ABSTRACT

Title of Dissertation: Multimedia Fingerprinting for Multiuser Forensics and Security

Hong Zhao, Doctor of Philosophy, 2004

Dissertation directed by: Professor K. J. Ray Liu
Department of Electrical and Computer Engineering

Recent development in multimedia and network technologies has made possible the ubiquitous sharing and distribution of multimedia over networks. However, illegal alteration and unauthorized copying of multimedia data pose serious threats to multimedia security and intellectual property rights, especially considering the ease manipulation of digital data. Therefore, it is critical to secure and protect multimedia content, and to ensure the integrity of rights by authorized users solely for intended purpose. Digital fingerprinting is an emerging technology to address post-delivery content protection and to enforce digital rights. In digital fingerprinting, unique identification information is embedded in each distributed copy, and is used to trace and identify the source of illicit copies. Such a traitor tracing is a fundamental problem in multimedia forensics, as well as an important tool for enforcing digital rights.

This thesis addresses various issues in multimedia fingerprinting. We first investigate the order statistics based nonlinear collusion attacks on digital fingerprinting, and analyze their effectiveness in defeating the fingerprinting systems. We also compare the performance of several commonly used detection statistics under collusion. We then examine the impact of scalable video coding and transmission on digital fingerprinting systems and collusion attacks. We analyze the effectiveness of the collusion attacks under the constraints that all colluders have equal probability of detection, and analyze the collusion resistance of scalable fingerprinting systems. We then consider the problem of traitors within traitors in digital fingerprinting, in which some selfish colluders wish to minimize their own risk of being captured while still profiting from the illegal redistribution of multimedia. We investigate the possible strategy by the selfish colluders to reduce probability of detection, and analyze their performance under the quality constraints. We also investigate the possible countermeasures by other colluders to protect their own interest. Finally, we investigate the secure distribution of fingerprinted copies for video streaming applications, and propose two secure fingerprint multicast schemes. We analyze their performance, including the communication cost and the robustness against collusion attacks, and discuss the tradeoff between the bandwidth efficiency and computation complexity.

Multimedia Fingerprinting for Multiuser Forensics and Security

by

Hong Zhao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2004

Advisory Committee:

Professor K. J. Ray Liu, Chairman
Professor Carlos A. Berenstein
Professor Virgil D. Gligor
Professor Z. Jane Wang
Professor Lawrence C. Washington
Professor Min Wu

DEDICATION

To my parents: Mozi Chen and Qifa Zhao.

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my advisor, Prof. K. J. Ray Liu, for his guidance and support during my study in University of Maryland. He always encourages me to pursue my goal and work hard to achieve excellence. I especially appreciate his effort to help his students and give them advice whenever they need. He has played a significant role in both my professional and personal development in Maryland, and his vision, energy and desire for excellence have influenced me with lifetime benefits.

I am also indebt to Prof. Min Wu, Prof. Z. Jane Wang and Prof. Wade Trappe for their help and support during my early stage of research. Working with them has tremendously helped me to understand the broad area of signal processing, and most importantly, become a professional researcher.

I would like to take this chance to thank members in the CSPL group for their friendship, encouragement and help. I always feel lucky to be in such an energetic and excellent group, and their accompanying during my stay in Maryland has helped me to survive my Ph.D study. Special thanks to my officemates: Zoltan Safar, Yan Sun, Johannes Thorsteinsson, Wei Yu, Hongmei Gou, Zhu Ji, and Ahmed Sadek. Thank you for the time spent together, and the happy time in our office will be always in my memory.

I am grateful to Zhanfeng Yue, my husband, for his love and support during our study in University of Maryland. It is because of his help and encouragement

that my life in the past years is much more joyful and colorful. I would also like to thank my parents-in-law and sister-in-law for their care and support.

Finally, I give my heartfelt gratitude to my parents, my role model and the two most important persons in my life. Without their love, unconditional support and countless sacrifices, I could never accomplish so much and reach this milestone in my life. I dedicate this thesis to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Motivation

In the past decades, we have witnessed the revolution of digital information technology and its significant impact on our daily lives. The popularity of digital camera, digital camcorder, MP3 player and DVD player, have inspired people all over the world to create and enjoy multimedia in digital domain. Furthermore, the ubiquity of broadband networks and the advance in multimedia technologies have proliferated the delivery and sharing of multimedia data over networks.

However, illegal alteration, repackaging and unauthorized redistribution of multimedia have serious consequences on governmental and military operations as well as commercial applications. Attackers can easily alter the multimedia content, produce copies of high quality and redistribute without authorization, which threatens multimedia security and intellectual property rights. Consequently, it is critical to secure multimedia transmission and to protect the rights of content providers.

Take commercial applications as an example, the U.S. copyright industries, including pre-recorded records and tape, motion pictures and videos, play a key

role in the U.S. economy. In Year 2002, the estimated value added for the copyright industries was $514.4 billion and 4.91% of the U.S. Gross domestic Product (GDP). However, piracy drastically affected the sales revenue for these copyright industries in Year 2002. For example, total foreign sales revenue for the copyright industries grew by only 1.1% from 2001 to 2002 – a dramatic decline from 1999 where growth was at 14.5% from 1998, largely attributed to piracy [51]. Consequently, content protection and digital rights enforcement are crucial to safeguard this valuable economic resource of copyright industries.

Digital rights management systems incorporate encryption, conditional access, copy control, and media identification and tracing, and aim to protect the multimedia security and the intellectual property rights [5, 27, 46]. Some important standardization groups and bodies that have been working on DRM systems and the integration of security into multimedia frameworks are the International Organization for Standardization (ISO) MPEG, Secure Digital Music Initiative (SDMI), DVD/Copy Protection Technical Working Group (CPTWG), *C, Open Platform Initiative for Multimedia Access (OPIMA), Digital Video Broadcasting (DVB), Digital Audio-Visual Council (DAVIC), Bluetooth Special Interest Group, TV anytime, etc. [27]

Access control and multimedia forensics are two fundamental modules in digital rights management systems to protect content security and prohibit unauthorized alteration and distribution of multimedia data. First, *encryption and access control* protects the secure transmission of multimedia information over networks and controls access to multimedia content [21,39,49,57,70]. Secondly, *multimedia forensics* helps a digital rights enforcer to detect the illegal tampering on multimedia content and to identify the people who generate the illicit copies [5, 27, 46, 71]. These two

approaches are complementary to each other: access control prevents unauthorized users from accessing multimedia content, while multimedia forensics detects, and therefore thwarts, illegal manipulation and redistribution of multimedia by users who have access to the clear text representation.

Digital watermarking is one emerging technology in multimedia forensics and offers the protection of multimedia content after the data are decrypted into clear text [12, 35, 43, 69]. In digital watermarking, a secondary information, often called *watermark*, is seamlessly attached to the primary multimedia data (also called *host signal*), and can be used for various purposes (e.g., ownership protection and authentication) depending on the applications and requirements. Compared with other possible solutions, digital watermarking has the advantage that the embedded watermark is seamlessly bounded to and travels with the host signal, which is desirable in many applications.

Digital fingerprinting is one application of digital watermarking, whose purpose is to trace the distribution of multimedia and identify the source of illicit copies [13, 58, 71]. Such a traitor tracing technique forces culprits to be responsible for their behavior, and is a fundamental tool in multimedia forensics. In digital fingerprinting, unique identification information is embedded in each distributed copy and serves as a digital fingerprint. Digital fingerprinting applications require that the embedded fingerprints can survive both common signal processing and intentional attacks, and therefore, the content owner can still detect the identities of the attackers with little ambiguity even if the data have been severely distorted.

In addition to the civilian usage in digital rights enforcement, digital fingerprinting can also be used in military applications. One example is to protect digital maps that contain classified and important information for military and

intelligence agencies [24, 25].

Due to the uniqueness of each distributed copy, there is a cost effective attack against digital fingerprinting systems, *collusion* attacks, by several users who receive copies of the same content but embedded with different fingerprints [3,13,52]. During collusion, the attackers (colluders) gather together, combine information from different copies and generate a new copy where the original fingerprints are removed or attenuated. If not properly designed, a fingerprinting system might fail to detect the traces of any fingerprints under collusion attacks with only a few colluders. Consequently, multiuser collusion poses new challenges on multimedia forensics, and a digital fingerprinting system should not only survive attacks on a single copy [14, 28, 42], but also be robust against multiuser collusion attacks.

In addition, the uniqueness of each distributed copy also challenges the secure and efficient distribution of the uniquely fingerprinted copies over networks, especially for video streaming applications where a large volume of data have to be transmitted to a large number of users under stringent delay constraints [1, 31, 65]. A simple solution of unicasting each fingerprinted copy to the corresponding user is inefficient, since the required bandwidth grows linearly as the number of users increases while the difference between fingerprinted copies is small. Multicast technology provides a bandwidth advantage when distributing the same content to multiple users [7, 11]. It reduces the overall communication cost by duplicating packages only when routing paths to multiple receivers diverge [41, 59]. However, traditional multicast technology is designed to transmit the same data to multiple users, and it cannot be directly applied to fingerprinting applications where different users receive slightly different copies. This calls for new distribution schemes for multimedia fingerprinting, in particular, for networked video applications.

This thesis addresses the issues regarding traitor tracing in multimedia forensics and studies various aspects of multimedia fingerprinting.

1.2 Prior Art

The prior work in digital fingerprinting for multimedia forensics can be roughly divided into three major areas: analysis of the effectiveness of the collusion attacks and the collusion resistance of fingerprinting systems; design of anti-collusion multimedia fingerprinting systems that jointly consider the multimedia fingerprint code design, embedding and detection; and investigation of the secure and efficient distribution of anti-collusion fingerprinted copies over networks.

Analysis of the Collusion Attacks

An important research area in digital fingerprinting is to study the effectiveness of collusion attacks. It helps to understand the collusion resistance of a digital fingerprinting system and plays an important role in the design of anti-collusion fingerprinting systems.

An early work on collusion attack and digital fingerprint code design for generic data was proposed in [3], which assumed that the colluders can detect a specific fingerprint code bit if it takes different values between their fingerprinted copies and can change it to any value. For those bits where different copies have the same value, it was assumed that the colluders cannot change an undetected bit without rendering the object useless.

Unlike generic data, multimedia has the unique characteristics that minor variations on the values will not introduce perceptually noticeable distortion. This robustness makes it feasible and desirable to embed fingerprints seamlessly into the

host multimedia data. Fingerprint codes designed based on the above assumptions are usually too long to be reliably embedded into and extracted from multimedia data. For generic data, colluders can easily detect a fingerprint code bit if it differs between different copies and change it to any value. However, for multimedia data such as images, the embedding is capable of spreading each fingerprint code bit over the entire content. Thus, different bits embedded additively over the same region are not distinguishable, neither can they be changed to any value due to the perceptual quality constrain. Consequently, the above assumptions of the collusion attacks are not always suitable for multimedia data. Instead, the average attack and those order statistics based nonlinear collusion attacks in [52] are more common when colluding multimedia data.

In [20], the collusion attack was modeled as averaging different copies followed by an additive noise, and $O(\sqrt{N/\log N})$ colluders were shown to be enough to break the fingerprinting system where N is the fingerprint length. Similar results were given in [32]. The work in [64] studied the relationships between the maximum allowable colluders by a fingerprinting system and other parameters, e.g., the fingerprint length, the total number of user and system performance requirements. The collusion attack model was generalized to linear shift invariant filtering followed by an additive noise in [54].

In [52], several types of collusion attacks were studied, including a few order statistics based nonlinear attacks. For uniformly distributed fingerprints, nonlinear collusion attacks were shown to defeat the fingerprinting system more effectively than the average attack [52]. Simulation results in [52] also showed that normally distributed fingerprints are more robust against nonlinear collusion attacks than uniform fingerprints, but analytical study on the Gaussian fingerprint's perfor-

mance was not provided.

Anti-Collusion Fingerprint Design

The ultimate goal of analyzing collusion attacks is to design anti-collusion fingerprinting systems for multimedia forensics. In an early work on collusion secure fingerprint code design for generic data [3], fingerprint codes of length $O(K^4 \log K)$ were proposed to catch at least one out of at most K colluders with an arbitrarily high probability. Similar work was presented in [9], which focused on tracing the leakage of decryption keys in broadcast instead of tracing multimedia content.

Improvement was made upon the fingerprint code in [3] by replacing the lower layer code with direct spread spectrum sequence in [73]. It relaxed the assumptions in [3] and increased the total number of users that can be supported by three times. In [22] and [44], new features were introduced in the fingerprint code in [3], such as dynamic code design and asymmetric fingerprinting.

To address the unique characteristics of multimedia where it is feasible and desired to embed the fingerprints seamlessly into the host signal, a two-layer fingerprinting design scheme for multimedia was proposed in [75] where the inner code from spread spectrum embedding [13,47] is combined with an outer error-correcting code (ECC). The work in [29] jointly considered the fingerprint code design and multimedia fingerprint embedding and studied the performance of error correction code (ECC) based fingerprinting systems. In [19], the finite projective geometry was used to generate codes whose overlap with each other can identify colluding users. In [58], combinatorial theories were used to design the Anti Collusion Code (ACC), and several colluder identification schemes were proposed with different performance tradeoff. In [62], group oriented fingerprinting was proposed where

prior knowledge of the possible collusion patterns was used to improve the collusion resistance of the fingerprinting systems. Observing that some colluders are more likely to collude with other due to social or geographical reasons, the group oriented fingerprint design introduced a well-controlled amount of correlation into the fingerprints assigned to different users to enhance the traitor tracing capability.

Secure Fingerprint Multicast

To address the secure fingerprint multicast issue, in [10], a two layer fingerprint design was used where the inner layer of spread spectrum embedding [13] was combined with the outer fingerprint code of [3]. Two uniquely fingerprinted copies were generated, encrypted and multicasted, where each frame in the two copies was encrypted with a unique key. Each user was given a unique set of keys for decryption and reconstructed a unique sequence. Although their scheme reduced the bandwidth requirement, their fingerprinting system was vulnerable to collusion attacks. From their reported results, for a two hour video distributed to $10,000$ users, only when no more than three users colluded could their system detect at least one colluder correctly with probability 0.9. Similar work was presented in [6, 33, 40].

In [4], the fingerprint design was similar to that of [10], and the sender generated and multicasted several uniquely fingerprinted copies. In their work, trusted routers in the multicast tree forwarded differently fingerprinted packets to different users. In [30], a hierarchy of trusted intermediaries was introduced into the network. All intermediaries embedded their unique IDs as fingerprints into the content as they forwarded the packets through the network, and a user was identified by all the IDs of the intermediaries that were embedded in his received copy.

In [72], fingerprints were embedded in the DC coefficients of the luminance component in I frames using spread spectrum embedding. For each fingerprinted copy, a small portion of the MPEG stream, including the fingerprinted DC coefficients, was encrypted and unicasted to the corresponding user. Their distribution scheme achieved the bandwidth efficiency by multicasting the rest of the video content to all users. Since the fingerprints were only embedded in a small number of coefficients and were of short length, the robustness against collusion attacks was limited.

A joint fingerprinting and decryption scheme was proposed In [34]. In their work, the content owner or the service provider encrypted the perceptually relevant features extracted from the host signal with a secret key K_S known to the content owner/service provider only, multicasted the encrypted content to all users, and transmitted to each user i a unique decryption key $K_i \neq K_S$. At the receiver's side, user i could only partially decrypt the received encrypted bit stream, and each user reconstructed a different version of the original host signal due to the uniqueness of the decryption key K_i. In [34], the fingerprint information was essentially the asymmetric key pair (K_S, K_i), and the unique signature from the partial decryption was used to identify the attacker/colluders.

1.3 Thesis Overview and Contributions

This thesis focuses on the study of the cost effective multiuser collusion on multimedia fingerprinting as well as secure fingerprint multicast in networked video applications.

We begin, in Chapter 2, with an introduction of digital fingerprinting for multimedia forensics. We first introduce the general framework of multimedia finger-

printing systems, and then discuss the multimedia fingerprint design and embedding. In particular, we focus on spread spectrum embedding that is widely used in multimedia fingerprinting, and review multimedia fingerprint design that jointly considers fingerprint code design and embedding. We also consider several possible applications of multimedia fingerprinting and study their system requirements.

In Chapter 3, we address nonlinear collusion attacks on Gaussian fingerprints. We first consider from the colluders' point of view and compare various nonlinear collusion attacks on independent Gaussian fingerprints. We analyze the effectiveness of the collusion attacks and the perceptual quality of the colluded signals. We then shift our role to desinger/detector and analyze the performance of several commonly used detection statistics [48, 52, 76] in the literature under collusion attacks. We also propose a preprocessing technique that improves the detection performance by utilizing the statistical features of the extracted fingerprints.

Most previous work on fingerprint code design and collusion attacks for multimedia assumed that all the colluders receive fingerprinted copies of the same quality. In practice, due to the heterogeneity of the networks and the end users, it is often required to have *scalability* during video encoding and transmission, which enables the users to recover physically meaningful information by partially decoding compressed bit streams [60]. In Chapter 4, taking temporal scalability as an example, we examine the impact of the scalability on multimedia fingerprinting and collusion attacks. We consider fair collusion attacks where all colluders have equal probability of detection, and analyze the effectiveness of the collusion under the fairness constraints when different colluders receive copies of different quality. We also investigate the collusion resistance of the scalable fingerprinting systems, and analyze the number of colluders that are required to undermine the tracing

capability of the scalable fingerprinting systems.

In addition, most prior work relied on the assumption that all colluders tell each other the true information of their received copies during collusion, and they are willing to share the same risk of being captured. However, some colluders might be selfish and wish to minimize their own risk while still profiting from collusion. To reduce their probability of detection, they process their received copies before multiuser collusion, and hide information of their fingerprinted copies from other colluders. In Chapter 5, we investigate the possible strategy that the selfish colluders can use to reduce probability of detection, analyze their performance, and find the optimal pre-collusion processing that minimizes the selfish colluder's risk under the quality constraints. We will also investigate the possible countermeasures by other colluders to protect their own interests and prevent those selfish colluders from processing their copies before collusion.

