BEHAVIOR FORENSICS IN TRAITORS WITHIN TRAITORS FOR SCALABLE MULTIMEDIA

K. J. Ray Liu and H. Vicky Zhao

Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742

ABSTRACT

A cost effective attack against multimedia forensics is the multi-user collusion attack, in which several attackers mount attacks collectively to remove traces of the identifying fingerprints and hinder traitor tracing. An important issue in collusion is to ensure that all colluders have the same probability of being detected. While they might agree so, some selfish colluders may wish to further lower their own risk of being caught. This paper investigates this problem of *"traitors within traitors"*, in an effort to formulate the dynamics among attackers during collusion. We consider scalable multimedia forensic systems where users receive fingerprinted copies of different quality due to network and device heterogeneity, and explore the techniques that a selfish colluder can use to minimize his/her probability of being detected. Our results show that changing the resolution of their received copies before multi-user collusion can help selfish colluders further reduce their risk, especially when the colluded copy has high resolution.

Index Terms— security, multimedia systems, video signal processing

1. INTRODUCTION

Behavior forensics formulates the dynamics in multimedia security and forensic systems, investigates how users with different (and usually conflicting) objectives interact with and respond to each other, and analyzes how they influence each other's decisions and performance. Such investigation enables a thorough understanding of multimedia security and forensic systems, e.g., how attackers behave, which information about attacks can help improve the system's robustness, etc. It helps the digital rights enforcer offer stronger protection of multimedia.

This paper considers digital fingerprinting for traitor tracing [1] and analyzes the dynamics among users in multimedia forensic systems. Multi-user collusion is a powerful attack against multimedia forensics, in which a group of attackers come together and attack collectively to undermine the traitor tracing capability. During collusion, attackers 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 taking a higher risk than the others, the colluders often demand a fair play during collusion and require that all colluders have the same probability of being captured. The work in [2] studied how to achieve fairness of collusion when attackers receive copies of different quality due to network and device heterogeneity, and analyzed the constraints on, and the effectiveness of, fair collusion.

Most prior work assumed that all colluders keep their agreement to share the same risk during collusion. In reality, the assumption of fair play may not always hold. Some colluders might be selfish and wish to further lower their risk of being caught. It was shown in [3] that a simple temporal filtering of the fingerprinted copies before collusion can help selfish colluders further reduce their risk. In this paper, we study this problem of traitors within traitors in scalable fingerprinting systems where users receive copies of different quality, investigate the possible techniques for selfish colluders to minimize their risk, and analyze their performance.

This paper is organized as follows. We begin in Section 2 with the introduction of scalable multimedia fingerprinting systems and the model of traitors within traitors. Section 3 investigates the possible strategies that the selfish colluders can use to further reduce their risk in scalable fingerprinting systems and analyzes their performance. Section 4 compares the performance of different risk minimization techniques, and conclusions are drawn in Section 5.

2. SYSTEM MODEL

2.1. Temporally Scalable Video Coding System

To address the network and device heterogeneity, in layered video coding, the video content is decomposed into non-overlapping parts of different priority: the base layer contains the most important information of the video and is received by all users; while the enhancement layers gradually refine the reconstructed sequence and are only received by users with sufficient bandwidth [4]. Without loss of generality, we consider a temporally scalable video coding system with three-layer scalability: the base layer with the highest priority, the enhancement layer 1 with medium priority, and the enhancement layer 2 with the lowest priority. Same as in [2], we encode different frames in different layers. For example, with MPEG-2 video coding, the base layer may contain all the I frames, the enhancement layer 1 contains all the P frames, and the enhancement layer 2 contains all the B frames. 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. As an example, $F_b = \{1, 5, 9, \cdots\}$, $F_{e1} = \{3, 7, 11, \cdots\}$ and $F_{e2} = \{2, 4, 6, \cdots\}$.

