FAIR COLLUSION ATTACKS ON SCALABLE VIDEO FINGERPRINTING SYSTEMS

H. Vicky Zhao and K. J. Ray Liu

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

ABSTRACT

Digital fingerprinting inserts identification information in the content to track the usage of digital data and protect content security. To trace traitors for multimedia over heterogeneous networks, this paper studies scalable multimedia fingerprinting systems in which users receive fingerprinted multimedia of different quality. We investigate the cost-effective multi-user collusion on fingerprinting systems and focus on fair collusion attacks in which colluders share the same risk of being captured. In this paper, we examine the fairness constraints on collusion when attackers receive copies of different quality and analyze the performance of scalable fingerprinting systems under fair collusion attacks.

1. INTRODUCTION

To enforce the proper usage of multimedia content after delivery, digital fingerprinting labels each distributed copy with unique identification information. It enables to trace traitors who use their copies for unauthorized purposes. There is a powerful attack against digital fingerprinting systems, called *collusion attack*. During collusion, a group of users receive differently fingerprinted versions of the same content, and they work together to attenuate the original fingerprints. To support multimedia forensics, a digital fingerprinting system should be robust against such multi-user collusion attacks as well as attacks by a single adversary.

Analysis of collusion attacks provides the basis for collusion secure fingerprint design and is an important research area in digital fingerprinting. Collusion attacks on fingerprints for generic data were studied in [1]. Considering the uniqueness of multimedia that fingerprints can be seamlessly embedded into host signals, in [2] and [3], collusion attacks were modeled as the averaging attack followed by an additive noise. The collusion attack was generalized to linear shift invariant filtering followed by an additive noise in [4]. In [5] and [6], several types of collusion were studied, including a few order statistics based nonlinear attacks.

Most prior work on multimedia fingerprint design and collusion attacks assumed that users receive copies of the same quality. However, *scalability* is often required for video transmission over heterogenous networks to users with different processing capability, and it enables the receivers to partially decode the compressed bit stream and reconstruct meaningful information of the content. To fully understand the challenges in multimedia fingerprinting, it is important to consider the general scenario and study the impact of scalability on digital fingerprinting systems and collusion attacks. Taking temporal scalability as an example, this paper studies multi-user collusion when attackers receive copies of different quality. We focus on fair collusion attacks in which colluders share the same risk of being captured, and we analyze the constraints on and the effectiveness of collusion in this scenario.

This paper is organized as follows. Section 2 introduces the system model, including the temporally scalable video coding systems and the digital fingerprinting systems. In Section 3, we investigate the collusion attacks when colluders receive copies of different quality, and provide statistical analysis on the effectiveness of the collusion attacks. Section 4 shows the simulation results on real video sequences. Conclusions are drawn in Section 5.

2. SYSTEM MODEL

2.1 Temporally Scalable Video Coding Systems

Layered video coding is widely used in the literature to accommodate heterogenous networks, and it decomposes the video into non-overlapping bit streams of different priority. Figure 1 shows the block diagrams of a two-layer scalable codec. 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. We use the simple frame skipping and frame copying to implement temporal down-sampling and up-sampling, respectively. In such a video coding system, different frames are encoded in different layers[1]. As an example, frame $1, 5, 9, \cdots$ are encoded in the base layer, frame $3, 7, 11, \cdots$ are encoded in the enhancement layer 1, and frame $2, 4, 6, 8, \cdots$ are encoded in the enhancement layer 2.

Assume that F_b, F_{e1} and F_{e2} contain the indices of the frames that are encoded in the base layer, enhancement layer 1 and enhancement layer 2, respectively. Define $F^{(i)}$ as the set containing the indices of the frames that user $\mathbf{u}^{(i)}$ receives from the content owner. We further define $\mathbf{U}^b \triangleq \{\mathbf{u}^{(i)} : F^{(i)} = F_b\}$ as the subgroup of users who receive the base layer only; $\mathbf{U}^{b,e1} \triangleq \{\mathbf{u}^{(i)} : F^{(i)} = F_b \cup F_{e1}\}$ is the subgroup of users who receive the base layer and enhancement layer 1; and $\mathbf{U}^{all} \triangleq \{\mathbf{u}^{(i)} : F^{(i)} = F_b \cup F_{e1} \cup F_{e2}\}$ is the subgroup of users who receive all three layers.

