the knowledge of the host image. To generally represent the performance, the 256 × 256 Lena and Baboon images with quite different natures are used as the host images for fingerprinting. The fingerprinted images have an average PSNR of 44.6 dB for Lena and 41.9 dB for Baboon. We compare the performance of the thresholding detector under average, minimum and minmax collusion attacks, respectively. We show in Fig. 12 the original host images, the colluded images, and the difference images. With $K = 50$, an average PSNR of 37.3 dB for Lena and 34.6 dB for Baboon are resulted after collusion attacks.

Denoting $\mathbf{s}_j$ as the Gaussian fingerprint, we note that the ith component of the jth fingerprint is actually embedded as

$$s_j(i)^t = \alpha(i) s_j(i) \tag{36}$$

where the superscript t means *actual*, with $\{\alpha(i)\}$ being the just-noticeable-difference (JND) parameters from human visual model to achieve the imperceptibility of the embedded fingerprint. Therefore, the composite embedded fingerprint $\mathbf{y}^t$ after attack is represented as

$$\begin{aligned} y(i)^t &= g(y_j(i)^t, j \in S_c) + d(i) \\ &= \alpha(i) g(s_j(i), j \in S_c) + x(i) + d(i) \end{aligned} \tag{37}$$

where $g(\cdot)$ is the collusion function discussed in Section IV, and the noise d is independently distributed. Under nonblind detection, $\alpha(i)$ are known in the detector side, and, thus, the effects of real images can be partially compensated by computing

$$w(i) = \frac{(y(i)^t - x(i))}{\alpha(i)} = g(s_j(i), j \in S_c) + \frac{d(i)}{\alpha(i)} \tag{38}$$

for $i = 1, \ldots, N$. In practice, the variance of $d(i)$ is often proportional to the value of $\alpha(i)^2$, for example, in image compression attack. As such, $d(i)/\alpha(i)$ can be approximately modeled as *iid* $\mathcal{N}\left(0, \sigma_d^2\right)$ distributed. Therefore, the test statistic $T_N(j)$ used in the thresholding detector is now defined as

$$T_N(j) = \frac{\mathbf{w}^T \mathbf{s}_j}{\sqrt{\| \mathbf{s}_j \|^2}} \tag{39}$$

for $j = 1, \ldots, n$.

We present the results in Figs. 10 and 11 based on 10^5 simulations using real images. The number of total users n is set to 100. We ignored the round-off error introduced by DCT/IDCT transform in simulations. The fingerprint $\mathbf{s}_j$ is assumed to be $\mathcal{N}(0, \mathbf{I})$. To make a fair comparison between the experimental and analytical results, we first demonstrate the results for Lena image under the average attack in Fig. 10, where the additive noise is with variance $\sigma_d^2 = 1$ and $P_{fp} = 10^{-3}$ is required. We note that the result from the real image is comparable to that based on analysis in Section II-B.

We further compare the performance of the thresholding detector under different types of attacks in Fig. 11. The threshold for each K is chosen to satisfy $P_{fp} = 10^{-2}$ by simulation runs. σ_d^2 is set as 1 for the minimum attack case, and the corresponding σ_d^2 is properly adjusted for the cases of the average and minmax attacks to ensure the attacked images have the same MSE level (thus, PSNR) with respect to the host image. The level of MSE is larger as K increases. It is noted that the detection performance is better under the minimum attack than under the other two attacks. This suggests that the minimum attack is less efficient from the attacker point of view, an observation that matches with the analysis. It is also noticed that a better performance is observed in the Baboon example than in Lena. One possible explanation for this is that, in Lena, the efficient length of the fingerprint is $N = 13\,691$, while a longer $N = 19\,497$ is allowed in Baboon. Different characteristics such as the amount of edges and smooth regions of these two images also contribute to the difference in the performance. It is worth mentioning that, for Gaussian watermarking, if K is large, the minimum attack is likely to produce noticeable distortion even with no additive noise is added. For instance, under the minimum attack, the MSE is as large as 13.3 for Lena when $K = 70$. In order to have the same MSE under the average attack, we need to have a corresponding WNR as low as -7.5 dB. With such a low WNR, noticeable distortion is introduced to the host signal and the quality of the image may not be acceptable. Thus, the minimum attack is not favored in practice because it generates noticeable distortion.

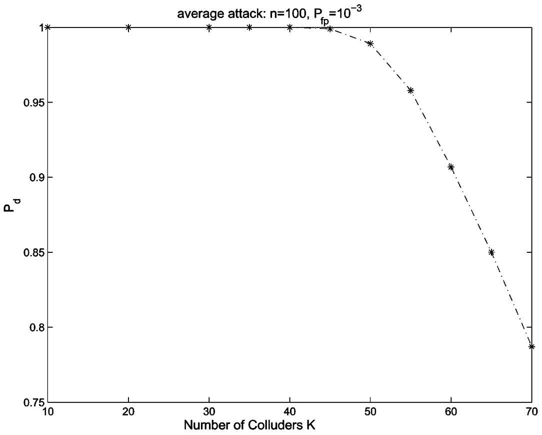


Fig. 10. Detection performance of the thresholding detector on Lena images under the average attack, where, equivalently, $N = 13\,691$. Here, $\sigma_d^2 = 1$, $n = 100$, and $P_{fp} = 10^{-3}$.