Define $F^{(i)}$ as the set containing the indices of the frames that user $\mathbf{u}^{(i)}$ receives from the content owner. $\mathbf{U}^b = \{\mathbf{u}^{(i)} : F^{(i)} = F_b\}$ is the subgroup of users who receive the base layer only; $\mathbf{U}^{b,e1} = \{\mathbf{u}^{(i)} : F^{(i)} = F_b \cup F_{e1}\}$ is the subgroup of users who receive both the base layer and the enhancement layer 1; and $\mathbf{U}^{all} = \{\mathbf{u}^{(i)} : F^{(i)} = F_b \cup F_{e1} \cup F_{e2}\}$ is the subgroup of users who receive all three layers.

2.2. Scalable Multimedia Fingerprinting System

Fingerprint Embedding With the above scalable coding systems, for the jth frame in the video represented by a vector $\mathbf{S}_j$ of length

The authors can be reached at kjrliu and hzhao@eng.umd.edu.

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 $\mathbf{X}_j^{(i)} = \mathbf{S}_j + JND_j\mathbf{W}_j^{(i)}$, where JND is from human visual models [5] to control the energy of the embedded fingerprints.

We use Gaussian distributed fingerprints due to its robustness against many attacks and let $\{\mathbf{W}_j^{(i)}\}$ follow distribution $\mathcal{N}(0, \sigma_W^2)$. To resist intra-content collusion attacks, in each distributed copy, correlated fingerprints are embedded into adjacent frames and the correlation depends on the similarity between the two host frames [6]. Fingerprints for different users are independent of each other.

Multi-user Collusion It was shown in [7] that if all collusions generate colluded copies of the same quality, nonlinear collusions have approximately the same performance as the averaging collusion. Thus, it suffices to consider averaging based collusion only.

The colluders first divide themselves into three non-overlapping subgroups: $SC^b = \{i : F^{(i)} = F_b\}$ contains the indices of the colluders who receive the base layer only; $SC^{b,e1} = \{i : F^{(i)} = F_b \cup F_{e1}\}$ contains the indices of the colluders who receive base layer and enhancement layer 1; and $SC^{all} = \{i : F^{(i)} = F_b \cup F_{e1} \cup F_{e2}\}$ contains the indices of the colluders who receive all three layers. K^b, $K^{b,e1}$ and K^{all} are the number of colluders in SC^b, $SC^{b,e1}$ and SC^{all}, respectively.

During collusion, following the work in [2], for each frame $j_1 \in F_b$ in the base layer, the colluded frame j_1 is

$$\mathbf{V}_{j_1} = \beta_1 \sum_{i \in SC^b} \frac{\mathbf{X}_{j_1}^{(i)}}{K^b} + \beta_2 \sum_{i \in SC^{b,e1}} \frac{\mathbf{X}_{j_1}^{(i)}}{K^{b,e1}} + \beta_3 \sum_{i \in SC^{all}} \frac{\mathbf{X}_{j_1}^{(i)}}{K^{all}} + \mathbf{n}_{j_1}, \quad (1)$$

where $0 \le \beta_1, \beta_2, \beta_3 \le 1$ and $\beta_1 + \beta_2 + \beta_3 = 1$. For each frame $j_2 \in F_{e1}$ in the enhancement layer 1, the colluded copy is

$$\mathbf{V}_{j_2} = \alpha_1 \sum_{i \in SC^{b,e1}} \frac{\mathbf{X}_{j_2}^{(i)}}{K^{b,e1}} + \alpha_2 \sum_{i \in SC^{all}} \frac{\mathbf{X}_{j_2}^{(i)}}{K^{all}} + \mathbf{n}_{j_2}, \quad (2)$$

where $0 \le \alpha_1, \alpha_2 \le 1$ and $\alpha_1 + \alpha_2 = 1$. For each frame $j_3 \in F_{e2}$ in the enhancement layer 2,

$$\mathbf{V}_{j_3} = \sum_{i \in SC^{all}} \frac{\mathbf{X}_{j_3}^{(i)}}{K^{all}} + \mathbf{n}_{j_3}. \quad (3)$$

In (1)-(3), $\mathbf{n}_j$ is additive noise to further hinder detection. The colluders seek $\{\alpha_k, \beta_l\}$ to ensure that all colluders take the same risk of being detected. Details of the collusion parameter selection and the constraints on collusion to achieve fairness are in [2].