2.2 Digital Fingerprinting Systems

We consider a digital fingerprinting system including fingerprint embedding, collusion attacks and colluder identification.

Fingerprint Embedding Spread spectrum embedding has been widely used in the literature due to its robustness against many attacks [7]. 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 copy 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

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

[1] For example, with MPEG-2 video coding, the base layer contains all the I frames, the enhancement layer 1 contains all the P frames, and the enhancement layer 2 contains all the B frames.

0-7803-8874-7/05/$20.00 ©2005 IEEE

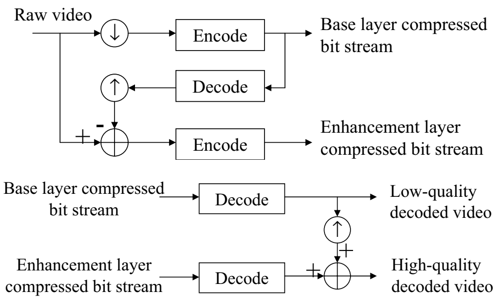

Raw video
Encode
Base layer compressed bit stream
Decode
-
+
Encode
Enhancement layer compressed bit stream
Base layer compressed bit stream
Decode
Low-quality decoded video
+
Enhancement layer compressed bit stream
Decode
+
High-quality decoded video

Fig. 1. A two-layer scalable codec. (Top): encoder, (bottom): decoder.

$\mathbf{X}_j^{(i)}$, 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 [7], and is used to control the energy of the fingerprints.

In this paper, we assume that $\{\mathbf{W}_j^{(i)}\}$ follow normal distribution with zero mean and variance σ_W^2, and assume that fingerprints for different users are independent of each other. To combat the intra-content collusion attacks including frame swapping and frame averaging, in each distributed copy, similar to the work in [8], we embed correlated fingerprints in adjacent frames, and the correlation between two fingerprints embedded in different frames depends on the similarity between the two host frames.

Collusion Attacks In a recent investigation [3], it was shown that if all collusion attacks generate colluded copies of the same quality, order statistics based nonlinear attacks have approximately the same performance as the averaging collusion. So we only consider the averaging collusion attacks in this paper.

Given that the colluders receive fingerprinted copies of different quality due to network heterogeneity, we assume that they wish to generate a colluded copy of high resolution and good quality under the constraints that every attacker has equal probability of detection. During collusion, the colluders first divide themselves into three non-overlapping subgroups:

- $SC^b \triangleq \{i : F^{(i)} = F_b\}$ contains the indices of the colluders who receive the base layer bit streams only;
- $SC^{b,e1} \triangleq \{i : F^{(i)} = F_b \cup F_{e1}\}$ contains the indices of the colluders who receive base layer and enhancement layer 1;
- and $SC^{all} \triangleq \{i : F^{(i)} = F_b \cup F_{e1} \cup F_{e2}\}$ contains the indices of the colluders who receive all three layers.

Assume that $K^b, K^{b,e1}$ and K^{all} are the number of colluders in subgroups $SC^b, SC^{b,e1}$ and SC^{all}, respectively. Secondly, the colluders apply the *intra-group 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 $SC^{b,e1}$ generate $\mathbf{Z}_j^{b,e1} = \sum_{i\in SC^{b,e1}} \mathbf{X}_j^{(i)}/K^{b,e1}$.
- For each frame $j \in F_b \cup F_{e1} \cup F_{e1}$ that they received, the colluders in SC^{all} generate $\mathbf{Z}_j^{all} = \sum_{i\in SC^{all}} \mathbf{X}_j^{(i)}/K^{all}$.

Define F^c as the set containing the indices of the frames in the colluded copy. For simplicity, we let $F^c \in \{F_b,\ F_b \cup F_{e1},\ F_b \cup F_{e1} \cup F_{e2}\}$. Finally, as shown in Figure 2, the colluders apply the *inter-group collusion attacks* to generate the colluded copy $\{\mathbf{V}_j\}$:

- For each frame $j \in F_b$ in the base layer, $\mathbf{V}_j = \beta_1 \mathbf{Z}_j^b + \beta_2 \mathbf{Z}_j^{b,e1} + \beta_3 \mathbf{Z}_j^{all} + \mathbf{n}_j$ where $\beta_1 + \beta_2 + \beta_3 = 1$. To ensure that the energy of each of the original fingerprints is reduced, we let $0 \le \beta_1,\ \beta_2,\ \beta_3 \le 1$ in this paper. $\mathbf{n}_j$ is an additive noise to further hinder the detection.

