Behavior Forensics for Scalable Multiuser Collusion: Fairness Versus Effectiveness

H. Vicky Zhao, *Member, IEEE*, and K. J. Ray Liu, *Fellow, IEEE*

***Abstract*—Multimedia security systems involve many users with different objectives and users influence each other's performance. To have a better understanding of multimedia security systems and offer stronger protection of multimedia, behavior forensics formulate the dynamics among users and investigate how they interact with and respond to each other. This paper analyzes the behavior forensics in multimedia fingerprinting and formulates the dynamics among attackers during multi-user collusion. In particular, this paper focuses on how colluders achieve the fair play of collusion and guarantee that all attackers share the same risk (i.e., the probability of being detected). We first analyze how to distribute the risk evenly among colluders when they receive fingerprinted copies of scalable resolutions due to network and device heterogeneity. We show that generating a colluded copy of higher resolution puts more severe constraints on achieving fairness. We then analyze the effectiveness of fair collusion. Our results indicate that the attackers take a larger risk of being captured when the colluded copy has higher resolution, and they have to take this tradeoff into consideration during collusion. Finally, we analyze the collusion resistance of the scalable fingerprinting systems in various scenarios with different system requirements, and evaluate the maximum number of colluders that the fingerprinting systems can withstand.**

***Index Terms*—Behavior forensics, collusion resistance, fairness, scalable multiuser collusion, traitor tracing.**

I. Introduction

RECENT development in multimedia processing and network technologies has facilitated the distribution and sharing of multimedia through networks. It is critical to protect multimedia from illegal alteration, repackaging, and unauthorized redistribution. Digital fingerprinting is such a forensic tool to identify the source of the illicit copies and trace traitors. It embeds a unique label, also known as the digital fingerprint, in each distributed copy before distribution. The unique fingerprint is seamlessly embedded into the host signal using traditional data hiding techniques [1] (e.g., the spread-spectrum embedding method [2]), and travels with the host signal. There is a cost effective attack against digital fingerprinting, the collusion attack, in which several attackers combine information from differently fingerprinted copies to remove traces of the embedded fingerprints [2]. To support multimedia forensics, there has been a lot of work on the design of anticollusion multimedia fingerprints [3]–[6], which can resist such multiuser collusion as well as common signal processing and attacks on a single copy [7], [8].

In multimedia security systems, different users have different goals and objectives, and they influence each other's decisions and performance. Therefore, it is important to study this behavior dynamics in multimedia fingerprinting. Behavior forensics formulate the dynamics among attackers during collusion and the dynamics between the colluders and the detector, and investigate how users interact with and respond to each other. Such investigation enables the digital rights enforcer to have a better understanding of the multimedia security systems (e.g., how attackers behave during collusion, which information of collusion can help improve the detection performance, etc.). This investigation helps the digital rights enforcer offer stronger protection of multimedia content.

We investigate the dynamics among colluders in this paper. During multiuser collusion, colluders not only share the profit from the illegal alteration and redistribution of multimedia, they also share the risk of being detected. Since no one is willing to take a higher risk than the others, the colluders demand a fair play during collusion and require that all colluders have the same probability of being captured. Achieving fairness of collusion is an important issue that the colluders need to address.

Most previous work on collusion attacks on multimedia fingerprinting assumed that all users receive fingerprinted copies of the same resolution. In this simple scenario, achieving fairness of collusion is trivial. For example, averaging all fingerprinted copies with equal weights reduces the energy of each contributing fingerprint by the same ratio, and guarantees that all colluders have the same probability of being detected [9]. For spread-spectrum embedding based multimedia fingerprinting, the collusion attack was modeled as averaging different copies with equal weights followed by an additive noise in [10]. The collusion attack model was generalized to multiple-input–single-output linear shift invariant filtering followed by an additive Gaussian noise in [11]. Nonlinear collusion attacks were examined and analyzed in [12] and [13]. Assuming that colluders receive fingerprinted copies of the same resolution, all these collusion attacks ensure fairness of collusion and guarantee the equal risk of all colluders.

In practice, due to the heterogeneity of the networks and that of the end users' devices, it is often required to have scalability for rich multimedia access from anywhere using any devices. Scalable coding and transmission enables users to recover physically meaningful information of the content even if they receive only part of the compressed bit streams [14]. This paper investigates how colluders distribute the risk evenly among themselves and achieve fairness of collusion when they receive copies of

Manuscript received March 15, 2005; revised April 2, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gaurav Sharma.

H. Vicky Zhao is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4 Canada (e-mail: zhaohong@ieee.org).

K. J. Ray Liu is with the Department of Electrical and Computer Engineering, and Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: kjrliu@eng.umd.edu).

Digital Object Identifier 10.1109/TIFS.2006.879279

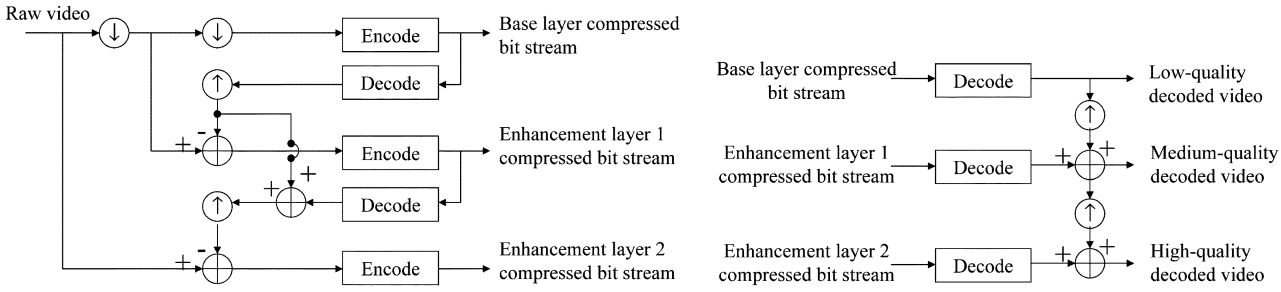

Fig. 1. Three-layer scalable codec. Left: encoder, right: decoder.

different resolutions due to network and device heterogeneity. We also analyze the effectiveness of such fair collusion in defeating the fingerprinting systems. We then switch our role to the digital rights enforcer's side and study the collusion resistance of the scalable fingerprinting system. We evaluate the maximum number of colluders that the embedded fingerprints can withstand in various scenarios with different requirements. We use video to demonstrate a typical multimedia system and take temporal scalability as an example.

The rest of the paper is organized as follows. We begin in Section II with the introduction of the scalable video coding and the digital fingerprinting system model. Section III investigates how to achieve fairness of collusion when attackers receive copies of different resolutions. We analyze the effectiveness of fair collusion in removing the embedded fingerprints in Section IV. Section V quantifies the collusion resistance of scalable fingerprinting systems and studies how many colluders are enough to undermine the tracing capability of multimedia fingerprints. Section VI shows the simulation results on video sequences, and conclusions are drawn in Section VII.

II. System Model

A. Temporally Scalable Video Coding Systems

In the literature, scalable video coding is widely used to accommodate heterogenous networks and devices with different computational capability. As an example, we use layered video coding and decompose the video content into non-overlapping streams (layers) with different priorities [14]. The base layer contains the most important information of the video sequence and is received by all users in the system. The enhancement layers gradually refine the resolution of the reconstructed copy at the decoder's side and are only received by those who have sufficient bandwidth.

Fig. 1 shows the block diagrams of a three-layer scalable codec. The encoder first down-samples the raw video and performs lossy compression to generate the base layer bit stream. Then, the encoder calculates the difference between the original video sequence and the up-sampled base layer, and applies lossy compression to this residue to generate the enhancement layer bit streams. At the receiver's side, to reconstruct a high-resolution video, the decoder has to first receive and decode both the base layer and the enhancement layer bit streams. Then the up-sampled base layer is combined with the enhancement layer refinements to form the high-resolution decoded video.

In this paper, we use temporally scalable video coding as an example, which provides multiple versions of the same video with different frame rates. Our analysis can also be applied to other types of scalability since the scalable codec in Fig. 1 is generic and can be used to achieve different types of scalability. The simplest way to perform temporal decimation and temporal interpolation is by frame skipping and by frame copying, respectively. For example, temporal decimation with a ratio of 2:1 can be achieved by discarding one frame from every two frames; and temporal interpolation with a ratio of 1:2 can be realized by making a copy of each frame and transmitting the two frames to the next stage.

We consider a temporally scalable video coding system with three-layer scalability, and use frame skipping and frame copying to implement temporal decimation and interpolation, respectively. In such a video coding system, different frames in the video sequence are encoded in different layers. Define F_b, F_{e1}, and F_{e2} as the sets containing the indices of the frames that are encoded in the base layer, enhancement layer 1 and enhancement layer 2, respectively. For example, with MPEG-2 video encoding, the base layer may contain all the I frames; the enhancement layer 1 consists of all the P frames; and the enhancement layer 2 includes all the B frames.[1]

Define $F^{(i)}$ as the set containing the indices of the frames that user $\mathbf{u}^{(i)}$ receives. Define $\mathbf{U}^b \triangleq \left\{\mathbf{u}^{(i)} : F^{(i)} = F_b\right\}$ as the subgroup of users who subscribe to the lowest resolution and receive the base layer bit stream only; $\mathbf{U}^{b,e1} \triangleq \left\{\mathbf{u}^{(i)} : F^{(i)} = F_b \cup F_{e1}\right\}$ is the subgroup of users who subscribe to the medium resolution and receive both the base layer and the enhancement layer 1; and $\mathbf{U}^{\text{all}} \triangleq \left\{\mathbf{u}^{(i)} : F^{(i)} = F_b \cup F_{e1} \cup F_{e2}\right\}$ is the subgroup of users who subscribe to the highest resolution and receive all three layers. $\mathbf{U}^b$, $\mathbf{U}^{b,e1}$ and $\mathbf{U}^{\text{all}}$ are mutually exclusive, and $M = |\mathbf{U}^b| + |\mathbf{U}^{b,e1}| + |\mathbf{U}^{\text{all}}|$ is the total number of users.

B. Digital Fingerprinting System and Collusion Attacks

We consider a digital fingerprinting system that consists of three parts: fingerprint embedding, collusion attacks and fingerprint detection. We use temporal scalability as an example and analyze the fairness issue during collusion. In this scenario, fingerprints embedded at different layers will not interfere with

[1]In this example, some users can only receive the I frames due to bandwidth and computation constraints; some users might have sufficient bandwidth and computation capability to receive and decode both I and P frames; while some users have enough bandwidth to receive all I, P, and B frames and reconstruct a sequence including every frame in the video.

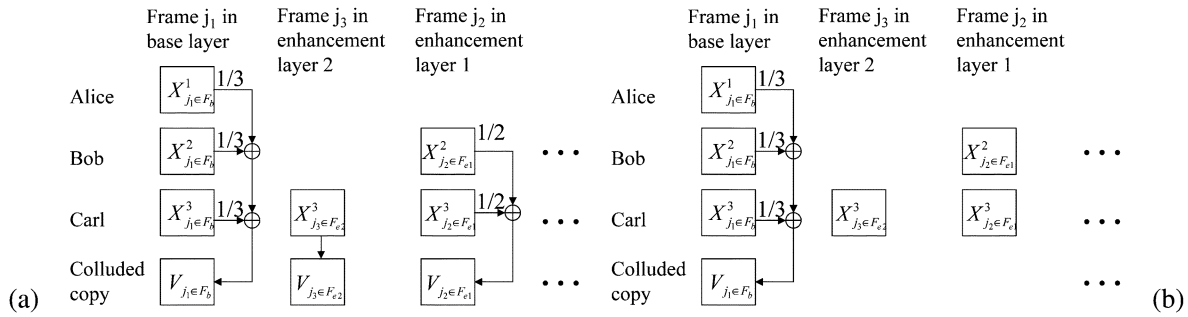

Fig. 2. Two trivial solutions of collusion by averaging all fingerprinted copies. Assume that Alice receives the fingerprinted copy $\{\mathbf{X}_j^1\}_{j\in F_b}$ consisting of the base layer only; Bob receives the fingerprinted copy $\{\mathbf{X}_j^2\}_{j\in F_b\cup F_{e1}}$ with both the base layer and the enhancement layer 1; and Carl receives the fingerprinted copy $\{\mathbf{X}_j^3\}_{j\in F_b\cup F_{e1}\cup F_{e2}}$ including all three layers. (a) Colluded copy $\{\mathbf{V}_j\}$ contains all three layers. (b) Colluded copy $\{\mathbf{V}_j\}$ includes frames in the base layer only.

each other. Our model can also be applied to other types of scalability, e.g., spatial and SNR scalability. However, with spatial or SNR scalability, the content owner has to take special care during fingerprint design and embedding to prevent fingerprints at different layers from interfering each other. This issue of fingerprint design and embedding is beyond the scope of this paper.

1) Fingerprint Embedding: Spread-spectrum embedding is one of the popular data hiding techniques due to its robustness against many attacks [2], [15]. For the jth frame in the video sequence represented by a vector $\mathbf{S}_j$ of length N_j, and for each user $\mathbf{u}^{(i)}$ who subscribes to frame j, the content owner generates a unique fingerprint $\mathbf{W}_j^{(i)}$ of length N_j. The fingerprinted frame j that will be distributed to $\mathbf{u}^{(i)}$ is $X_j^{(i)}(k) = S_j(k)+JND_j(k)\cdot W_j^{(i)}(k)$, where $X_j^{(i)}(k), S_j(k)$ and $W_j^{(i)}(k)$ are the kth components of the fingerprinted frame $\mathbf{X}_j^{(i)}$, the host signal $\mathbf{S}_j$ and the fingerprint vector $\mathbf{W}_j^{(i)}$, respectively. JND_j is the just-noticeable-difference from human visual models [15], and it is used to control the energy and achieve the imperceptibility of the embedded fingerprints. Finally, the content owner transmits to each user $\mathbf{u}^{(i)}$ all the fingerprinted frames $\left\{\mathbf{X}_j^{(i)}\right\}$ that $\mathbf{u}^{(i)}$ subscribes to.