VII. Conclusion

In this paper, we have investigated the collusion resistance of a Gaussian fingerprinting system based upon orthogonal modulation. Specifically, assuming the host content is available on the detector side (nonblind scenario), we study the problem of determining how many independently marked copies of the same multimedia content is required for an attacker to cause a fingerprinting system to fail. We introduced the collusion problem for additive embedding and started with the average collusion attack where an average operation is performed by weighing marked copies equally. Since knowledge of the number of colluders (different marked copies) is normally not available in practice, a likelihood-based classifier approach was proposed to estimate the number of colluders K. Simulation results show that the collusion resistance based on the estimated K matches the analysis of the ideal case.

We introduced two detection approaches, and studied the collusion resistance of a fingerprinting system to the average attack when considering the performance criteria represented by P_{fp} and P_{np}. We derived lower and upper bounds of the maximum number of colluders $K_{\max}$. It is noted that $\sqrt{\eta N}$ is an important factor, where η is the watermark to noise ratio. Using the upper bound, an attacker can determine how many independent copies are required to guarantee the success of a collusion attack; on the other hand, an owner (detector) will benefit from these bounds in designing a fingerprinting system. For instance, in order to achieve a collusion-free fingerprinting system, a desirable security requirement is to have it very unlikely for a potential attacker to collect more copies than the lower bound, and further to have the distribution size limited by the maximum value of $K_{\max}$.

Our work was further extended to different attacks and different sets of performance criteria. From the detector point of view, the thresholding detector is robust to different attacks, since different attacks yield very close performance as long as the levels of MSE distortion introduced by different attacks are the same. Therefore, the average attack is most efficient from

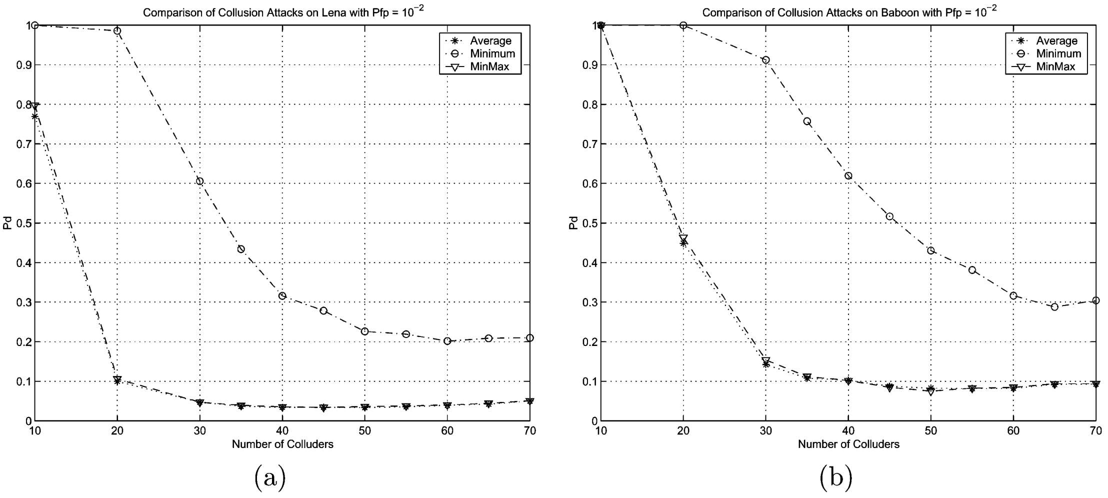


Fig. 11. Detection performance of the thresholding detector on real images under different kinds of attacks. Here, $n = 100$ and $P_{fp} = 10^{-2}$. (a) Lena image with equivalent $N = 13\,691$. (b) Baboon image with equivalent $N = 19\,497$.

the attacker side. We also evaluated the performance on real images, and noted that the average attack is the most efficient from the attacker point of view under the same MSE (thus, PSNR) assumption. Different sets of performance criteria were explored to satisfy different concerns, and it seems that attacks based on a few dozen independent copies will confound a fingerprinting system accommodating as many as ten thousand users. This observation suggests that the number of independently marked copies of the same content that can be distributed should be determined by many concerns, such as the system requirements, and the cost of obtaining multiple independent copies. Furthermore, it suggests that tracing colluders via fingerprints should work in concert with other operations, for example, suspecting a user may lead the owner to more closely monitor that user and further gather additional evidence. As fingerprinting is one of the many components in decision-making, it is the confidence in the fidelity of all technical and nontechnical evidences as a whole that allows a system to identify a colluder.