Fingerprint Detection and Colluder Identification During fingerprint detection, the host signal is first removed from the test copy before colluder identification. The detector then extracts the fingerprint $\mathbf{Y}_j$ from the jth frame $\mathbf{V}_j$ in the colluded copy, calculates the similarity between the extracted fingerprint and each of the original fingerprints, compares with a threshold h, and estimates the identities of the colluders $\widehat{SC}$.

In this paper, following the work in [2], we consider a simple detector that collectively uses fingerprints extracted from all layers to identify colluders. For each user $\mathbf{u}^{(i)}$, the detector calculates first $\breve{F}^{(i)} = F^{(i)} \cap F^c$, where $F^{(i)}$ contains the indices of the frames received by $\mathbf{u}^{(i)}$ and F^c contains the indices of the frames in the colluded copy. Then the detector calculates

$$TN^{(i)} = \left(\sum_{j \in \breve{F}^{(i)}} \langle \mathbf{Y}_j, \mathbf{W}_j^{(i)} \rangle \right) / \sqrt{\sum_{j \in \breve{F}^{(i)}} ||\mathbf{W}_j^{(i)}||^2}, \quad (4)$$

where $||\mathbf{W}_j^{(i)}||$ is the Euclidean norm of $\mathbf{W}_j^{(i)}$. Given a pre-determined threshold h, $\widehat{SC} = \{i : TN^{(i)} > h\}$.

2.3. Traitors within Traitors

Assume that $\mathbf{X}^{(i)}$ is the fingerprinted copy that $\mathbf{u}^{(i)}$ received from the content owner. If all colluders provide one another correct information about their fingerprinted signals, the multi-user collusion function $g(\cdot)$ is applied to the originally received copies, and the final colluded copy equals to $\mathbf{V} = g(\{\mathbf{X}^{(i)}\}_{i \in SC})$.

When there are selfish colluders, to further reduce his/her own probability of being captured, a selfish colluder $\mathbf{u}^{(i_1)}$ may process his/her received copy $\mathbf{X}^{(i_1)}$, generate a new copy $\widetilde{\mathbf{X}}^{(i_1)}$ that is perceptually similar to $\mathbf{X}^{(i_1)}$, and use $\widetilde{\mathbf{X}}^{(i_1)}$ during collusion. If other colluders do not discover this selfish behavior, the collusion function $g(\cdot)$ is applied to $\widetilde{\mathbf{X}}^{(i_1)}$ and $\{\mathbf{X}^{(i)}\}_{i \in SC, i \neq i_1}$, and the colluded copy equals to $\mathbf{V}' = g(\widetilde{\mathbf{X}}^{(i_1)}, \{\mathbf{X}^{(i)}\}_{i \in SC, i \neq i_1})$.

The selfish colluders wish to find the most effective pre-collusion processing techniques to minimize their chance of being captured. Meanwhile, in order to profit from the illegal redistribution of multimedia, the selfish colluders have to ensure that the newly generated copies are perceptually similar to the originally received ones, so that other colluders cannot detect their pre-collusion processing and exclude them from collusion.

3. PRE-COLLUSION PROCESSING IN SCALABLE FINGERPRINTING SYSTEMS

In [3], during pre-collusion processing, the selfish colluders temporally filter their received copies to attenuate the energies of the embedded fingerprints even before collusion. In scalable fingerprinting systems where users receive fingerprinted copies of different quality, in addition to temporal filtering, the selfish colluders can also change the resolution of their copies before collusion.