We apply orthogonal fingerprint modulation [3], [9] and assume that the total number of users is much smaller than the length of the embedded fingerprints. For each frame j in the video sequence, with orthogonal modulation, fingerprints for different users are orthogonal to each other and have the same energy, i.e., for user $\mathbf{u}^{(i_1)}$ and $\mathbf{u}^{(i_2)}$

$$\left\langle \mathbf{W}_j^{(i_1)}, \mathbf{W}_j^{(i_2)} \right\rangle = \|\mathbf{w}_j\|^2 \delta_{i_1,i_2} \tag{1}$$

where δ_{i_1,i_2} is the Dirac-Delta function. δ_{i_1,i_2} equals to 1 if and only if $i_1 = i_2$ and 0 otherwise. $\|\mathbf{w}_j\|^2$ depends on the fingerprint's length N_j and $\|\mathbf{w}_j\|^2 = N_j \cdot \xi^2$ where ξ is a constant where xi is a constant. To combat the intracontent collusion attacks [16]–[19] in each distributed copy $\left\{\mathbf{X}_j^{(i)}\right\}$, we embed correlated fingerprints $\mathbf{W}_{j_1}^{(i)}$ and $\mathbf{W}_{j_2}^{(i)}$ in adjacent frames $\mathbf{S}_{j_1}$ and $\mathbf{S}_{j_2}$, respectively. The correlation between the two fingerprints $\mathbf{W}_{j_1}^{(i)}$ and $\mathbf{W}_{j_2}^{(i)}$ depends on the similarity between the two host frames $\mathbf{S}_{j_1}$ and $\mathbf{S}_{j_2}$, similar to the work in [20], [21].

2) Collusion Attacks: The attackers apply multiuser collusion attacks to remove traces of the embedded fingerprints. In a recent investigation [9], [22], we have shown that nonlinear collusion attacks can be modeled as the averaging attack followed by an additive noise. Under the constraint that the colluded copies from different collusion attacks have the same perceptual quality, different collusion attacks have approximately identical performance. Therefore, it suffices to consider the averaging based collusion only.

We consider in this paper fair collusion in which all colluders share the same risk and have the same probability of being caught. When colluders receive copies of the same quality, averaging all copies with the same weight reduces the energy of each contributing fingerprint by an equal amount, and therefore, gives each colluder the same probability of being detected. However, achieving fairness of collusion is much more complicated when colluders receive copies of different resolutions due to network and device heterogeneity, especially when the attackers wish to generate a copy of high resolution.

With the temporally scalable fingerprinting system in Section II-B1, we consider a simple example of collusion including three attackers: Alice who receives the base layer only, Bob who receives the base layer and the enhancement layer 1, and Carl who receives all three layers. Fig. 2 shows two trivial solutions of collusion by averaging the three fingerprinted copies. In Fig. 2(a), the colluded copy includes all three layers and is generated as follows.

- For each frame $j_1 \in F_b$ in the base layer, the colluders average the three copies of fingerprinted frame j_1 that they have and generate $\mathbf{V}_{j_1\in F_b} = (1/3)\left(\mathbf{X}_{j_1}^1 + \mathbf{X}_{j_1}^2 + \mathbf{X}_{j_1}^3\right)$.
- For each frame $j_2 \in F_{e1}$ in the enhancement layer 1, the colluders average the fingerprinted frame j_2 from Bob and Carl, respectively, and $\mathbf{V}_{j_2\in F_{e1}} = (1/2)\left(\mathbf{X}_{j_2}^2 + \mathbf{X}_{j_2}^3\right)$.
- For each frame $j_3 \in F_{e2}$ in the enhancement layer 2, frame j_3 in the colluded copy equals to that in the copy from Carl and let $\mathbf{V}_{j_3\in F_{e2}} = \mathbf{X}_{j_3}^3$.

In the colluded copy in Fig. 2(a), the three fingerprints corresponding to the three attackers have the same energy in the base layer; while the enhancement layers contain only Bob and Carl's fingerprints, not the fingerprint identifying Alice. It is obvious that among the three, Carl has the largest probability of being

caught and Alice takes the smallest risk. Consequently, the collusion in Fig. 2(a) is not fair.

In Fig. 2(b), the collusion outputs an attacked copy consisting of the base layer only, and the colluded copy equals to $\mathbf{V}_{j_1 \in F_b} = (1/3)\left(\mathbf{X}_{j_1}^1 + \mathbf{X}_{j_1}^2 + \mathbf{X}_{j_1}^3\right)$ for each frame $j_1 \in F_b$ in the base layer. Under the collusion in Fig. 2(b), the fingerprints corresponding to the three attackers have the same energy in the colluded copy, and therefore, the three attackers have the same probability of being detected. Although the collusion in Fig. 2(b) ensures fairness, the attacked copy has low resolution.

So the question is when there is difference in the resolution of fingerprinted copies due to network and device heterogeneity, how can colluders conduct fair multiuser collusion that guarantees the collective equal risk among all attackers while still generating an attacked copy of high resolution. Assume that there are a total of K colluders, and SC is the set containing their indices. During collusion, the colluders first divide themselves into three non-overlapping subgroups: $SC^b \triangleq \left\{i \in SC : F^{(i)} = F_b\right\}$ contains the indices of the colluders who receive the base layer only; $SC^{b,e1} \triangleq \left\{i \in SC : F^{(i)} = F_b \cup F_{e1}\right\}$ contains the indices of the colluders who receive the base layer and the enhancement layer 1; and $SC^{\text{all}} \triangleq \left\{i \in SC : F^{(i)} = F_b \cup F_{e1} \cup F_{e2}\right\}$ contains the indices of the colluders who receive all three layers. Define K^b, $K^{b,e1}$ and K^{all} as the number of colluders in SC^b, $SC^{b,e1}$ and SC^{all}, respectively.

Then, the colluders apply the intragroup collusion followed by the intergroup collusion to generate the colluded copy $\{\mathbf{V}_j\}$, as shown in Fig. 3.[2] The colluders first apply the **intragroup collusion attacks**.

- For each frame $j \in F_b$ that they received, the colluders in the subgroup SC^b generate $\mathbf{Z}_j^b = \sum_{i \in SC^b} \mathbf{X}_j^{(i)} / K^b$.
- For each frame $j \in F_b \cup F_{e1}$ that they received, the colluders in the subgroup $SC^{b,e1}$ generate $\mathbf{Z}_j^{b,e1} = \sum_{i \in SC^{b,e1}} \mathbf{X}_j^{(i)} / K^{b,e1}$.
- For each frame $j \in F_b \cup F_{e1} \cup F_{e2}$ that they received, the colluders in the subgroup SC^{all} generate $\mathbf{Z}_j^{\text{all}} = \sum_{i \in SC^{\text{all}}} \mathbf{X}_j^{(i)} / K^{\text{all}}$.

Define F^c as the set containing the indices of the frames that are in the colluded copy, and $F^c \in \{F_b, F_b \cup F_{e1}, F_b \cup F_{e1} \cup F_{e2}\}$. Then, the colluders apply the **intergroup collusion attacks** to generate the colluded copy $\{\mathbf{V}_j\}_{j \in F^c}$.

- For each frame $j_1 \in F_b$ in the base layer, $\mathbf{V}_{j_1} = \beta_1 \mathbf{Z}_{j_1}^b + \beta_2 \mathbf{Z}_{j_1}^{b,e1} + \beta_3 \mathbf{Z}_{j_1}^{\text{all}} + \mathbf{n}_{j_1}$. To maintain the average intensity of the original host signal and ensure the quality of the colluded copy, we let $\beta_1 + \beta_2 + \beta_3 = 1$. Our analysis can also be applied to other scenarios where $\beta_1 + \beta_2 + \beta_3 \neq 1$. To guarantee that the energy of each of the original fingerprints is reduced, we select $0 \leq \beta_1, \beta_2, \beta_3 \leq 1$. $\mathbf{n}_{j_1}$ is the additive noise that the colluders add to $\mathbf{V}_{j_1}$ to further hinder detection.

[2]Note that the intragroup and intergroup collusion attacks should be adjusted according to the type of scalability used in video coding, and they should be applied to each individual layer. For example, with SNR scalability, intragroup and intergroup collusion should be applied to different layers which are the fingerprinted video sequences quantized with different step sizes. Then the colluders combine the newly generated base layer and enhancement layers to produce the final colluded copy.

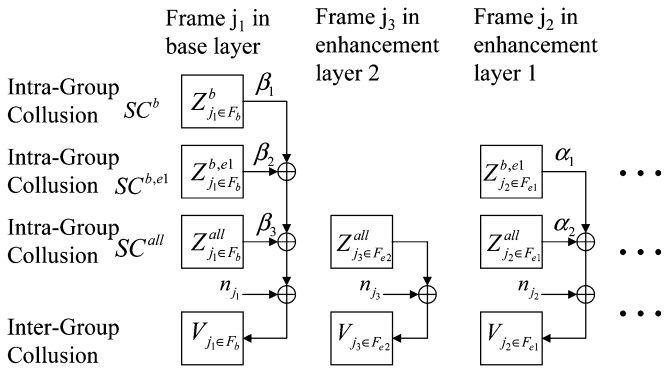

Fig. 3. Intragroup and the intergroup collusion attacks.

- If $F_{e1} \subset F^c$ and the colluded copy contains frames in the enhancement layers, then for each frame $j_2 \in F_{e1}$ in the enhancement layer 1, $\mathbf{V}_{j_2} = \alpha_1 \mathbf{Z}_{j_2}^{b,e1} + \alpha_2 \mathbf{Z}_{j_2}^{\text{all}} + \mathbf{n}_{j_2}$, where $0 \leq \alpha_1, \alpha_2 \leq \alpha_1 + \alpha_2 = 1$. $\mathbf{n}_{j_2}$ is an additive noise. Our analysis can also be extended to the more general case of $\alpha_1 + \alpha_2 \neq 1$.
- If $F_{e2} \subset F^c$ and the colluded copy contains frames in all three layers, then for each frame $j_3 \in F_{e2}$ in the enhancement layer 2, $\mathbf{V}_{j_3} = \mathbf{Z}_{j_3}^{\text{all}} + \mathbf{n}_{j_3}$, where $\mathbf{n}_{j_3}$ is an additive noise.

The colluders adjust the energy of the additive noises to ensure that frames of similar content at different layers in the colluded copy have approximately the same perceptual quality. We consider challenging scenarios with a large number of colluders (e.g., more than 100 attackers). In addition, we consider scenarios where the energy of the additive noise $\mathbf{n}_j$ is comparable with that of the originally embedded fingerprints and the final colluded copy has good quality. For frame j_1 in the base layer, frame j_2 in the enhancement layer 1, and frame j_3 in the enhancement layer 2 that have similar content, we can show that this requirement can be simplified to $\|\mathbf{n}_{j_1}\|^2 \approx \|\mathbf{n}_{j_2}\|^2 \approx \|\mathbf{n}_{j_3}\|^2$ in the scenarios that we are interested in.

The colluders seek the *collusion parameters*, F^c, $\{\beta_k\}_{k=1,2,3}$ and $\{\alpha_l\}_{l=1,2}$, to ensure that all colluders have the same probability to be captured. The detailed analysis is given in Section III.

3) Fingerprint Detection and Colluder Identification: When the content owner discovers the unauthorized redistribution of $\{\mathbf{V}_j\}_{j \in F^c}$, he/she applies a fingerprint detection process to identify the colluders.

With spread-spectrum embedding, depending on the absence or presence of the host signal during the detection process, there are two main detection scenarios, blind and non-blind detection, respectively. In the blind detection scenario, the host signal is not available to the detector and serves as an additional noise during detection; while in the non-blind scenario, the host signal is available to the detector and is first removed from the test copy before detection. Different from other data hiding applications where blind detection is preferred or required, in many fingerprinting applications, the fingerprint verification and colluder identification process is usually handled by the content owner or an authorized forensic party who can have access to the original host signal. Therefore, a non-blind detection scenario is fea-

sible and often preferred in multimedia fingerprinting applications [3], [9], [22].

For each frame $\mathbf{V}_j$ in the colluded copy, the detector first extracts the fingerprint $\mathbf{Y}_j = (\mathbf{V}_j - \mathbf{S}_j)/JND_j$. Then, following the thresholding detection in [9], the detector calculates the similarity between the extracted fingerprint $\{\mathbf{Y}_j\}_{j\in F^c}$ and each of the M original fingerprints $\left\{\mathbf{W}_j^{(i)}\right\}_{j\in F^{(i)}}$, compares with a threshold and outputs a set $\widehat{SC}$ containing the estimated indices of the colluders. Following the prior art [3], [9], [22], we use the correlation based detection statistics to measure the similarity between the extracted fingerprint and the original fingerprint. We consider a detector that uses fingerprints extracted from all layers collectively to identify colluders. For each user $\mathbf{u}^{(i)}$, the detector first calculates $\breve{F}^{(i)} \triangleq F^{(i)} \cap F^c$, where $F^{(i)}$ contains the indices of the frames received by user $\mathbf{u}^{(i)}$ and F^c contains the indices of the frames in the colluded copy. Then, the detector applies the thresholding detection in [9] and calculates

$$TN^{(i)} = \frac{\left(\sum_{j\in\breve{F}^{(i)}} \left\langle \mathbf{Y}_j, \mathbf{W}_j^{(i)} \right\rangle\right)}{\sqrt{\sum_{j\in\breve{F}^{(i)}} \left\|\mathbf{W}_j^{(i)}\right\|^2}} \tag{2}$$

where $\left\|\mathbf{W}_j^{(i)}\right\|$ is the Euclidean norm of $\mathbf{W}_j^{(i)}$. Given the M detection statistics $\{TN^{(i)}\}_{i=1,\ldots,M}$ and a pre-determined threshold h, the estimated colluder set is $\widehat{SC} = \{i : TN^{(i)} > h\}$.

C. Performance Criteria

Digital fingerprinting can be used in different scenarios with different goals and different requirements [9], [22]. To evaluate the effectiveness of the collusion attacks and the performance of the detection statistics, we adopt the commonly used criteria in the literature and use the following measurements.

- P_d: probability of capturing at least one colluder.
- P_{fp}: probability of accusing at least one innocent user.
- $E[F_d]$: expected fraction of colluders that are successfully captured.
- $E[F_{fp}]$: expected fraction of innocent users that are falsely accused.

To measure the temporal resolution of the colluded copy, we use the total number of frames in the colluded copy $L^c = |F^c|$ (or equivalently the frame rate of the colluded copy). $L^c = |F_b|$, $L^c = |F_b| + |F_{e1}|$, and $L^c = |F_b| + |F_{e1}| + |F_{e2}|$ correspond to the three scenarios where the colluded copy has the lowest, medium and highest temporal resolution, respectively.

III. Selection of the Collusion Parameters in Fair Collusion

In this section, given the system model as in Section II, we investigate how the colluders should select the collusion parameters to achieve fairness of collusion and still generate a high-resolution attacked copy in scalable fingerprinting systems. We consider the simple detector in Section II-B3 that uses the fingerprints extracted from all layers collectively to identify colluders, and study how to guarantee that all colluders have the same probability of being detected accordingly.

A. Analysis of the Detection Statistics

To study the selection of collusion parameters in fair collusion, we first need to analyze the detection statistics and calculate each attacker's probability of being detected.