Appendix I
Derivation of (15) and (16)

Recall that N represents the sample length, n is the number of total users and K means the number of colluders.

Since we assume $\epsilon \ll 1$, meaning a false positive should be unlikely to occur, it immediately implies that the threshold h should yield $Q(h) < 1/(n-K)$ for a fingerprinting system accommodating n users. We provide an intuitive proof for this observation, defining

$$\gamma_j = \begin{cases} 1, & \text{if } j\text{th user is falsely accused} \\ 0, & \text{otherwise} \end{cases} \tag{42}$$

then the expectation of the number of innocents falsely accused is

$$\begin{aligned} E\left(\sum_{j \notin S_c} \gamma_j\right) &= \sum_{j \notin S_c} E(\gamma_j) \\ &= \sum_{j \notin S_c} P_r\{\gamma_j = 1\} = (n-K)Q(h). \end{aligned} \tag{43}$$

Thus, if $Q(h) > 1/(n-K)$, then a false positive almost always happens, which is against our assumption. Therefore, it gives the observation $Q(h) < 1/(n-K)$. We further note that $\epsilon \ll 1$ and K is normally small compared to n; therefore, it is fair to claim $Q(h) \ll 1/n$ in most situations. Since $Q(h) \ll 1/n$, it is safe to assume $h > 1$ due to the fact that $Q(1) \approx 1/6$ and that $Q(h)$ is a monotonously decreasing function for $h > 0$. We summarize these useful observations as follows:

$$Q(h) < \frac{1}{(n-K)}; \quad Q(h) \ll \frac{1}{n}; \quad h > 1 \tag{44}$$

to help our derivation. By applying Lemma 1 and 2, we have

$$\begin{aligned} &1-(n-K)Q(h) < (1-Q(h))^{n-K} < 1-(n-K)Q(h) \\ &\quad + \frac{(n-K)(n-K-1)}{2}Q^2(h), \text{by Lemma 2} \\ &(n-K)Q(h) - \frac{(n-K)(n-K-1)}{2}Q^2(h) < P_{fp} \\ &\quad = 1-(1-Q(h))^{n-K} < (n-K)Q(h) \\ &(n-K)Q(h) < (n-K)Q_b(h), \text{by Lemma 1} \\ &(n-K)Q(h) - \frac{(n-K)(n-K-1)}{2}Q^2(h) \\ &\quad > (n-K)Q_a(h) - \frac{(n-K)(n-K-1)}{2}Q_b^2(h) \end{aligned} \tag{45}$$

therefore, inequalities in (15) follow.

The observations (44) could be used to find a lower bound for h. Since $h > 1$

$$Q(h) < \frac{1}{\sqrt{2\pi}h}\exp\left(-\frac{h^2}{2}\right) < \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{h^2}{2}\right).$$

Suppose we let the last term be equal to $1/n$

$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{h^2}{2}\right) = \frac{1}{n}, \text{thus } \ h = \sqrt{\log\left(\frac{0.5n^2}{\pi}\right)} \triangleq h_{L1} \tag{46}$$

the corresponding h_{L1} serves as a lower bound of the threshold to guarantee $Q(h) \ll 1/n$.

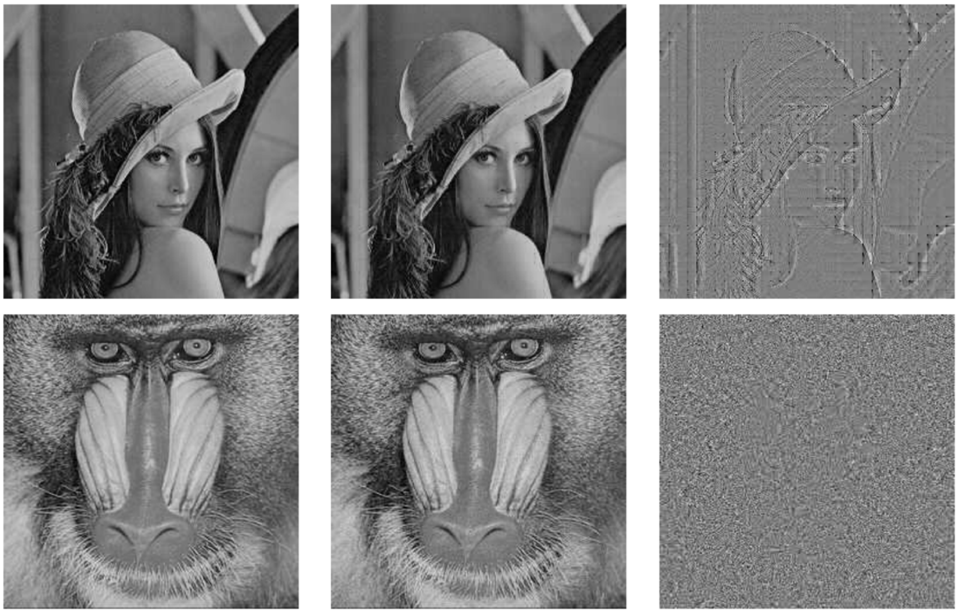


Fig. 12. (Left) Host images, (middle) colluded images with $K = 50$, and (right) difference images for Lena and Baboon. The min attack is illustrated for Lena, and the average attack for Baboon.