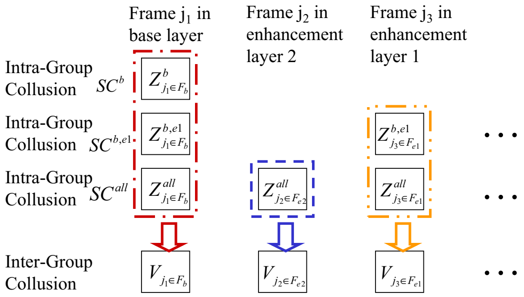

Frame j$_1$ in base layer
Frame j$_2$ in enhancement layer 2
Frame j$_3$ in enhancement layer 1
Intra-Group Collusion SC^b
Intra-Group Collusion $SC^{b,e1}$
Intra-Group Collusion SC^{all}
$Z^b_{j_1\in F_b}$
$Z^{b,e1}_{j_1\in F_b}$
$Z^{all}_{j_1\in F_b}$
$Z^{all}_{j_2\in F_{e2}}$
$Z^{b,e1}_{j_3\in F_{e1}}$
$Z^{all}_{j_3\in F_{e1}}$
Inter-Group Collusion
$V_{j_1\in F_b}$
$V_{j_2\in F_{e2}}$
$V_{j_3\in F_{e1}}$
. . .

Fig. 2. The intra-group and inter-group collusion attacks.

- If $F_{e1} \subset F^c$ and the colluded copy contains frames in the enhancement layers, then for each frame $j \in F_{e1}$ in the enhancement layer 1, $\mathbf{V}_j = \alpha_1 \mathbf{Z}_j^{b,e1} + \alpha_2 \mathbf{Z}_j^{all} + \mathbf{n}_j$ where $0 \le \alpha_1,\ \alpha_2 \le \alpha_1 + \alpha_2 = 1$ and $\mathbf{n}_j$ is an additive noise.
- If $F_{e2} \subset F^c$ and the colluded copy contains all the frames, then for each frame $j \in F_{e2}$ in the enhancement layer 2, $\mathbf{V}_j = \mathbf{Z}_j^{all} + \mathbf{n}_j$ where $\mathbf{n}_j$ is an additive noise.

Define $n_j(k)$ as the kth component of the additive noise vector $\mathbf{n}_j$. In practice, the variance of $n_j(k)$ is usually proportional to $JND_j(k)$, the corresponding just-noticeable-difference. This is because from human visual models [7], a larger $JND_j(k)$ implies that a noise with larger energy can be added to the corresponding host signal component without introducing perceptually distinguishable distortion; and the colluders usually maximize the energy of the noise $\mathbf{n}_j$ under the perceptual constraints in order to maximize the effectiveness of the collusion attacks. In this paper, we model $\{\frac{\mathbf{n}_j}{JND_j}\}$ as i.i.d. following distribution $\mathcal{N}(0, \sigma_n^2)$.

During collusion, the colluders seek the *collusion parameters*, F^c, $\{\beta_k\}_{k=1,2,3}$ and $\{\alpha_l\}_{l=1,2}$, to satisfy the fairness constraints. Detailed analysis of the fairness constraints is in Section 3.

Colluder Identification For better detection performance [3], we consider a non-blind detection scenario where the host signal is first removed from the test copy before colluder identification. Then for each frame $j \in F^c$ in the test copy, the detector extracts the fingerprint $\mathbf{Y}_j = (\mathbf{V}_j - \mathbf{S}_j)\,/JND_j$. Finally, the detector calculates the similarity between the extracted fingerprint and each of the original fingerprints, compares with a threshold and estimates the identities of the colluders $\widehat{SC}$.

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. During detection, we use a simple detector that considers fingerprints extracted from all layers collectively and calculates

$$T_N^{(i)} = \left(\sum_{j\in\breve{F}^{(i)}} \langle \mathbf{Y}_j, \mathbf{W}_j^{(i)} \rangle\right) \Big/ \sqrt{\sum_{j\in\breve{F}^{(i)}} ||\mathbf{W}_j^{(i)}||^2}, \qquad (1)$$

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

2.3 Performance Criteria

To evaluate the effectiveness of the collusion attacks and the robustness of the embedded fingerprints, we use the commonly used criteria in the literature [3]: the probability of capturing at

Table 1. Fairness Constraints on Collusion Attacks and The Selection of Collusion Parameters.