For each frame $j_1 \in F_b$ in the base layer, the extracted fingerprint $\mathbf{Y}_j$ can be rewritten as

$$\mathbf{Y}_{j_1} = \frac{\beta_1}{K^b}\sum_{i\in SC^b}\mathbf{W}_{j_1}^{(i)} + \frac{\beta_2}{K^{b,e1}}\sum_{i\in SC^{b,e1}}\mathbf{W}_{j_1}^{(i)} + \frac{\beta_3}{K^{\text{all}}}\sum_{i\in SC^{\text{all}}}\mathbf{W}_{j_1}^{(i)} + \mathbf{d}_{j_1} \tag{3}$$

where K^b, $K^{b,e1}$, and K^{all} are the number of colluders who receive copies of low, medium, and high resolution, respectively, and $\mathbf{d}_{j_1} = \mathbf{n}_{j_1}/JND_{j_1}$ is the detection noise. If the colluded copy contains frames in the enhancement layers, for each frame $j_2 \in F_{e1}$ in the enhancement layer 1

$$\mathbf{Y}_{j_2} = \frac{\alpha_1}{K^{b,e1}}\sum_{i\in SC^{b,e1}}\mathbf{W}_{j_2}^{(i)} + \frac{\alpha_2}{K^{\text{all}}}\sum_{i\in S^{\text{all}}}\mathbf{W}_{j_2}^{(i)} + \mathbf{d}_{j_2} \tag{4}$$

where $\mathbf{d}_{j_2} = \mathbf{n}_{j_2}/JND_{j_2}$ is the detection noise. If the colluded copy contains all three layers, for each frame $j_3 \in F_{e2}$ in the enhancement layer 2

$$\mathbf{Y}_{j_3} = \frac{1}{K^{\text{all}}}\sum_{i\in SC^{\text{all}}}\mathbf{W}_{j_3}^{(i)} + \mathbf{d}_{j_3} \tag{5}$$

where $\mathbf{d}_{j_3} = \mathbf{n}_{j_3}/JND_{j_3}$ is the detection noise.

With orthogonal fingerprint modulation as in Section II-B1, since the M originally embedded fingerprints are considered as known signals during fingerprint detection, under the assumption that the colluders have reasonably good estimates of JND_j and $\{\mathbf{d}_j\}_{j\in F^c}$ are i.i.d. Gaussian $\mathcal{N}\left(0,\sigma_n^2\right)$, it follows that given the colluder set SC, the detection statistics follow Gaussian distribution $p\left(TN^{(i)}|SC\right) \sim \mathcal{N}\left(\mu^{(i)},\sigma_n^2\right)$ [23]. $\mu^{(i)} = 0$ when user $\mathbf{u}^{(i)}$ is innocent, and $\mu^{(i)} > 0$ when $\mathbf{u}^{(i)}$ is guilty. For a guilty colluder $i \in SC$, $\mu^{(i)}$ depends on the number of frames in the colluded copy and the number frames that $\mathbf{u}^{(i)}$ receives.

1) $\mathbf{F^c = F_b \cup F_{e1} \cup F_{e2}}$: When the colluded copy contains all three layers, we can show that (see (6) at the bottom of the next page).
   Define $N_b = \sum_{j\in F_b} N_j$, $N_{e1} = \sum_{j\in F_{e1}} N_j$ and $N_{e2} = \sum_{j\in F_{e2}} N_j$ as the lengths of the fingerprints that are embedded in the base layer, enhancement layer 1 and enhancement layer 2, respectively. With orthogonal fingerprint modulation in Section II-B1, we have

$\sum_{j_1\in F_b}\left\|\mathbf{W}_{j_1}^{(i)}\right\|^2 = N_b\cdot\xi^2$, $\sum_{j_2\in F_{e1}}\left\|\mathbf{W}_{j_2}^{(i)}\right\|^2 = N_{e1}\cdot\xi^2$, and $\sum_{j_3\in F_{e2}}\|\mathbf{W}_{j_3}^{(i)}\|^2 = N_{e2}\cdot\xi^2$. Therefore,

$$\mu^{(i)}=\begin{cases}\frac{\beta_1\sqrt{N_b}}{K^b}\xi, & \text{if } i\in SC^b\\ \frac{\beta_2N_b+\alpha_1N_{e1}}{K^{b,e1}\sqrt{N_b+N_{e1}}}\xi, & \text{if } i\in SC^{b,e1}\\ \frac{\beta_3N_b+\alpha_2N_{e1}+N_{e2}}{K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}}\xi, & \text{if } i\in SC^{\text{all}}\end{cases}. \tag{7}$$

2) $\mathbf{F^c}=\mathbf{F_b}\cup\mathbf{F_{e1}}$: When the colluded copy contains frames in the base layer and the enhancement layer 1, similar to the above analysis,

$$\mu^{(i)}=\begin{cases}\frac{\beta_1\sqrt{N_b}}{K^b}\xi, & \text{if } i\in SC^b\\ \frac{\beta_2N_b+\alpha_1N_{e1}}{K^{b,e1}\sqrt{N_b+N_{e1}}}\xi, & \text{if } i\in SC^{b,e1}\\ \frac{\beta_3N_b+\alpha_2N_{e1}}{K^{\text{all}}\sqrt{N_b+N_{e1}}}\xi, & \text{if } i\in SC^{\text{all}}\end{cases}. \tag{8}$$

3) $\mathbf{F^c}=\mathbf{F_b}$: When the colluded copy contains frames in the base layer only, we have

$$\mu^{(i)}=\begin{cases}\frac{\beta_1\sqrt{N_b}}{K^b}\xi, & \text{if } i\in SC^b\\ \frac{\beta_2\sqrt{N_b}}{K^{b,e1}}\xi, & \text{if } i\in SC^{b,e1}\\ \frac{\beta_3\sqrt{N_b}}{K^{\text{all}}}\xi, & \text{if } i\in SC^{\text{all}}\end{cases}. \tag{9}$$

B. Selection of the Collusion Parameters

With the above analysis of the detection statistics, given a threshold h, for colluder $\mathbf{u}^{(i)}$ whose detection statistics follow distribution $\mathcal{N}\left(\mu^{(i)},\sigma_n^2\right)$, the probability that $\mathbf{u}^{(i)}$ is captured is $P^{(i)}=P\left[TN^{(i)}>h\right]=Q\left((h-\mu^{(i)})/\sigma_n\right)$, where $Q(x)=\int_x^\infty(1/\sqrt{2\pi})e^{-(t^2/2)dt}$ is the Gaussian tail function. Therefore, all colluders share the same risk and are equally likely to be detected if and only if their detection statistics have the same mean.

1) $\mathbf{F^c}=\mathbf{F_b}\cup\mathbf{F_{e1}}\cup\mathbf{F_{e2}}$: When the colluded copy contains frames in all three layers, from (7), the colluders seek $\{0\le\beta_k\le1\}_{k=1,2,3}$ and $\{0\le\alpha_l\le1\}_{l=1,2}$ to satisfy

$$\begin{aligned}\frac{\beta_1\sqrt{N_b}}{K^b}\xi&=\frac{\beta_2N_b+\alpha_1N_{e1}}{K^{b,e1}\sqrt{N_b+N_{e1}}}\xi\\&=\frac{\beta_3N_b+\alpha_2N_{e1}+N_{e2}}{K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}}\xi,\\ \text{s.t.}\quad&\beta_1+\beta_2+\beta_3=1,\quad\alpha_1+\alpha_2=1.\end{aligned} \tag{10}$$

Note that

$$\begin{aligned}\frac{\beta_1\sqrt{N_b}}{K^b}\xi=\frac{\beta_2N_b+\alpha_1N_{e1}}{K^{b,e1}\sqrt{N_b+N_{e1}}}\xi&\Longleftrightarrow\frac{K^{b,e1}\sqrt{N_b+N_{e1}}}{K^b\sqrt{N_b}}\\&=\frac{\beta_2N_b+\alpha_1N_{e1}}{\beta_1N_b}.\end{aligned} \tag{11}$$

In addition, let $\beta_3=1-\beta_1-\beta_2$ and $\alpha_2=1-\alpha_1$, we have

$$\begin{aligned}\frac{\beta_1\sqrt{N_b}}{K^b}\xi&=\frac{\beta_3N_b+\alpha_2N_{e1}+N_{e2}}{K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}}\xi\\&\Longleftrightarrow\frac{K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}}{K^b\sqrt{N_b}}\\&=\frac{N_b+N_{e1}+N_{e2}}{\beta_1N_b}\\&\quad-1-\frac{\beta_2N_b+\alpha_1N_{e1}}{\beta_1N_b}.\end{aligned} \tag{12}$$

Plugging (11) into (12), we have

$$\begin{aligned}\frac{N_b+N_{e1}+N_{e2}}{\beta_1N_b}&=\frac{K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}}{K^b\sqrt{N_b}}+1\\&\quad+\frac{K^{b,e1}\sqrt{N_b+N_{e1}}}{K^b\sqrt{N_b}}.\end{aligned} \tag{13}$$

Therefore, from (11) and (13), the colluders should choose (14), shown at the bottom of the next page.

$$\mu^{(i)}=\begin{cases}\frac{\beta_1}{K^b}\sqrt{\sum_{j\in F^{(i)}}\left\|\mathbf{W}_j^{(i)}\right\|^2}, & \text{if } i\in SC^b\\ \frac{\beta_2\sum_{j\in F_b}\left\|\mathbf{W}_j^{(i)}\right\|^2+\alpha_1\sum_{j\in F_{e1}}\left\|\mathbf{W}_j^{(i)}\right\|^2}{K^{b,e1}\sqrt{\sum_{j\in F^{(i)}}\left\|\mathbf{W}_j^{(i)}\right\|^2}}, & \text{if } i\in SC^{b,e1}\\ \frac{\beta_3\sum_{j\in F_b}\left\|\mathbf{W}_j^{(i)}\right\|^2+\alpha_2\sum_{j\in F_{e1}}\left\|\mathbf{W}_j^{(i)}\right\|^2+\sum_{j\in F_{e2}}\left\|\mathbf{W}_j^{(i)}\right\|^2}{K^{\text{all}}\sqrt{\sum_{j\in F^{(i)}}\left\|\mathbf{W}_j^{(i)}\right\|^2}}, & \text{if } i\in SC^{\text{all}}\end{cases} \tag{6}$$

From Section II-B2, the collusion parameters are required to be in the range of $[0,1]$. From (14), $0 \le \beta_1 \le 1$ if and only if

$$\frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}} \le \frac{N_b}{N_b+N_{e1}+N_{e2}}. \quad (15)$$

Furthermore, from (14), see (16) at the bottom of the page. Given β_1 as in (14), $0 \le \beta_2 \le 1-\beta_1$. Consequently, from (16), we have $\underline{\alpha} \le \alpha_1 \le \overline{\alpha}$, where (see (17) at the bottom of the page). If $[0,1]\cap[\underline{\alpha},\overline{\alpha}]$ is not empty, then there exists at least one α_1^* such that $0 \le \alpha_1^* \le 1$ and $\underline{\alpha} \le \alpha_1^* \le \overline{\alpha}$. Note that $\overline{\alpha} > 0$, so $[0,1]\cap[\underline{\alpha},\overline{\alpha}] \ne \emptyset$ if and only if $\underline{\alpha} \le 1$, which is equivalent to

$$\frac{K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}} \ge \frac{N_{e2}}{N_b+N_{e1}+N_{e2}}. \quad (18)$$

To summarize, in order to generate a colluded copy with the highest temporal resolution under the fairness constraints, $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$ and (N_b, N_{e1}, N_{e2}) have to satisfy (15) and (18), and the colluders should choose the collusion parameters as in (14).

2) $\mathbf{F^c} = \mathbf{F_b} \cup \mathbf{F_{e1}}$: In this scenario, the colluded copy has medium resolution and contains frames in the base layer and the enhancement layer 1. For colluder $\mathbf{u}^{\left(i_1 \in SC^{\text{all}}\right)}$ and colluder $\mathbf{u}^{\left(i_2 \in SC^{b,e1}\right)}$ who receive copies of the highest and the medium resolution, respectively, the overall lengths of their fingerprints in the colluded copy are the same and equal to $N_b + N_{e1}$. In this scenario, the collusion attacks among colluders in subgroup $SC^{b,e1}$ and SC^{all} are the same as in the simple case in [13] where all attackers receive copies of the same resolution. Therefore, during the intergroup collusion in Fig. 3, $\mathbf{u}^{(i_1)}$ and $\mathbf{u}^{(i_2)}$ let $\alpha_1/K^{b,e1} = \alpha_2/K^{\text{all}}$ and $\beta_2/K^{b,e1} = \beta_3/K^{\text{all}}$. Such a parameter selection not only guarantees $\mu^{(i_1)} = \mu^{(i_2)}$, but also ensures that for each frame j in the colluded copy, the energies of these two colluders' fingerprints $\mathbf{W}_j^{(i_1)}$ and $\mathbf{W}_j^{(i_2)}$ are reduced by the same ratio. For a given $0 \le \beta_1 \le 1$, it is equivalent to

$$\begin{aligned} \alpha_1 &= \frac{K^{b,e1}}{K^{b,e1}+K^{\text{all}}}, \quad \alpha_2 = 1-\alpha_1 \\ \beta_2 &= \frac{K^{b,e1}}{K^{b,e1}+K^{\text{all}}}(1-\beta_1) \\ \text{and } \beta_3 &= 1-\beta_1-\beta_2. \end{aligned} \quad (19)$$

With the above selected parameters, for colluder $\mathbf{u}^{\left(i_1 \in SC^{\text{all}}\right)}$ and colluder $\mathbf{u}^{\left(i_2 \in SC^{b,e1}\right)}$

$$\mu^{(i_1)} = \mu^{(i_2)} = \frac{(1-\beta_1)N_b + N_{e1}}{(K^{b,e1}+K^{\text{all}})\sqrt{N_b+N_{e1}}}\xi. \quad (20)$$

The colluders seek $0 \le \beta_1 \le 1$ such that

$$\frac{\beta_1\sqrt{N_b}}{K^b}\xi = \frac{(1-\beta_1)N_b + N_{e1}}{(K^{b,e1}+K^{\text{all}})\sqrt{N_b+N_{e1}}}\xi \quad (21)$$

$$\begin{aligned} \beta_1 &= \frac{N_b+N_{e1}+N_{e2}}{N_b}\frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}} \\ \text{and } \beta_2 N_b + \alpha_1 N_{e1} &= \frac{(N_b+N_{e1}+N_{e2})\,K^{b,e1}\sqrt{N_b+N_{e1}}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}} \end{aligned} \quad (14)$$

$$\alpha_1 = \frac{N_b+N_{e1}+N_{e2}}{N_{e1}}\frac{K^{b,e1}\sqrt{N_b+N_{e1}}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}} - \beta_2\frac{N_b}{N_{e1}} \quad (16)$$

$$\begin{aligned} \underline{\alpha} &= \frac{N_b+N_{e1}+N_{e2}}{N_{e1}}\frac{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}} - \frac{N_b}{N_e1} \\ \text{and } \overline{\alpha} &= \frac{N_b+N_{e1}+N_{e2}}{N_{e1}}\frac{K^{b,e1}\sqrt{N_b+N_{e1}}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}} \end{aligned} \quad (17)$$