Recalling (15) and applying the lower bound h_{L1} result in

$$\begin{aligned} P_{fp} &< (n-K)Q_b(h) = (n-K)\frac{1}{\sqrt{2\pi}h}\exp\left(-\frac{h^2}{2}\right) \\ &< (n-K)\frac{1}{\sqrt{2\pi}h_{L1}}\exp\left(-\frac{h^2}{2}\right) \\ &< n\frac{1}{\sqrt{2\pi}h_{L1}}\exp\left(-\frac{h^2}{2}\right). \end{aligned} \tag{47}$$

To provide $P_{fp} \leq \epsilon$, we can require the last term yield ϵ. It gives

$$\begin{aligned} n\frac{1}{\sqrt{2\pi}h_{L1}}\exp\left(-\frac{h^2}{2}\right) &= \epsilon, \text{thus} \\ h &= \sqrt{\log\left(\frac{n^2}{2\pi\epsilon^2\log\left(\frac{0.5n^2}{\pi}\right)}\right)} \\ &\triangleq h_H \end{aligned} \tag{48}$$

this h_H serves as an upper bound of the threshold h to guarantee $P_{fp} \leq \epsilon$.

Since the tighter the bounds of the threshold h the better, we would like to further adjust the lower bound of h by considering (15) and the above upper bound h_H

$$\begin{aligned} P_{fp} &> (n-K)Q_a(h) - \frac{(n-K)(n-K-1)}{2}Q_b^2(h) \\ &= (n-K)\frac{1}{\sqrt{2\pi}h}\exp\left(-\frac{h^2}{2}\right) \\ &\quad \times\left(\left(1-\frac{1}{h^2}\right) - \frac{n-K-1}{2\sqrt{2\pi}h}\exp\left(-\frac{h^2}{2}\right)\right) \\ &> (n-K)\frac{1}{\sqrt{2\pi}h_H}\exp\left(-\frac{h^2}{2}\right) \\ &\quad \times\left(\left(1-\frac{1}{h_{L1}^2}\right) - \frac{n-K-1}{2\sqrt{2\pi}h_{L1}}\exp\left(-\frac{h_{L1}^2}{2}\right)\right) \\ &= (n-K)\frac{1}{\sqrt{2\pi}h_H} \\ &\quad \times\exp\left(-\frac{h^2}{2}\right)\left(1-\frac{1}{h_{L1}^2}-\frac{n-K-1}{2nh_{L1}}\right) \\ &> \frac{1}{\sqrt{2\pi}h_H}\exp\left(-\frac{h^2}{2}\right)\left(1-\frac{1}{h_{L1}^2}-\frac{1}{2h_{L1}}\right). \end{aligned} \tag{49}$$

By requiring the last term be equal to ϵ, we will obtain a lower bound to satisfy $P_{fp} = \epsilon$, such that

$$h = \sqrt{2\log\left(\frac{2h_{L1}^2 - h_{L1} - 2}{2\sqrt{2\pi}\epsilon h_H h_{L1}^2}\right)} \triangleq h_{L2}. \tag{50}$$

By combining together the lower bounds in (46) and (50), we shall determine a tighter lower bound as $h_L = \max\{h_{L1}, h_{L2}\}$. Therefore, it completes the derivation of (16). We would like to point out that the above derived $\{h_L, h_H\}$ is one, but not the only one, choice of bound pairs satisfying the inequalities in (14).

Appendix II
Derivation of (17)

We repeat the formula of P_d in (14) as below

$$P_d = 1 - \left(1 - Q\left(h - \frac{\sqrt{\eta N}}{K}\right)\right)^K \geq \beta \tag{51}$$

where β is close to 1. We first show a lower bound of $K_{\max}$ tolerated by a Gaussian fingerprinting system with n users under some specific WNR η. The lower bound K_L must be chosen such that the pair (K_L, h_H) satisfies the probability requirements. Since the tail integral $Q(t)$ monotonously decreases as t increases, we observe that

$$\begin{aligned} h_H - \frac{\sqrt{\eta N}}{K} &= 0 \text{ and } Q(0) = \frac{1}{2} \\ &\Rightarrow (1 - Q\left(h_H - \frac{\sqrt{\eta N}}{K}\right) = \frac{1}{2} \\ P_d &= 1 - \left(1 - Q\left(h - \frac{\sqrt{\eta N}}{K}\right)\right)^K \\ &= 1 - \left(\frac{1}{2}\right)^K \to 1 \end{aligned} \tag{52}$$