Assume that $\{\mathbf{X}_j^{(i_1)}\}$ are the fingerprinted frames that a selfish colluder $\mathbf{u}^{(i_1)}$ receives from the content owner. $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$ is the copy that $\mathbf{u}^{(i_1)}$ generates during pre-collusion processing, and its temporal resolution is different from that of $\{\mathbf{X}_j^{(i_1)}\}$. We define the processing parameter as $CP^{(i_1)} = (F^{(i_1)}, \widetilde{F}^{(i_1)})$, where $F^{(i_1)}$ contains the indices of the frames that $\mathbf{u}^{(i_1)}$ received from the content owner, and $\widetilde{F}^{(i_1)}$ contains the indices of the frames in the newly generated copy $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$. We consider a simple scenario where $\widetilde{F}^{(i_1)} \in \{F_b,\ F_b \cup F_{e1},\ F_b \cup F_{e1} \cup F_{e2}\}$.

3.1. Increasing the Resolution Before Collusion

In this type of pre-collusion processing, $F^{(i_1)} \subset \widetilde{F}^{(i_1)}$ and the selfish colluder receives a copy of low resolution and generates a copy of higher resolution before collusion. For example, $\mathbf{u}^{(i_1)}$ receives the fingerprinted base layer only and generates a copy including all three layers before multi-user collusion.

Pre-collusion Processing In the above example where $CP^{(i_1)} = (F^{(i_1)} = F_b, \widetilde{F}^{(i_1)} = F_b \cup F_{e1} \cup F_{e2})$, we assume that for every frame $j \in F^{(i_1)} = F_b$ in the base layer that $\mathbf{u}^{(i_1)}$ received, $\widetilde{\mathbf{X}}_j^{(i_1)} = \mathbf{X}_j^{(i_1)}$. Furthermore, $\mathbf{u}^{(i_1)}$ needs to forge frames in the enhancement layers that he/she did not receive. Assume that $\mathbf{X}_{j_1}^{(i_1)}$ and $\mathbf{X}_{j_3}^{(i_1)}$ are two adjacent frames in the base layer that $\mathbf{u}^{(i_1)}$ received. To forge a frame $\widetilde{\mathbf{X}}_{j_2}^{(i_1)}$ in the enhancement layers where $j_2 \in F_{e1} \cup F_{e2}$ and $j_1 < j_2 < j_3$, we consider a simple linear interpolation based method in this paper, and let

$$\widetilde{\mathbf{X}}_{j_2}^{(i_1)} = \lambda_1 \cdot \mathbf{X}_{j_1}^{(i_1)} + \lambda_2 \cdot \mathbf{X}_{j_3}^{(i_1)},$$

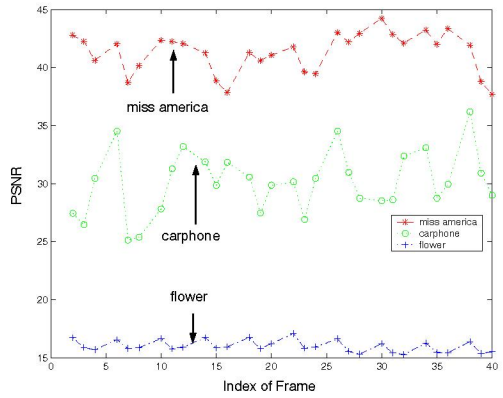


Fig. 1. The quality of the enhancement layers that are forged by the selfish colluder using (5) during pre-collusion processing. $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$, $F_b = \{1, 5, 9, \cdots\}$, $F_{e1} = \{3, 7, 11, \cdots\}$ and $F_{e2} = \{2, 4, 6, 8, \cdots\}$.

$$\text{where} \quad \lambda_1 = \frac{j_3 - j_2}{j_3 - j_1} \quad \text{and} \quad \lambda_2 = \frac{j_2 - j_1}{j_3 - j_1}. \tag{5}$$

Perceptual Quality Constraints To examine the perceptual quality of the forged enhancement layers, we consider the above example with $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$ and use (5) to generate the enhancement layers. For a selfish colluder $\mathbf{u}^{(i_1)}$ in subgroup SC^b and for a frame $j \in F_{e1} \cup F_{e2}$ in the enhancement layers, define $\mathbf{X}_j^{(i_1)}$ as the fingerprinted frame j that $\mathbf{u}^{(i_1)}$ would have received if he/she had subscribed to frame j. We use $\mathbf{X}_j^{(i_1)}$ as the ground truth and calculate the PSNR of $\widetilde{\mathbf{X}}_j^{(i_1)}$ when compared with $\mathbf{X}_j^{(i_1)}$.