TABLE I
CONSTRAINTS AND THE SELECTION OF COLLUSION PARAMETERS DURING COLLUSION TO ACHIEVE FAIRNESS

$F^c = F_b \cup F_{e1} \cup F_{e2}$ (Highest resolution)	Fairness Constraints	$\begin{cases} \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{all}\sqrt{N_b+N_{e1}+N_{e2}}} \le \frac{N_b}{N_b+N_{e1}+N_{e2}}, \\ \frac{K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{all}\sqrt{N_b+N_{e1}+N_{e2}}} \ge \frac{N_{e2}}{N_b+N_{e1}+N_{e2}}. \end{cases}$
	Parameter Selection	$\begin{cases} \beta_1 = \frac{N_b+N_{e1}+N_{e2}}{N_b} \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}, \\ \beta_2 N_b + \alpha_1 N_{e1} = \frac{(N_b+N_{e1}+N_{e2})K^{b,e1}\sqrt{N_b+N_{e1}}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}, \\ \beta_3 = 1-\beta_1-\beta_2,\ \alpha_2 = 1-\alpha_1. \end{cases}$
$F^c = F_b \cup F_{e1}$ (Medium resolution)	Fairness Constraints	$\frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b}+(K^{b,e1}+K^{all})\sqrt{N_b+N_{e1}}} \le \frac{N_b}{N_b+N_{e1}}$.
	Parameter Selection	$\begin{cases} \beta_1 = \frac{N_b+N_{e1}}{N_b} \frac{K_b\sqrt{N_b}}{K^b\sqrt{N_b}+(K^{b,e1}+K^{all})\sqrt{N_b+N_{e1}}}, \\ \beta_2 = \frac{K^{b,e1}}{K^{b,e1}+K^{all}}(1-\beta_1),\ \beta_3 = 1-\beta_1-\beta_2, \\ \alpha_1 = \frac{K^{b,e1}}{K^{b,e1}+K^{all}},\ \alpha_2 = 1-\alpha_1. \end{cases}$
$F^c = F_b$ (Lowest resolution)	Fairness Constraints	No constraints on $(K^b, K^{b,e1}, K^{all})$ and (N_b, N_{e1}, N_{e2}).
	Parameter Selection	$\beta_1 = \frac{K^b}{K^b+K^{b,e1}+K^{all}},\ \beta_2 = \frac{K^{b,e1}}{K^b+K^{b,e1}+K^{all}},\ \beta_3 = \frac{K^{all}}{K^b+K^{b,e1}+K^{all}}$.

and the solution is

$$\beta_1 = \frac{N_b+N_{e1}}{N_b} \frac{K_b\sqrt{N_b}}{K^b\sqrt{N_b}+(K^{b,e1}+K^{\text{all}})\sqrt{N_b+N_{e1}}}. \tag{22}$$

With β_1 as in (22), $0 \le \beta_1 \le 1$ if and only if

$$\frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b}+(K^{b,e1}+K^{\text{all}})\sqrt{N_b+N_{e1}}} \le \frac{N_b}{N_b+N_{e1}}. \tag{23}$$

Given $0 \le \beta_1 \le 1$, from (19), it is straightforward to show that $0 \le \beta_2, \beta_3, \alpha_1, \alpha_2 \le 1$.

To summarize, under the fairness constraints, $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$ and (N_b, N_{e1}, N_{e2}) have to satisfy (23) if the colluders wish to generate a colluded copy of medium temporal resolution. The colluders should choose the collusion parameters as in (19) and (22).

3) $\mathbf{F^c = F_b}$: When the colluded copy contains frames in the base layer only, the colluders choose $\{0 \le \beta_k \le 1\}_{k=1,2,3}$ with $\beta_1+\beta_2+\beta_3 = 1$ to satisfy

$$\frac{\beta_1\sqrt{N_b}}{K^b}\xi = \frac{\beta_2\sqrt{N_b}}{K^{b,e1}}\xi = \frac{\beta_3\sqrt{N_b}}{K^{\text{all}}}\xi \tag{24}$$

and the solution is

$$\begin{aligned} \beta_1 &= \frac{K^b}{K^b+K^{b,e1}+K^{\text{all}}} \\ \beta_2 &= \frac{K^{b,e1}}{K^b+K^{b,e1}+K^{\text{all}}} \\ \text{and } \beta_3 &= \frac{K^{\text{all}}}{K^b+K^{b,e1}+K^{\text{all}}}. \end{aligned} \tag{25}$$

In this scenario, there are no constraints on $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$ and (N_b, N_{e1}, N_{e2}), and the colluders can always generate a colluded copy containing frames in the base layer only.

C. Summary of the Parameter Selection to Achieve Fairness During Collusion

Table I summarizes the constraints and the parameter selection during collusion to ensure fairness in three scenarios, where the colluded copy has the highest, medium and lowest temporal resolution, respectively. From Table I, if the colluders want to generate a colluded copy of higher resolution, the constraints are more severe in order to distribute the risk of being detected evenly among all attackers.

Note that to select the collusion parameters, the colluders need to estimate $N_b : N_{e1} : N_{e2}$, the ratio of the lengths of the fingerprints embedded in different layers. Since adjacent frames in a video sequence are similar to each other and have approximately the same number of embeddable coefficients, the colluders can use the following approximation $N_b : N_{e1} : N_{e2} \approx |F_b| : |F_{e1}| : |F_{e2}|$.

IV. EFFECTIVENESS OF FAIR COLLUSION IN UNDERMINING THE TRAITOR TRACING CAPABILITY

In this section, we investigate the effectiveness of collusion in defeating the scalable fingerprinting systems, assuming that the attackers choose the collusion parameters as in Table I.

A. Statistical Analysis

Assume that there are a total of M users. From the analysis in the previous section, if the colluders select the collusion parameters as in Table I, then given a colluder set SC, for each user $\mathbf{u}^{(i)}$

$$p\left(TN^{(i)}|SC\right) \sim \begin{cases} \mathcal{N}\left(\mu, \sigma_n^2\right), & \text{if } i \in SC \\ \mathcal{N}\left(0, \sigma_n^2\right), & \text{if } i \notin SC \end{cases} \tag{26}$$

where σ_n^2 is the variance of the detection noise $\mathbf{d}_j$, and the M detection statistics $\{TN^{(i)}\}_{i=1,\ldots,M}$ are independent of each

other due to the orthogonality of the fingerprints. In addition, for $i \in SC$, see (27) at the bottom of the page. Note that

$$\begin{aligned}&\frac{N_b+N_{e1}}{K^b\sqrt{N_b}+(K^{b,e1}+K^{\text{all}})\sqrt{N_b+N_{e1}}}\\&\quad=\frac{\sqrt{N_b+N_{e1}}}{K^b\sqrt{\frac{N_b}{N_b+N_{e1}}}+K^{b,e1}+K^{\text{all}}}\\&\quad\geq\frac{\sqrt{N_b+N_{e1}}}{K^b+K^{b,e1}+K^{\text{all}}}\geq\frac{\sqrt{N_b}}{K^b+K^{b,e1}+K^{\text{all}}}.\end{aligned}\tag{28}$$

Similarly, we can also show that

$$\begin{aligned}&\frac{N_b+N_{e1}+N_{e2}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}}\\&\qquad\geq\frac{N_b+N_{e1}}{K^b\sqrt{N_b}+(K^{b,e1}+K^{\text{all}})\sqrt{N_b+N_{e1}}}.\end{aligned}\tag{29}$$

Therefore, under the fairness constraints, μ in (27) is larger when the colluded copy has higher resolution.

Given a threshold h, from (26), we can have

$$\begin{aligned}P_d&=P\left[\max_{i\in SC}TN^{(i)}>h\right]\\&=1-\left[1-Q\left(\frac{h-\mu}{\sigma_n}\right)\right]^K\\P_{fp}&=P\left[\max_{i\notin SC}TN^{(i)}>h\right]\\&=1-\left[1-Q\left(\frac{h}{\sigma_n}\right)\right]^{M-K}\\E[F_d]&=\frac{\sum_{i\in SC}P\left[TN^{(i)}>h\right]}{K}\\&=Q\left(\frac{h-\mu}{\sigma_n}\right)\\\text{and }E[F_{fp}]&=\frac{\sum_{i\notin SC}P\left[TN^{(i)}>h\right]}{(M-K)}\\&=Q\left(\frac{h}{\sigma_n}\right).\end{aligned}\tag{30}$$

From (27) and (30), the effectiveness of fair collusion in defeating the scalable fingerprinting systems depends on the total number of colluders K as well as the temporal resolution of the colluded copy L^c. For a fixed resolution of the colluded copy $L^c=|F^c|$, when there are more colluders in the systems, the colluders are less likely to be captured and the collusion attack is more effective. For a fixed total number of colluders K, when the colluded copy has a higher resolution, the extracted fingerprint is longer and provides more information of the colluders' identities to the detector. Therefore, the colluders have a larger probability of being detected. During collusion, the colluders have to take into consideration the tradeoff between the risk of being detected and the resolution of the colluded copy.

B. Simulation Results With Ideal Gaussian Models

When simulating the scalable fingerprinting systems and collusion attacks using ideal Gaussian models, we test on a total of 40 frames as an example. Following the example in Section II-A, we consider a temporally scalable coding system where frame $F_b=\{1,5,\ldots,37\}$ are encoded in the base layer, frame $F_{e1}=\{3,7,\ldots,39\}$ are in the enhancement layer 1, and the enhancement layer 2 consists of frame $F_{e2}=\{2,4,\ldots,40\}$. For user $\mathbf{u}^{(i_1)}\in\mathbf{U}^b$, he receives the base layer only and reconstructs a fingerprinted copy of 10 frames including frame $1,5,\ldots$, and frame 37. For user $\mathbf{u}^{(i_2)}\in\mathbf{U}^{b,e1}$ who receives the base layer and the enhancement layer 1, his fingerprinted copy includes all the 20 odd frames. User $\mathbf{u}^{(i_3)}\in\mathbf{U}^{\text{all}}$ subscribes to all three layers and receives a fingerprinted copy of all 40 frames.

From the human visual models [15], not all coefficients are embeddable due to imperceptibility constraints. For real video sequences like "akiyo" and "carphone," the number of embeddable coefficients in each frame varies from 3000 to 7000, depending on the characteristics of the video sequences. In our simulations, we assume that the length of the fingerprints embedded in each frame is 5000, and the lengths of the fingerprints embedded in the base layer, enhancement layer 1 and enhancement layer 2 are $N_b=50\,000$, $N_{e1}=50\,000$ and $N_{e2}=100\,000$, respectively. We assume that there are a total of $M=450$ users and $|\mathbf{U}^b|=|\mathbf{U}^{b,e1}|=|\mathbf{U}^{\text{all}}|=150$. We first generate independent vectors following Gaussian distribution $\mathcal{N}\left(0,\sigma_W^2\right)$ with $\sigma_W^2=1/9$, and then apply Gram–Schmidt orthogonalization to produce fingerprints that satisfy (1). In each fingerprinted copy, fingerprints embedded in adjacent frames are correlated with each other.

We assume that $0\leq K^b, K^{b,e1}, K^{\text{all}}\leq 150$ are the number of colluders in subgroups SC^b, $SC^{b,e1}$ and SC^{all}, respectively. During collusion, the colluders apply the intragroup collusion followed by the intergroup collusion as in Fig. 3. Furthermore, we assume that the detection noise follows Gaussian distribution with zero mean and variance $\sigma_n^2=2\sigma_W^2$.

In Fig. 4, we fix the ratio $K^b:K^{b,e1}:K^{\text{all}}=1:1:1$, and assume that the colluded copy has medium resolution and in-

$$\mu=\frac{\beta_1\sqrt{N_b}}{K^b}\xi=\begin{cases}\frac{N_b+N_{e1}+N_{e2}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{\text{all}}\sqrt{N_b+N_{e1}+N_{e2}}}\xi, & \text{if } F^c=F_b\cup F_{e1}\cup F_{e2}\\\frac{N_b+N_{e1}}{K^b\sqrt{N_b}+(K^{b,e1}+K^{\text{all}})\sqrt{N_b+N_{e1}}}\xi, & \text{if } F^c=F_b\cup F_{e1}\\\frac{\sqrt{N_b}}{K^b+K^{b,e1}+K^{\text{all}}}\xi, & \text{if } F^c=F_b\end{cases}\tag{27}$$

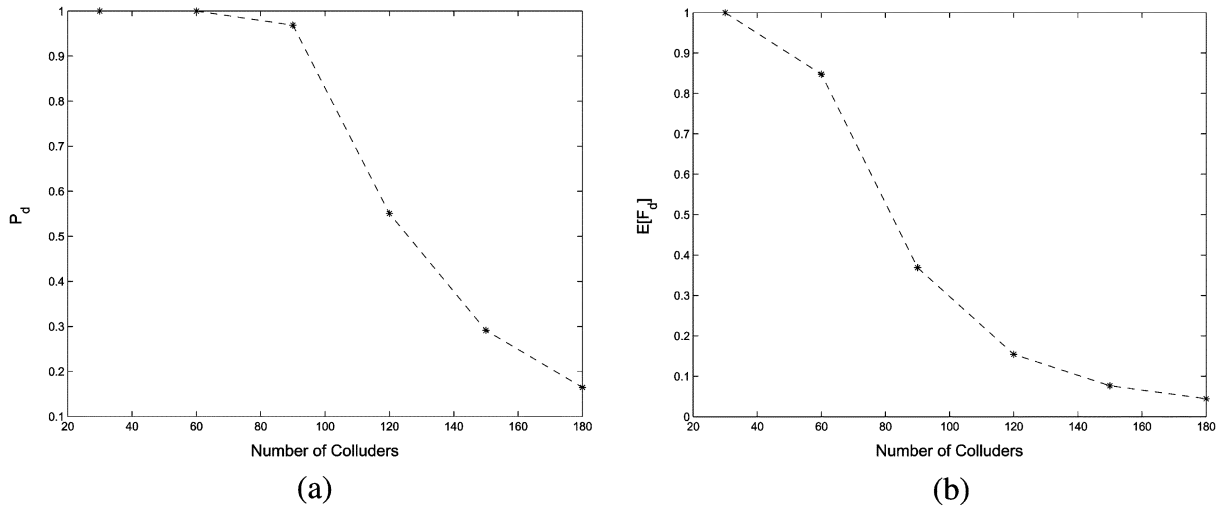

Fig. 4. Effectiveness of the collusion attacks on scalable fingerprinting systems. Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{\text{all}}| = 150$. $N_b = 50\,000$, $N_{e1} = 50\,000$ and $N_{e2} = 100\,000$. $K^b : K^{b,e1} : K^{\text{all}} = 1 : 1 : 1$ and $F^c = F_b \cup F_{e1}$. $\sigma_n^2/\sigma_W^2 = 2$. $P_{fp} = 10^{-3}$ in (a), and $E[F_{fp}] = 10^{-3}$ in (b).

cludes all the 20 odd frames. In Fig. 4(a), we select the threshold h to fix the probability of accusing at least one innocent user as 10^{-3} and plot the probability of capturing at least one colluder P_d when the total number of colluders K increases. In Fig. 4(b), $E[F_{fp}] = 10^{-3}$ and we plot the expected fraction of the colluders that are captured when K increases. From Fig. 4, the collusion is more effective in removing traces of the fingerprints when there are more colluders.