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

least one colluder (P_d) and the probability of accusing at least one innocent user (P_{fp}). Other criteria give the same trend.

To measure the perceptual quality of the colluded copy, we use $|F^c|$ that is the total number of frames in the colluded copy[2]. For simplicity, $|F^c| \in \{|F_b|, |F_b|+|F_{e1}|, |F_b|+|F_{e1}|+|F_{e2}|\}$. When $|F^c|$ is larger, the colluded copy has more frames and higher temporal resolution, and therefore, better quality.

3. FAIRNESS CONSTRAINTS AND PERFORMANCE ANALYSIS OF COLLUSION ATTACKS

3.1 Fairness Constraints on Collusion Attacks

Assume that 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, and SC is set containing indices of the colluders. We can show that for each user $\mathbf{u}^{(i)}$, the detection statistics of (1) follow Gaussian distribution $p(T_N^{(i)}|SC) \sim \mathcal{N}(\mu^{(i)}, \sigma_n^2)$, where σ_n^2 is the variance of the additive noise $\frac{\mathbf{n}_j}{JND_j}$. $\mu^{(i)} = 0$ when $\mathbf{u}^{(i)}$ is innocent, and $\mu^{(i)} > 0$ when $\mathbf{u}^{(i)}$ is guilty. For a guilty colluder $i \in SC$, $\mu^{(i)}$ depends on F^c and $F^{(i)}$.

$\mathbf{F^c = F_b \cup F_{e1} \cup F_{e2}}$ When the colluded copy contains all frames in the video, we can show that for colluder $\mathbf{u}^{(i \in SC)}$,

$$\mu^{(i)} \approx \begin{cases} \frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W, & \text{if } i \in SC^b, \\ \frac{\beta_2 N_b+\alpha_1 N_{e1}}{K^{b,e1}\sqrt{N_b+N_{e1}}}\sigma_W, & \text{if } i \in SC^{b,e1}, \\ \frac{\beta_3 N_b+\alpha_2 N_{e1}+N_{e2}}{K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}\sigma_W, & \text{if } i \in SC^{all}. \end{cases} \quad (2)$$

Detailed derivation of $\mu^{(i)}$ is available in [9].

$\mathbf{F^c = F_b \cup F_{e1}}$ When the colluded copy contains frames in the base layer and enhancement layer 1, we can approximate $\mu^{(i)}$ by

$$\mu^{(i)} \approx \begin{cases} \frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W, & \text{if } i \in SC^b, \\ \frac{\beta_2 N_b+\alpha_1 N_{e1}}{K^{b,e1}\sqrt{N_b+N_{e1}}}\sigma_W, & \text{if } i \in SC^{b,e1}, \\ \frac{\beta_3 N_b+\alpha_2 N_{e1}}{K^{all}\sqrt{N_b+N_{e1}}}\sigma_W, & \text{if } i \in SC^{all}. \end{cases} \quad (3)$$

$\mathbf{F^c = F_b}$ In this scenario, we have

$$\mu^{(i)} \approx \begin{cases} \frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W, & \text{if } i \in SC^b, \\ \frac{\beta_2\sqrt{N_b}}{K^{b,e1}}\sigma_W, & \text{if } i \in SC^{b,e1}, \\ \frac{\beta_3\sqrt{N_b}}{K^{all}}\sigma_W, & \text{if } i \in SC^{all}. \end{cases} \quad (4)$$

From the above analysis, all colluders have the same probability to be detected if their detection statistics have the same mean. Therefore, for colluders $\mathbf{u}^{(i_1)}$, $\mathbf{u}^{(i_2)}$ and $\mathbf{u}^{(i_3)}$ where $i_1 \in SC^b$, $i_2 \in SC^{b,e1}$ and $i_3 \in SC^{all}$, they seek the parameters F^c, $\{\beta_k\}_{k=1,2,3}$ and $\{\alpha_l\}_{l=1,2}$ to satisfy $\mu^{(i_1)} = \mu^{(i_2)} = \mu^{(i_3)}$.

Table 1 summarizes the constraints on the collusion attacks and the selection of the collusion parameters for three different scenarios, where the colluded copy has the highest, medium and lowest frame rates, respectively. Detailed analysis is in [9].