In Chapter 6 and 7, we address secure fingerprint multicast in networked video applications. Most prior work in fingerprint multicast considered applications where the goal of the fingerprinting system is to be resistant to collusion attacks with a few colluders, e.g., seven or ten traitors, and designed the efficient distribution schemes accordingly. In many video applications, there are a large number of users (e.g., several thousand users), and therefore, potentially a large number of colluders (e.g., a few dozen or maybe even one hundred colluders). Some prior work [58, 61, 62] has shown that with proper design and embedding of the fingerprints, the fingerprinting systems can resist collusion attacks with dozens of colluders, e.g., up to 60 colluders. In Chapter 6, we consider video applications where the fingerprinting system aims to survive collusion attacks with dozens of or even a hundred colluders, adopt the fingerprinting systems with strong traitor

tracing capability [58, 62], and investigate the secure and efficient distribution of fingerprinted copies. In particular, we propose two secure fingerprint multicast scheme: a general fingerprint multicast scheme that can be used with most spread spectrum embedding based fingerprinting systems, and a joint fingerprint design and distribution scheme that utilizes the special structure of the fingerprint design to further reduce the communication cost. In Chapter 7, we analyze the performance of these two fingerprint multicast schemes, including the bandwidth efficiency and the robustness of the embedded fingerprints against collusion attacks. We also analyze the quality of the reconstructed sequences, and propose a fingerprint drift compensation scheme to improve the quality of the reconstructed frames at the receiver's side without extra communication overhead.

Finally, we draw conclusions and discuss some possible future directions in Chapter 8.

Chapter 2

Multimedia Fingerprinting System Overview

2.1 General Framework of Digital Fingerprinting Systems

Figure 2.1 shows a general framework for digital fingerprinting, which consists of three parts: fingerprint embedding, multiuser collusion attacks and colluder identification.

Fingerprint Embedding

Starting with an original copy of the host signal $\mathbf{S}$, the content owner or the service provider generates a unique fingerprint $\mathbf{W}^{(i)}$ for each user $\mathbf{u}^{(i)}$ in the system, and embeds it into the fingerprinted copy $\mathbf{X}^{(i)}$ that will be distributed to $\mathbf{u}^{(i)}$. For the purpose of traitor tracing in digital fingerprinting applications, the fingerprint embedding has to satisfy the following requirements:

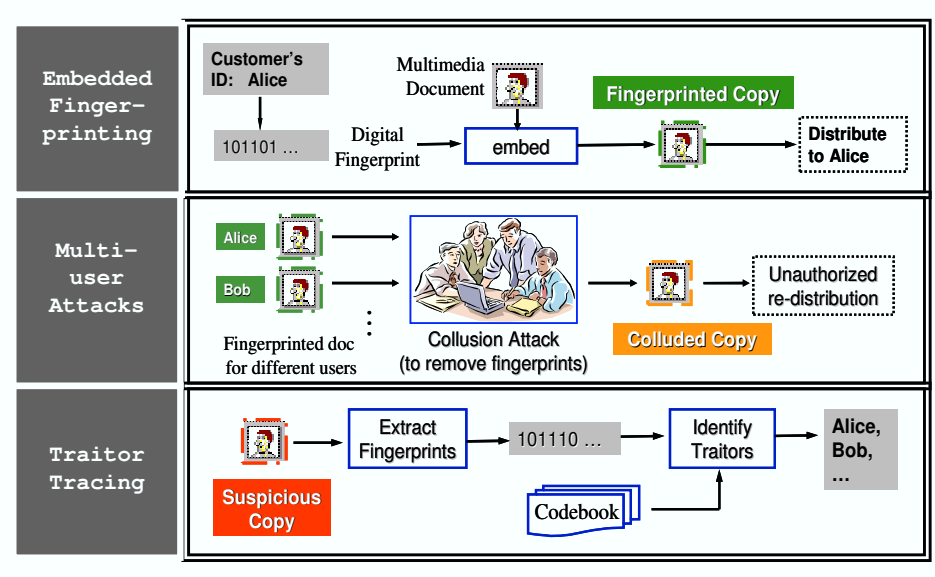


Figure 2.1: The general framework for digital fingerprinting.

- *Imperceptibility*: The fingerprinted copy $\mathbf{X}^{(i)}$ that is distributed user $\mathbf{u}^{(i)}$ is perceptually the same as the original host signal $\mathbf{S}$, and the embedded fingerprint $\mathbf{W}^{(i)}$ should not introduce perceptually noticeable distortion into the host signal $\mathbf{S}$.

- *Security*: The embedded fingerprint $\mathbf{W}^{(i)}$ should only be known to and accessed by authorized party. According to the Kerckhoff's assumption in cryptography [39], for a fingerprinting system that requires a very high level of security, the fingerprinting system designer must assume that the adversary has complete knowledge of the fingerprinting algorithm, and the secrecy of the embedded fingerprints relies only on the secret keys that are used to generate the unique fingerprints.

- *Robustness*: The fingerprints must persist in the host data after manipulation, including both unintentional signal processing (e.g., compression) and

intentional attacks to remove/attenuate the fingerprints (e.g., collusion attacks).

After the fingerprint embedding, the content owner or the service provider distributes the fingerprinted copy $\mathbf{X}^{(i)}$ to $\mathbf{u}^{(i)}$.

Multiuser Collusion Attacks

At the attackers' side, the colluders apply multiuser collusion attacks to the fingerprinted copies that they receive, and try to remove or attenuate the embedded fingerprints. A simple example of the collusion attack is to average all the fingerprinted copies, and each fingerprint's energy is reduced by a factor of $\frac{1}{K^2}$, where K is the total number of colluders. The colluders can also apply order statistics based nonlinear collusion attacks, e.g., taking the minimum values of the corresponding components in the K copies. In this thesis, we consider *fair* multiuser collusion attacks, where all colluders share the same risk and have the same probability to be captured.

In addition to the multiuser collusion, the colluders can also apply single-copy attacks, e.g., low pass filtering and compression, to further hinder the detection process. Then, the newly generated colluded copy is redistributed without authorization.

Fingerprint Detection and Colluder Identification

When the content owner discovers the existence of the illegally redistributed colluded copy, he applies a fingerprint detection and colluder identification process to the suspicious copy. The detector first extracts the fingerprint $\mathbf{Y}$ from the suspicious copy, compares this extracted fingerprint $\mathbf{Y}$ with each of the original

fingerprints $\{\mathbf{W}^{(i)}\}$, and estimates the identities of the colluders.

Depending on the presence of the host signal $\mathbf{S}$ during the colluder identification process, there are two main detection scenarios in data hiding applications, blind and non-blind detection, respectively [46, 58]. 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 signal before detection. Compared with the blind detection, previous work has shown that non-blind detection has better detection performance due to the following two reasons. First, compared with blind detection, the non-blind detection first removes the host signal from the test copy before fingerprint detection, and therefore, significantly reduces the energy of the noise during the detection process [46, 58, 64]. In addition, in the non-blind detection scenario, the detector can use the host signal to estimate the possible modifications by the attackers, and therefore, compensate accordingly. For example, by registering the test copy with respect to the original host signal, the detector can successfully undo the geometric attacks with a very small alignment noise [38, 46].

In many data hiding applications, the host signal is often not available to the detector and blind detection is preferred or even required [46]. For example, when proving ownership of multimedia data, the host signal itself is questionable and the blind detection must be applied [16, 76]. However, for many fingerprinting applications, the fingerprint verification and colluder identification process is usually handled by the content owner or an authorized third party who can have access to the original host signal. Therefore, the host signal can be regarded as available to the detector and the non-blind detection is feasible for fingerprinting applications.

To improve the detection performance, in this thesis, non-blind detection is chosen, and we further assume that the test copy has been registered to the original host signal before the detection process.

2.2 Multimedia Fingerprint Design and Embedding

In this section, we first introduce spread spectrum embedding that is widely used in digital fingerprinting systems, and then discuss multimedia fingerprint design that jointly considers the encoding, embedding, and detection of fingerprints in multimedia fingerprinting systems.

2.2.1 Spread Spectrum Embedding

Spread spectrum watermark/fingerprint embedding borrows the idea of spread spectrum modulation in communication systems, and is widely used in digital watermarking and fingerprinting systems due to its robustness against many attacks [12, 13, 47]. It fits watermarking into the traditional model of a communication system, where the watermark is regarded as the message that is to be sent from the watermark embedder to the watermark detector, and the modifications to the watermarked copies (both unintentionally and intentionally) are modeled as the noise in the channel during transmission. If blind detection is applied at the detector's side, then the host signal is also considered as one source of the noise.

In additive spread spectrum embedding, depending on the applications and the design requirement, the watermark can be embedded in the spatial domain [26], the frequency domain (e.g., DCT or DWT) [13, 47], or the feature points selected

from the host signal [34]. Assume that $\mathbf{S}$ is the host signal represented by a vector of length N, and $\mathbf{W}$ is the watermark of the same length N to embed. The watermarked copy $\mathbf{X}$ is generated by

$$\mathbf{X}_j = \mathbf{S}_j + JND_j\mathbf{W}_j, \tag{2.1}$$

where $\mathbf{X}_j$, $\mathbf{S}_j$ and $\mathbf{W}_j$ are the jth components of the watermarked copy, the host signal and the watermark, respectively. JND is the *just-noticeable-distortion* from human visual models [12,47], and it controls the energy and achieve the imperceptibility of the embedded watermarks.

At the detector's side, given the suspicious copy $\mathbf{Y}$ of length N, to test the presence of the watermark $\mathbf{W}$ in $\mathbf{Y}$, the detection process can be modeled as a hypothesis testing problem [68]:

$$\begin{cases} H_0 : \mathbf{Y}_j = \mathbf{n}_j & (j = 1, \cdots, N) \quad \text{if watermark is absent,} \\ H_1 : \mathbf{Y}_j = \mathbf{W}_j + \mathbf{n}_j & (j = 1, \cdots, N) \quad \text{if watermark is present.} \end{cases} \tag{2.2}$$

In (2.2), the deterministic signal $\mathbf{W}$ is the watermark to test, $\mathbf{n}$ is an additive noise that comes from signal processing as well as attacks on the watermarked copy $\mathbf{X}$, and N is the number of the coefficients that carry the watermark information. If $\mathbf{n}$ is modeled as i.i.d. Gaussian $\mathcal{N}(0, \sigma_n^2)$, then from the detection theory [48], the optimum detector is the matched filter

$$T_N = < \mathbf{Y}, \mathbf{W} > / ||\mathbf{W}||, \tag{2.3}$$

where $||\mathbf{W}||$ is the Euclidean norm of $\mathbf{W}$. The detection statistics T_N follow Gaussian distribution

$$T_N \sim \begin{cases} \mathcal{N}(0, \sigma_n^2) & \text{if watermark is absent,} \\ \mathcal{N}(||\mathbf{W}||, \sigma_n^2) & \text{if watermark is present.} \end{cases} \tag{2.4}$$

Then, T_N is compared with a threshold h and the detector decides H_1 if $T_N > h$ and H_0 otherwise. The threshold h can be set according to the Bayesian rule or the Neyman-Pearson rule [48], depending on the requirement of the applications.

The above hypothesis testing is to test if a watermark $\mathbf{W}$ is present or absent. Another popular model considers the scenario where an one bit information is embedded using antipodal model [58, 68]. Assume that $\mathbf{d}$ is a deterministic sequence, and $b = \{-1, +1\}$ is the one bit information to embed, then the detection problem can be modeled as:

$$\begin{cases} H_0 : \mathbf{Y}_j = -\mathbf{d}_j + \mathbf{n}_j \quad (j = 1, \cdots, N) \quad \text{if } b = -1, \\ H_1 : \mathbf{Y}_j = +\mathbf{d}_j + \mathbf{n}_j \quad (j = 1, \cdots, N) \quad \text{if } b = +1. \end{cases} \tag{2.5}$$

The analysis of the detection statistics is similar to that for the model in 2.2.

2.2.2 Fingerprint Design for Multimedia Forensics

Orthogonal Fingerprint Design

A straightforward way of extending spread spectrum embedding to digital fingerprint is to assign users mutually orthogonal fingerprints [20, 32]. The advantage of the orthogonal fingerprint modulation is the simplicity of the fingerprint design and embedding. From the prior work in [64], orthogonal fingerprinting systems can survive collusion attacks with up to a few dozen colluders, and are preferred for applications with a small group of users. Given a total of M orthogonal basis, orthogonal fingerprinting systems have limited capacity and can support no more than a total of M users.

Group Oriented Fingerprint Design

To improve the traitor tracing capability of multimedia fingerprinting systems, group oriented fingerprinting systems take advantage of the prior knowledge of the possible collusion patterns during the design of the multimedia fingerprints [62]. Observing that adversaries are more likely to collude with some users than others due to geographic or social circumstances, in group oriented fingerprint design, some users who are more likely to collude with each other are assigned correlated fingerprints to enhance the collusion resistance performance.

Coded Fingerprint Design

Compared with orthogonal fingerprint design, given a limited cardinality of the orthogonal basis, coded fingerprint design has the advantage that it can accommodate more users in the fingerprinting systems [58]. In coded fingerprint design, given ν orthogonal basis signals $\{\mathbf{d}^{(k)}\}_{k=1,\cdots,\nu}$, each user in the system is assigned a unique code $\vec{b}^{(i)} = [b_{i,1}, \cdots, b_{i,\nu}]$ where $b_{i,k} = \{-1, +1\}$. To generate the fingerprint $\mathbf{W}^{(i)}$ for user $\mathbf{u}^{(i)}$, there are two types of fingerprint modulation schemes: *the CDMA based modulation* and *the TDMA based modulation* [68]. In the CDMA based fingerprint modulation, for user $\mathbf{u}^{(i)}$,

$$\mathbf{W}^{(i)} = \sum_{k=1}^{\nu} b_{i,k} \cdot \mathbf{d}^{(k)}. \tag{2.6}$$

In the TDMA based fingerprint modulation, the host signal (audio, image or video) is first partitioned into ν non-overlapping regions. For user $\mathbf{u}^{(i)}$, the signal $b_{i,k} \cdot \mathbf{d}^{(k)}$ is embedded into the kth region of the host signal.

A designer of the coded fingerprinting systems should design the fingerprint code $\vec{b}$ with good collusion resistance property while supporting as many users as possible. Prior work in the literature uses technologies from different areas to

design collusion resistant fingerprint code for multimedia, including the projective geometry [19], error correction code [29, 75], combinatorial theory [58], etc.

2.3 Performance Criteria for Digital Fingerprinting Systems

Although the overall goal of the digital fingerprinting system designer is to trace traitors and prevent information leakage, different applications of multimedia fingerprinting systems may have different concerns, and therefore, different requirements [64]. The digital fingerprinting systems should be designed according to the requirements of the applications and the appropriate performance criteria. This section analyzes the possible requirements of different applications and the corresponding performance criteria.

Catch One

In the *catch one* applications, the goal is to maximize the chance to capture one colluder while minimizing the probability of falsely accusing any innocent users. An example of such applications is to provide digital evidence in the court of law. In such applications, the performance criteria 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 fails to capture any of the colluders or it falsely accuse an innocent user as a colluder. Consequently, the system requirements are

$$P_d \geq \gamma_d, \quad \text{and} \quad P_{fp} \leq \gamma_{fp}, \tag{2.7}$$

where the parameters γ_d and γ_{fp} are determined by the requirements of the applications and are properly chosen by the designer of the fingerprinting systems.

Catch More

In the *catch more* fingerprinting applications, the goal is to capture as many colluders as possible, though possibly at a cost of accusing more innocent users. In these applications, the detection process is combined with other components in the decision making system and other evidences to make the final decision. The set of performance criteria consists of the fraction of colluders that are successfully captured $E[F_d]$, and the fraction of innocent users that are falsely placed under suspicion $E[F_{fp}]$. The system requirements for such applications are

$$E[F_d] \geq \lambda_d, \quad \text{and} \quad E[F_{fp}] \leq \lambda_{fp}, \tag{2.8}$$

where λ_d and λ_{fp} are the parameters determined by the requirements of the applications.

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 damage. An example of this scenario is to protect the highly classified documents in military applications. 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 efficiency rate

$$R = \frac{(M-K) \cdot E[F_{fp}]}{K \cdot E[F_d]} \tag{2.9}$$

that describes the number of innocents accused per colluder, and the probability of capturing all colluders

$$P_{d,all} = P\left[\min_{i \in SC} T_N^{(i)} > h\right]. \tag{2.10}$$

The system requirements for these applications are

$$R \leq \theta_r, \quad \text{and} \quad P_{d,all} \geq \theta_d. \tag{2.11}$$

θ_r and θ_d are determined by the requirements of the applications.

Chapter 3

Nonlinear Collusion Attacks on Multimedia Fingerprinting

Most prior works on digital fingerprinting and collusion attacks for multimedia employ the watermark embedding method in [13] and use a linear collusion attack model. In [52], several types of collusion attacks were studied, including a few order statistics based nonlinear attacks. For uniformly distributed fingerprints, nonlinear collusion attacks were shown to defeat the fingerprinting system more effectively than the averaging attack [52]. Simulation results in [52] also showed that normally distributed fingerprints are more robust against nonlinear collusion attacks than uniform fingerprints, but analytical study on the Gaussian fingerprints' performance was not provided. In addition to the robustness against collusion attacks, compared with discrete watermarks and uniform watermarks, Gaussian watermarks are proven to be resistant to statistical and histogram attacks [15]. Therefore, Gaussian distributed fingerprints should be used in multimedia fingerprinting systems for robustness against various types of attacks.

In this chapter, we mainly address the analysis of order statistics based nonlin-

ear collusion attacks on independent Gaussian fingerprints. We first consider from the colluders' point of view and compare various nonlinear collusion attacks on independent Gaussian fingerprints. We analyze the effectiveness of the collusion attacks and the perceptual quality of the colluded signals under different collusion attacks. We then shift our role to desinger/detector and analyze the performance of several commonly used detection statistics [48, 52, 76] in the literature under collusion attacks. The analysis of different detection statistics provides a guideline for the selection of the detector in a multimedia forensic system.

This chapter is organized as follows. We begin, in Section 3.1, with a system model of digital fingerprinting and collusion attacks. Then in Section 3.2, we analyze the effectiveness and the perceptual quality of different nonlinear collusion attacks, and investigate the detection performance of different detection statistics. In Section 3.3, we first study the resistance of independent unbounded Gaussian fingerprints to different collusion attacks. We then introduce bounded Gaussian-like fingerprints to achieve both the robustness against collusion attacks and the imperceptibility of the embedded fingerprints, and analyze their performance. In Section 3.4, we propose a pre-processing technique of the extracted fingerprints to improve the detection performance. Section 3.5 shows the simulation results on real images. A few more nonlinear collusion attacks are discussed in Section 3.6.

3.1 System Model

3.1.1 System Model and Assumptions

We consider a digital fingerprinting and collusion attack system that consists of three parts: fingerprint embedding, collusion attacks and fingerprint detection.