We then fix the total number of colluders $K = 150$, and compare the effectiveness of the collusion attacks when the temporal resolution of the colluded copy changes. Define the lines $\overline{AB}$ and $\overline{CD}$ as (see (31) and (32) at the bottom of the page), respectively, as shown in Fig. 5(a). Line $\overline{AB}$ and Line $\overline{CD}$ are the boundaries of the two constraints to achieve fairness, respectively, when generating an attacked copy of the highest resolution. For a fixed $K = 150$, we study the effectiveness of collusion when $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$ takes different values on Line $\overline{AB}$ and Line $\overline{CD}$, respectively. In our simulations, we assume that the colluders generate a colluded copy of the highest possible resolution under the constraints in Table I. Fig. 5(b) plots the regions where the colluders can generate a colluded copy of high resolution and regions where the colluders can generate a medium resolution copy under the fairness constraints in Table I.

Fig. 6 shows the simulation results when $K = 150$ is fixed and $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$ takes different values on Line $\overline{AB}$ (31). In Fig. 6, a given value of K^{all} corresponds to a unique point on Line $\overline{AB}$ and, therefore, a unique triplet $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$. Fig. 6(a) shows the number of frames in the colluded copy L^c. $L^c = 20$ when the attacked copy has medium resolution and $L^c = 40$ when attackers generate a copy including all three layers. Fig. 6(b) shows the means of the detection statistics of the guilty colluders. In Fig. 6(c), we select the threshold used to fix $P_{fp} = 10^{-3}$ and we compare P_d of the collusion attacks when the triplet $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$ takes different values on Line $\overline{AB}$. In Fig. 6(d), $E[F_{fp}] = 10^{-3}$ by selecting the threshold in the simulation runs and we compare $E[F_d]$ of the fair collusion for different triplets $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$ on Line $\overline{AB}$.

Similarly, Fig. 7 shows the simulation results when K is fixed as 150 and $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$ moves on Line $\overline{CD}$ (32). In Fig. 7, each K^b represents one point on Line $\overline{CD}$ and a unique $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$. Fig. 7(a) plots the total number of frames in

$$\overline{AB} \triangleq \left\{ \left(K^b, K^{b,e1}, K^{\text{all}}\right) : \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}} = \frac{N_b}{N_b + N_{e1} + N_{e2}}, \right.$$
$$\left. 0 \le K^b \le |\mathbf{U}^b|, 0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|, 0 \le K^{\text{all}} \le |\mathbf{U}^{\text{all}}|, K^b + K^{b,e1} + K^{\text{all}} = K \right\} \quad \text{and} \tag{31}$$

$$\overline{CD} \triangleq \left\{ \left(K^b, K^{b,e1}, K^{\text{all}}\right) : \frac{K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}} = \frac{N_{e2}}{N_b + N_{e1} + N_{e2}} \right.$$
$$\left. 0 \le K^b \le |\mathbf{U}^b|, 0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|, 0 \le K^{\text{all}} \le |\mathbf{U}^{\text{all}}|, K^b + K^{b,e1} + K^{\text{all}} = K \right\} \tag{32}$$

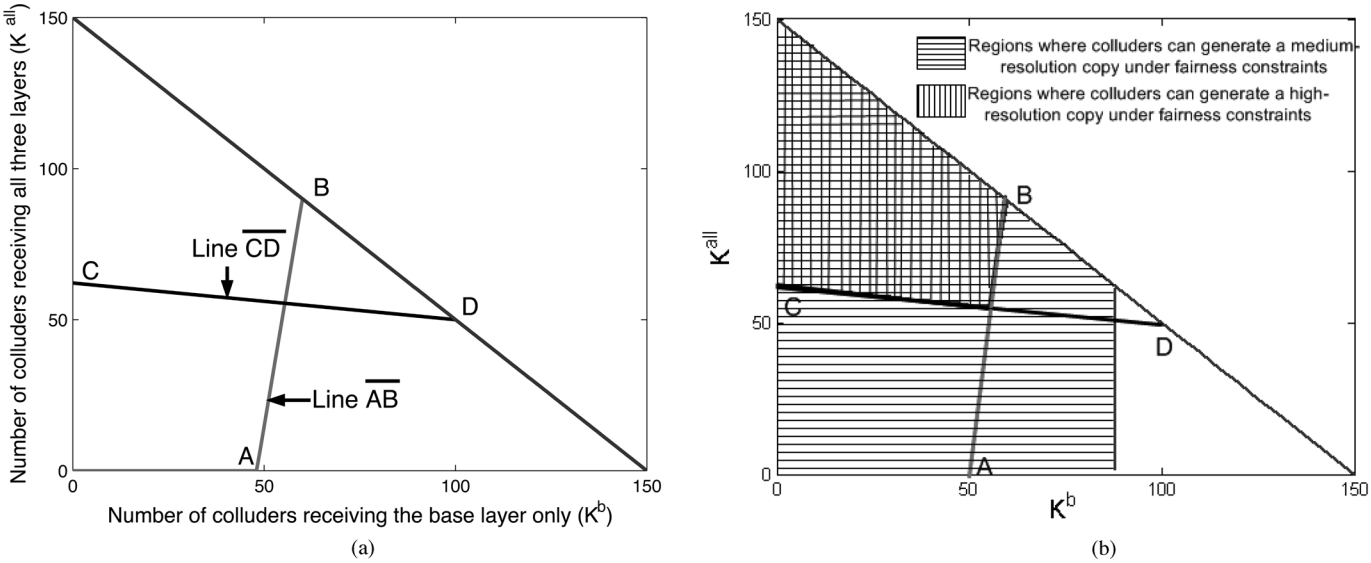

Fig. 5. (a) Line $\overline{AB}$ of (31) and Line $\overline{CD}$ of (32), and (b) regions where colluders can generate a medium-resolution or a high-resolution copy while still ensuring fairness of collusion. Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{\text{all}}| = 150$. $(N_b, N_{e1}, N_{e2}) = (50\,000, 50\,000, 100\,000)$. The total number of colluders is fixed as $K = 150$. The x axis is the number of colluders who receive the base layer only, and the y axis is the number of colluders who receive all three layers. Each point in the figure represents a unique triplet $(K^b, K^{b,e1}, K^{\text{all}})$ with $K^{b,e1} = K - K^b - K^{\text{all}}$.

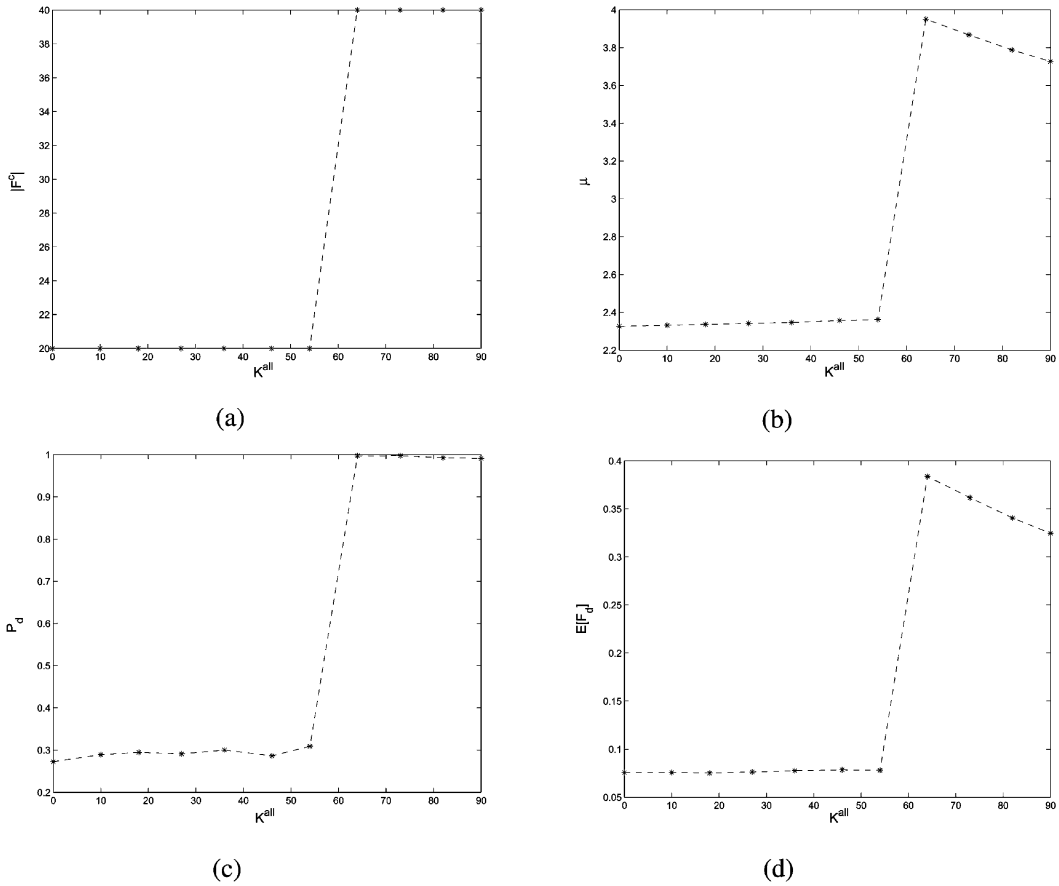

Fig. 6. Simulation results of fair collusion when $(K^b, K^{b,e1}, K^{\text{all}})$ takes different values on Line $\overline{AB}$ (31). The x axis is the number of colluders who receive all three layers K^{all}. Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{\text{all}}| = 150$. $(N_b, N_{e1}, N_{e2}) = (50\,000, 50\,000, 100\,000)$. The total number of colluders is fixed as $K = 150$. $\sigma_n^2/\sigma_W^2 = 2$. $P_{fp} = 10^{-3}$ in (c), and $E[F_{fp}] = 10^{-3}$ in (d).

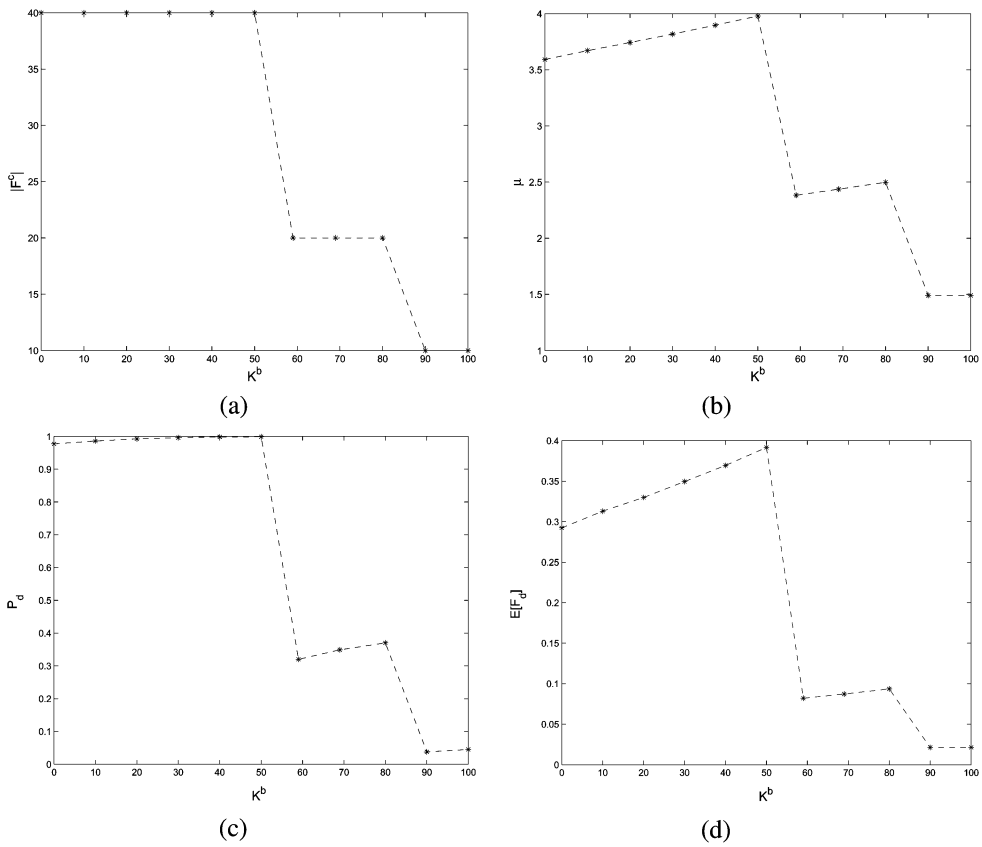

Fig. 7. Simulation results of fair collusion when $(K^b, K^{b,e1}, K^{\text{all}})$ takes different values on Line $\overline{CD}$ (32). The x axis is the number of colluders who receive the base layer only K^b. $M = 450$ and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{\text{all}}| = 150$. $(N_b, N_{e1}, N_{e2}) = (50\,000, 50\,000, 100\,000)$. $K = 150$. $\sigma_n^2/\sigma_W^2 = 2$. $P_{fp} = 10^{-3}$ in (c), and $E[F_{fp}] = 10^{-3}$ in (d).

the colluded copy. $L^c = 10$, $L^c = 20$, and $L^c = 40$ correspond to the scenario where the colluded copy has low, medium, and high resolution, respectively. Fig. 7(b) shows the mean of the guilty colluders' detection statistics. In Fig. 7(c), P_{fp} is fixed as 10^{-3} and we compare P_d when $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$ moves from left to right on Line $\overline{CD}$. Fig. 7(d) fixes $E[F_{fp}] = 10^{-3}$ and plots $E[F_d]$ for different $\left(K^b, K^{b,e1}, K^{\text{all}}\right)$ on Line $\overline{CD}$.

From Figs. 6 and 7, when the colluded copy has higher temporal resolution, the attacked copy contains more information of the attackers' fingerprints, and the colluders have a larger probability to be captured. It is in agreement with our statistical analysis in Section IV-A. The colluders have to consider the tradeoff between the probability of being detected and the resolution of the attacked copy during collusion.