Figure 1 shows the results on the first 40 frames of sequence "miss america", "carphone" and "flower". For sequence with flat regions and slow motion ("miss america"), the forged enhancement layers have high quality; while for sequence that has fast movement ("flower"), the selfish colluder can only generate low-quality and blurred enhancement layers. Although motion based interpolation [8] can help improve the quality, for these sequences with fast movement and complex scene composition, the selfish colluder may still not be able to forge enhancement layers of good enough quality to use. Therefore, for complicated sequences with fast movement, the selfish colluder might not be able to apply this type of pre-collusion processing due to the quality constraints.

Selfish Colluder's Probability of Being Detected To analyze the effectiveness of this pre-collusion processing technique in reducing a selfish colluder's risk, we compare his/her probability of being detected when the selfish colluder increases the temporal resolution with that when he/she does not apply pre-collusion processing. We assume that there is only one selfish colluder $\mathbf{u}^{(i_1)}$ in the system and $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$. We further assume that the colluders generate a colluded copy of high quality with $F^c = F_b \cup F_{e1} \cup F_{e2}$ under the fairness constraints in [2]. Our analysis can be extended to other scenarios.

WITHOUT PRE-COLLUSION PROCESSING In this scenario, colluders give each other correct information about their received copies. They select the parameters $\{\alpha_k, \beta_l\}$ and generate the colluded copy $\mathbf{V}$ in the same way as in [2]. The analysis of $\mathbf{u}^{(i_1)}$'s detection statistics ($TN^{(i_1)}$) and that of his/her probability of being detected ($P_d^{(i_1)}$) are the same as in [2].

Under the assumption that the detection noises are i.i.d. and follow distribution $\mathcal{N}(0, \sigma_n^2)$, from the analysis in [2], we have

$$TN^{(i_1)} \sim \mathcal{N}\left(\mu^{(i_1)}, \sigma_n^2\right) \text{ and } P_d^{(i_1)} \approx Q\left((h - \mu^{(i_1)})/\sigma_n\right),$$

$$\text{where} \quad \mu^{(i_1)} \approx \beta_1 \sqrt{N_b}\sigma_W / K^b. \tag{6}$$

$Q(\cdot)$ is the Gaussian tail function, h is the threshold, and N_b, N_{e1} and N_{e2} are the lengths of the fingerprints embedded in the base layer, enhancement layer 1 and enhancement layer 2, respectively.

WITH PRE-COLLUSION PROCESSING In this scenario, the selfish colluder $\mathbf{u}^{(i_1)}$ increases the resolution of his/her copy before collusion using parameter $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$. Assume that $\mathbf{u}^{(i_1)}$ is the only selfish colluder and others do not discover his/her selfish behavior. Thus, the colluders believe that $\widetilde{K}^b = K^b - 1$ colluders receive the base layer only, $\widetilde{K}^{b,e1} = K^{b,e1}$ colluders receive both the base layer and the enhancement layer 1, and $\widetilde{K}^{all} = K^{all} + 1$ colluders receive all three layers.