We use the spread spectrum embedding [13, 47] to hide fingerprints in the host signal. Assume that there are a total of M users in the system. Given a host signal represented by a vector $\mathbf{S}$ of length N, the owner generates a unique fingerprint $\mathbf{W}^{(i)}$ of length N for each user $\mathbf{u}^{(i)}$, $i = 1, 2, \cdots, M$. We assume that the M fingerprints $\{\mathbf{W}^{(i)}\}_{i=1}^{M}$ are independent of each other. The fingerprinted copy $\mathbf{X}^{(i)}$ that is distributed to user $\mathbf{u}^{(i)}$ is generated by $\mathbf{X}_j^{(i)} = \mathbf{S}_j + \alpha_j \mathbf{W}_j^{(i)}$. Here $\mathbf{X}_j^{(i)}$, $\mathbf{S}_j$ and $\mathbf{W}_j^{(i)}$ are the jth components of the fingerprinted copy, the original signal, and the fingerprint, respectively, and α is the *just-noticeable-difference* (JND) from human visual models [47] to control the energy and achieve the imperceptibility of the embedded fingerprints. Then, the fingerprinted copy $\mathbf{X}^{(i)}$ is distributed to user $\mathbf{u}^{(i)}$.

Assume that K out of M users collude, and $S_C = \{i_1, i_2, \cdots, i_K\}$ is the set containing the indices of the colluders. We further assume that the collusion attack is in the same domain as the fingerprint embedding. With K different copies $\{\mathbf{X}^{(k)}\}_{k \in S_C}$, the colluders generate the jth component of the attacked copy V_j using one of the collusion functions shown in (3.1).

$$
\begin{aligned}
\text{average attack:} \quad & V_j^{ave} && = \sum_{k \in S_C} X_j^{(k)} / K, \\
\text{minimum attack:} \quad & V_j^{min} && = \min\left(\{X_j^k\}_{k \in S_C}\right), \\
\text{maximum attack:} \quad & V_j^{max} && = \max\left(\{X_j^{(k)}\}_{k \in S_C}\right), \\
\text{median attack:} \quad & V_j^{med} && = \text{median}\left(\{X_j^{(k)}\}_{k \in S_C}\right), \\
\text{minmax attack:} \quad & V_j^{minmax} && = (V_j^{min} + V_j^{max})/2, \\
\text{modified negative attack:} \quad & V_j^{modneg} && = V_j^{min} + V_j^{max} - V_j^{med}, \\
\text{randomized negative attack:} \quad & V_j^{randneg} && = \begin{cases} V_j^{min} & \text{with prob. } p, \\ V_j^{max} & \text{with prob. } 1-p. \end{cases}
\end{aligned}
\tag{3.1}
$$

In (3.1), $\min\left(\{X_j^k\}_{k \in S_C}\right)$, $\max\left(\{X_j^k\}_{k \in S_C}\right)$ and $\text{median}\left(\{X_j^k\}_{k \in S_C}\right)$ return the

minimum, the maximum and the median values of $\{X_j^k\}_{k\in S_C}$, respectively. The colluded copy is $\mathbf{V} = [V_1, V_2, \cdots, V_N]$. For our model, applying the collusion attacks to the fingerprinted copies is equivalent to applying the collusion attacks to the embedded fingerprints. For example, $\mathbf{V}_j^{min} = \min\left(\{\mathbf{S}_j + \alpha \cdot \mathbf{W}_j^{(k)}\}_{k\in S_C}\right) = \mathbf{S}_j + \alpha \cdot \min\left(\{\mathbf{W}_j^{(k)}\}_{k\in S_C}\right)$.

In fingerprinting applications, the original signal $\mathbf{S}$ is often available to detectors. To improve the detection performance [58], the detector first removes the host signal from the attacked copy and extracts the fingerprint $\mathbf{Y} = g(\{\mathbf{W}^{(k)}\}_{k\in S_C})$ where $g(\cdot)$ is a collusion function defined in (3.1). The detector analyzes the similarity between $\mathbf{Y}$ and each of the M original fingerprints $\{\mathbf{W}^{(i)}\}$, and outputs the estimated colluder set.

In the literature, there are three detection statistics available to test the presence of the original fingerprint $\mathbf{W}^{(i)}$ in the extracted fingerprint $\mathbf{Y}$ [48, 52, 76].

$$
\begin{aligned}
T_N^{(i)} &= \langle \mathbf{Y}, \mathbf{W}^{(i)} \rangle / \sqrt{\|\mathbf{W}^{(i)}\|^2}, \qquad (3.2)\\
Z^{(i)} &= \frac{1}{2}\sqrt{N-3}\log\frac{1+\rho^{(i)}}{1-\rho^{(i)}}, \text{ where } \rho^{(i)} = \frac{\frac{1}{N}\sum_{j=1}^N Y_j W_j^{(i)} - \widetilde{Y}\cdot\widetilde{W}^{(i)}}{\sqrt{\hat{\sigma}_W^2\hat{\sigma}_Y^2}},\\
\text{and} \quad q^{(i)} &= \sqrt{N} M_y / \sqrt{V_y^2}, \text{ where}\\
M_y &= \sum_{j=1}^N \frac{Y_j W_j^{(i)}}{N} \text{ and } V_y^2 = \sum_{j=1}^N \frac{(Y_j W_j^{(i)} - M_y)^2}{N-1}.
\end{aligned}
$$

In (3.2), $\|\mathbf{W}^{(i)}\|$ is the Euclidean norm of $\mathbf{W}^{(i)}$; N is the length of the fingerprint; $\rho^{(i)}$ is the estimated correlation coefficient between $\mathbf{Y}$ and $\mathbf{W}^{(i)}$; $\widetilde{Y} = \frac{1}{N}\sum_{j=1}^N Y_j$ and $\widetilde{W}^{(i)} = \frac{1}{N}\sum_{j=1}^N W_j^{(i)}$ are the sample means of $\mathbf{Y}$ and $\mathbf{W}^{(i)}$, respectively; $\hat{\sigma}_W^2 = \frac{1}{N-1}\sum_j (W_j^{(i)} - \widetilde{W}^{(i)})$ and $\hat{\sigma}_Y^2 = \frac{1}{N-1}\sum_j (Y_j - \widetilde{Y})$ are the unbiased estimates of the original fingerprint's variance and the extracted fingerprint's variance, respectively; and M_y and V_y^2 are the sample mean and sample variance of $\{Y_j W_j^{(i)}\}$. Note that all three detection statistics are correlation based in which the correlation between

the extracted fingerprint $\mathbf{Y}$ and the original fingerprint $\mathbf{W}^{(i)}$ is the kernel term, and they differ primarily in the way of normalization.

3.1.2 Performance Criteria

We consider the following performance criteria to analyze different collusion attacks and different detection statistics.

Effectiveness of Collusion Attacks and Detection Performance of Detection Statistics

To study the effectiveness of collusion attacks and the performance of detection statistics, different criteria were used to address different applications in the literature. One set of criteria is the probability of falsely accusing at least one innocent user and the probability of not identifying any of the colluders [20,32]. The second set of criteria is the fraction of colluders that are successfully captured and the fraction of innocent users that are falsely accused, as considered in [58] and [63]. We adopt these criteria and use the following measurements:

- P_d: the probability of capturing at least one colluder,
- P_{fp}: the probability of falsely accusing at least one innocent user,
- F_d: the fraction of colluders that are successfully captured, and
- F_{fp}: the fraction of innocent users that are falsely accused.

Perceptual Quality

When considering the perceptual quality, one of the commonly used objective measurements on perceptual distortion is the mean square error (MSE) and equiv-

alently PSNR for image applications. A major weakness of MSE is that it ignores the unique characteristic of multimedia data: minor perturbations on the data values will not cause noticeable distortion as long as they do not exceed the *just-noticeable-difference* [47]. Furthermore, MSE only measures the average energy of the noise introduced and does not consider the local constraints on each noise component.

We take JND into consideration and define the following two new measurements,

- $F_{JND} \triangleq \sum_{j=1}^{N} I_{[|n_j| > JND_j]}/N$, and
- the redefined mean square error $MSE_{JND} \triangleq \sum_{j=1}^{N} n_j'^2$ where n_j' is defined as

$$n_j' = \begin{cases} n_j + JND_j & \text{if } n_j < -JND_j, \\ 0 & \text{if } -JND_j \leq n_j \leq JND_j, \\ n_j - JND_j & \text{if } n_j > JND_j. \end{cases} \tag{3.3}$$

MSE_{JND} calculates the power of the noise components that introduce perceptual distortion and F_{JND} reflects the percentage of the noise components that exceed JND. A large MSE_{JND} or a large F_{JND} indicates large perceptual distortion introduced.

3.2 Statistical Analysis of Collusion Attacks and Detection Statistics

In this section, we will analyze the statistical behavior of three detection statistics under different collusion attacks.

3.2.1 Analysis of the Correlation Term under Different Collusion Attacks

In our system model, the extracted fingerprint is $\mathbf{Y} = g\left(\{\mathbf{W}^{(k)}\}_{k\in S_C}\right)$. As discussed in the previous section, when measuring the similarity between $\mathbf{Y}$ and $\mathbf{W}^{(i)}$, all three statistics are correlation based, and the common kernel term is the linear correlation

$$T_N^{'(i)} \triangleq \frac{1}{N} < \mathbf{Y}, \mathbf{W}^{(i)} >= \frac{1}{N}\sum_{j=1}^{N} g(\{W_j^{(k)}\}_{k\in S_c})W_j^{(i)}, \tag{3.4}$$

where N is the length of the fingerprint. For different collusion attacks, $T_N^{'(i)}$ follows different distributions. This section analyzes the statistical behavior of this correlation term under different collusion attacks.

Under the assumption that $\{W_j^{(k)}, k = 1, \cdots, M\}_{j=1}^{N}$ are i.i.d. distributed with zero mean and variance σ_W^2, $\{g(\{W_j^{(k)}\}_{k\in S_c})W_j^{(i)}\}_{j=1}^{N}$ are also i.i.d. distributed. From central limit theorem, if $\{g(\{W_j^{(k)}\}_{k\in S_c})W_j^{(i)}\}_{j=1}^{N}$ have finite mean $\mu_{T_N^{'(i)}}$ and finite variance $\sigma^2_{T_N^{'(i)}}$, then $T_N^{'(i)}$ can be approximated by:

$$T_N^{'(i)} \sim \mathcal{N}\left(\mu_{T_N^{'(i)}}, \sigma^2_{T_N^{'(i)}}/N\right). \tag{3.5}$$

The problem is reduced to find the terms $\mu_{T_N^{'(i)}} = E\left[g(\{W^{(k)}\}_{k\in S_c})W^{(i)}\right]$ and $\sigma^2_{T_N^{'(i)}} = var\left[g(\{W^{(k)}\}_{k\in S_c})W^{(i)}\right]$. We simplify the notation by dropping the subscript j. For a given K and a given collusion function $g(\cdot)$, due to the symmetry of $g(\{W^{(k)}\}_{k\in S_c})W^{(i)}$ with respect to the user index i, all $g(\{W^{(k)}\}_{k\in S_c})W^{(i)}$ where $i \in S_C$ have the same mean and variance, and similarly, all $g(\{W^{(k)}\}_{k\in S_c})W^{(i)}$ where $i \notin S_C$ have the same mean and variance.

For $i \in S_C$, define

$$\begin{aligned} \mu_{g,H_1} &\triangleq E\left[g(\{W^{(k)}\}_{k\in S_c})W^{(i)}\right], \qquad (3.6)\\ \text{and } \sigma^2_{g,H_1} &\triangleq var\left[g(\{W^{(k)}\}_{k\in S_c})W^{(i)}\right] = E\left[\left(g(\{W^{(k)}\}_{k\in S_c})W^{(i)}\right)^2\right] - \left(\mu_{g,H_1}\right)^2. \end{aligned}$$

For $i \notin S_C$, because $\{W^{(i)}\}_{i=1}^M$ are i.i.d. distributed with zero mean and variance σ^2_W, we have

$$\begin{aligned} \mu_{g,H_0} &\triangleq E\left[g(\{W^{(k)}\}_{k\in S_c})W^{(i)}\right] = 0, \qquad (3.7)\\ \text{and } \sigma^2_{g,H_0} &\triangleq var\left[g(\{W^{(k)}\}_{k\in S_c})W^{(i)}\right] = E\left[\left(g(\{W^{(k)}\}_{k\in S_c})\right)^2\right]\sigma^2_W. \end{aligned}$$

Therefore, the three terms $E\left[g(\{W^{(k)}\}_{k\in S_c})W^{(i)}\right]$, $E\left[\left(g(\{W^{(k)}\}_{k\in S_c})W^{(i)}\right)^2\right]$ for $i \in S_C$ and $E\left[\left(g(\{W^{(k)}\}_{k\in S_c})\right)^2\right]$ are needed for analyzing the correlation term under each collusion attack.

Under the average attack, if $i \in S_C$, we have

$$\begin{aligned} E\left[\left(\frac{1}{K}\sum_{k\in S_C} W^{(k)}\right)W^{(i)}\right] &= \frac{1}{K}\sigma^2_W,\\ E\left[\left(\frac{1}{K}\sum_{k\in S_C} W^{(k)}W^{(i)}\right)^2\right] &= \frac{1}{K^2}E\left[\left(W^{(i)}\right)^4\right] + \frac{K-1}{K^2}\sigma^4_W,\\ \text{and} \quad E\left[\left(\frac{1}{K}\sum_{k\in S_C} W^{(k)}\right)^2\right] &= \frac{1}{K}\sigma^2_W. \qquad (3.8) \end{aligned}$$

Under the minimum attack, given the total number of colluders K, if $f(\cdot)$ and $F(\cdot)$ are the pdf and cdf of $W^{(i)}$, respectively, from the probability and order statistics theory [18], we can get the pdf of $W^{min} \triangleq \min\left(\{W^{(k)}\}_{k\in S_C}\right)$

$$f_{W^{min}}(W^{min} = w') = Kf(w')[1 - F(w')]^{K-1}. \qquad (3.9)$$

From (3.9), we can calculate the second moment of W^{min}. For $i \in S_C$, we can express the joint pdf of W^{min} and $W^{(i)}$ as follows by noticing that $f_{W^{min},W^{(i)}}(w', w)$

breaks into two nonzero regions

$$
\begin{aligned}
& f_{W^{min},W^{(i)}}(W^{min}=w',W^{(i)}=w) \qquad (3.10)\\
= & \begin{cases} f(w')[1-F(w')]^{K-1} & \text{if } W^{min}=W^{(i)}, \\ (K-1)f(w')f(w)[1-F(w')]^{K-2} & \text{if } W^{min}<W^{(i)}. \end{cases}
\end{aligned}
$$

Consequently, $E\left[W^{min}W^{(i)}\right]=E\left[W^{min}W^{(i)}\right]_1+E\left[W^{min}W^{(i)}\right]_2$, where

$$
\begin{aligned}
E\left[W^{min}W^{(i)}\right]_1 &= \int_{-\infty}^{\infty} w'^2 f(w')[1-F(w')]^{K-1}dw' \qquad (3.11)\\
E[W^{min}W^{(i)}]_2 &= \int_{-\infty}^{\infty} w'(K-1)f(w')[1-F(w')]^{K-2}\left(\int_{w'}^{\infty} wf(w)dw\right)dw'.
\end{aligned}
$$

The calculation of $E\left[\left(W^{min}W^{(i)}\right)^2\right]$ is similar.

The analysis of the maximum and median attacks follows the same approach. For the maximum attack, the pdf of $W^{max}\triangleq\max\left(\{W^{(k)}\}_{k\in S_C}\right)$ is:

$$
f_{W^{max}}(W^{max}=w')=Kf(w')F^{K-1}(w'), \qquad (3.12)
$$

and the joint pdf of W^{max} and $W^{(i)}$ for $i\in S_C$ is:

$$
\begin{aligned}
& f_{W^{max},W^{(i)}}(W^{max}=w',W^{(i)}=w) \qquad (3.13)\\
= & \begin{cases} f(w')F^{K-1}(w') & \text{if } W^{max}=W^{(i)}, \\ (K-1)f(w')f(w)F^{K-2}(w') & \text{if } W^{max}>W^{(i)}. \end{cases}
\end{aligned}
$$

Under the median attack, define $W^{med}\triangleq\text{median}\left(\{W^{(k)}\}_{k\in S_C}\right)$. If $K=2l+1$, the pdf of W^{med} is:

$$
f_{W^{med}}(W^{med}=w')=K\binom{2l}{l}f(w')F^l(w')[1-F(w')]^l, \qquad (3.14)
$$

and the joint pdf of W^{med} and $W^{(i)}$ for $i \in S_C$ is

$$
\begin{aligned}
& f_{W^{med},W^{(i)}}(W^{med} = w', W^{(i)} = w) \qquad (3.15)\\
= & \begin{cases}
\binom{2l}{l} f(w')F^l(w')[1-F(w')]^l & \text{if } W_{med} = W^{(i)},\\
(K-1)\binom{2l-1}{l} f(w')f(w)F^l(w')[1-F(w')]^{l-1} & \text{if } W^{med} < W^{(i)},\\
(K-1)\binom{2l-1}{l} f(w')f(w)F^{l-1}(w')[1-F(w')]^l & \text{if } W^{med} > W^{(i)}.
\end{cases}
\end{aligned}
$$

Under the minmax attack $W^{minmax} \triangleq \frac{1}{2}(W^{min} + W^{max})$, if $i \in S_C$, we have

$$
\begin{aligned}
E\left[W^{minmax}W^{(i)}\right] &= \left(E\left[W^{min}W^{(i)}\right] + E\left[W^{max}W^{(i)}\right]\right)/2,\\
E\left[\left(W^{minmax}W^{(i)}\right)^2\right] &= \left\{E\left[\left(W^{min}W^{(i)}\right)^2\right] + E\left[\left(W^{max}W^{(i)}\right)^2\right]\right\}/4\\
&\quad + E\left[W^{min}W^{max}\left(W^{(i)}\right)^2\right]/2,\\
\text{and} \quad E\left[\left(W^{minmax}\right)^2\right] &= \left\{E\left[\left(W^{min}\right)^2\right] + E\left[(W^{max})^2\right]\right\}/4\\
&\quad + E\left[W^{min}W^{max}\right]/2. \qquad (3.16)
\end{aligned}
$$

The results from the previous analysis on the minimum and the maximum attacks can be applied to (3.16). In addition, we can find the correlation between W^{min} and W^{max} from their joint pdf

$$
\begin{aligned}
& f_{W^{min},W^{max}}(W^{min} = w', W^{max} = w'') \qquad (3.17)\\
= \; & K(K-1)f(w')f(w'')[F(w'') - F(w')]^{K-2},
\end{aligned}
$$

thus, we have

$$
E\left[W^{min} \cdot W^{max}\right] = \int_{-\infty}^{\infty}\int_{w'}^{\infty} w'w'' f_{W^{min},W^{max}}(w', w'')\, dw''\, dw'. \qquad (3.18)
$$