Note that from Figs. 6 and 7, if we fix the total number of colluders K and the resolution of the colluded copy $L^c = |F^c|$, P_d and $E[F_d]$ have larger values when K^b is smaller (or equivalently, when K^{all} is larger). This is because, with fixed $K = K^b + K^{b,e1} + K^{\text{all}}$ and fixed $F^c = F_b \cup F_{e1}$, from (27)

$$\begin{aligned}\mu &= \frac{N_b + N_{e1}}{K^b\sqrt{N_b} + (K^{b,e1} + K^{\text{all}})\sqrt{N_b + N_{e1}}}\xi \\ &= \frac{N_b + N_{e1}}{K^b\sqrt{N_b} + (K - K^b)\sqrt{N_b + N_{e1}}}\xi \\ &= \frac{N_b + N_{e1}}{\sqrt{N_b + N_{e1}} + K^b(\sqrt{N_b} - \sqrt{N_b + N_{e1}})}\xi \end{aligned} \quad (33)$$

is an increasing function of K^b. Therefore, μ takes larger values when K^b increases, and the fair collusion attacks are less effective. The analysis is similar with fixed $F^c = F_b \cup F_{e1} \cup F_{e2}$ and fixed K.

V. Resistance of the Scalable Fingerprinting Systems to Collusion Attacks

Analysis of the collusion attacks helps evaluate the traitor tracing capacity of digital fingerprinting systems, and provide guidance to the digital rights enforcers on the design of collusion resistant fingerprinting systems [9], [10], [24]. In this section, we analyze the collusion resistance of the scalable fingerprinting systems in Section II-B, and quantify the traitor tracing capacity by studying $K_{\max}$, the maximum number of colluders that the fingerprinting systems can successfully resist under the system requirements.

A. Catch One

In the catch one scenario, the fingerprinting systems wish to maximize the chance to capture one colluder while minimizing the probability of falsely accusing an innocent user. An example of such a scenario is to provide trustworthy digital evidence in the court of law. The performance criteria in this scenario are the probability of capturing at least one colluder P_d and the probability of accusing at least one innocent user P_{fp}. From the detector's point of view, the detector fails if either it does

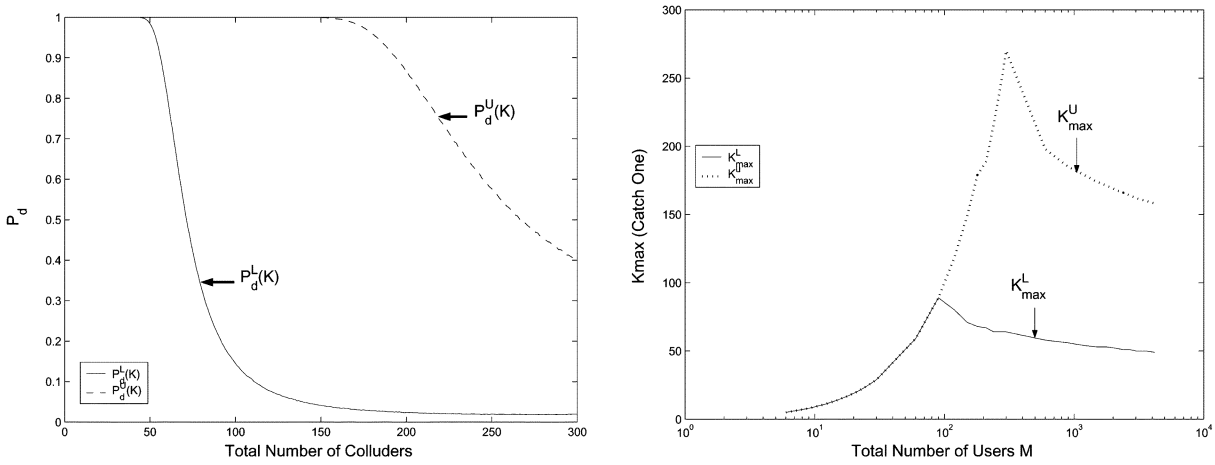

Fig. 8. Collusion resistance in the catch one scenario. $|\mathbf{U}^b| : |\mathbf{U}^{b,e1}| : |\mathbf{U}^{\text{all}}| = 1 : 1 : 1$ and $(N_b, N_{e1}, N_{e2}) = (50\,000, 50\,000, 100\,000)$. $\sigma_n^2/\xi^2 = 2$. $\gamma_d = 0.8$ and $\gamma_{fp} = 10^{-3}$. In (a), there are a total of $M = 450$ users in the system and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{\text{all}}| = 150$. We plot P_d^U and P_d^L versus the total number of colluders K. (b) illustrates $K_{\max}^U$ and $K_{\max}^L$ versus the total number of users M.

not capture any of the colluders or it falsely accuses an innocent user as a colluder. Consequently, the system requirements are

$$P_d \geq \gamma_d \text{ and } P_{fp} \leq \gamma_{fp}. \tag{34}$$

1) Upper and Lower Bounds of $k_{\max}$: To quantify the collusion resistance of the scalable fingerprinting system in Section II-B and analyze $K_{\max}$, we first need to analyze P_d and P_{fp}. From (27) and (30), if we fix the probability of accusing at least one innocent user $P_{fp} = \gamma_{fp}$, given the system parameters $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{\text{all}}|)$ and (N_b, N_{e1}, N_{e2}), the performance of the detector in Section II-B3 depends on the number of colluders in different subgroups $(K^b, K^{b,e1}, K^{\text{all}})$ and the temporal resolution of the colluded copy L^c. For a fixed total number of colluders K, we define

$$\begin{aligned} & P_d^U(K) \triangleq \max_{L^c,(K^b,K^{b,e1},K^{\text{all}})} P_d \\ \text{s.t.} \quad & K^b + K^{b,e1} + K^{\text{all}} = K, \quad 0 \leq K^b \leq |\mathbf{U}^b| \\ & 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|, \quad 0 \leq K^{\text{all}} \leq |\mathbf{U}^{\text{all}}| \\ & \text{fairness constraints in Table I are satisfied;} \end{aligned} \tag{35}$$

$$\begin{aligned} \text{and} \quad & P_d^L(K) \triangleq \min_{L^c,(K^b,K^{b,e1},K^{\text{all}})} P_d \\ \text{s.t.} \quad & K^b + K^{b,e1} + K^{\text{all}} = K, \quad 0 \leq K^b \leq |\mathbf{U}^b| \\ & 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|, \quad 0 \leq K^{\text{all}} \leq |\mathbf{U}^{\text{all}}| \\ & \text{fairness constraints in Table I are satisfied.} \end{aligned} \tag{36}$$

P_d reaches the upper bound $P_d^U(k)$ when the colluders generate a colluded copy of the highest resolution; while P_d is equal to $P_d^L(K)$ when the colluded copy contains the base layer only. Fig. 8(a) shows an example of $P_d^U(K)$ and $P_d^L(K)$ when there are a total of $M = 450$ users and $\gamma_{fp} = 10^{-3}$. From Fig. 8(a), the fingerprinting system's performance degrades when K becomes larger. Under the requirements $P_d \geq 0.8$ and $P_{fp} \leq 10^{-3}$, we can see from Fig. 8(a) that when the total number of colluders is larger than 210, $P_d^U(K) < 0.8$ and the fingerprinting systems will always fail no matter which resolution the colluded copy has. When there are fewer than 60 attackers, $P_d^L(k) \geq 0.8$ and the colluders can never bypass the detector without being detected, even if they only generate a colluded copy of low resolution.

In the catch one scenario, given the system parameters $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{\text{all}}|)$ and the total number of users M, we further define

$$\begin{aligned} K_{\max}^U &\triangleq \arg\max_K \left\{P_d^U(K) \geq \gamma_d\right\} \\ \text{and } K_{\max}^L &\triangleq \arg\max_K \left\{P_d^L(K) \geq \gamma_d\right\}. \end{aligned} \tag{37}$$

Given the parameters $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{\text{all}}|)$ and (N_b, N_{e1}, N_{e2}), when the total number of colluders K is smaller than $K_{\max}^L$, no matter what values L^c and $(K^b, K^{b,e1}, K^{\text{all}})$ take, the system requirements of (34) are always satisfied. On the contrary, if the total number of colluders K is larger than $K_{\max}^U$, for all possible values of L^c and $(K^b, K^{b,e1}, K^{\text{all}})$, the detector will always fail under the system requirements. Therefore, $K_{\max}^U$ and $K_{\max}^L$ provide the upper and lower bounds of $K_{\max}$, respectively.

From the colluders' point of view, if colluders can collect no more than $K_{\max}^L$ independent copies, no matter how they collude, the collusion will always fail. However, if they manage to collect more than $K_{\max}^U$ copies, they can be guaranteed success even if they generate a colluded copy of the highest resolution. From the content owner's point of view, if he/she can ensure that potential colluders cannot collect more than $K_{\max}^L$ independent copies, the fingerprinting system is essentially collusion resistant.

Fig. 8(b) shows $K_{\max}^U$ and $K_{\max}^L$ as functions of the total number of users M under the system requirements $\gamma_{fp} = 10^{-3}$ and $\gamma_d = 0.8$. In Fig. 8(b), $|\mathbf{U}^b| : |\mathbf{U}^{b,e1}| : |\mathbf{U}^{\text{all}}| = 1 : 1 : 1$ and $(N_b, N_{e1}, N_{e2}) = (50\,000, 50\,000, 100\,000)$. From Fig. 8(b), with thousands of users, the fingerprinting system can withstand 50 colluders if the colluded copy has low resolution,

and it can resist attacks with up to 150 colluders if the colluded copy has high resolution. Furthermore, if the content owner distributes no more than 100 copies, the detection performance will always satisfy the requirement (34) even if all users participate in collusion. Consequently, the fingerprinting system is also collusion-secure if $M \leq 100$.

In Fig. 8(b), $K_{\max}$ first increases and then decreases, as the total number of users M increases. The intuitive explanation of this behavior is the same as in [9]. When the total number of users is small (e.g., $M \leq 20$), even if all users participate in collusion, the fingerprinting system can still successfully capture them with $P_d = 1$, as shown in Fig. 8(a). Therefore, when M is small, $K_{\max} = M$ and it increases as M increases. When M continues to increase, due to the energy reduction of the embedded fingerprints during collusion, P_d starts to drop when there are more colluders, and the fingerprinting system is more likely to make errors when identifying colluders: either it fails to detect any colluders or falsely accuses innocents. Thus, $K_{\max}$ drops as M increases when the total number of users is sufficiently large.

2) Calculation of $K_{\max}^U$ and $K_{\max}^L$: To calculate $K_{\max}^U$ and $K_{\max}^L$, we need to first find $P_d^U(K)$ and $P_d^L(K)$. From the analysis in Section IV-A, the detector has the worst performance when the colluded copy contains frames in the base layer only and $F^c = F_b$. In this scenario, for a guilty colluder $i \in SC$, the mean of his/her detection statistics is $\mu = \sqrt{N_b} \cdot \xi / K$, where N_b is the length of the fingerprints embedded in the base layer and σ_W^2 is the variance of the fingerprint. Therefore, from (30), for a given K, the lower bound of P_d is

$$P_d^L(K) = 1 - \left[1 - Q\left(\frac{h - \frac{\sqrt{N_b}\cdot\xi}{K}}{\sigma_n}\right)\right]^K \tag{38}$$

where σ_n^2 is the variance of the detection noise and the detection threshold h is chosen to satisfy $P_{fp} = \gamma_{fp}$.

To calculate the upper bound of P_d, given (N_b, N_{e1}, N_{e2}) and K, we define (39) and (40), shown at the bottom of the page.

From Section IV-A, for a given K, P_d is maximized when the colluded copy has the highest possible temporal resolution under the fairness constraints. If $\mathbb{RC}^3 \neq \emptyset$, then there exists at least one triplet $(K^b, K^{b,e1}, K^{\text{all}})$ that satisfies the fairness constraints in Table I for generating an attacked copy of the highest resolution with $F^c = F_b \cup F_{e1} \cup F_{e2}$. Therefore, see (41), shown at the bottom of the next page. From (30), maximizing P_d when $F^c = F_b \cup F_{e1} \cup F_{e2}$ is equivalent to maximizing the corresponding mean of the detection statistics $\mu = \left((N_b + N_{e1} + N_{e2})/\left(K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}\right)\right)\xi$. It is also equivalent to minimizing the denominator of μ, which is $\vartheta(K) \triangleq K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}$. Consequently, the optimization problem of (41) can be simplified to

$$\vartheta^L(K) \triangleq \min_{(K^b, K^{b,e1}, K^{\text{all}})} K^b\sqrt{N_b} + K^{b,e1} \cdot\sqrt{N_b + N_{e1}} + K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}} \tag{42}$$

with the same constraints as in (41). We can use linear programming [25] to solve the optimization problem of (42), and then calculate

$$P_d^U(K) = 1 - \left[1 - Q\left(\frac{h - \frac{(N_b + N_{e1} + N_{e2})\xi}{\vartheta^L(K)}}{\sigma_n}\right)\right]^K. \tag{43}$$

If $\mathbb{RC}^3 = \emptyset$ and $\mathbb{RC}^2 \neq \emptyset$, no matter what value the triplet $(K^b, K^{b,e1}, K^{\text{all}})$ takes, the colluders cannot generate a colluded copy of the highest resolution while still achieving fairness of collusion. However, there exists at least one $(K^b, K^{b,e1}, K^{\text{all}})$ with which the colluders can generate an attacked copy of medium resolution with $F^c = F_b \cup F_{e1}$ and still guarantee the equal risk of all colluders. In this scenario, the calculation of $P_d^L(K)$ is similar to that when $\mathbb{RC}^3 \neq \emptyset$ and not repeated here.

If $\mathbb{RC}^3 = \emptyset$ and $\mathbb{RC}^2 = \emptyset$, to ensure that all attackers have the same risk, the colluders can only generate a colluded copy of the lowest resolution with $F^c = F_b$. In this scenario, $P_d^U(K) = P_d^L(K)$.