if K is reasonably large, for instance, $K = 4$ gives $P_d = 15/16$, which is close to 1. Therefore, $K = \sqrt{\eta N}/h_H$ serves as a loose lower bound

$$K_L = \frac{\sqrt{\eta N}}{h_H}. \tag{53}$$

Next, we find an upper bound K_H such that the pair (K_H, h_L) results in a smaller P_d than the requirement β, and a larger P_{fp} than the requirement ϵ. The smaller the gap, the tighter the bound. Similar as in the above observation, if the number of colluders $K \leq \sqrt{\eta N}/h_L$, then the resulting $P_d \to 1$, thus the bound $K_{HL} = (\sqrt{\eta N}/h_L)$ is actually a lower bound of the upper bound K_H and we have $h_L - \sqrt{\eta N}/K > 0$ assumed for searching K_H. We further note that

$$\begin{aligned} P_d &= 1 - \left(1 - Q\left(h_L - \frac{\sqrt{\eta N}}{K}\right)\right)^K \\ &< 1 - \left(1 - Q\left(h_L - \frac{\sqrt{\eta N}}{K}\right)\right)^n \end{aligned} \tag{54}$$

since $(1 - Q(h_L - \sqrt{\eta N}/K)) \in (0, 1)$ and $K \leq n$ are assumed by definition. By setting the last term to be β, we obtain the solution

$$K = \frac{\sqrt{\eta N}}{h_L - Q^{-1}(1 - \sqrt[n]{1-\beta})} \triangleq \tilde{K}. \tag{55}$$

Clearly, this $\tilde{K}$ can serve as an upper bound of the upper bound K_H. Therefore, we have

$$\begin{aligned} 1 - &\left(1 - Q\left(h_L - \frac{\sqrt{\eta N}}{K}\right)\right)^K \\ &< 1 - \left(1 - Q\left(h_L - \frac{\sqrt{\eta N}}{K}\right)\right)^{\tilde{K}} \end{aligned} \tag{56}$$

and calculate the corresponding K_H via letting

$$\begin{aligned} 1 - \left(1 - Q\left(h_L - \frac{\sqrt{\eta N}}{K}\right)\right)^{\tilde{K}} &= \beta, \text{ thus} \\ K &= \frac{\sqrt{\eta N}}{h_L - Q^{-1}(1 - \sqrt[\tilde{K}]{1-\beta})} \\ &\triangleq K_H. \end{aligned} \tag{57}$$

Clearly $P_d < \beta$ is met with this choice of K_H. Recall that $K \leq n$ by definition, it is straightforward that

$$K_{\max} \geq \min\{n, K_L\}; \quad K_{\max} \leq \min\{n, K_H\}. \tag{58}$$

It completes the derivation of (17). It is worth mentioning that the bound K_H can be further tightened by letting $\tilde{K} = K_H$ and then updating K_H according to (57) iteratively, until $\tilde{K}$ is very close to K_H.

APPENDIX III
DERIVATION OF THE MEAN AND THE VARIANCE OF $T_N(j)$ FOR THE MINIMUM ATTACK

Denote the pdf of each Gaussian fingerprint component as $f(x)$, i.e., $f(x) = \mathcal{N}\left(0, \sigma_s^2\right)$, and the cdf as $F(x)$. Now, under the minimum attack, the correlator $T_N(j)$ is

$$T_N(j) = \frac{1}{||\mathbf{s}||} \sum_{i=1}^{N} \left(\min_{l \in S_c}\{s_l(i)\} + d_{\min}(i)\right) s_j(i). \tag{59}$$

Define $s_{\min}(i) = \min_{l \in S_c}\{s_l(i)\}$, we have the pdf of $s_{\min}(i)$ as

$$f_{\min}(x) = K f(x)[1 - F(x)]^{K-1}. \tag{60}$$

For $j \notin S_c$, it is easy to show that $E\{T_N(j)\} = 0$. For $j \in S_c$, we can express the joint pdf of $s_{\min}(i)$ and $s_j(i)$ as in (61), shown at the bottom of the page. By employing the rule of integration by parts, we have

$$\begin{aligned} E\{s_{\min}(i)s_j(i)\} &= \int_{-\infty}^{\infty} x'^2 f(x')[1 - F(x')]^{K-1} dx' \\ &\quad + \int_{-\infty}^{\infty} x'(K-1) f(x')[1 - F(x')]^{K-2} \\ &\quad \times \left(\int_{x'}^{\infty} x f(x) dx\right) dx' \\ &= \sigma_s^2 \int_{-\infty}^{\infty} f(x')[1 - F(x')]^{K-1} dx' \\ &= \frac{\sigma_s^2}{K} \int_{-\infty}^{\infty} f_{\min}(x') dx' = \frac{\sigma_s^2}{K}. \end{aligned} \tag{62}$$