The calculation of $E\left[(W^{min}W^{max})^2\right]$ is similar. $E\left[W^{min}W^{max}\left(W^{(i)}\right)^2\right]$is ob-

tained based on the joint pdf of W^{min}, W^{max}, and $W^{(i)}$, which is

$$
\begin{aligned}
& f_{W^{min},W^{max},W^{(i)}}(W^{min} = w', W^{max} = w'', W^{(i)} = w) \qquad (3.19) \\
= & \begin{cases}
(K-1)f(w')f(w'')[F(w'') - F(w')]^{K-2} \\
\qquad\qquad\qquad\qquad\qquad\qquad \text{if } W^{min} = W^{(i)}, \\
(K-1)f(w')f(w'')[F(w'') - F(w')]^{K-2} \\
\qquad\qquad\qquad\qquad\qquad\qquad \text{if } W^{max} = W^{(i)}, \\
(K-1)(K-2)f(w')f(w'')f(w)[F(w') - F(w'')]^{K-3} \\
\qquad\qquad\qquad\qquad\qquad\qquad \text{if } W^{min} < W^{(i)} < W^{max}.
\end{cases}
\end{aligned}
$$

The analysis of the modified negative (ModNeg) attack is similar to that of the minmax attack. If $K = 2l + 1$, then the joint pdf of W^{min} and W^{max} is

$$
\begin{aligned}
& f_{W^{min},W^{med}}(W^{min} = w', W^{med} = w'') \qquad (3.20) \\
= \; & (2l+1)2l\binom{2l-1}{l} f(w')f(w'')[F(w'') - F(w')]^{l-2}[1 - F(w'')]^{l},
\end{aligned}
$$

and the joint pdf of W^{med} and W^{max} is

$$
\begin{aligned}
& f_{W^{med},W^{max}}(W^{med} = w', W^{max}w'') \qquad (3.21) \\
= \; & (2l+1)2l\binom{2l-1}{l} f(w')f(w'')[F(w'') - F(w')]^{l-1}F^{l}(w').
\end{aligned}
$$

For $i \in S_C$, the joint pdf of W^{min}, W^{med} and $W^{(i)}$ is

$$f_{W^{min},W^{med},W^{(i)}}(W^{min} = w', W^{med} = w'', W^{(i)} = w) \tag{3.22}$$

$$= \begin{cases} 2l\binom{2l-1}{l} f(w')f(w'')[F(w'') - F(w')]^{l-1}[1 - F(w'')]^l \\ \qquad\qquad \text{if } W^{min} = W^{(i)}, \\ 2l\binom{2l-1}{l} f(w')f(w'')[F(w'') - F(w')]^{l-1}[1 - F(w'')]^l \\ \qquad\qquad \text{if } W^{med} = W^{(i)}, \\ 2l(2l-1)\binom{2l-2}{l-1} f(w')f(w'')f(w)[F(w'') - F(w')]^{l-1}[1 - F(w'')]^{l-1} \\ \qquad\qquad \text{if } W^{med} < W^{(i)} < W^{max}, \\ 2l(2l-1)\binom{2l-2}{l} f(w')f(w'')f(w)[F(w'') - F(w')]^{l-2}[1 - F(w'')]^l \\ \qquad\qquad \text{if } W^{min} < W^{(i)} < W^{med}, \end{cases} \tag{3.23}$$

and the joint pdf of W^{max}, W^{med} and $W^{(i)}$ is

$$f_{W^{med},W^{max},W^{(i)}}(W^{med} = w', W^{max} = w'', W^{(i)} = w) \tag{3.24}$$

$$= \begin{cases} 2l\binom{2l-1}{l} f(w')f(w'')[F(w'') - F(w')]^{l-1}F^l(w'') \\ \qquad\qquad \text{if } W^{med} = W^{(i)}, \\ 2l\binom{2l-1}{l} f(w')f(w'')[F(w'') - F(w')]^{l-1}F^l(w'') \\ \qquad\qquad \text{if } W^{max} = W^{(i)}, \\ 2l(2l-1)\binom{2l-2}{l-1} f(w')f(w'')f(w)[F(w'') - F(w')]^{l-1}F^{l-1}(w'') \\ \qquad\qquad \text{if } W^{min} < W^{(i)} < W^{med}, \\ 2l(2l-1)\binom{2l-2}{l} f(w')f(w'')f(w)[F(w'') - F(w')]^{l-2}F^l(w'') \\ \qquad\qquad \text{if } W^{med} < W^{(i)} < W^{max}. \end{cases}$$

Under the randomized negative (RandNeg) attack, we assume that p is independent of $\{W^{(i)}\}$. The colluded fingerprint can be written as $W^{randneg} =$

$W^{min} \cdot B_p + W^{max} \cdot (1 - B_p)$, where B_p is a Bernoulli random variable with parameter p and is independent of $\{W^{(i)}\}$. The m-th moment ($m = 1, 2, \cdots$) of $W^{randneg} W^{(i)}$ for $i \in S_C$ and the m-th moment of $W^{randneg}$ are

$$\begin{aligned} E\left[\left(W^{randneg} W^{(i)}\right)^m\right] &= p \cdot E\left[\left(W^{min} W^{(i)}\right)^m\right] + (1-p) \cdot E\left[\left(W^{max} W^{(i)}\right)^m\right], \\ \text{and } E\left[\left(W^{randneg}\right)^m\right] &= p \cdot E\left[\left(W^{min}\right)^m\right] + (1-p) \cdot E\left[(W^{max})^m\right]. \end{aligned} \tag{3.25}$$

From all the above analysis, the correlation kernel term $T_N^{'(i)}$ can be approximated by the following Gaussian distribution

$$T_N^{'(i)} \sim \begin{cases} \mathcal{N}\left(0,\ \frac{\sigma_{g,H_0}^2}{N}\right) & \text{if } i \notin S_C, \\ \mathcal{N}\left(\mu_{g,H_1},\ \frac{\sigma_{g,H_1}^2}{N}\right) & \text{if } i \in S_C. \end{cases} \tag{3.26}$$

3.2.2 Analysis of the Detection Statistics

From (3.26), we can approximate the detection statistics $T_N^{(i)}$ by a Gaussian random variable

$$T_N^{(i)} = \frac{N T_N^{'(i)}}{\sqrt{\|\mathbf{W}^i\|^2}} \sim \begin{cases} \mathcal{N}\left(0,\ \frac{\sigma_{g,H_0}^2}{\sigma_W^2}\right) & \text{if } i \notin S_C, \\ \mathcal{N}\left(\frac{\sqrt{N}\mu_{g,H_1}}{\sigma_W},\ \frac{\sigma_{g,H_1}^2}{\sigma_W^2}\right) & \text{if } i \in S_C. \end{cases} \tag{3.27}$$

The Z statistics can be approximated by a Gaussian random variable $\mathcal{N}(\mu_Z^{(i)}, 1)$ with mean $\mu_Z^{(i)} = \frac{1}{2}\sqrt{N-3}\log\frac{1+E[\rho^{(i)}]}{1-E[\rho^{(i)}]}$, where $E[\rho^{(i)}]$ is the mean of $\rho^{(i)}$ defined in (3.2) and is the estimated correlation coefficient of the extracted fingerprint $\mathbf{Y}$ and the original fingerprint $\mathbf{W}^{(i)}$ [52]. We can show that

$$Z^{(i)} \sim \begin{cases} \mathcal{N}\left(0,\ 1\right) & \text{if } i \notin S_C, \\ \mathcal{N}\left(\frac{1}{2}\sqrt{N-3}\log\frac{1+E[\rho^{(i)}]}{1-E[\rho^{(i)}]},\ 1\right) & \text{if } i \in S_C. \end{cases}$$

Here, for $i \in S_C$,

$$E[\rho^{(i)}] \approx \frac{cov\left[g(\{W^{(k)}\}_{k \in S_C}), W^{(i)}\right]}{\sqrt{\sigma_W^2 \sigma_{g,Y}^2}} = \frac{\mu_{g,H_1}}{\sqrt{\sigma_W^2 \sigma_{g,Y}^2}}, \tag{3.28}$$

where $\sigma^2_{g,Y}$ is the variance of the extracted fingerprint.

The q statistics normalize the correlation term with the unbiased estimate of its variance. So we have

$$q^{(i)} \sim \begin{cases} \mathcal{N}(0,\ 1) & \text{if } i \notin S_C, \\ \mathcal{N}\left(\frac{\sqrt{N}\mu_{g,H_1}}{\sqrt{\sigma^2_{g,H_1}}},\ 1\right) & \text{if } i \in S_C. \end{cases} \tag{3.29}$$

3.2.3 Analysis of the Performance of Collusion Attacks and Detection Statistics

Analysis of P_d, P_{fp}, $E[F_d]$, and $E[F_{fp}]$

In our system model with a total of M users and K colluders, given a signal to be tested and given one detection statistics, K out of the M statistics $\{T_N^{(i)}\}_{i=1}^M$ are normally distributed with a positive mean and the others are normally distributed with a zero mean, as analyzed in the previous section.

Take the T_N statistics as an example, define $\mu_1 \stackrel{\triangle}{=} \frac{\sqrt{N}\mu_{g,H_1}}{\sigma_W}$, $\sigma_1^2 \stackrel{\triangle}{=} \frac{\sigma^2_{g,H_1}}{\sigma^2_W}$, and $\sigma_0^2 \stackrel{\triangle}{=} \frac{\sigma^2_{g,H_0}}{\sigma^2_W}$. If $\{T_N^{(i)}\}_{i=1}^M$ are uncorrelated with each other or the correlation is very small, then for a given threshold h, we can approximate P_d and P_{fp} by

$$\begin{aligned} P_d &= P\left[\max_{i\in S_C} T_N^{(i)} > h\right] \approx 1 - \left[1 - Q\left(\frac{h-\mu_1}{\sigma_1}\right)\right]^K, \\ \text{and} \quad P_{fp} &= P\left[\max_{i\notin S_C} T_N^{(i)} > h\right] \approx 1 - \left[1 - Q\left(\frac{h}{\sigma_0}\right)\right]^{M-K}, \end{aligned} \tag{3.30}$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$ is the Gaussian tail function.

To calculate $E[F_d]$ and $E[F_{fp}]$, we can have the following approximations

$$\begin{aligned} E[F_d] &= P\left[T_N^{(i\in S_C)} > h\right] \approx Q\left(\frac{h-\mu_1}{\sigma_1}\right), \\ \text{and} \quad E[F_{fp}] &= P\left[T_N^{(i\notin S_C)} > h\right] \approx Q\left(\frac{h}{\sigma_0}\right). \end{aligned} \tag{3.31}$$

The analysis of P_d, P_{fp}, F_d and F_{fp} for the Z and q statistics are the same.

Perceptual Quality

In our system model, the distortion introduced to the host signal by the colluded fingerprint is $n_j = JND_j \cdot g(\{W_j^{(k)}\}_{k\in S_C})$, $j = 1, 2, \cdots, N$. Given the collusion attack $g(\cdot)$ and the number of colluders K, if $\mathbf{A} \triangleq g(\{W^{(k)}\}_{k\in S_C})$ has the pdf $f_{g,K}(w)$, we can simplify the MSE_{JND} and $E[F_{JND}]$ to

$$\begin{aligned} & MSE_{JND} \approx N \times E\left[(|\mathbf{A}|-1)^2 \mid |\mathbf{A}|>1\right] \\ = \quad & N\int_{-\infty}^{-1}(w+1)^2 f_{g,K}(w)\,dw + N\int_{1}^{\infty}(w-1)^2 f_{g,K}(w)\,dw, \\ \text{and} \quad & E[F_{JND}] = P\left[|\mathbf{A}|>1\right] = \int_{-\infty}^{-1} f_{g,K}(w)\,dw + \int_{1}^{\infty} f_{g,K}(w)\,dw. \quad (3.32) \end{aligned}$$

3.3 Effectiveness of Collusion Attacks on Gaussian Based Fingerprints

It has been shown in [52] that the uniform fingerprints can be easily defeated by nonlinear collusion attacks, and the simulation results there also showed that the Gaussian fingerprints are more resistant to nonlinear collusion attacks than the uniform fingerprints. However, no analytic study was provided in the literature on the resistance of Gaussian fingerprints to nonlinear collusion attacks. In this section, we study the effectiveness of nonlinear collusion attacks on Gaussian based fingerprints.

3.3.1 Unbounded Gaussian Fingerprints

Statistical Analysis

We first study the resistance of unbounded Gaussian fingerprints to collusion attacks. As before, we assume that there are a total of M users and the fingerprints

$\{W_j^{(i)}\}$ are i.i.d. Gaussian with zero mean and variance σ_W^2. Usually we take $\sigma_W^2 \approx 1/9$ to ensure that around 99.9% of fingerprint components are in the range of $[-1, 1]$ and are imperceptible after being scaled by a JND factor.

Under the assumption that the Bernoulli random variable B_p in the randomized negative attack is independent of the zero mean Gaussian fingerprints, we have $E\left[\left(W^{randneg}\right)^2\right] = E\left[\left(W^{min}\right)^2\right] = E\left[\left(W^{max}\right)^2\right]$ for all possible $p \in [0, 1]$. Consequently, we have

$$\sigma_{randneg,Y}^2 = E\left[\left(W^{randneg}\right)^2\right] - \left(E\left[W^{randneg}\right]\right)^2 \leq E\left[\left(W^{min}\right)^2\right], \qquad (3.33)$$

and the upper bound of the variance in (3.33) is achieved when $p = 0.5$ and $E\left[W^{randneg}\right] = 0$. From (3.30) and (3.31), the larger the variance, the more effective the attack. Consequently, we take $p = 0.5$ and consider the most effective attack.

Given the analysis in the previous section, we can calculate the parameters μ_{g,H_1}, σ_{g,H_1}^2, σ_{g,H_0}^2 and $\sigma_{g,Y}^2$ for Gaussian distribution with zero mean and variance σ_W^2. Due to the existence of the $Q(\cdot)$ terms in the pdfs and joint pdfs, analytical expressions are not available. We use the recursive adaptive Simpson quadrature method [23] to numerically evaluate the integrals with an absolute error tolerance of 10^{-6} and the results for $\sigma_W^2 = 1/9$ are plotted in Figure 3.1.

From Figure 3.1, we find that, for a given number of colluders K, μ_{g,H_1} are the same for all collusion attacks and equal to σ_W^2/K. Different collusion attacks have different σ_{g,H_1}^2, σ_{g,H_0}^2 and $\sigma_{g,Y}^2$. The relationship of σ_{g,H_1}^2 and σ_{g,H_0}^2 for different collusion attacks are

$$\begin{aligned} & \sigma_{randneg,H_1}^2 = \sigma_{min,H_1}^2 = \sigma_{max,H_1}^2 > \sigma_{modneg,H_1}^2 \\ & > \sigma_{ave,H_1}^2 \approx \sigma_{med,H_1}^2 \approx \sigma_{minmax,H_1}^2, \\ \text{and} \quad & \sigma_{randneg,H_0}^2 = \sigma_{min,H_0}^2 = \sigma_{max,H_0}^2 > \sigma_{modneg,H_0}^2 \end{aligned}$$

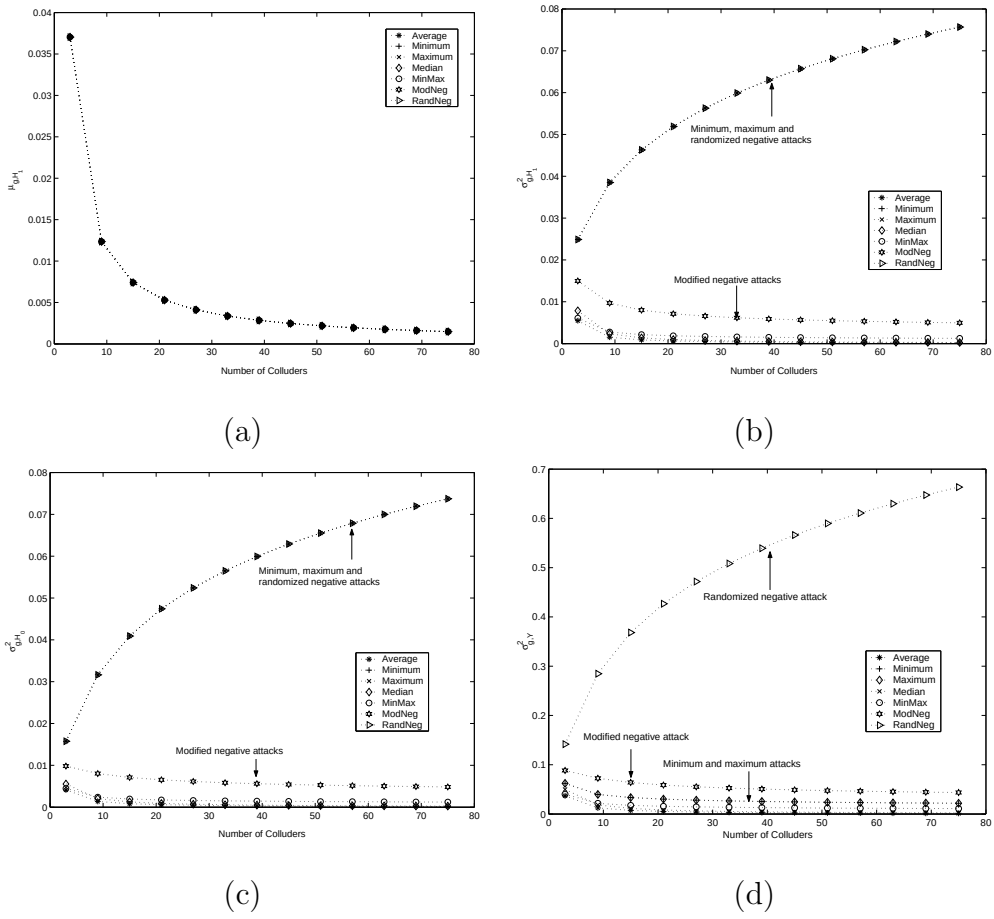


Figure 3.1: (a) μ_{g,H_1}, (b) σ^2_{g,H_1}, (c) σ^2_{g,H_0}, and (d) $\sigma^2_{g,Y}$ of the unbounded Gaussian fingerprints with $\sigma^2_W = 1/9$.

$$> \sigma^2_{ave,H_0} \approx \sigma^2_{med,H_0} \approx \sigma^2_{minmax,H_0}, \tag{3.34}$$

and that of $\sigma^2_{g,Y}$ is

$$\begin{aligned} &\sigma^2_{randneg,Y} > \sigma^2_{modneg,Y} > \sigma^2_{min,Y} = \sigma^2_{max,Y} \\ &> \sigma^2_{ave,Y} \approx \sigma^2_{med,Y} \approx \sigma^2_{minmax,Y}. \end{aligned} \tag{3.35}$$

Note that the extracted fingerprint $\mathbf{Y}$ under the minimum or maximum attack is not zero mean. σ^2_{g,H_0} is proportional to the second moment of $\mathbf{Y}$, and is the largest under the minimum, maximum, and randomized negative attacks. However, the variance of $\mathbf{Y}$ under the minimum or maximum attacks is small and comparable with $\sigma^2_{g,Y}$ under the average, median, and minmax attacks.