Once we obtain $P_d^U(K)$ and $P_d^L(K)$, the analysis of $K_{\max}^U$ and $K_{\max}^L$ is the same as in [9] and omitted.

$$\mathbb{RC}^3 \triangleq \left\{(K^b, K^{b,e1}, K^{\text{all}}) : \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}} \leq \frac{N_b}{N_b + N_{e1} + N_{e2}}, \right.$$
$$\frac{K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}} \geq \frac{N_{e2}}{N_b + N_{e1} + N_{e2}},$$
$$\left. 0 \leq K^b \leq |\mathbf{U}^b|, 0 \leq K^{b,e1} \leq |\mathbf{U}^{\text{all}}|, 0 \leq K^{\text{all}} \leq |\mathbf{U}^{\text{all}}|, K^b + K^{b,e1} + K^{\text{all}} = K, \right\} \tag{39}$$

$$\text{and } \mathbb{RC}^2 \triangleq \left\{(K^b, K^{b,e1}, K^{\text{all}}) : \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + (K^{b,e1} + K^{\text{all}})\sqrt{N_b + N_{e1}}} \leq \frac{N_b}{N_b + N_{e1}}, \right.$$
$$\left. 0 \leq K^b \leq |\mathbf{U}^b|, 0 \leq K^{b,e1} \leq |\mathbf{U}^{\text{all}}|, 0 \leq K^{\text{all}} \leq |\mathbf{U}^{\text{all}}|, K^b + K^{b,e1} + K^{\text{all}} = K \right\} \tag{40}$$

B. Catch More

In the catch more scenario, the goal of the fingerprinting system is to capture as many colluders as possible, though possibly at a cost of accusing more innocent users. For this scenario, the set of performance criteria consists of the expected fraction of colluders that are successfully captured $E[F_d]$, and the expected fraction of innocent users that are falsely placed under suspicion $E[F_{fp}]$. The system requirements for such applications are $E[F_d] \geq \lambda_d$ and $E[F_{fp}] \leq \lambda_{fp}$.

Similar to the catch one scenario, if we fix $E[F_{fp}]$ as λ_{fp}, given $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{\text{all}}|)$, (N_b, N_{e1}, N_{e2}), and the total number of colluders K, we define

$$\begin{aligned} & F_d^U(K) \triangleq \max_{L^c,(K^b,K^{b,e1},K^{\text{all}})} E[F_d], \\ \text{s.t.} \quad & K^b + K^{b,e1} + K^{\text{all}} = K, \quad 0 \leq K^b \leq |\mathbf{U}^b| \\ & 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|, \quad 0 \leq K^{\text{all}} \leq |\mathbf{U}^{\text{all}}| \\ & \text{fairness constraints in Table I are satisfied} \end{aligned} \tag{44}$$

$$\begin{aligned} \text{and} \quad & F_d^L(K) \triangleq \min_{L^c,(K^b,K^{b,e1},K^{\text{all}})} E[F_d], \\ \text{s.t.} \quad & K^b + K^{b,e1} + K^{\text{all}} = K, \quad 0 \leq K^b \leq |\mathbf{U}^b| \\ & 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|, \quad 0 \leq K^{\text{all}} \leq |\mathbf{U}^{\text{all}}| \\ & \text{fairness constraints in Table I are satisfied} \end{aligned} \tag{45}$$

which are the upper and lower bounds of $E[F_d]$, respectively. $F_d^U(K)$ and $F_d^L(K)$ are decreasing functions of K since the collusion is more effective in undermining the tracing capacity with larger number of attackers. Then, we define

$$\begin{aligned} K_{\max}^U &\triangleq \arg\max_K \left\{F_d^U(K) \geq \lambda_d\right\} \\ \text{and} \quad K_{\max}^L &\triangleq \arg\max_K \left\{F_d^L(K) \geq \lambda_d\right\} \end{aligned} \tag{46}$$

which are the upper and lower bounds of $K_{\max}$ in the catch more scenario, respectively. The analysis of $\left(F_d^U(K), F_d^L(K)\right)$ and $\left(K_{\max}^U, K_{\max}^L\right)$ in the catch more scenario is similar to that in the catch one scenario and thus omitted. It is worth mentioning that similar to the scenario where users receive copies of the same resolution [22], in scalable fingerprinting systems, the detection threshold h is only determined by λ_{fp}, and $K_{\max}$ is not affected by the total number of users in the catch more scenario.

Fig. 9 shows the simulation results on the collusion resistance of the fingerprinting systems in the catch more scenario. In our simulation, $(N_b, N_{e1}, N_{e2}) = (50\,000, 50\,000, 100\,000)$ and $\sigma_n^2 = 2\sigma_W^2$. Fig. 9(a) plots $F_d^U(K)$ and $F_d^L(K)$ versus the total number of colluders K when $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{\text{all}}| = 300$ and $\lambda_{fp} = 0.01$. Under the requirements that $E[F_d] \geq 0.5$ and $E[F_{fp}] \leq 0.01$, from Fig. 9(a), $K_{\max}^U$ is approximately 180 and $K_{\max}^L$ is around 70. Fig. 9(b) plots $K_{\max}^U$ and $K_{\max}^L$ versus λ_{fp} with fixed $\lambda_d = 0.5$. From Fig. 9(b), the fingerprinting system can resist a few dozen to hundreds of colluders, depending on the resolution of the colluded copy as well as the system requirements. If the fingerprinting system can afford to put a large fraction of innocents under suspicion, it can withstand more colluders.

C. Catch All

In this scenario, the fingerprints are designed to maximize the probability of capturing all colluders, while maintaining an acceptable amount of innocents being falsely accused. This goal arises when the data's security is of great concern and any information leakage could result in serious damages. Assume that there are a total of M users and a total K colluders in the system. This set of performance criteria consists of measuring the probability of capturing all colluders $P_{d,\text{all}} = P\left[\min_{i \in SC} T_N^{(i)} > h\right]$, and the efficiency rate $R = ((M - K) \cdot E[F_{(fp)}]/K \cdot E[F_d])$ that describes the number of innocents falsely accused per colluder successfully captured. The system requirements for these applications are $R \leq \theta_r$ and $P_{d,\text{all}} \geq \theta_d$.

Similar to the catch one scenario, given $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{\text{all}}|)$ and (N_b, N_{e1}, N_{e2}), for a fixed total number of colluders K and fixed $P_{d,\text{all}} = \theta_d$, define

$$\begin{aligned} & R^U(K) \triangleq \max_{L^c,(K^b,K^{b,e1},K^{\text{all}})} R \\ \text{s.t.} \quad & K^b + K^{b,e1} + K^{\text{all}} = K, \quad 0 \leq K^b \leq |\mathbf{U}^b| \\ & 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|, \quad 0 \leq K^{\text{all}} \leq |\mathbf{U}^{\text{all}}| \\ & \text{fairness constraints in Table I are satisfied} \end{aligned} \tag{47}$$

$$\begin{aligned} \text{and} \quad & {}^L(K) \triangleq \min_{L^c,(K^b,K^{b,e1},K^{\text{all}})} R \\ \text{s.t.} \quad & K^b + K^{b,e1} + K^{\text{all}} = K, \quad 0 \leq K^b \leq |\mathbf{U}^b| \\ & 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|, \quad 0 \leq K^{\text{all}} \leq |\mathbf{U}^{\text{all}}| \\ & \text{fairness constraints in Table I are satisfied} \end{aligned} \tag{48}$$

which are the upper and lower bounds of R, respectively. We further define

$$\begin{aligned} K_{\max}^U &\triangleq \arg\max_K \left\{R^L(K) \leq \theta_r\right\} \\ \text{and} \quad K_{\max}^L &\triangleq \arg\max_K \left\{R^U(K) \leq \theta_r\right\} \end{aligned} \tag{49}$$

$$\begin{aligned} & P_d^U(K) = \max_{F^c = F_b \cup F_{e1} \cup F_{e2}, (K^b, K^{b,e1}, K^{\text{all}})} P_d \\ \text{s.t.} \quad & K^b + K^{b,e1} + K^{\text{all}} = K, \quad 0 \leq K^b \leq |\mathbf{U}^b|, \quad 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|, \quad 0 \leq K^{\text{all}} \leq |\mathbf{U}^{\text{all}}| \\ & \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}} \leq \frac{N_b}{N_b + N_{e1} + N_{e2}} \\ & \frac{K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{\text{all}}\sqrt{N_b + N_{e1} + N_{e2}}} \geq \frac{N_{e2}}{N_b + N_{e1} + N_{e2}} \end{aligned} \tag{41}$$

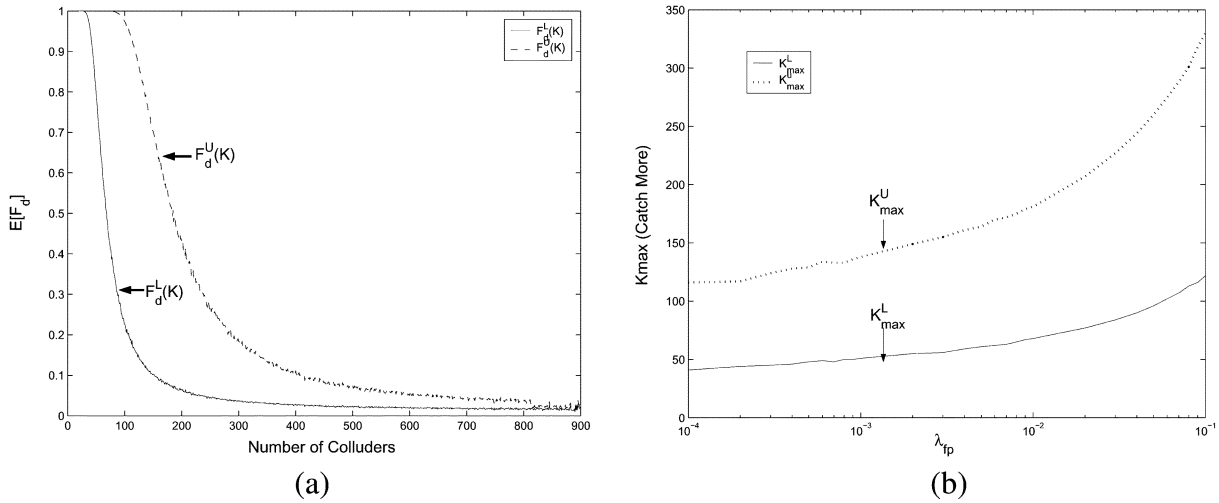

Fig. 9. Collusion resistance in the catch more scenario. $(N_b, N_{e1}, N_{e2}) = (50\,000, 50\,000, 100\,000)$. $\sigma_n^2/\xi^2 = 2$. In (a), $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{\text{all}}| = 300$ and $\lambda_{fp} = 0.01$. We plot F_d^U and F_d^L versus the total number of colluders. In (b), $\lambda_d = 0.5$, and we plot $K_{\max}^U$ and $K_{\max}^L$ under different requirements of λ_{fp}.

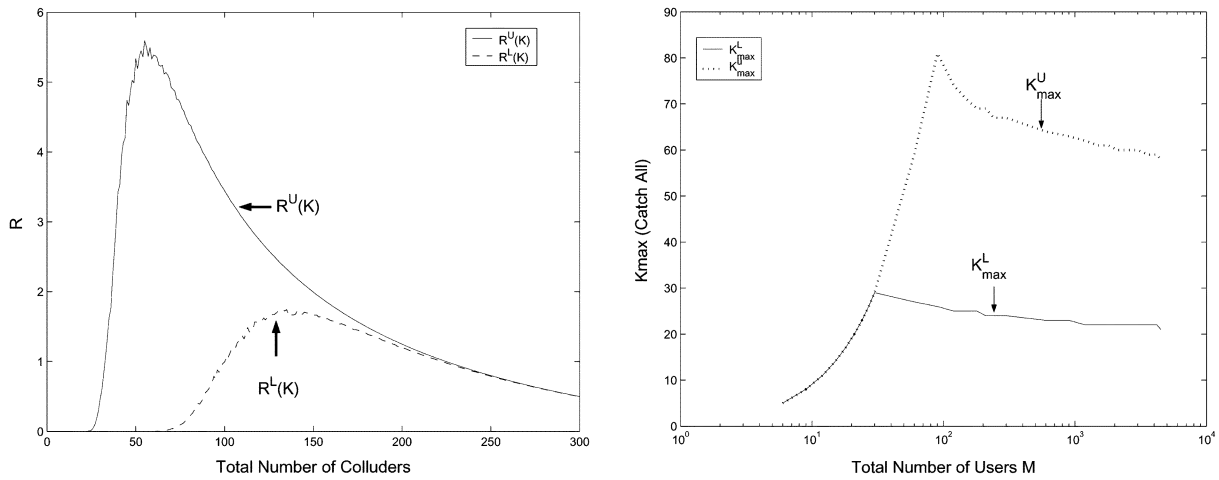

Fig. 10. Collusion resistance in the catch all scenario. $|\mathbf{U}^b| : |\mathbf{U}^{b,e1}| : |\mathbf{U}^{\text{all}}| = 1 : 1 : 1$ and $(N_b, N_{e1}, N_{e2}) = (50\,000, 50\,000, 100\,000)$. $\sigma_n^2/\xi^2 = 2$. $\theta_d = 0.99$ and $\theta_r = 0.01$. In (a), $M = 450$ and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{\text{all}}| = 150$. We plot R^U and R^L versus the total number of colluders. (b) shows $K_{\max}^U$ and $K_{\max}^L$ versus the total number of users M.

as the upper and lower bounds of $K_{\max}$, respectively. The analysis of $K_{\max}^U$ and $K_{\max}^L$ in the catch all scenario is similar to that in the catch one scenario and not repeated.

In our simulations of the catch one scenario, we let $|\mathbf{U}^b| : |\mathbf{U}^{b,e1}| : |\mathbf{U}^{\text{all}}| = 1 : 1 : 1$ and $(N_b, N_{e1}, N_{e2}) = (50\,000, 50\,000, 100\,000)$. Fig. 10(a) plots $R^U(K)$ and $R^L(K)$ versus the total number of colluders K when there are $M = 450$ users and $\theta_d = 0.99$. We consider a scenario that is required to catch all colluders with probability larger than 0.99 ($P_{d,\text{all}} \geq 0.99$) and accuse no more than on innocent for every 100 colluders captured ($R \leq 0.01$). Under these requirements, from Fig. 10(a), the attacker should collect more than $K_{\max}^U = 65$ different copies to ensure the success of collusion, and the scalable fingerprinting system is collusion free when there are fewer than $K_{\max}^L = 25$ colluders. Fig. 10(b) shows $K_{\max}^U$ and $K_{\max}^L$ versus the total number of users M when $\theta_d = 0.99$ and $\theta_r = 0.01$. From Fig. 10(b), in the catch all scenario with thousands of users, the scalable fingerprinting systems can survive collusion by 20 to 60 attackers, depending on the resolution of the colluded copy. It is collusion-secure if the content owner distributes no more than 30 different copies. The non-monotonic behavior in Fig. 10 can be explained in the same way as in the catch one scenario.