From Table 1, generating a colluded copy of higher resolution and better quality puts more severe constraints on collusion. When the colluded copy has higher resolution, the fairness constraints require that there are more attackers in subgroups $SC^{b,e1}$ and SC^{all} and more colluders receive enhancement layers.

To check the fairness constraints and 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 $N_b : N_{e1} : N_{e2} \approx |F_b| : |F_{e1}| : |F_{e2}|$.

3.2 Performance Analysis

Assume that there are a total of M users and a total of K colluders. If the colluders choose the collusion parameters as in Table 1, then given a colluder set SC, for each user $\mathbf{u}^{(i)}$,

$$p(T_N^{(i)}|SC) \sim \begin{cases} \mathcal{N}(\mu,\ \sigma_n^2) & \text{if } i \in SC, \\ \mathcal{N}(0,\ \sigma_n^2) & \text{if } i \notin SC, \end{cases} \quad (5)$$

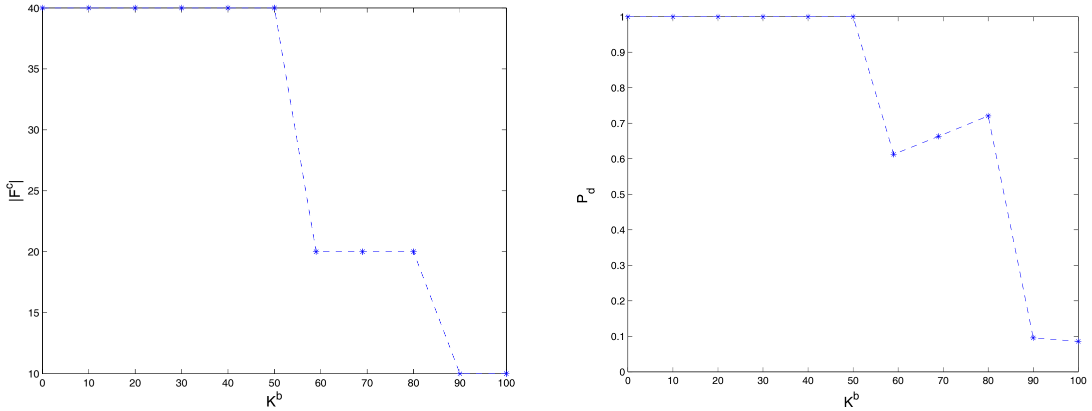
where σ_n^2 is the variance of $\mathbf{n}_j/JND_j$. The M detection statistics $\{T_N^{(i)}\}_{i=1,\cdots,M}$ are independent of each other since the M fingerprints assigned to different users are generated independently. We can show that for $i \in SC$, μ in (5) can be approximated by

$$\mu \approx \begin{cases} \frac{N_b+N_{e1}+N_{e2}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}\sigma_W \\ \qquad\qquad \text{if } F^c = F_b \cup F_{e1} \cup F_{e2}, \\ \frac{N_b+N_{e1}}{K^b\sqrt{N_b}+(K^{b,e1}+K^{all})\sqrt{N_b+N_{e1}}}\sigma_W \quad \text{if } F^c = F_b \cup F_{e1}, \\ \frac{\sqrt{N_b}}{K^b+K^{b,e1}+K^{all}}\sigma_W \qquad \text{if } F^c = F_b. \end{cases}$$

Given a threshold h, we can approximate P_d and P_{fp} by $P_d \approx 1-(1-Q(\frac{h-\mu}{\sigma_n}))^K$ and $P_{fp} \approx 1-(1-Q(\frac{h}{\sigma_n}))^{M-K}$.

From the above analysis, the effectiveness of the collusion attacks depends on the resolution of the colluded copy. When the

[2] $|A|$ denotes the size of the set A.



(a) $|F^c|$: the total number of frames in the colluded copy (b) P_d: the probability of capturing at least one colluder

Fig. 3. Simulation results on the first 40 frames of sequence "carphone". Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. $(N_b, N_{e1}, N_{e2}) = (72222, 71926, 143820)$. $K = 150$ and $(K^b, K^{b,e1}, K^{all})$ are on the line (6). $P_{fp} = 10^{-3}$.

colluded copy has a higher frame rate and better quality, the extracted fingerprint is longer and provides more information of the colluders' identities. Thus, the colluders have a larger probability to be captured. The colluders have to consider the tradeoff between the probability of detection and the quality of the colluded copy.