$$f_{\min,1}(s_{\min}(i) = x', s_j(i) = x) = \begin{cases} f(x')[1 - F(x')]^{K-1}, & \text{if } s_{\min}(i) = s_j(i) \\ (K-1) f(x') f(x)[1 - F(x')]^{K-2}, & \text{if } s_{\min}(i) < s_j(i). \end{cases} \tag{61}$$

$$
\begin{aligned}
(\hat{K},\hat{S}_c) &= \arg\max_K \left\{ \max_{S_c} p(\mathbf{T}_N|H_K,S_c) \right\} = \arg\max_K \left\{ \max_{S_c} \frac{p(\mathbf{T}_N|H_K)}{p(\mathbf{T}_N|H_0)} \right\} \\
&= \arg\max_K \left\{ \max_{S_c} \sum_{j\in S_c} \left(\frac{2\parallel \mathbf{s} \parallel}{K} T_N(j) - \frac{\parallel \mathbf{s} \parallel^2}{K^2} \right) \right\}, \text{ by applying log operation} \\
&= \arg\max_K \left\{ \max_{S_c} \frac{2\parallel \mathbf{s} \parallel}{K} \sum_{j\in S_c} T_N(j) - \frac{\parallel \mathbf{s} \parallel^2}{K} \right\} \\
\text{thus, } \hat{K} &= \arg\max_K \left\{ \frac{2\parallel \mathbf{s} \parallel}{K} \sum_{j\in \hat{S}_c} T_N(j) - \frac{\parallel \mathbf{s} \parallel^2}{K} \right\} = \arg\max_K \left\{ \frac{2\parallel \mathbf{s} \parallel}{K} \sum_{j=1}^{K} T_N^{(j)} - \frac{\parallel \mathbf{s} \parallel^2}{K} \right\} \qquad (65)
\end{aligned}
$$

It is clear that, for $j \in S_c$, the mean of $T_N(j)$ under the minimum attack is the same as that of the average attack. We can calculate $E\{(s_{\min}(i)s_j(i))^2\}$ and $\operatorname{var}\{s_{\min}(i)s_j(i)\}$ numerically. Therefore, we can calculate the mean and variance of $T_N(j)$ correspondingly as

$$
\begin{aligned}
E\{T_N(j)\} &= \frac{N}{\|\mathbf{s}\|} E\{s_{\min}(i)s_j(i)\} = \frac{\sqrt{N\sigma_s^2}}{K} \\
\operatorname{var}\{T_N(j)\} &= \frac{\operatorname{var}\{s_{\min}(i)s_j(i)\}}{\sigma_s^2}. \qquad (63)
\end{aligned}
$$

Appendix IV
Derivation of (35)

Recall that the estimator is

$$(\hat{K},\hat{S}_c) = \arg\max_{K,S_c} p(\mathbf{T}_N|H_K,S_c). \qquad (64)$$

By introducing an additional dummy class H_0 as $p(\mathbf{T}_N|H_0) = \mathcal{N}\left(0,\sigma_d^2\mathbf{I}_n\right)$, we have (65), shown at the top of the page, where $T_N^{(j)}$ are the ordered as $T_N^{(1)} \geq T_N^{(2)} \geq \cdots \geq T_N^{(n)}$. The last equation is due to the ML estimate

$$
\begin{aligned}
\hat{S}_c &= \arg\max_{|S_c|=K} \left\{ \frac{2\parallel \mathbf{s} \parallel}{K} \sum_{j\in S_c} T_N(j) - \frac{\parallel \mathbf{s} \parallel^2}{K} \right\} \\
&= \arg\max_{|S_c|=K} \sum_{j\in S_c} T_N(j) \\
&= \text{the index of } K\text{largest } T_N(j). \qquad (66)
\end{aligned}
$$

References

[1] N. Balakrishnan and C. Rao, Eds., *Order Statistics: Theory & Methods*. New York: Elesvier Science, 1998.

[2] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, “Techniques for data hiding,” *IBM Syst. J.*, vol. 35, no. 3–4, pp. 313–336, 1996.

[3] F. Boland, J. O. Ruanaidl JJK, and C. Dautzenberg, “Watermarking digital images for copyright protection,” in *Proc. 5th Int. Conf. Image Processing and its Application*, Jul. 1995, pp. 326–330.

[4] D. Boneh and J. Shaw, “Collusion-secure fingerprinting for digital data,” in *Proc. Advances in Cryptology*, vol. LNCS 963, 1995, pp. 452–65.

[5] ——, “Collusion-secure fingerprinting for digital data,” *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1897–1905, Sep. 1998.

[6] B. Chor, A. Fiat, M. Naor, and B. Pinkas, “Tracing traitors,” *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 893–910, May 2000.

[7] H. Chu, L. Qiao, and K. Nahrstedt, “A secure multicast protocol with copyright protection,” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 32, no. 2, pp. 42–60, Apr. 2002.