During collusion, other colluders believe that the extracted fingerprints from all three layers will be used by the detector to determine if $\mathbf{u}^{(i_1)}$ is a colluder. Using the same analysis as in [2], if $F^c = F_b \cup F_{e1} \cup F_{e2}$ and the detection noises are i.i.d. and follow distribution $\mathcal{N}(0, \sigma_n^2)$, the colluders estimate that

$$TN^{(i_1)} \sim \mathcal{N}\left(\bar{\mu}^{(i_1)}, \sigma_n^2\right) \text{ where } \bar{\mu}^{(i_1)} \approx \frac{\widetilde{\beta}_3 N_b + \widetilde{\alpha}_2 N_{e1} + N_{e2}}{\widetilde{K}^{all}\sqrt{N_b + N_{e1} + N_{e2}}}. \tag{7}$$

Then, they select $\{\widetilde{\alpha}_k, \widetilde{\beta}_l\}$ to ensure that $\bar{\mu}^{(i_1)}$ equals to the means of other colluders' detection statistics. From [2],

$$\widetilde{\beta}_1 = \frac{N\widetilde{K}^b\sqrt{N_b}}{N_b(\widetilde{K}^b\sqrt{N_b} + \widetilde{K}^{b,e1}\sqrt{N_b + N_{e1}} + \widetilde{K}^{all}\sqrt{N})}, \tag{8}$$

$$\widetilde{\beta}_2 N_b + \widetilde{\alpha}_1 N_{e1} = \frac{N\widetilde{K}^{b,e1}\sqrt{N_b + N_{e1}}}{\widetilde{K}^b\sqrt{N_b} + \widetilde{K}^{b,e1}\sqrt{N_b + N_{e1}} + \widetilde{K}^{all}\sqrt{N}},$$

$$\widetilde{\beta}_3 = 1 - \widetilde{\beta}_1 - \widetilde{\beta}_2, \ \widetilde{\alpha}_2 = 1 - \widetilde{\alpha}_1, \text{ where } N = N_b + N_{e1} + N_{e2}.$$

At the detector's side, since $\mathbf{u}^{(i_1)}$ only received the base layer, the detector uses fingerprints extracted from the base layer only to decide if $\mathbf{u}^{(i_1)}$ is a traitor. Following the same analysis as in [2],

$$TN^{(i_1)} \sim \mathcal{N}\left(\widetilde{\mu}^{(i_1)}, \sigma_n^2\right) \text{ and } P_d^{(i_1)} \approx Q\left((h - \widetilde{\mu}^{(i_1)})/\sigma_n\right),$$

$$\text{where} \quad \widetilde{\mu}^{(i_1)} \approx \widetilde{\beta}_3\sqrt{N_b}\sigma_W / \widetilde{K}^{all}. \tag{9}$$

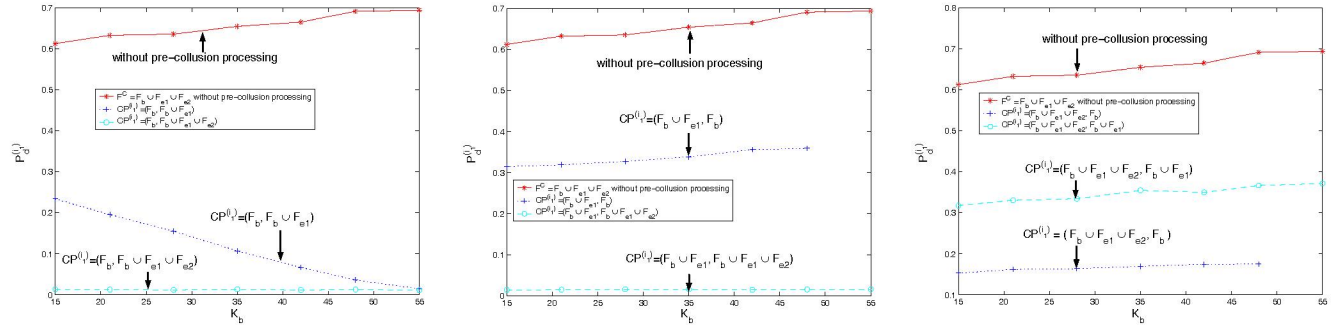
$\bar{\mu}^{(i_1)}$ in (7) does not equal to $\widetilde{\mu}^{(i_1)}$ in (9), and the colluders make an error in estimating $TN^{(i_1)}$'s mean. This is due to $\mathbf{u}^{(i_1)}$'s pre-collusion processing, and this estimation error helps the selfish colluder further lower his/her risk of being caught.