In order to compare the effectiveness of different collusion attacks, we define the following notations:

- " attack A > attack B ": attack A is more effective than attack B in defeating the system,
- " attack A = attack B ": attack A and attack B have the same performance in defeating the system,
- " attack A $\approx$ attack B ": attack A and attack B have similar performance in defeating the system.

From (3.30), (3.31), (3.34), and (3.35), with the T_N statistics or the q statistics, we can sort different collusion attacks in the descending order of their effectiveness as:

$$\begin{aligned} &Minimum = Maximum = RandNeg > ModNeg \\ &> Average \approx Median \approx MinMax; \end{aligned} \tag{3.36}$$

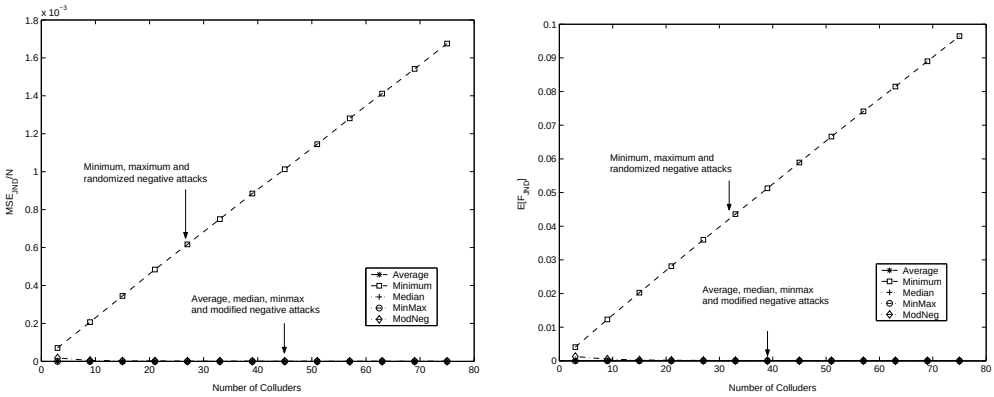


Figure 3.2: Perceptual quality of the attacked copy under different attacks with unbounded Gaussian fingerprints. Here $\sigma_W^2 = 1/9$. (Left) MSE_{JND}/N. (Right) $E[F_{JND}]$.

and with the Z statistics, we can sort different attacks in the descending order of their effectiveness as:

$$\begin{aligned} &RandNeg > ModNeg > Minimum = Maximum \\ &> Average \approx Median \approx MinMax. \end{aligned} \tag{3.37}$$

Therefore, the randomized negative attack is the most effective attack.

So far we have studied the effectiveness of different collusion attacks. As for the perceptual quality, Figure 3.2 shows the MSE_{JND} and $E[F_{JND}]$ of different collusion attacks with i.i.d. $\mathcal{N}(0, 1/9)$ fingerprints. As we can see from Figure 3.2, although the minimum, maximum, and randomized negative attacks are more effective in defeating the fingerprinting system, they also introduce larger noticeable distortion that is proportional to the number of colluders.

Simulation Results

Our simulation is set up as follows. Since the number of embeddable coefficients in 256×256 and 512×512 images is usually $O(10^4)$, we assume that the length of the fingerprints is $10,000$. To accommodate a total of $M = 100$ users, we generate

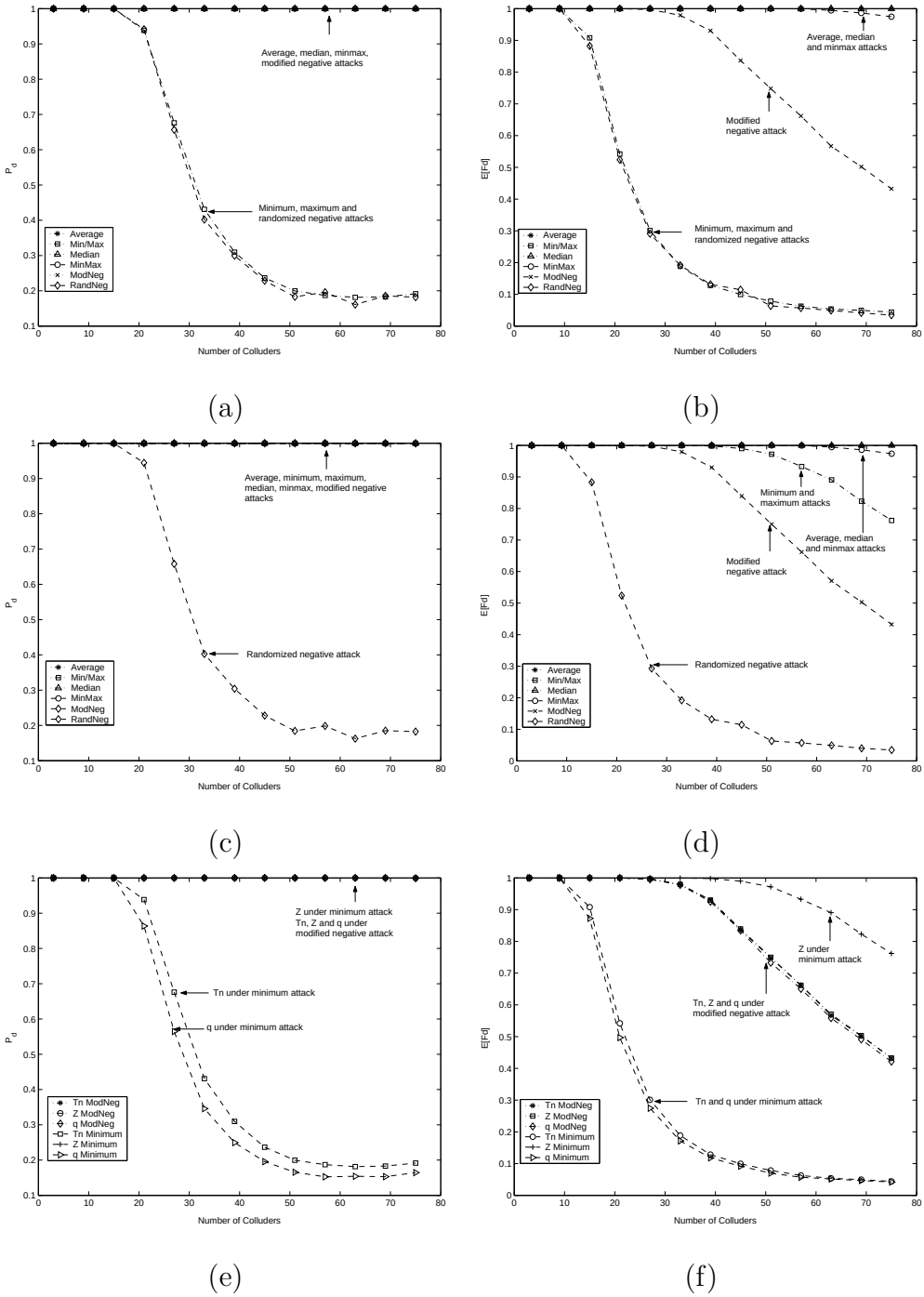


Figure 3.3: (a) P_d of the T_N statistics under different attacks, (b) $E[F_d]$ of the T_N statistics under different attacks, (c) P_d of the Z statistics under different attacks, (d) $E[F_d]$ of the Z statistics under different attacks, (e) P_d of different statistics, and (f) $E[F_d]$ of different statistics with unbounded Gaussian fingerprints. Here $\sigma_W^2 = 1/9,\ M = 100$, and $N = 10^4$. In (a), (c) and (e), $P_{fp} = 10^{-2}$. In (b), (d) and (f), $E[F_{fp}] = 10^{-2}$.

100 independent fingerprints of length $10,000$. Every fingerprint component is independent of each other and follows the $\mathcal{N}(0, 1/9)$ Gaussian distribution. Our results are based on a total of 2000 simulation runs.

In Figure 3.3 (a) and (c), P_{fp} is fixed as 10^{-2} and we compare P_d of the T_N and Z statistics, respectively, under different collusion attacks. In Figure 3.3 (b) and (d), $E[F_{fp}]$ is fixed as 10^{-2} and we compare $E[F_d]$ of the T_N and Z statistics, respectively, under different attacks. The performance of the q statistics is similar to that of T_N and is not shown here. We compare different detection statistics with $P_{fp} = 10^{-2}$ in Figure 3.3 (e) and $E[F_{fp}] = 10^{-2}$ in Figure 3.3 (f). Note that in Figure 3.3 (e) and (f), we only plot the performance of the minimum and that of the modified negative attacks since the maximum attack yield the same result as the minimum attack and all other attacks have a similar trend.

The simulation results shown in Figure 3.3 agree with our analysis. From Figure 3.3 (a) and (b), with the T_N or q statistics, the minimum, maximum, and randomized negative attack are the most effective attacks followed by the modified negative attack. The average, median, and minmax attacks are the least effective attacks. From Figure 3.3 (c) and (d), with the Z statistics, the randomized negative attack is the most effective attack followed by the modified negative attack. The average, median, and minmax attacks have similar performance and they are the least efficient attacks. The minimum and maximum attacks are the second least effective attacks. From Figure 3.3 (e) and (f), the Z statistics are more resistant to the minimum and maximum attacks than the T_N and q statistics while the three statistics have similar performance under other collusion attacks. Therefore, from the colluders' point view, the best strategy for them is to choose the randomized negative attack. From the detector's point of view, the Z statistics should be used

to be more robust against the minimum and maximum attacks.

In Figure 3.4, we show the attacked images after the average and the minimum attacks with 75 colluders. Although the minimum, maximum and randomized negative attacks are more effective, they also introduce much larger noticeable distortion in the host image. This is because the fingerprints are not bounded, and in fact, such unbounded fingerprints can introduce noticeable distortion in the fingerprinted copies even when without collusion.

3.3.2 Bounded Gaussian-like Fingerprints

Compared with uniform fingerprints, Gaussian fingerprints improve the detector's resistance to nonlinear collusion attacks [52] and are resilient to statistical and histogram attacks [15]. Because Gaussian distribution is unbounded, it is possible that the embedded fingerprints exceed the JND and introduce perceptually distinguishable distortion. However, imperceptibility is a requirement of digital fingerprinting and the owner has to guarantee the perceptual quality of the fingerprinted copies. In order to remove the perceptual distortion while maintaining the robustness against collusion attacks, we introduce the bounded Gaussian-like fingerprints and study their performance under collusion attacks.

Assume that $f_X(\cdot)$ and $F_X(\cdot)$ are the pdf and cdf of a Gaussian random variable with zero mean and variance σ_W^2, respectively. The pdf of a bounded Gaussian-like distribution $\tilde{f}_X(\cdot)$ is:

$$\tilde{f}_X(x) = \begin{cases} \frac{f_X(x)}{F_X(1)-F_X(-1)} & \text{if } -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.38}$$

We can show that the variance of fingerprints following pdf (3.38) is σ_W^2, and the embedded fingerprints introduce no perceptual distortion since $MSE_{JND} = 0$

Figure 3.4: Comparison of perceptual quality of the attacked images under different attacks with 75 colluders. Fingerprints are generated from unbounded Gaussian distribution with $\sigma_W^2 = 1/9$. (Left) Lena. (Right) Baboon. (Top) The zoomed-in region of the original 256×256 images. (Middle) The colluded images under the average attack. (Bottom) The colluded images under the minimum attack.

and $F_{JND} = 0$. By bounding the fingerprints in the range of $[-1, 1]$, we maintain the energy of the embedded fingerprints while achieving the imperceptibility.

For fingerprints following distribution (3.38), the analyses of the collusion attacks and the detection statistics are similar to the unbounded case and thus omitted. If we sort different collusion attacks according to their effectiveness, the result is the same as that of the unbounded Gaussian fingerprints.

The simulation of the bounded Gaussian-like fingerprints under collusion attacks is set up similarly to that in Section 3.3.1. Assume that there are a total of $M = 100$ users and the host signal has $N = 10^4$ embeddable coefficients. The i.i.d. fingerprints are generated from the distribution (3.38) with $\sigma_W^2 = 1/9$. In Figure 3.5 (a) and (c), $P_{fp} = 10^{-2}$ and we compare P_d of the T_N and Z statistics, respectively, under different collusion attacks. In Figure 3.5 (b) and (d), $E[F_{fp}] = 10^{-2}$ and we compare $E[F_d]$ of the T_N and Z statistics, respectively, under different collusion attacks. The performance of the q statistics is similar to that of T_N. We compare the performance of different detection statistics under the minimum and the modified negative attacks with $P_{fp} = 10^{-2}$ in Figure 3.5 (e) and $E[F_{fp}] = 10^{-2}$ in Figure 3.5 (f), respectively. The simulation results agree with the analysis and we have the same observations as in the unbounded case. From the colluders' point of view, the most efficient attack is the randomized negative attack, and from the detector's point of view, the Z statistics are more robust.

3.4 Pre-Processing of the Extracted Fingerprints

The three detection statistics we have studied so far are not specifically designed for collusion scenarios, and therefore do not take into account the characteristics of the newly generated copies after the collusion attacks. Intuitively, utilizing the

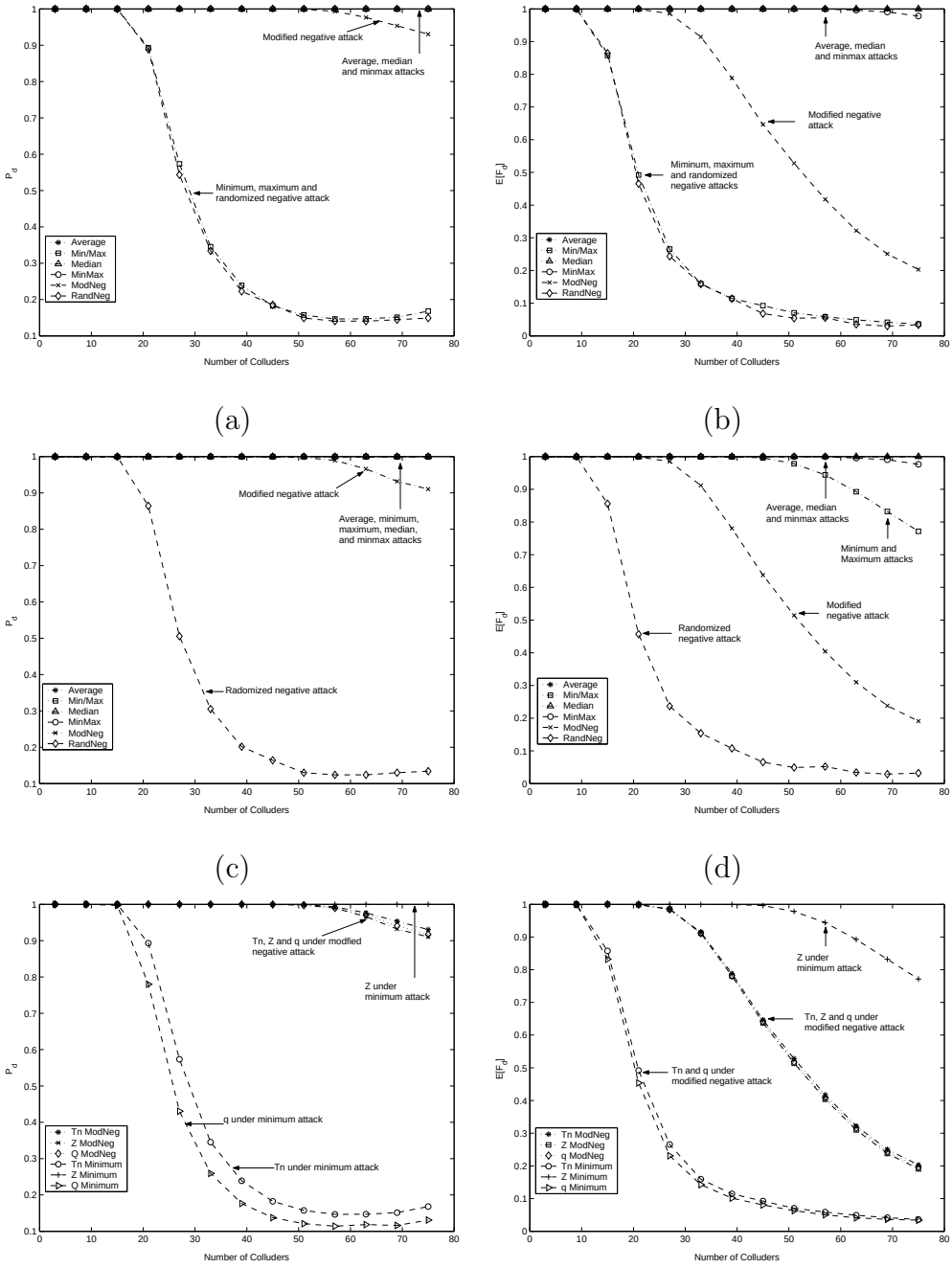


Figure 3.5: (a) P_d of the T_N statistics under different attacks, (b) $E[F_d]$ of the T_N statistics under different attacks, (c) P_d of the Z statistics under different attacks, (d) $E[F_d]$ of the Z statistics under different attacks, (e) P_d of different statistics, and (f) $E[F_d]$ of different statistics with bounded Gaussian-like fingerprints. Here $\sigma_W^2 = 1/9$, $M = 100$, and $N = 10^4$. In (a), (c) and (e), $P_{fp} = 10^{-2}$. In (b), (d) and (f), $E[F_{fp}] = 10^{-2}$.