[8] I. Cox, J. Bloom, and M. Miller, *Digital Watermarking: Principles & Practice*. San Mateo, CA: Morgan Kaufman, 2001.

[9] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–87, Dec. 1997.

[10] J. Dittmann, P. Schmitt, E. Saar, J. Schwenk, and J. Ueberberg, “Combining digital watermarkds and collusion secure fingerprints fro digital images,” *SPIE J. Electron. Imag.*, vol. 9, pp. 456–67, 2000.

[11] J. Domingo-Ferrer and J. Herrera-Joancomartí, “Simple collusion-secure fingerprinting schemes for images,” in *Proc. IEEE Int. Conf. Information Technology: Coding and Computing*, 2000, pp. 128–32.

[12] F. Ergun, J. Kilian, and R. Kumar, “A note on the limits of collusion-resistant watermarks,” in *Proc. Eurocrypt*, 1999, pp. 140–49.

[13] F. Hartung and M. Kutter, “Multimedia watermarking techniqures,” *Proc. IEEE*, vol. 87, no. 7, pp. 1079–1107, Jul. 1999.

[14] N. Johnson, Z. Duric, and S. Jajodia, “Recovery of watermarks from distorted images,” in *Proc. 3rd Int. Information Hiding Workshop*, 1999, pp. 361–375.

[15] J. Kilian, T. Leighton, L. Matheson, T. Shamoon, R. Tarjan, and F. Zane, “Resistance of digital watermarks to collusive attacks,” in *Proc. IEEE Int. Symp. Information Theory*, Aug. 1998, pp. 271–271.

[16] E. Lehmann, *Theory of Point Estimation*. New York: Wiley, 1983.

[17] C. Lin, M. Wu, J. Bloom, M. Miller, I. Cox, and Y. Lui, “Rotation, scale, and translation resilient public watermarking for images,” *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 767–782, May 2001.

[18] J. Linnartz, A. Kalker, G. Depovere, and R. Beuker, “A reliability model for detection of electronic watermarks in digital images,” in *Proc. Benelux Symp. Communication Theory*, Oct. 1997, pp. 202–209.

[19] J. Lubin, J. Bloom, and H. Cheng, “Robust, content-dependent, high-fidelity watermark for tracking in digital cinema,” presented at the Security and Watermarking of Multimedia Contents V, P. W. Wong and E. J. Delp, Eds., 2003, http://www.geocities.com/Jeffrey_Bloom/research/lubin03-spie.pdf.

[20] B. Pfitzmann and M. Waidner, “Anonymous fingerprinting,” IBM Research, Res. Rep. RZ 2881, 1996.

[21] ——, “Asymmetric fingerprinting for larger collusions,” in *Proc. 4th ACM Conf. Coumputer and Communications Security*, 1997, pp. 151–160.

[22] C. Podilchuk and W. Zeng, “Image adaptive watermarking using visual models,” *IEEE J. Sel. Areas Commun.*, vol. 16, no. 5, pp. 525–540, May 1998.

[23] M. Simon, S. Hinedi, and W. Lindsey, “Appendix 3B: The Gaussian integral $Q(x)$,” in *Digital Communication Techniques: Signal Design and Detection*. Upper Saddle River, NJ: Prentice-Hall, 1995.

[24] H. Stone, “Analysis of attacks on image watermarks with randomized coefficients,” NEC Res. Inst., Princeton, NJ, Tech. Rep. 96-045, 1996.

[25] J. Su, J. Eggers, and B. Girod, “Capacity of digital watermarks subjected to an optimal collusion attack,” presented at the Eur. Signal Processing Conf., Sep. 2000.

[26] W. Trappe, M. Wu, Z. J. Wang, and K. J. R. Liu, “Anti-collusion fingerprinting for multimedia,” *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1069–1087, Apr. 2003.

[27] M. Wu, W. Trappe, Z. J. Wang, and K. J. R. Liu, “Review paper: Collusion resistant fingerprinting for multimedia,” *IEEE Signal Process. Mag.*, vol. 21, no. 2, pp. 15–27, Feb. 2004.

[28] Z. J. Wang, M. Wu, W. Trappe, and K. J. R. Liu, “Group-oriented fingerprinting for multimedia forensics,” *EURASIP J. Appl. Signal Process.*, vol. 14, pp. 2153–2173, 2004.

[29] M. Wu and B. Liu, *Multimedia Data Hiding*. New York: Springer-Verlag, 2002.

[30] Y. Yacobi, “Improved Boneh-Shaw content fingerprinting,” in *CT-RSA 2001*. Berlin, Germany: Springer-Verlag, 2001, LNCS 2020, pp. 378–91.

[31] F. Zane, “Efficient watermark detection and collusion security,” in *Proc. 4th Int. Conf. Financial Cryptography*, Feb. 2000, pp. 21–32.