Note that when $F^c = F_b$, no matter how many frames that $\mathbf{u}^{(i_1)}$ claims that he/she has received, other colluders can always correctly estimate $TN^{(i_1)}$'s mean during collusion, and increasing the frame rate cannot help $\mathbf{u}^{(i_1)}$ further reduce his/her risk. To generalize, increasing the temporal resolution is effective in reducing $\mathbf{u}^{(i_1)}$'s probability of being captured only if $F^c \supset F^{(i_1)}$.

3.2. Reducing the Resolution Before Collusion

In this type of pre-collusion processing, a selfish colluder receives a copy of higher resolution and tells other colluders that he/she only has a copy of lower quality. For example, $\mathbf{u}^{(i_1)}$ subscribes to all three layers while claiming that he/she only has the fingerprinted base layer. In this example, $\mathbf{u}^{(i_1)}$ simply drops frames in the two enhancement layers during pre-collusion processing.

When reducing the frame rate of the received copy, the selfish colluder does not need to forge any frames and, therefore, he/she does not need to worry about the quality constraints. The analysis of the selfish colluder's probability of being detected is similar to that in Section 3.1, and reducing the frame rate of the fingerprinted copy can help the selfish colluder further lower his/her risk only if $F^c \supset \widetilde{F}^{(i_1)}$.



(a) $i_1 \in SC^b$ (b) $i_1 \in SC^{b,e1}$ (c) $i_1 \in SC^{all}$

Fig. 2. Comparison of different pre-collusion processing strategies. $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. There are a total of 450 users in the system, and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. Each K^b in the X axis corresponds to a unique triplet $(K^b, K^{b,e1}, K^{all})$ on Line (10). The colluded copy has high quality and $F^c = F_b \cup F_{e1} \cup F_{e2}$. The probability of falsely accusing an innocent is fixed at 0.01.

4. COMPARISON OF DIFFERENT STRATEGIES

This section compares the effectiveness of different strategies in reducing the selfish colluder's risk, assuming that the quality constraints are satisfied and other colluders do not discover the pre-collusion processing. As an example, we consider the scenario where the colluders generate a colluded copy of high quality with $F^c = F_b \cup F_{e1} \cup F_{e2}$ under the fairness constraints in [2].

Figure 2 shows the simulation results. In our simulations, we assume that each frame has 5000 embeddable coefficients and we test on a total of 40 frames. We consider a temporally scalable video coding system with $F_b = \{1, 5, 9, \cdots\}$, $F_{e1} = \{3, 7, 11, \cdots\}$ and $F_{e2} = \{2, 4, 8, \cdots\}$. The fingerprints follow Gaussian distribution $\mathcal{N}(0, 1/9)$, and are embedded in the DCT domain using human visual model based spread spectrum embedding [5].

During collusion, we assume that there are a total of $K = 150$ colluders and $(K^b, K^{b,e1}, K^{all})$ are on the line

$$\Big\{(K^b, K^{b,e1}, K^{all}) : K^b + K^{b,e1} + K^{all} = K, \qquad (10)$$
$$\frac{K^{all}\sqrt{N}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N}} = \frac{N_{e2}}{N},$$
$$0 \le K^b \le |\mathbf{U}^b|,\ 0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|,\ 0 \le K^{all} \le |\mathbf{U}^{all}|,\Big\}$$

where $N = N_b + N_{b,e1} + N_{e2}$. Line (10) is the boundary of one of the fairness constraints in [2]. In our simulations, we adjust the power of the additive noise such that $||\mathbf{n}_j||^2 = 2||\mathbf{W}_j^{(i)}||^2$, and assume that there is only one selfish colluder $\mathbf{u}^{(i_1)}$.