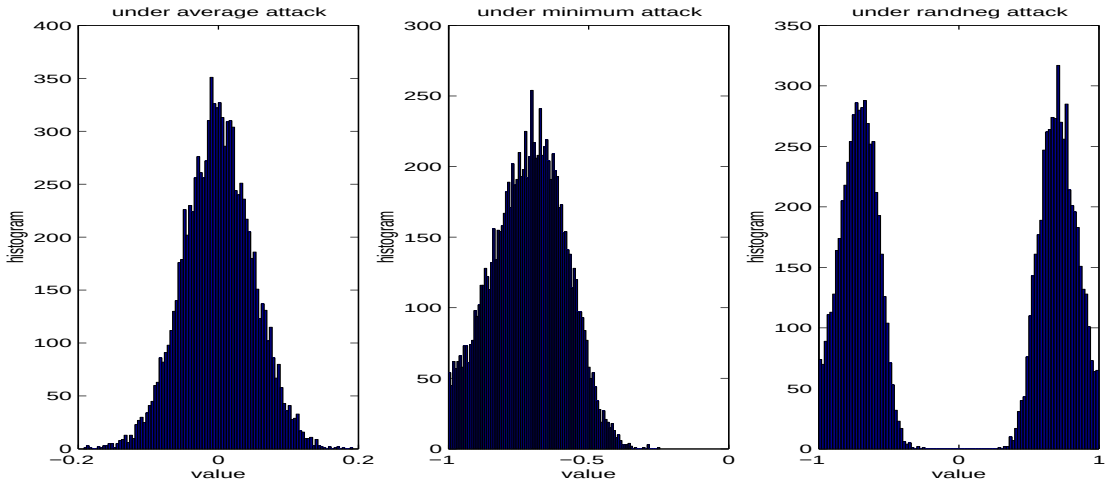


Figure 3.6: Histograms of the extracted fingerprints under the average, minimum and randomized negative attacks, respectively. The original fingerprints follow the distribution in (3.38) with $\sigma_W^2 = 1/9$. $N = 10^4$ and $K = 45$.

statistical features of the attacked copies may improve the detection performance, and one of such features is the sample mean of the extracted fingerprint under the collusion attacks. From the histogram plots of the extracted fingerprints under different attacks as shown in Figure 3.6, we observe different patterns of the sample means of the extracted fingerprints: the extracted fingerprints have approximately zero sample mean under the average, median, minmax and modified negative attacks; the minimum attack yields a negative sample mean, and the maximum attack yields a positive sample mean; and under the randomized negative attack, the histogram of the extracted fingerprint components have two clusters, one with a negative mean and the other with a positive mean.

Recall from Section 3.2.1 that σ_{g,H_0}^2 is proportional to the second moment of the extracted fingerprint, subtracting the sample mean from the extracted fingerprint will reduce its second-order moment, thus help improve the detection performance.

Similarly, the detection performance under the randomized negative attack can be improved by decreasing σ^2_{g,H_0} and $\sigma^2_{g,Y}$.

Motivated by this analysis, we propose a pre-processing stage in the detection process: given the extracted fingerprint $\{g(\{W_j^k\}_{k\in S_c})\}_{j=1}^N$, we first investigate its histogram. If a single non-zero sample mean is observed, we subtract it from the extracted fingerprint, and then apply the detection statistics. If the fingerprint components are merged from two (or more) distributions that have distinct mean values, we need to cluster components and then subtract from each colluded fingerprint component the sample mean of the corresponding cluster. In the later case, the means can be estimated using a Gaussian-mixture approximation, and the clustering is based on the nearest-neighbor principle. In our problem, under the randomized negative attack, a simple solution is to first observe the bi-modality in the histogram of $\{Y_j\}$, and then cluster all negative components into one distribution and cluster all positive components into the other distribution. Given the extracted fingerprint $\{Y_j\}_{j=1}^N$, define $\mu_{neg} \stackrel{\triangle}{=} \sum_j Y_j \cdot I\left[Y_l < 0\right] / \sum_l I\left[Y_l < 0\right]$ as the sample mean of the negative extracted fingerprint components where $I[\cdot]$ is the indication function, and $\mu_{pos} \stackrel{\triangle}{=} \sum_j Y_j \cdot I\left[Y_j > 0\right] / \sum_l I\left[Y_l > 0\right]$ as the sample mean of the positive extracted fingerprint components. Then the pre-processing stage generates

$$Y_j' = \begin{cases} Y_j - \mu_{neg} & \text{if } Y_j < 0, \\ Y_j - \mu_{pos} & \text{if } Y_j > 0, \end{cases} \tag{3.39}$$

and the detector applies the detection statistics to $\{Y_j'\}_{j=1}^N$. The analysis of the detection statistics with the pre-processing is the same as in Section 3.2 and is not repeated.

The simulation is set up the same as before and the fingerprint components are

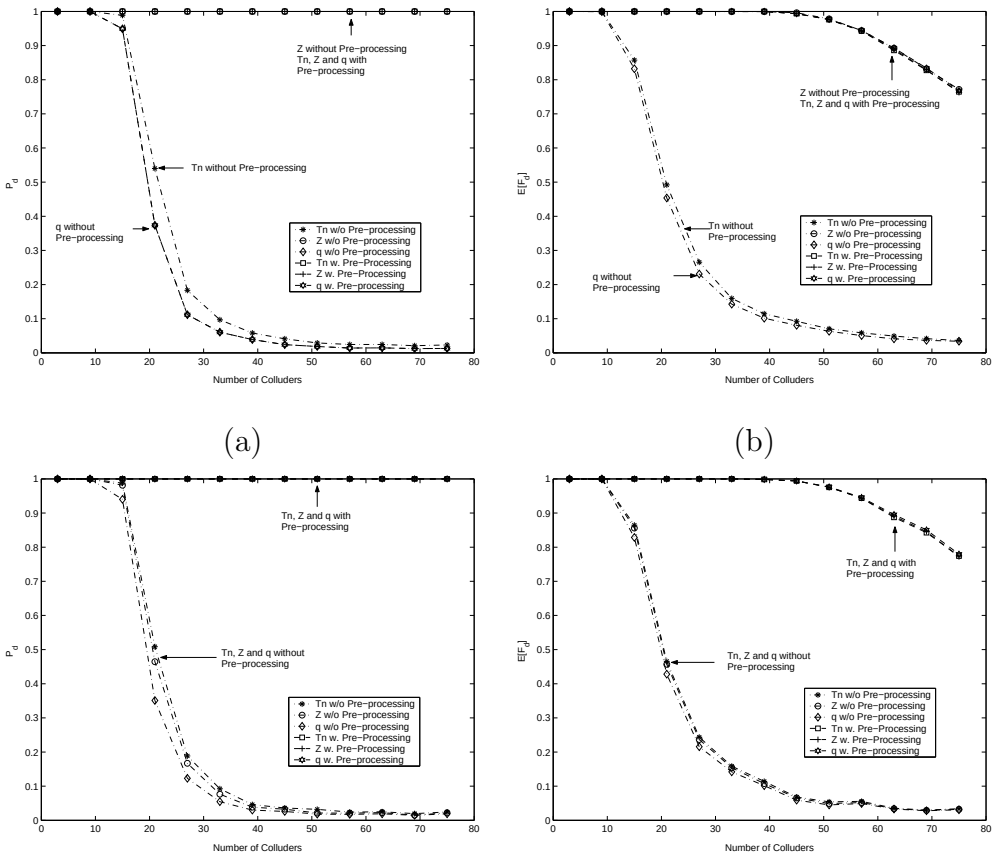


Figure 3.7: (a) P_d under the minimum attack, (b) $E[F_d]$ under the minimum attack, (c) P_d under the randomized negative attack, and (d) $E[F_d]$ under the randomized negative attack with and without pre-processing. Fingerprints are generated from bounded Gaussian-like distribution (3.38) with $\sigma_W^2 = 1/9$. $M = 100$ and $N = 10^4$. In (a) and (c), $P_{fp} = 10^{-2}$. In (b) and (d), $E[F_{fp}] = 10^{-2}$.

generated from the bounded Gaussian-like distribution (3.38) with $\sigma_W^2 = 1/9$. In Figure 3.7 (a) and (c), with $P_{fp} = 10^{-2}$, we compare P_d of the three statistics with and without the pre-processing under the minimum and the randomized negative attacks, respectively. In Figure 3.7 (b) and (d), with $E[F_{fp}] = 10^{-2}$, we compare $E[F_d]$ of the three statistics with and without the pre-processing under the minimum and the randomized negative attacks, respectively. The detection performance under the maximum attack is the same as that of the minimum attack and is not shown here. We can see that the pre-processing substantially improves the detection performance of the detector, and the three statistics have similar performance under the minimum, maximum, and randomized negative attacks.

Note that the estimated correlation coefficient $\rho^{(i)}$ in the Z statistics removes the mean of the extracted fingerprint before calculating the correlation between the extracted fingerprint and the original fingerprint. This explains why the Z statistics perform better than the T_N and q statistics without pre-processing under the minimum and maximum attacks, whereby the mean of the colluded fingerprint components is substantially deviated from zero.

3.5 Simulation Results on Real Images

To study the performance of Gaussian based fingerprints under different nonlinear collusion attacks on real images, we choose two 256×256 host images, Lena and Baboon, which have a variety of representative visual features such as the texture, sharp edges, and smooth areas. We use the human visual model based spread spectrum embedding in [47], and embed the fingerprints in the DCT domain. The generated fingerprints follow the bounded Gaussian-like distribution (3.38) with $\sigma_W^2 = 1/9$. We assume that the collusion attacks are also in the DCT domain.

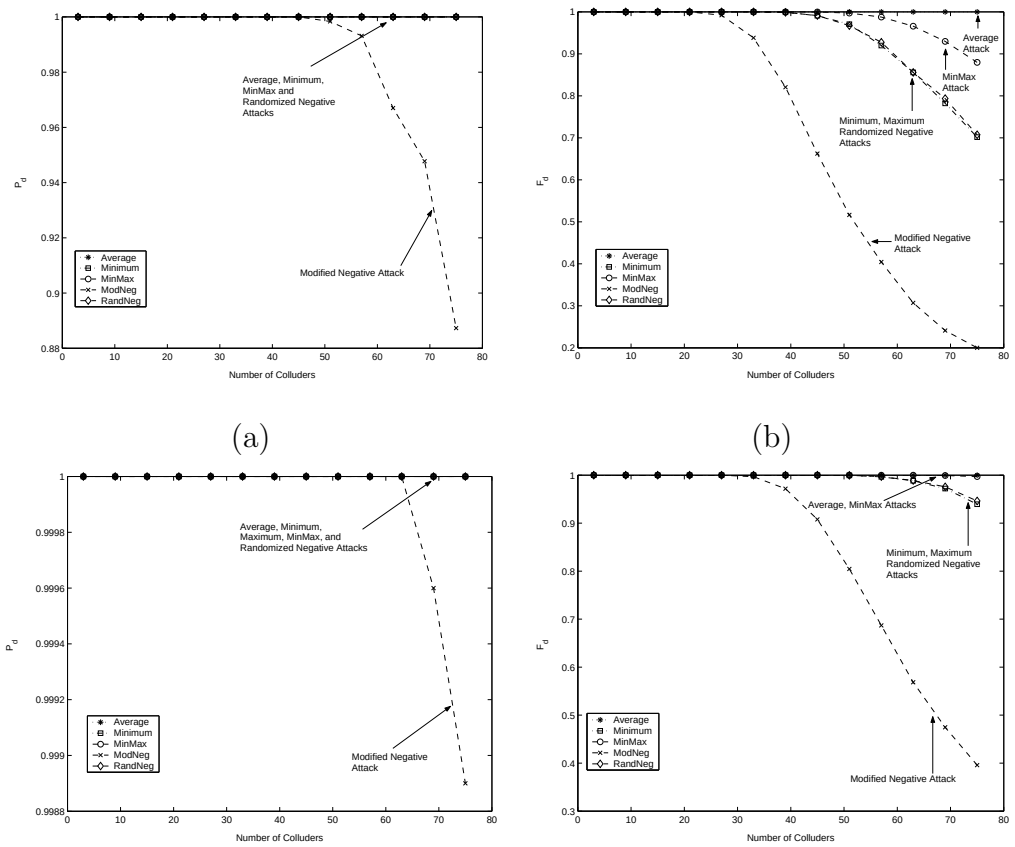


Figure 3.8: (a) P_d of Lena, (b) $E[F_d]$ of Lena, (c) P_d of Baboon, and (d) $E[F_d]$ of Baboon with the Z statistics under different collusion attacks. The original fingerprints follow the distribution in (3.38) with $\sigma_W^2 = 1/9$. $M = 100$. In (a) and (b), the length of the embedded fingerprints is $N = 13691$. In (c) and (d), the length of the embedded fingerprints is $N = 19497$. In (a) and (c), $P_{fp} = 10^{-2}$ and simulation results are based on 10,000 simulation runs. In (b) and (d), $E[F_{fp}] = 10^{-2}$ and simulation results are based on 1,000 simulation runs.

At the detector's side, a non-blind detection is performed where the host signal is first removed from the colluded copy. Then the detector applies the pre-processing to the extracted fingerprint if a non-zero sample mean is observed. Finally, the detector uses the detection statistics to identify the colluders.

Figure 3.8 shows the simulation results of the Z statistics. The T_N and q statistics have similar performance and are not shown here. We assume that there are a total of $M = 100$ users. In Figure 3.8 (a) and (c), we fix $P_{fp} = 10^{-2}$ and compare P_d of Lena and Baboon, respectively, under different nonlinear collusion attacks. In Figure 3.8 (b) and (d), we fix $E[F_{fp}] = 10^{-2}$ and compare $E[F_d]$ of Lena and Baboon, respectively, under different nonlinear collusion attacks. The simulation results from real images agree with our analysis in Section 3.2, and are comparable to the simulation results in Section 3.3 and 3.4. In addition, a better performance is observed in the Baboon example than in Lena. This is because the length of the embedded fingerprints in Baboon, which is $N = 19497$, is larger than that in Lena, which is $N = 13691$. Different characteristics of the two images, e.g., smooth regions and the texture, also contribute to the difference in performance.

3.6 A Few More Collusion Attacks

Besides of the attacks listed in (3.1), we further consider a few other possible collusion attacks. One of them is the *copy and paste attack* where in generating each component of the attacked copy V_j, the colluders equiprobably choose one of the K different copies $\{X_j^{(k)}\}_{k \in S_C}$ and take that value as V_j. In terms of the effects on the energy reduction of the original fingerprints and the effect it has upon the detection performance, this attack and the average attack have similar performance.

Another possible attack is on bounded fingerprints. Since all the K embedded fingerprints are within the range of $[-JND, JND]$, so are the minimum and the maximum of these K copies. The minimum and the maximum values also tell the colluders the lower and upper bounds of the possible fingerprints that will not introduce noticeable distortion. The colluders can randomly choose any value between the minimum and the maximum as the colluded copy without introducing perceptual distortion. We call it the *uniform attack*, which can be modelled as the minmax attack followed by an additive noise $\mathbf{n}$. The extracted fingerprint is $\{Y_j = \frac{1}{2}(W_j^{min} + W_J^{max}) + n_j\}_{j=1}^N$ where n_j is uniformly distributed in $[-\frac{W_j^{max}-W_j^{min}}{2}, \frac{W_j^{max}-W_j^{min}}{2}]$. When K is large, $\{n_j\}$ are approximately uniformly distributed in $[-1, 1]$. Note that in addition to the collusion functions listed in (3.1), the colluders can also add another additive noise to the attacked copy, as long as the overall distortion introduced in the host signal (the extracted fingerprint plus the additive noise in this case) is bounded by JND. This additional noise will hinder the detection performance without degrading the perceptual quality of the attacked signal. We can show that given a fixed power of the overall noise introduced in the host signal, different collusion attacks have comparable performance in defeating the fingerprinting systems.

3.7 Chapter Summary

In this chapter, we have provided theoretical analysis on the effectiveness of different collusion attacks, and studied the perceptual quality of the attacked signals under different collusion attacks. We have also studied several commonly used detection statistics and compared their performance under collusion attacks. Furthermore, we have proposed the pre-processing techniques specifically for collusion

scenarios to improve the detection performance.

We first studied the effectiveness of average and various basic nonlinear collusion attacks with unbounded Gaussian fingerprints. From both our analytical and simulation results, we found that with the three detection statistics as defined in the literature and without any modification, the randomized negative attack is the most effective attack against the fingerprinting system. We showed that the Z statistics are more robust against the minimum and maximum attacks than the other two statistics by implicitly removing the mean of the extracted fingerprint. We also showed that all three statistics have similar performance under other collusion attacks. However, the unbounded Gaussian fingerprints may exceed JND and introduce perceptual distortion in the host signal even when without collusion, and the minimum, maximum, and randomized negative attacks introduce much larger distortion in the attacked copies than others.

In order to remove the noticeable distortion introduced by the unbounded fingerprints, we proposed the bounded Gaussian-like fingerprints, which maintain the robustness against the collusion attacks. With the bounded Gaussian-like fingerprints, the randomized negative attack is still the most effective attack, and the Z statistic are more robust against the minimum and maximum attacks than the other two statistics. The bounding improves the perceptual quality of the fingerprinted copies and that of the attacked copies, and both the fingerprint designer and the colluders do not introduce noticeable distortion.

Observing that the extracted fingerprints under the minimum and the maximum attacks do not have a zero mean, we proposed the pre-processing of the extracted fingerprints, which removes the mean from the extracted fingerprints before applying the detection statistics. We also applied pre-processing to the

extracted fingerprints after the randomized negative attacks, which have distinct bimodal distribution as opposed to the single modality under other collusions. We showed that these pre-processing techniques improve the detection performance, and all detection statistics give similar performance after pre-processing.

We have also studied the effectiveness of different collusion attacks and the performance of different statistics on real images. Our real image simulation results agree with our analysis and are comparable with the ideal case simulation results.

Chapter 4

Fair Collusion Attacks on Scalable Fingerprinting Systems

All prior work on multimedia fingerprinting and collusion attacks assumed that all users receive copies of the same quality. In practice, users access the multimedia content through different communication links and have different bandwidth available. In addition, different users have various processing capability and different computation constraints. To address the heterogeneity of the networks and the end users, it is often required to have *scalability* during video coding and transmission. "As we move to the convergence of wireless, Internet and multimedia, scalability becomes increasingly important for rich media access from anywhere, by anyone, at any time, with any device, and in any form." [60]

In this chapter, we study the impact of scalability on multimedia fingerprinting and collusion attacks. We first study the collusion attacks when different colluders receive copies of different quality. In particular, we consider *fair* collusion attacks where all colluders share the same risk and have equal probability of detection, and analyze the effectiveness of the collusion attacks under the fairness constraints. We

then investigate the collusion resistance of the scalable fingerprinting systems, and evaluate the number of colluders that are required to undermine the tracing capability of the scalable fingerprinting systems under different system requirements.