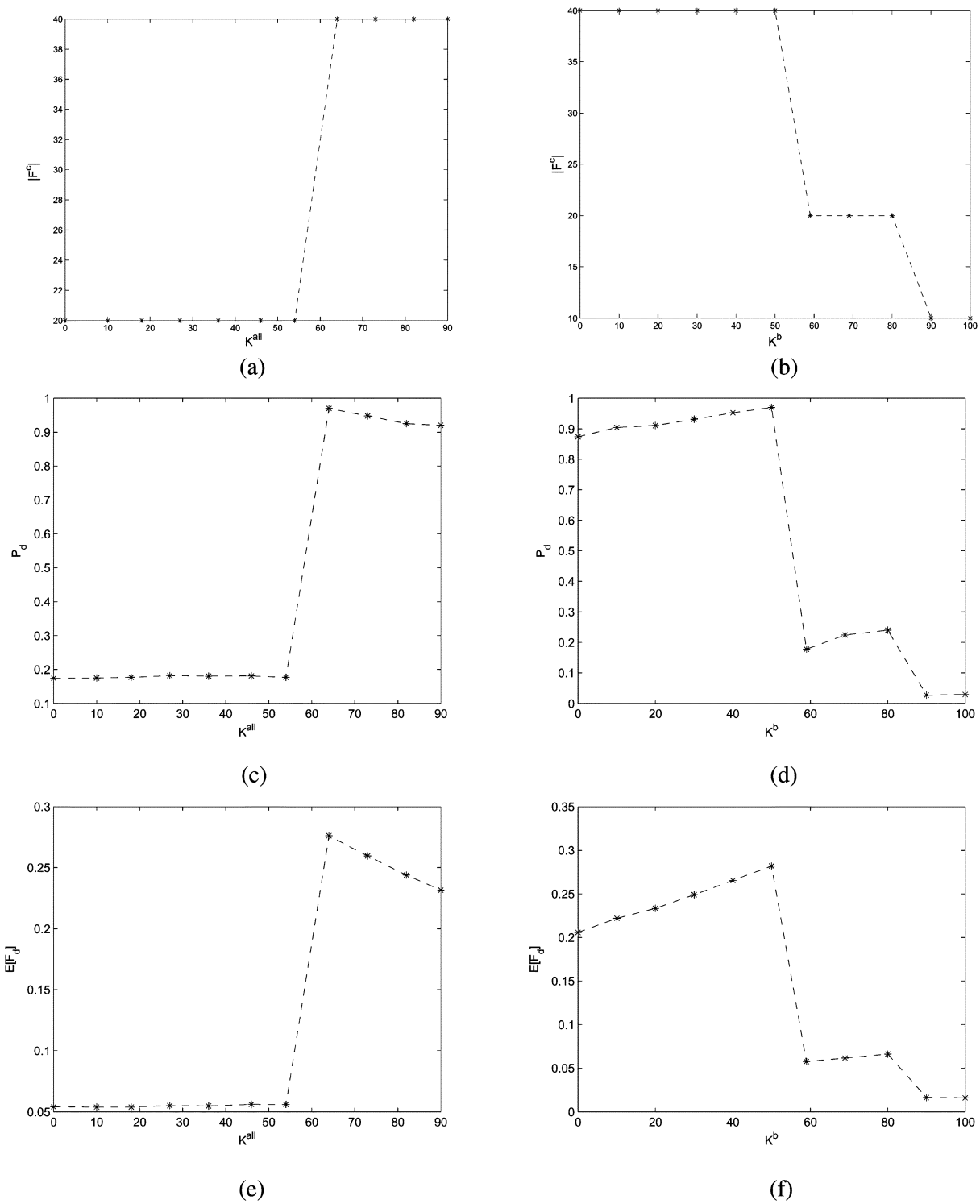

Fig. 11. Simulation results on the first 40 frames of sequence "carphone." The base layer contains frame $F_b = \{1, 5, \ldots, 37\}$, the enhancement layer 1 contains frame $F_{e1} = \{3, 7, \ldots, 39\}$, and the enhancement layer 2 contains frame $F_{e2} = \{2, 4, \ldots, 40\}$. Assume that there are a total of $M = 450$ users and a fixed $K = 150$ colluders. $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{\text{all}}| = 150$. In (a), (c), and (e), each value of K^{all} corresponds to a unique triplet $(K^b, K^{b,e1}, K^{\text{all}})$ on Line $\overline{AB}$ (31). In (b), (d), and (f), each value of K^b represents a unique triplet $(K^b, K^{b,e1}, K^{\text{all}})$ on Line $\overline{CD}$ (32). $P_{fp} = 10^{-3}$ in (c) and (d), and $E[F_{fp}] = 10^{-3}$ in (e) and (f).

VI. Simulation Results on Video Sequences

In our simulations on real videos, we test on the first 40 frames of sequence "carphone" as an example. Following Section II-A, we choose $F_b = \{1, 5, \ldots, 37\}$, $F_{e1} = \{3, 7, \ldots, 39\}$ and $F_{e2} = \{2, 4, \ldots, 40\}$ as an example of the temporal scalability. Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{\text{all}}| = 150$. We adopt the human visual model based spread-spectrum embedding in [15], and embed the fingerprints in the DCT domain. The lengths of the embedded fingerprints in the base layer, enhancement layer 1 and enhancement layer 2 are $N_b = 39\,988$, $N_{e1} = 39\,934$ and $N_{e2} = 79\,686$, respectively. We first generate independent

vectors following Gaussian distribution $\mathcal{N}(0, 1/9)$ and apply Gram-Schmidt orthogonalization to produce fingerprints satisfying the strict orthogonality and equal energy requirements in (1). In each fingerprinted copy, the fingerprints embedded in different frames are correlated with each other, depending on the similarity between the host frames.

During collusion, we fix the total number of colluders as $K = 150$ and assume that the collusion attack is also in the DCT domain. In our simulations, the colluders apply the intragroup collusion attacks followed by the intergroup attacks as in Section II-B2. We adjust the power of the additive noise such that $\|\mathbf{n}_j\|^2 / \left\| JND_j \mathbf{W}_j^{(i)} \right\|^2 = 2$ for every frame $j \in F^c$ in the colluded copy. In our simulations, we assume that the colluders generate a colluded copy of the highest possible resolution under the fairness constraints.

At the detector's side, we consider a non-blind detection scenario where the host signal is removed from the colluded copy before colluder identification process. The detector follows the detection process in Section II-B3 and estimates the indices of the colluders $\widehat{SC}$.

Fig. 11 shows the simulation results. In Fig. 11(a), (c), and (e), the same as in Fig. 6, the x axis is the number of colluders who receive all three layers K^{all}, and each value of K^{all} represents a unique triplet $(K^b, K^{\text{all}}, K^{\text{all}})$ on Line $\overline{AB}$ (31). In Fig. 11(b), (d), and (f), the same as in Fig. 7, the x axis is the number of colluders who receive the base layer only, and a given K^b corresponds to a triplet $(K^b, K^{b,e1}, K^{\text{all}})$ on Line $\overline{CD}$ (32). Fig. 11(a) and (b) show the total number of frames in the colluded copy L^c, and $L^c = 10$, $L^c = 20$ and $L^c = 40$ when the colluded copy has the lowest, medium and highest resolution, respectively. In Fig. 11(c) and (d), we select the threshold to fix $P_{fp} = 10^{-3}$ and compare P_d when $(K^b, K^{b,e1}, K^{\text{all}})$ takes different values. In Fig. 11(e) and (f), $E[F_{fp}]$ is fixed as 10^{-3} by selecting the threshold in the simulation runs, and we compare $E[F_d]$ of the collusion attacks with different $(K^b, K^{b,e1}, K^{\text{all}})$.

From Fig. 11, the effectiveness of collusion in defeating the scalable fingerprinting systems depends on the resolution of the colluded copy. When the colluded copy has higher resolution, the extracted fingerprint gives the detector more information about the colluders' identities, and the attackers take a larger risk of being detected. The simulation results on real videos agree with our analytical results and are comparable with those simulation results in Section IV-B.

VII. Conclusion

In this paper, we have studied the behavior forensics in multimedia fingerprinting and analyzed the dynamics among colluders to ensure fairness of collusion. We have investigated how to achieve fairness of collusion when fingerprinted copies used in collusion have different resolutions, and analyzed the effectiveness of such fair collusion in removing the fingerprints. We have also examined the collusion resistance of the scalable fingerprinting systems and evaluated the maximum number of colluders that they can withstand.

We first investigated how to distribute the risk of being detected evenly to all colluders when they receive copies of different resolutions due to network and device heterogeneity. We showed that higher resolution of the colluded copy puts more severe constraints on achieving fairness of collusion. We then analyzed the effectiveness of such fair collusion attacks. Both our analytical and simulation results showed that the colluders are more likely to be captured when the colluded copy has higher resolution. The colluders have to take into consideration the tradeoff between the probability of being detected and the resolution of the colluded copy during collusion.

We also analyzed the collusion resistance of the scalable fingerprinting systems for various fingerprinting scenarios with different requirements. We evaluated the maximum number of colluders that the fingerprinting systems can resist, and showed that the scalable fingerprinting systems can withstand dozens to hundreds of colluders, depending on the resolution of the colluded copy as well as the system requirements. We also provided the lower and upper bounds of $K_{\max}$. From the colluders' point of view, $K_{\max}^U$ tells attackers how many independent copies are required to guarantee the success of collusion under all circumstances. From the content owner's point of view, to achieve collusion free, a desired security requirement is to make the potential colluders very unlikely to collect more than $K_{\max}^L$ copies.

References

[1] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, no. 7, pp. 1079–1107, Jul. 1999.

[2] I. Cox, J. Killian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

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

[4] Z. J. Wang, M. Wu, W. Trappe, and K. J. R. Liu, "Group-oriented fingerprinting for multimedia forensics," *EURASIP J. Applied Signal Processing, Special Issue on Multimedia Security and Rights Management*, vol. 2004, no. 14, pp. 2142–2162, Nov. 2004.

[5] F. Zane, "Efficient watermark detection and collusion security," in *Proc. 4th Int. Conf. Financial Cryptography*, Feb. 2000, vol. 1962, Lecture of Notes in Computer Science, pp. 21–32.

[6] J. Dittmann, P. Schmitt, E. Saar, J. Schwenk, and J. Ueberberg, "Combining digital watermarks and collusion secure fingerprints for digital images," *SPIE J. Electron. Imag.*, vol. 9, no. 4, pp. 456–467, Oct. 2000.

[7] I. Cox and J. P. Linnartz, "Some general methods for tampering with watermarking," *IEEE J. Select. Areas Commun.*, vol. 16, no. 4, pp. 587–593, May 1998.

[8] F. Petitcolas, R. Anderson, and M. Kuhn, "Attacks on copyright marking systems," in *Proc. 2nd Workshop Information Hiding*, Apr. 1998, Lecture Notes in Computer Science, pp. 218–238.

[9] Z. J. Wang, M. Wu, H. Zhao, W. Trappe, and K. J. R. Liu, "Anti-collusion forensics of multimedia fingerprinting using orthogonal modulation," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 804–821, Jun. 2005.

[10] F. Ergun, J. Killian, and R. Kumar, "A note on the limits of collusion-resistant watermarks," in *Proc. Advances in Cryptology—EuroCrypto*, 2001, vol. 1592, Lecture Notes in Computer Science, pp. 140–149.

[11] J. Su, J. Eggers, and B. Girod, "Capacity of digital watermarks subject to an optimal collusion attacks," in *Proc. Eur. Signal Processing Conf.*, 2000.

[12] H. Stone, Analysis of Attacks on Image Watermarks With Randomized Coefficients NEC Res. Inst., 1996, Tech. Rep. 96-045.

[13] H. Zhao, M. Wu, Z. J. Wang, and K. J. R. Liu, "Forensic analysis of nonlinear collusion attacks for multimedia fingerprinting," *IEEE Trans. Image Process.*, vol. 14, no. 5, pp. 646–661, May 2005.

[14] Y. Wang, J. Ostermann, and Y. Zhang, *Video Processing and Communications*, 1st ed. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[15] C. Podilchuk and W. Zeng, "Image adaptive watermarking using visual models," *IEEE J. Select. Areas Commun.*, vol. 16, no. 4, pp. 525–540, May 1998.

[16] M. Swanson, B. Zhu, and A. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," *IEEE J. Select. Areas Commun.*, vol. 16, no. 4, pp. 540–550, May 1998.

[17] M. Holliman and N. Memon, "Counterfeiting attacks and blockwise independent watermarking techniques," *IEEE Trans. Image Process.*, vol. 9, pp. 432–441, Mar. 2000.
[18] D. Kiroski and F. A. P. Petitcolas, "Blind pattern matching attack on watermarking systems," *IEEE Trans. Signal Process.*, vol. 51, pp. 1045–1053, 2003.
[19] G. Doerr, J. L. Dugelay, and L. Grange, "Exploiting self-similarities to defeat digital watermarking systems: A case study on still images," in *Proc. ACM Multimedia and Security Workshop*, 2004.
[20] K. Su, D. Kundur, and D. Hatzinakos, "Statistical invisibility for collusion-resistant digital video watermarking," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 43–51, Feb. 2005.
[21] ——, "Spatially localized image-dependent watermarking for statistical invisibility and collusion resistance," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 52–66, Feb. 2005.
[22] Z. J. Wang, M. Wu, H. Zhao, W. Trappe, and K. J. R. Liu, "Resistance of orthogonal Gaussian fingerprints to collusion attacks," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Apr. 2003.
[23] H. V. Poor, *An Introducton to Signal Detection and Estimation*, 2nd ed. New York: Springer Verlag, 1999.
[24] J. Killian, T. Leighton, L. R. Matheson, T. G. Shamoon, R. Tajan, and F. Zane, Resistance of Digital Watermarks to Collusive Attacks Dept. Computer Science, Princeton Univ., Princeton, NJ, 1998, Tech. Rep. TR-585-98.
[25] G. Dantzig, *Linear Programming and Extensions*. Princeton, NJ: Princeton Univ. Press, 1963.

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

She has been a Research Associate with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland. Since 2006, she has been an Assistant Professor with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. She coauthored the book *Multimedia Fingerprinting Forensics for Traitor Tracing* (Hindawi, 2005). Her research interests include information security and forensics, multimedia, digital communications, and signal processing.

K. J. Ray Liu (F'03) is Professor and Associate Chair, Graduate Studies and Research of the Graduate Studies and Research of Electrical and Computer Engineering Department, University of Maryland, College Park. His research contributions encompass broad aspects of wireless communications and networking, information forensics and security, multimedia communications and signal processing, bioinformatics and biomedical imaging, and signal processing algorithms and architectures.

Dr. Liu is the recipient of best paper awards from the IEEE Signal Processing Society (twice), IEEE Vehicular Technology Society, and EURASIP, IEEE Signal Processing Society Distinguished Lecturer, EURASIP Meritorious Service Award, and the National Science Foundation Young Investigator Award. He also received Poole and Kent Company Senior Faculty Teaching Award and Invention of the Year Award, both from the University of Maryland. He is Vice President—Publications and on the Board of Governor of IEEE Signal Processing Society. He was the Editor-in-Chief of *IEEE Signal Processing Magazine* and the founding Editor-in-Chief of *EURASIP Journal on Applied Signal Processing*.