If $i_1 \in SC^b$, $\mathbf{u}^{(i_1)}$ can choose two different parameters during pre-collusion processing: $CP_1^{(i_1)} = (F_b, F_b \cup F_{e1})$ and $CP_2^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$. From Figure 2 (a), $CP_2^{(i_1)}$ gives $\mathbf{u}^{(i_1)}$ a smaller probability of being detected than $CP_1^{(i_1)}$. Thus, under the quality constraints, a selfish colluder in SC^b should pretend to have received all three layers in order to minimize his/her risk.

Similarly, for a selfish colluder in $SC^{b,e1}$, he/she can increase the resolution of the received copy with $CP_1^{(i_1)} = (F_b \cup F_{e1}, F_b \cup F_{e1} \cup F_{e2})$; and $\mathbf{u}^{(i_1)}$ can also drop the enhancement layer 1 with $CP_2^{(i_1)} = (F_b \cup F_{e1}, F_b)$. From Figure 2 (b), $CP_1^{(i_1)}$ gives $\mathbf{u}^{(i_1)}$ a smaller chance of being detected than $CP_2^{(i_1)}$ and, therefore, a selfish colluder in $SC^{b,e1}$ should increase the frame rate of his/her copy during pre-collusion processing if the colluders decide to generate a colluded copy with all three layers.

For a selfish colluder in SC^{all}, he/she can reduce the frame rate of the received copy with two different parameters before collusion: $CP_1^{(i_1)} = (F_b \cup F_{e1} \cup F_{e2}, F_b)$ and $CP_2^{(i_1)} = (F_b \cup F_{e1} \cup F_{e2}, F_b \cup F_{e1})$. Figure 2 (c) shows that $P_d^{(i_1)}$ of $CP_1^{(i_1)}$ is smaller than $P_d^{(i_1)}$ of $CP_2^{(i_1)}$. Consequently, dropping both enhancement layers before multi-user collusion is preferred for a selfish colluder in subgroup SC^{all} to minimize his/her risk of being detected.

5. CONCLUSIONS

This paper studies the problem of traitors within traitors in scalable multimedia fingerprinting, and investigates the techniques for selfish colluders to minimize their own probability of being detected when users receive copies of different quality. We show that under the quality constraints, changing the resolution of the received copy can help a selfish colluder further reduce his/her risk, especially when the colluded copy has good quality and high resolution. We also investigate the optimal processing strategy for selfish colluders in scalable fingerprinting systems to minimize their risk of being detected under the quality constraints.

6. REFERENCES

[1] M. Wu, W. Trappe, Z. J. Wang, and K. J. R. Liu, "Collusion resistant fingerprint for multimedia," *IEEE Signal Processing Magazine*, vol. 21, no. 2, pp. 15–27, March 2004.

[2] H. Zhao and K. J. R. Liu, "Fair collusion attacks on scalable video fingerprinting systems," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 2, pp. 1045–1048, March 2005.

[3] H. V. Zhao and K. J. R. Liu, "Risk minimization in traitors within traitors in multimedia forensics," *IEEE Int. Conf. on Image Processing*, vol. 3, pp. 89–92, Sept. 2005.

[4] Y. Wang, J. Ostermann, and Y. Zhang, *Video Processing and Communications*, Prentice Hall, 1st edition, 2001.

[5] C. Podilchuk and W. Zeng, "Image adaptive watermarking using visual models," *IEEE Journal on Sel. Area in Comm.*, vol. 16, no. 4, pp. 525–540, May 1998.

[6] K. Su, D. Kundur, and D. Hatzinakos, "Statistical invisibility for collusion-resistant digital video watermarking," *IEEE Tran. on Multimedia*, vol. 7, no. 1, pp. 43–51, Feb. 2005.

[7] 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. on Image Proc.*, vol. 14, no. 6, June 2005.

[8] T. Chen, "Adaptive temporal interpolation using bidirectional motion estimation and compensation," *IEEE Int. Conf. on Image Proc.*, Sept. 2002.