4. SIMULATION RESULTS

In our simulations, we test on the first 40 frames of the sequence "carphone" and choose $F_b = \{1, 5, 9, \cdots\}$, $F_{e1} = \{3, 7, 11, \cdots\}$, and $F_{e2} = \{2, 4, 6, \cdots\}$ as an example of the temporal scalability. At the content owner's side, we use the human visual model based spread spectrum embedding in [7] and embed fingerprints in the DCT domain. The fingerprints follow distribution $\mathcal{N}(0, \sigma_W^2)$ with $\sigma_W^2 = 1/9$. The length of the fingerprints embedded in the base layer, enhancement layer 1 and enhancement layer 2 are $N_b = 72222$, $N_{e1} = 71926$ and $N_{e2} = 143820$, respectively. We assume that there are a total of $M = 450$ users, and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$.

During collusion, we assume that the collusion attack is also in the DCT domain and we fix the total number of colluders $K = 150$. $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and they are on the line

$$\frac{K^{all}\sqrt{N_b + N_{e1} + N_{e2}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} = \frac{N_{e2}}{N_b + N_{e1} + N_{e2}}, \qquad (6)$$

which is the boundary of the fairness constraints in Table 1. For each frame j in the colluded copy, We adjust the power of the additive noise $\mathbf{n}_j$ such that $||\mathbf{n}_j / JND_j||^2 = 2||\mathbf{W}_j^{(i)}||^2$, where JND_j is the just-noticeable-difference from human visual models. We assume that the colluders generate a colluded copy of the highest possible quality under the fairness constraints.

We consider a non-blind detector where the host signal is first removed from the colluded copy. The detector then follows the detection procedure in Section 2 to identify the colluders.

Figure 3 shows the simulation results. Figure 3 (a) plots the total number of frames in the colluded copy when $(K^b, K^{b,e1}, K^{all})$ changes on the line (6), and Figure 3 (b) shows the corresponding probability of capturing at least one colluder. From Figure 3, when the colluded copy contains more frames and higher resolution, the detector has more information of the colluders' identities, and therefore, the colluders have a larger probability to be detected. This is consistent with our analysis in Section 3.

5. CONCLUSIONS

In this paper, we have studied fair collusion attacks on scalable fingerprinting systems in which users receive fingerprinted copies of different quality. We have shown that the fairness constraints on collusion are more severe when generating a colluded copy of higher resolution. In addition, both our analytical and simulation results have shown that the colluders have a larger risk and are more likely to be detected when they generate a colluded copy of higher resolution and better quality. During collusion, the colluders have to take into consideration this tradeoff between the probability of detection and the perceptual quality.

6. REFERENCES

[1] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Tran. on Information Theory*, vol. 44, no. 5, pp. 1897–1905, Sept. 1998.

[2] F. Ergun, J. Killian, and R. Kumar, "A note on the limits of collusion-resistant watermarks," *Advances in Cryptology – EuroCrypto '99, Lecture Notes in Computer Science*, vol. 1592, pp. 140–149, 2001.

[3] Z. J. Wang, M. Wu, H. Zhao, W. Trappe, and K. J. R. Liu, "Resistance of orthogonal gaussian fingerprints to collusion attacks," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, April 2003.

[4] J. Su, J. Eggers, and B. Girod, "Capacity of digital watermarks subject to an optimal collusion attacks," *European Signal Processing Conference (EUSIPCO 2000)*, 2000.

[5] H. Stone, "Analysis of attacks on image watermarks with randomized coefficients," Tech. Rep. 96-045, NEC Research Insistute, 1996.

[6] H. Zhao, M. Wu, Z. J. Wang, and K. J. R. Liu, "Nonlinear collusion attacks on independent fingerprints for multimedia," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, April 2003.

[7] 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.

[8] K. Su, D. Kundur, and D. Hatzinakos, "Statistical invisibility for collusion-resistant digital video watermarking," *to appear in IEEE Tran. on Multimedia*.

[9] H. Zhao and K. J. R. Liu, "Multi-user collusion on scalable fingerprinting systems in multimedia forensics: Fairness constraints and effectiveness analysis," *submitted to IEEE Tran. on Information Forensics and Security*.