This chapter is organized as follows. Section 4.1 introduces the system model of the scalable video coding systems and the digital fingerprinting systems. Section 4.2 analyzes the fairness constraints on the collusion attacks when different colluders receive copies of different quality. We study the effectiveness of the collusion attacks on scalable fingerprinting systems in Section 4.3, and analyze the collusion resistance of the fingerprinting systems in Section 4.4. Section 4.5 shows the simulation results on real video sequences.

4.1 System Model

In this section, we first review temporally scalable video coding systems, and then we introduce a digital fingerprinting system that consists of three parts: fingerprint embedding, collusion attacks and fingerprint detection and colluder identification. Finally, we discuss the performance criteria that we use in this chapter to measure the effectiveness of the collusion attacks and the performance of the detectors.

4.1.1 Temporally Scalable Video Coding Systems

In the literature, scalable video coding is widely used to accommodate heterogenous networks and users with different computation capabilities. One example of scalable coding is the layered coding, where the video content is decomposed into non-overlapping streams (layers) with different priorities [60]. The base layer contains the most important information of the video content, provides the roughest

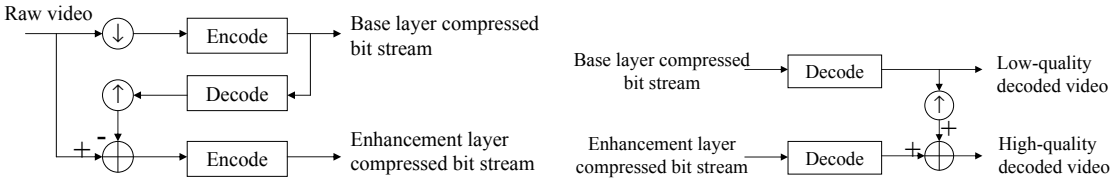


Figure 4.1: A two-layer temporally scalable codec. Left: encoder, right: decoder.

resolution of the video, and is received by all users in the systems. The enhancement layers contain less important information, gradually refine the reconstructed video at the decoder's side, and are only received by the users who have sufficient bandwidth and computation capability.

In this chapter, we consider temporally scalable video coding, which provides multiple versions of the same video with different temporal resolutions or frame rates. Figure 1 shows the block diagrams of a two-layer temporally scalable codec. At the encoder's side, the raw video is temporally down-sampled and encoded to generate the base layer bit stream. Then, the encoder calculates the difference between the temporally up-sampled base layer and the original video sequence, and encodes this residue to generate the enhancement layer bit stream. In a temporally scalable decoder, to reconstruct a high-quality video, both the base layer and the enhancement layer bit streams have to be received and decoded. Then the temporally up-sampled base layer is combined with the enhancement layer refinements to form the high-quality decoded video.

The simplest way to perform temporal down-sampling and temporal up-sampling is by frame skipping and frame copying, respectively. For example, temporal down-sampling with a ratio of 2:1 can be achieved by discarding one frame from every two frames; and temporal up-sampling 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.

In this chapter, we consider a temporally scalable video coding system with three-layer scalability, and use 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. Assume that F_b, F_{e1} and F_{e2} are the sets containing the indices of the frames that are are encoded in the base layer, enhancement layer 1 and enhancement layer 2, respectively. As an example, if $|F_b| : |F_{e1}| : |F_{e2}| = 1 : 1 : 2$ where $|A|$ denotes the size of the set A, $F_b = \{j = 4k + 1 : k = 0, 1, \cdots\}$, $F_{e1} = \{j = 4k + 3 : k = 0, 1, \cdots\}$ and $F_{e2} = \{j = 2k : k = 0, 1, \cdots\}$.

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

4.1.2 Digital Fingerprinting System and Collusion Attacks

We consider a digital fingerprinting system that consists of three parts: fingerprint embedding, collusion attacks and fingerprint detection.

Fingerprint Embedding

Spread spectrum embedding has been widely used in the literature due to its robustness against many attacks [13], [47]. For the jth frame in the video se-

quence 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 $\mathbf{X}_j^{(i)}$, 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 [47], 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.

Previous works have shown that Gaussian distributed fingerprints are more robust against the nonlinear collusion attacks [52] and are resilient to the statistical/histogram attacks [15]. Therefore, we consider Gaussian fingerprints and assume that $\left\{\mathbf{W}_j^{(i)}\right\}$ follow normal distribution with zero mean and variance σ_W^2. In addition, to combat the intra-content collusion attacks [56], in each distributed copy $\{\mathbf{X}_j^{(i)}\}$, similar to the work in [55], 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 $\mathbf{S}_{j_2}$. Finally, in this chapter, we use orthogonal fingerprint modulation [58] and assign independent fingerprints to different users.

Collusion Attacks

Assume that K out of M users collude, and SC is the set containing the indices of these colluders. The colluders apply collusion attacks to remove or attenuate the original fingerprints. In a recent investigation [64], we have shown that order statistics based nonlinear collusion attacks can be modeled as the averaging attack

followed by an additive noise. Under the constraint that the colluded copies under different collusion attacks have the same perceptual quality, different collusion attacks have approximately identical performance. Therefore, in this chapter, we focus on the averaging collusion attack.

During collusion, the colluders first divide themselves into three non-overlapping subgroups:

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

Then, the colluders apply the *intra-group collusion attacks* followed by the *inter-group collusion attacks* to generate the colluded copy $\{\mathbf{V}_j\}$, as shown in Figure 4.2. The colluders first 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 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 $\mathbf{Z}_j^{all} = \sum_{i \in SC^{all}} \mathbf{X}_j^{(i)} / K^{all}$.

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

- For each frame $j \in F_b$ in the base layer,

$$\mathbf{V}_j = \beta_1 \mathbf{Z}_j^b + \beta_2 \mathbf{Z}_j^{b,e1} + \beta_3 \mathbf{Z}_j^{all} + \mathbf{n}_j, \tag{4.1}$$

  where $\beta_1 + \beta_2 + \beta_3 = 1$ to maintain the average intensity of the colluded copy. To guarantee that the energy of each of the original fingerprints is reduced, we let $0 \leq \beta_1,\ \beta_2,\ \beta_3 \leq 1$. In (4.1), $\mathbf{n}_j$ is the additive noise that the colluders add to $\mathbf{V}_j$ to further hinder the detection.

- 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,

$$\mathbf{V}_j = \alpha_1 \mathbf{Z}_j^{b,e1} + \alpha_2 \mathbf{Z}_j^{all} + \mathbf{n}_j, \tag{4.2}$$

  where $0 \leq \alpha_1,\ \alpha_2 \leq \alpha_1 + \alpha_2 = 1$ and $\mathbf{n}_j$ is an additive noise.

- If $F_{e2} \subset F^c$ and the colluded copy contains frames in all three layers, then for each frame $j \in F_{e2}$ in the enhancement layer 2,

$$\mathbf{V}_j = \mathbf{Z}_j^{all} + \mathbf{n}_j, \tag{4.3}$$

  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 [47], a larger $JND_j(k)$ implies that a noise of larger energy can be added to the corresponding

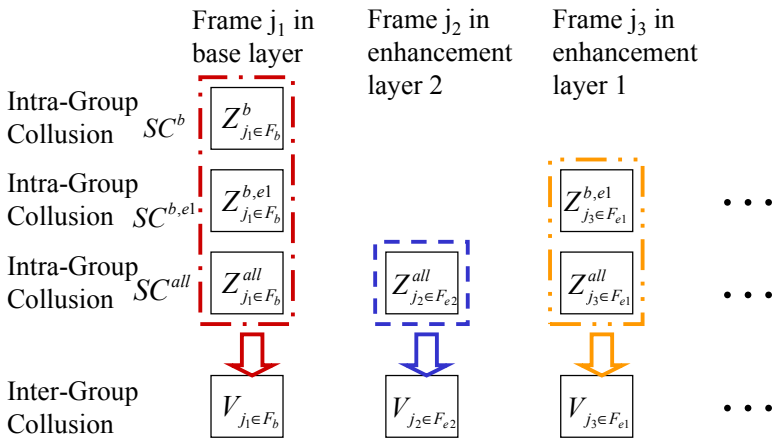


Figure 4.2: The intra-group and the inter-group collusion attacks.

host signal component without introducing perceptually distinguishable distortion; and the colluders usually maximize the energy of the additive noise $\mathbf{n}_j$ under the perceptual quality constraints in order to maximize the effectiveness of the collusion attacks. In this chapter, we model $\{\frac{\mathbf{n}_j}{JND_j}\}$ as i.i.d. following distribution $\mathcal{N}(0, \sigma_n^2)$.

In addition, we assume that the collusion attacks is a fair attack where all colluders share the same risk and are equally likely to be detected. 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. The analysis of the fairness constraints and the selection of the parameters are in Section 4.2.

Fingerprint Detection and Colluder Identification

When the content owner discovers the unauthorized redistribution of $\{\mathbf{V}_j\}_{j\in F^c}$, he applies a fingerprint detection process to identify the colluders.

In digital fingerprinting applications, the host signal $\{\mathbf{S}_j\}$ is often made available to the detector. To improve the detection performance [58], [64], we consider a non-blind detection scenario where the host signal is first removed from the col-

luded copy before colluder identification. We assume that the detector has a frame synchronization module that finds the corresponding original frame $\mathbf{S}_j$ for each frame $\mathbf{V}_j$ in the colluded copy. Then for each frame j in the colluded 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 $\{\mathbf{Y}_j\}_{j\in F^c}$ and each of the M original fingerprints $\{\mathbf{W}^{(i)}_{j\in F^{(i)}}\}$, compares with a threshold and outputs a set $\widehat{SC}$ containing the estimated indices of the colluders.

We use the correlation based detection statistics [48] that have been widely adopted in the literature. 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. For example, if $F^c = F_b \cup F_{e1}$, then $\breve{F}^{(i_1)} = F_b$ for $\mathbf{u}^{(i_1)} \in \mathbf{U}^b$; $\breve{F}^{(i_2)} = F_b \cup F_{e1}$ for $\mathbf{u}^{(i_2)} \in \mathbf{U}^{b,e1}$; and $\breve{F}^{(i_3)} = F_b \cup F_{e1}$ for $\mathbf{u}^{(i_3)} \in \mathbf{U}^{all}$. Then the detector calculates the detection statistics

$$T_N^{(i)} = \left(\sum_{j\in\breve{F}^{(i)}} \langle \mathbf{Y}_j, \mathbf{W}_j^{(i)} \rangle \right) / \sqrt{\sum_{j\in\breve{F}^{(i)}} ||\mathbf{W}_j^{(i)}||^2}, \tag{4.4}$$

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

4.1.3 Performance Criteria

To evaluate the effectiveness of the collusion attacks and the performance of the detection statistics, we adopt the commonly used criteria in the literature [58, 64] and use the following measurements:

- P_d: the probability of capturing at least one colluder;
- P_{fp}: the probability of accusing at least one innocent user;

- F_d: the fraction of colluders that are successfully captured; and
- F_{fp}: the fraction of innocent users that are falsely accused.

To measure the quality of the colluded copy, we use the total number of frames in the colluded copy $L^c = |F^c|$. We let $L^c \in \{|F_b|,\ |F_b| + |F_{e1}|,\ |F_b| + |F_{e1}| + |F_{e2}|\}$ for simplicity, which correspond to the three scenarios where the colluded copy has the lowest, medium and highest temporal resolution, respectively. When L^c is larger, the colluded copy has higher temporal resolution and better quality.

4.2 Fairness Constraints on the Collusion Attacks

In this section, given the system model as in Section 4.1, we analyze the fairness constraints on the collusion attacks and study the selection of collusion parameters during collusion.

4.2.1 Analysis of the Detection Statistics

For each frame $j \in F_b$ in the base layer, from (4.1), the extracted fingerprint $\mathbf{Y}_j$ can be rewritten as

$$\mathbf{Y}_j = \frac{\beta_1}{K^b} \sum_{i \in SC^b} \mathbf{W}_j^{(i)} + \frac{\beta_2}{K^{b,e1}} \sum_{i \in SC^{b,e1}} \mathbf{W}_j^{(i)} + \frac{\beta_3}{K^{all}} \sum_{i \in SC^{all}} \mathbf{W}_j^{(i)} + \mathbf{n}_j / JND_j. \quad (4.5)$$

If the colluded copy contains frames in the enhancement layers, from (4.2), for each frame $j \in F_{e1}$ in the enhancement layer 1,

$$\mathbf{Y}_j = \frac{\alpha_1}{K^{b,e1}} \sum_{i \in SC^{b,e1}} \mathbf{W}_j^{(i)} + \frac{\alpha_2}{K^{all}} \sum_{i \in S^{all}} \mathbf{W}_j^{(i)} + \mathbf{n}_j / JND_j. \quad (4.6)$$

If the colluded copy contains frames in all three layers, from (4.3), for each frame $j \in F_{e2}$ in the enhancement layer 2,

$$\mathbf{Y}_j = \frac{1}{K^{all}} \sum_{j \in SC^{all}} \mathbf{W}_j^{(i)} + \mathbf{n}_j / JND_j. \tag{4.7}$$

It is straightforward to show that given the colluder set SC, for each user $\mathbf{u}^{(i)}$, the detection statistics follows a Gaussian distribution with mean $\mu^{(i)}$ and variance σ_n^2, i.e.,

$$p\left(T_N^{(i)} | SC\right) \sim \mathcal{N}\left(\mu^{(i)}, \sigma_n^2\right), \tag{4.8}$$

where σ_n^2 is the variance of the additive noise $\mathbf{n}_j / JND_j$. For user $\mathbf{u}^{(i)}$, $\mu^{(i)} = 0$ when he is innocent, and $\mu^{(i)} > 0$ when he is guilty. For $i \in SC$, $\mu^{(i)}$ depends on the number of frames in the colluded copy and the number frames that colluder $\mathbf{u}^{(i)}$ receives.

$\mathbf{F^c = F_b \cup F_{e1} \cup F_{e2}}$

When the colluded copy contains frames in all three layers, we can show that for $i \in SC$,

$$\mu^{(i)} = \begin{cases} \frac{\beta_1}{K^b} \sqrt{\sum_{j \in F^{(i)}} ||\mathbf{W}_j^{(i)}||^2} & \text{if } i \in SC^b, \\ \frac{\beta_2 \sum_{j \in F_b} ||\mathbf{W}_j^{(i)}||^2 + \alpha_1 \sum_{j \in F_{e1}} ||\mathbf{W}_j^{(i)}||^2}{K^{b,e1} \sqrt{\sum_{j \in F^{(i)}} ||\mathbf{W}_j^{(i)}||^2}} & \text{if } i \in SC^{b,e1}, \\ \frac{\beta_3 \sum_{j \in F_b} ||\mathbf{W}_j^{(i)}||^2 + \alpha_2 \sum_{j \in F_{e1}} ||\mathbf{W}_j^{(i)}||^2 + \sum_{j \in F_{e2}} ||\mathbf{W}_j^{(i)}||^2}{K^{all} \sqrt{\sum_{j \in F^{(i)}} ||\mathbf{W}_j^{(i)}||^2}} & \text{if } i \in SC^{all}. \end{cases} \tag{4.9}$$

Define N_b, N_{e1} and N_{e2} as the lengths of the fingerprints that are embedded in the frames in the base layer, enhancement layer 1 and enhancement layer 2, respectively. Since $\{\mathbf{W}_j^{(i)}\}$ follow Gaussian distribution with zero mean and variance σ_W^2, we can have the approximation that

$$\sum_{j \in F_b} ||\mathbf{W}_j^{(i)}||^2 \approx N_b \sigma_W^2 \quad \text{for} \quad i \in SC,$$

$$\begin{aligned} \sum_{j\in F_{e1}} ||\mathbf{W}_{e1}^{(i)}||^2 &\approx N_{e1}\sigma_W^2 \quad \text{for} \quad i \in SC^{b,e1} \cup SC^{all}, \\ \text{and} \quad \sum_{j\in F_{e2}} ||\mathbf{W}_{e1}^{(i)}||^2 &\approx N_{e2}\sigma_W^2 \quad \text{for} \quad i \in SC^{all}. \end{aligned} \tag{4.10}$$

Therefore, we can approximate the mean of the detection statistics by

$$\mu^{(i)} \approx \begin{cases} \frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W & \text{if } i \in SC^b, \\ \frac{\beta_2 N_b + \alpha_1 N_{e1}}{K^{b,e1}\sqrt{N_b+N_{e1}}}\sigma_W & \text{if } i \in SC^{b,e1}, \\ \frac{\beta_3 N_b + \alpha_2 N_{e1} + N_{e2}}{K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}\sigma_W & \text{if } i \in SC^{all}. \end{cases} \tag{4.11}$$

$\mathbf{F^c = F_b \cup F_{e1}}$

When the colluded copy contains frames in the base layer and the enhancement layer 1, similar to the above analysis, we can have the approximation that

$$\mu^{(i)} \approx \begin{cases} \frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W & \text{if } i \in SC^b, \\ \frac{\beta_2 N_b + \alpha_1 N_{e1}}{K^{b,e1}\sqrt{N_b+N_{e1}}}\sigma_W & \text{if } i \in SC^{b,e1}, \\ \frac{\beta_3 N_b + \alpha_2 N_{e1}}{K^{all}\sqrt{N_b+N_{e1}}}\sigma_W & \text{if } i \in SC^{all}. \end{cases} \tag{4.12}$$

$\mathbf{F^c = F_b}$

When the colluded copy contains frames in the base layer only, we have

$$\mu^{(i)} \approx \begin{cases} \frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W & \text{if } i \in SC^b, \\ \frac{\beta_2\sqrt{N_b}}{K^{b,e1}}\sigma_W & \text{if } i \in SC^{b,e1}, \\ \frac{\beta_3\sqrt{N_b}}{K^{all}}\sigma_W & \text{if } i \in SC^{all}. \end{cases} \tag{4.13}$$

4.2.2 Analysis of the Fairness Constraints

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

$$P^{(i)} = P\left[T_N^{(i)} > h\right] = Q\left(\frac{h-\mu^{(i)}}{\sigma_n}\right), \tag{4.14}$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ is the Gaussian tail function. Therefore, for a given σ_n^2 and a given threshold h, all colluders share the same risk and are equally likely to be detected if their detection statistics have the same mean. In this section, we will study the fairness constraints on the collusion attacks.