[32] H. V. Zhao, M. Wu, Z. J. Wang, and K. J. R. Liu, “Forensic Analysis of Nonlinear Collusion on Independent Multimedia Fingerprints,” *IEEE Trans. Image Process.*, vol. 14, no. 5, pp. 646–661, May 2005.

Z. Jane Wang (M’02) received the B.Sc. degree (with the highest honors) from Tsinghua University, Beijing, China, in 1996 and the M.Sc. and Ph.D. degrees from the University of Connecticut, Storrs, in 2000 and 2002, respectively, all in electrical engineering.

She has been a Research Associate with the Electrical and Computer Engineering Department, Institute for Systems Research at the University of Maryland, College Park. Since August 2004, she has been an Assistant Professor with the Department Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada. Her research interests are in the broad areas of statistical signal processing, with applications to information security, biomedical imaging, genomic, and wireless communications.

Dr. Wang received the Outstanding Engineering Doctoral Student Award while at the University of Connecticut.

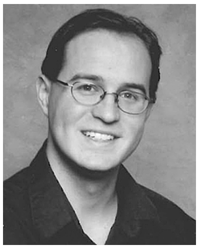
Min Wu (S’95–M’01) received the B.Eng. degree in electrical engineering and the B.A. degree in economics (both with the highest honors) from Tsinghua University, Beijing, China, in 1996 and the M.A. degree and Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1998 and 2001, respectively.

She was with NEC Research Institute, Princeton, NJ, in 1998 and Panasonic Information and Networking Laboratories, Princeton, in 1999. Since 2001, she has been an Assistant Professor with the Department of Electrical and Computer Engineering, Institute of Advanced Computer Studies and the Institute of Systems Research, University of Maryland, College Park. She is a Guest Editor of the Special Issue on Media Security and Rights Management for the *EURASIP Journal on Applied Signal Processing*. She co-authored the book *Multimedia Data Hiding* (New York: Springer-Verlag, 2003) and holds four U.S. patents on multimedia security. Her research interests include information security, multimedia signal processing, and multimedia communications.

Dr. Wu received a CAREER award from the U.S. National Science Foundation in 2002, a George Corcoran Faculty Award from University of Maryland in 2003, a TR100 Young Innovator Award from MIT Technology Review Magazine in 2004, and a Young Investigator Award from U.S. Office of Naval Research in 2005. She is a member of the IEEE Technical Committee on Multimedia Signal Processing and was the Publicity Chair of the 2003 IEEE International Conference on Multimedia and Expo.

Hong Vicky Zhao (S’02–M’05) received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 1997 and 1999, respectively, and the Ph.D. degree from the University of Maryland, College Park, in 2004, all in electrical engineering.

Since 2005, she has been a Research Associate with the Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland. Her research interests include multimedia security, digital rights management, multimedia communication over networks, and multimedia signal processing.



Wade Trappe (M’02) received the B.A. degree in mathematics from The University of Texas at Austin in 1994, and the Ph.D. in applied mathematics and scientific computing from the University of Maryland, College Park, in 2002.

He is currently an Assistant Professor at the Wireless Information Network Laboratory (WINLAB) and the Electrical and Computer Engineering Department, Rutgers University, Piscataway, NJ. His research interests include wireless network security, wireless networking, and multimedia security. He is a coauthor of the textbook *Introduction to Cryptography with Coding Theory* (Upper Saddle River, NJ: Prentice-Hall, 2001).

Dr. Trappe received the George Harhalakis Outstanding Systems Engineering Graduate Student award while at the University of Maryland.

K. J. Ray Liu (F’03) received the B.S. degree from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1983 and the Ph.D. degree from the University of California, Los Angeles, in 1990, both in electrical engineering.

He is a Professor and Director of Communications and Signal Processing Laboratories of Electrical and Computer Engineering Department and Institute for Systems Research, University of Maryland, College Park. His research contributions encompass broad aspects of information forensics and security; wireless communications and networking; multimedia communications and signal processing; signal processing algorithms and architectures; and bioinformatics, in which he has published over 350 refereed papers.

Dr. Liu is the recipient of numerous honors and awards, including the IEEE Signal Processing Society’s 2004 Distinguished Lecturer; the 1994 National Science Foundation’s Young Investigator Award; the IEEE Signal Processing Society’s 1993 Senior Award (Best Paper Award); the IEEE 50th Vehicular Technology Conference Best Paper Award, Amsterdam, The Netherlands, 1999; and the EURASIP 2004 Meritorious Service Award. He also received the George Corcoran Award in 1994 for outstanding contributions to electrical engineering education and the Outstanding Systems Engineering Faculty Award in 1996 in recognition for outstanding contributions in interdisciplinary research, both from the University of Maryland. He is the Editor-in-Chief of *IEEE Signal Processing Magazine*, the prime proposer and architect of the new IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and was the founding Editor-in-Chief of *EURASIP Journal on Applied Signal Processing*. He is a member of Board of Governors of IEEE Signal Processing Society.