$\mathbf{F^c = F_b \cup F_{e1} \cup F_{e2}}$

When the colluded copy contains frames in all three layers, from (4.11), the colluders seek $\{0 \le \beta_k \le 1\}_{k=1,2,3}$ and $\{0 \le \alpha_l \le 1\}_{l=1,2}$ to satisfy

$$\begin{aligned} & \frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W = \frac{\beta_2 N_b + \alpha_1 N_{e1}}{K^{b,e1}\sqrt{N_b+N_{e1}}}\sigma_W = \frac{\beta_3 N_b + \alpha_2 N_{e1} + N_{e2}}{K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}\sigma_W, \\ s.t. \quad & \beta_1+\beta_2+\beta_3 = 1,\ \alpha_1+\alpha_2 = 1. \end{aligned} \tag{4.15}$$

Note that

$$\frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W = \frac{\beta_2 N_b + \alpha_1 N_{e1}}{K^{b,e1}\sqrt{N_b+N_{e1}}}\sigma_W \iff \frac{K^{b,e1}\sqrt{N_b+N_{e1}}}{K^b\sqrt{N_b}} = \frac{\beta_2 N_b + \alpha_1 N_{e1}}{\beta_1 N_b}. \tag{4.16}$$

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

$$\begin{aligned} & \frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W = \frac{\beta_3 N_b + \alpha_2 N_{e1} + N_{e2}}{K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}\sigma_W \\ \iff \quad & \frac{K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}{K^b\sqrt{N_b}} = \frac{N_b+N_{e1}+N_{e2}}{\beta_1 N_b} - 1 - \frac{\beta_2 N_b + \alpha_1 N_{e1}}{\beta_1 N_b}. \end{aligned} \tag{4.17}$$

Plug (4.16) into (4.17), we have

$$\frac{N_b+N_{e1}+N_{e2}}{\beta_1 N_b} = \frac{K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}{K^b\sqrt{N_b}} + 1 + \frac{K^{b,e1}\sqrt{N_b+N_{e1}}}{K^b\sqrt{N_b}}. \tag{4.18}$$

Therefore, to satisfy the fairness constraints, from (4.16) and (4.18), the colluders should choose

$$\beta_1 = \frac{N_b + N_{e1} + N_{e2}}{N_b} \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} \quad (4.19)$$

and

$$\beta_2 N_b + \alpha_1 N_{e1} = \frac{(N_b + N_{e1} + N_{e2})\, K^{b,e1}\sqrt{N_b + N_{e1}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}}. \quad (4.20)$$

From Section 4.1.2, the collusion parameters are required to be in the range of $[0, 1]$. From (4.19), $0 \leq \beta_1 \leq 1$ if and only if

$$\frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} \leq \frac{N_b}{N_b + N_{e1} + N_{e2}}. \quad (4.21)$$

Furthermore, from (4.20),

$$\alpha_1 = \frac{N_b + N_{e1} + N_{e2}}{N_{e1}} \frac{K^{b,e1}\sqrt{N_b + N_{e1}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} - \beta_2 \frac{N_b}{N_{e1}} \quad (4.22)$$

Given β_1 as in (4.19), $0 \leq \beta_2 \leq 1 - \beta_1$. Consequently, from (4.22), we have $\underline{\alpha} \leq \alpha_1 \leq \overline{\alpha}$, where

$$\underline{\alpha} = \frac{N_b + N_{e1} + N_{e2}}{N_{e1}} \frac{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} - \frac{N_b}{N_e 1} \quad (4.23)$$

and

$$\overline{\alpha} = \frac{N_b + N_{e1} + N_{e2}}{N_{e1}} \frac{K^{b,e1}\sqrt{N_b + N_{e1}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}}. \quad (4.24)$$

If $[0, 1] \cap [\underline{\alpha}, \overline{\alpha}]$ is not empty, then there exists at least one α_1^* such that $0 \leq \alpha_1^* \leq 1$ and $\underline{\alpha} \leq \alpha_1^* \leq \overline{\alpha}$. Note that $\overline{\alpha} > 0$, so $[0, 1] \cap [\underline{\alpha}, \overline{\alpha}] \neq \emptyset$ if and only if $\underline{\alpha} \leq 1$, which is equivalent to

$$\frac{K^{all}\sqrt{N_b + N_{e1} + N_{e2}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} \geq \frac{N_{e2}}{N_b + N_{e1} + N_{e2}}. \quad (4.25)$$

To summarize, in order to generate a colluded copy with the highest temporal resolution under the fairness constraints, $\left(K^b, K^{b,e1}, K^{all}\right)$ and (N_b, N_{e1}, N_{e2}) have to satisfy (4.21) and (4.25), and the colluders should choose the collusion parameters as in (4.19) and (4.20).

$\mathbf{F^c = F_b \cup F_{e1}}$

In this scenario, for colluder $\mathbf{u}^{(i_1 \in SC^{all})}$ and colluder $\mathbf{u}^{(i_2 \in SC^{b,e1})}$ who received 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 $N_b + N_{e1}$. To ensure that for each frame $j \in F^c$ in the colluded copy, the energies of these two colluders' fingerprints $\mathbf{X}_j^{(i_1)}$ and $\mathbf{X}_j^{(i_2)}$ are reduced by the same ratio, the colluders should choose $\alpha_1/K^{b,e1} = \alpha_2/K^{all}$ and $\beta_2/K^{b,e1} = \beta_3/K^{all}$. For a given $0 \leq \beta_1 \leq 1$, it is equivalent to

$$\begin{aligned} \alpha_1 &= \frac{K^{b,e1}}{K^{b,e1} + K^{all}}, \quad \alpha_2 = 1 - \alpha_1, \\ \beta_2 &= \frac{K^{b,e1}}{K^{b,e1} + K^{all}} (1 - \beta_1), \quad \text{and} \quad \beta_3 = 1 - \beta_1 - \beta_2. \end{aligned} \tag{4.26}$$

Consequently, for these two colluders,

$$\mu^{(i_1)} = \mu^{(i_2)} = \frac{(1-\beta_1)N_b + N_{e1}}{(K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}}\sigma_W. \tag{4.27}$$

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

$$\frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W = \frac{(1-\beta_1)N_b + N_{e1}}{(K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}}\sigma_W, \tag{4.28}$$

and the solution is

$$\beta_1 = \frac{N_b + N_{e1}}{N_b} \frac{K_b\sqrt{N_b}}{K^b\sqrt{N_b} + (K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}}. \tag{4.29}$$

With β_1 as in (4.29), $0 \leq \beta_1 \leq 1$ if and only if

$$\frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + (K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}} \leq \frac{N_b}{N_b + N_{e1}}. \tag{4.30}$$

Given $0 \leq \beta_1 \leq 1$, from (4.26), it is straightforward to show that $0 \leq \beta_2, \beta_3, \alpha_1, \alpha_2 \leq 1$.

To summarize, under the fairness constraints, $\left(K^b, K^{b,e1}, K^{all}\right)$ and (N_b, N_{e1}, N_{e2}) have to satisfy (4.30) if the colluders wish to generate a colluded copy of the medium temporal resolution. The colluders should choose the collusion parameters as in (4.26) and (4.29).

$\mathbf{F^c = F_b}$

When the colluded copy contains frames in the base layer only, the colluders choose $\{0 \leq \beta_k \leq 1\}_{k=1,2,3}$ with $\beta_1 + \beta_2 + \beta_3 = 1$ to satisfy

$$\frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W = \frac{\beta_2\sqrt{N_b}}{K^{b,e1}}\sigma_W = \frac{\beta_3\sqrt{N_b}}{K^{all}}\sigma_W, \tag{4.31}$$

and the solution is

$$\beta_1 = \frac{K^b}{K^b + K^{b,e1} + K^{all}},\ \beta_2 = \frac{K^{b,e1}}{K^b + K^{b,e1} + K^{all}},\ \text{and } \beta_3 = \frac{K^{all}}{K^b + K^{b,e1} + K^{all}}. \tag{4.32}$$

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

4.2.3 Summary of the Fairness Constraints and the Selection of Collusion Parameters

From (4.21), (4.25) and (4.30), 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. Note that the adjacent frames in a video sequence are similar to each other and have approximately

Table 4.1: Fairness constraints on collusion attacks and the selection of collusion parameters.

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

the same number of embeddable coefficients, the colluders can use the following approximation $N_b : N_{e1} : N_{e2} \approx |F_b| : |F_{e1}| : |F_{e2}|$.

Table 4.1 summarizes the fairness 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 temporal resolution, respectively. From Table 4.1, we can observe that given $\left(K^b, K^{b,e1}, K^{all}\right)$ and (N_b, N_{e1}, N_{e2}), the fairness constraints on the collusion attacks are the constraints on the best possible quality of the colluded copy. With increasing resolution of the colluded copy, the fairness constraints on $\left(K^b, K^{b,e1}, K^{all}\right)$ and (N_b, N_{e1}, N_{e2}) are more severe, and generating a colluded copy of higher quality requires that there are more colluders in subgroups $SC^{b,e1}$ and SC^{all} who receive the enhancement layer bit streams.

4.3 Effectiveness of the Collusion Attacks under the Fairness Constraints

4.3.1 Statistical Analysis

Assume that there are a total of M users in the system. From the analysis in the previous section, if the colluders select the collusion parameters as in Table 4.1, then given a colluder set SC, for each user $\mathbf{u}^{(i)}$,

$$p\left(T_N^{(i)}|SC\right) \sim \begin{cases} \mathcal{N}\left(\mu,\ \sigma_n^2\right) & \text{if } i \in SC, \\ \mathcal{N}\left(0,\ \sigma_n^2\right) & \text{if } i \notin SC, \end{cases} \tag{4.33}$$

where σ_n^2 is the variance of $\mathbf{n}_j/JND_j$, and 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. It is straightforward to show that for $i \in SC$, μ in

(4.33) can be approximated by

$$\begin{aligned}
\mu &= \frac{\beta_1\sqrt{N_b}}{K^b}\sigma_W \\
&\approx \begin{cases}
\frac{N_b+N_{e1}+N_{e2}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}\sigma_W & \text{if } F^c = F_b \cup F_{e1} \cup F_{e2}, \\
\frac{N_b+N_{e1}}{K^b\sqrt{N_b}+(K^{b,e1}+K^{all})\sqrt{N_b+N_{e1}}}\sigma_W & \text{if } F^c = F_b \cup F_{e1}, \\
\frac{\sqrt{N_b}}{K^b+K^{b,e1}+K^{all}}\sigma_W & \text{if } F^c = F_b.
\end{cases}
\end{aligned} \tag{4.34}$$

Given a threshold h, from (4.33), we can approximate P_d and P_{fp} by

$$\begin{aligned}
P_d &= P\left[\max_{i\in SC} T_N^{(i)} > h\right] \approx 1 - \left[1 - Q\left(\frac{h-\mu}{\sigma_n}\right)\right]^K, \\
\text{and} \quad P_{fp} &= P\left[\max_{i\notin SC} T_N^{(i)} > h\right] \approx 1 - \left[1 - Q\left(\frac{h}{\sigma_n}\right)\right]^{M-K}.
\end{aligned} \tag{4.35}$$

In addition, $E[F_d]$ and $E[F_{fp}]$ can be approximated by

$$\begin{aligned}
E[F_d] &= \sum_{i\in SC} P\left[T_N^{(i)} > h\right]/K \approx Q\left(\frac{h-\mu}{\sigma_n}\right), \\
\text{and} \quad E[F_{fp}] &= \sum_{i\notin SC} P\left[T_N^{(i)} > h\right]/(M-K) \approx Q\left(\frac{h}{\sigma_n}\right).
\end{aligned} \tag{4.36}$$

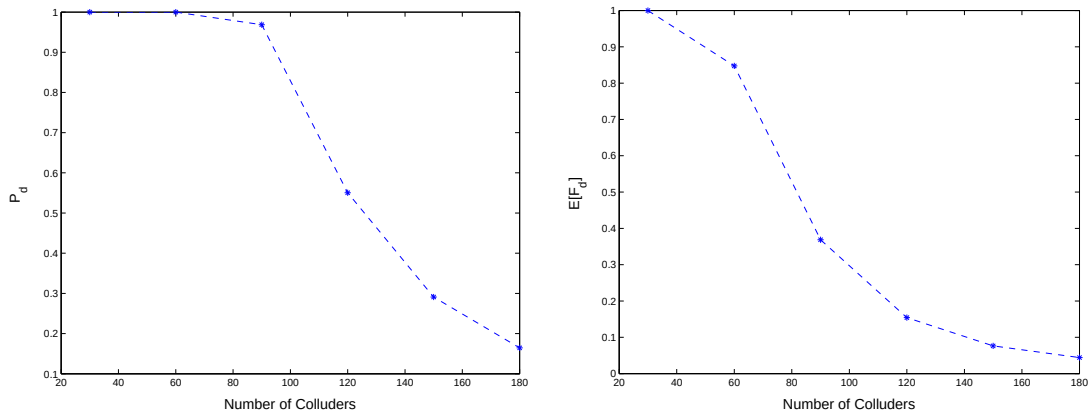
From (4.34-4.36), the effectiveness of the collusion attacks depends on the total number of colluders K as well as the temporal resolution of the colluded copy L^c. For a fixed $L^c = |F^c|$, the colluders have a smaller probability to be captured and the collusion attack is more effective when there are more colluders in the systems. If the total number of colluders K is fixed, the colluders have a larger probability of detection when the colluded copy has a higher temporal resolution, and therefore, better quality. This is because the extracted fingerprint is longer and provides more information of the colluders' identities to the detector. The colluders have to take into consideration the tradeoff between the probability of detection and the perceptual quality of the colluded copy during collusion.

4.3.2 Simulation Results

From human visual models [47], not all coefficients are embeddable due to imperceptibility constraints. For real video sequences like "akiyo", "foreman" 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 we test on a total of 40 frames. We choose $F_b = \{j : j = 4k + 1, k = 0, \cdots, 9\}$, $F_{e1} = \{j : j = 4k + 3, k = 0, \cdots, 9\}$ and $F_{e2} = \{j : j = 2k, k = 1, \cdots, 20\}$ as an example of the temporal scalability, and the lengths of the fingerprints embedded in the base layer, enhancement layer 1 and enhancement layer 2 are $N_b = 50000$, $N_{e1} = 50000$ and $N_{e2} = 100000$, respectively. We assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. Each user is assigned a unique fingerprint following Guassian distribution $\mathcal{N}(0, \sigma_W^2)$ with $\sigma_W^2 = 1/9$, and for each user, fingerprints embedded in adjacent frames are correlated with each other. The fingerprints for different users are generated independently.

We assume that $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ are the number of colluders in subgroups SC^b, $SC^{b,e1}$ and SC^{all}, respectively. During collusion, the colluders apply the intra-group collusion attacks followed by the inter-group collusion attacks. Furthermore, we assume that the additive noise $\mathbf{n}_j/JND_j$, which is introduced into the colluded copy by the colluders to further hinder the detection, has variance $\sigma_n^2 = 2\sigma_W^2$.

In Figure 4.3, we fix the ratio $K^b : K^{b,e1} : K^{all} = 1 : 1 : 1$, and assume that the colluded copy has temporal resolution $F^c = F_b \cup F_{e1}$, which satisfies the fairness constraints in Table 4.1. In Figure 4.3 (a), we fix the probability of accusing at

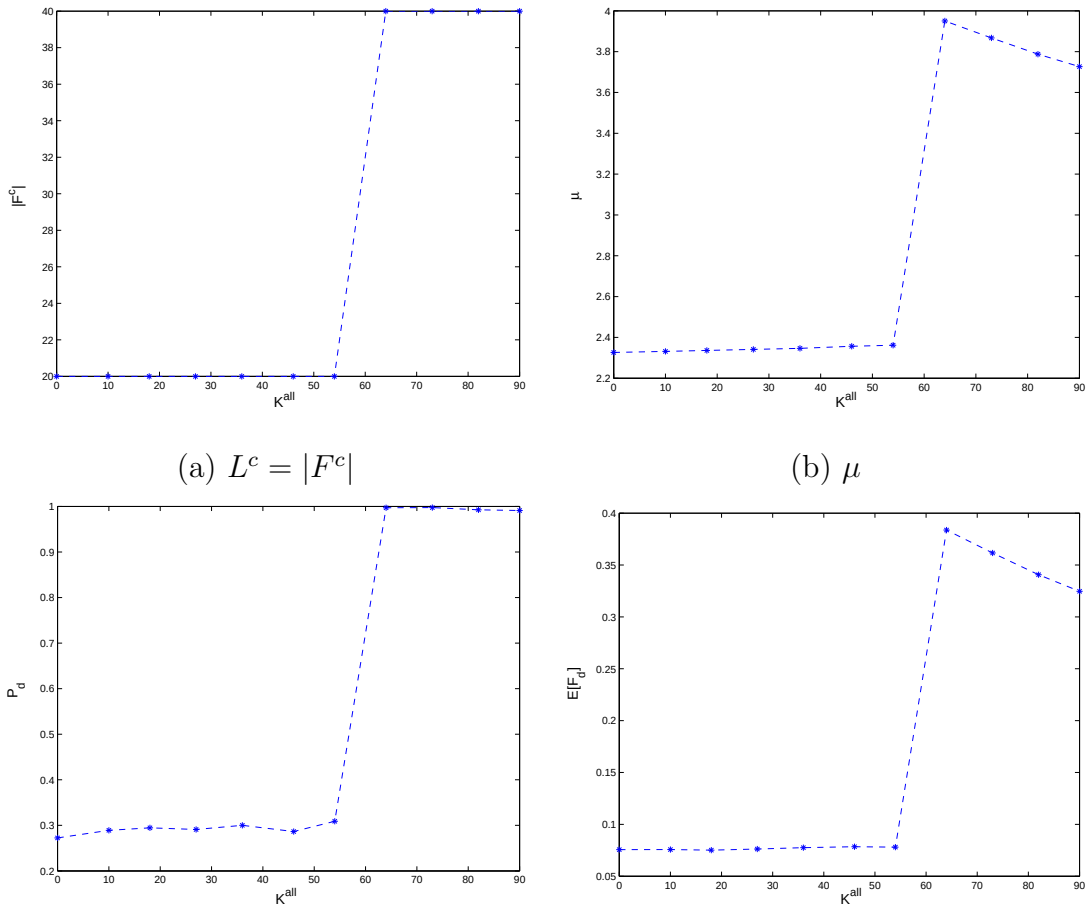


(a) P_d (b) $E[F_d]$

Figure 4.3: Effectiveness of the collusion attacks on scalable fingerprinting systems. Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. $N_b = 50,000$, $N_{e1} = 50,000$ and $N_{e2} = 100,000$. $K^b : K^{b,e1} : K^{all} = 1 : 1 : 1$ 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).

least one innocent user P_{fp} as 10^{-3} and plot the probability of capturing at least one colluder P_d when the total number of colluders K increases. In Figure 4.3 (b), the fraction of the innocent users that are accused is $E[F_{fp}] = 10^{-3}$ and we plot the fraction of the colluders that are captured $E[F_d]$ when K increases. From Figure 4.3, the collusion attacks are more effective when the total number of colluders K increases.

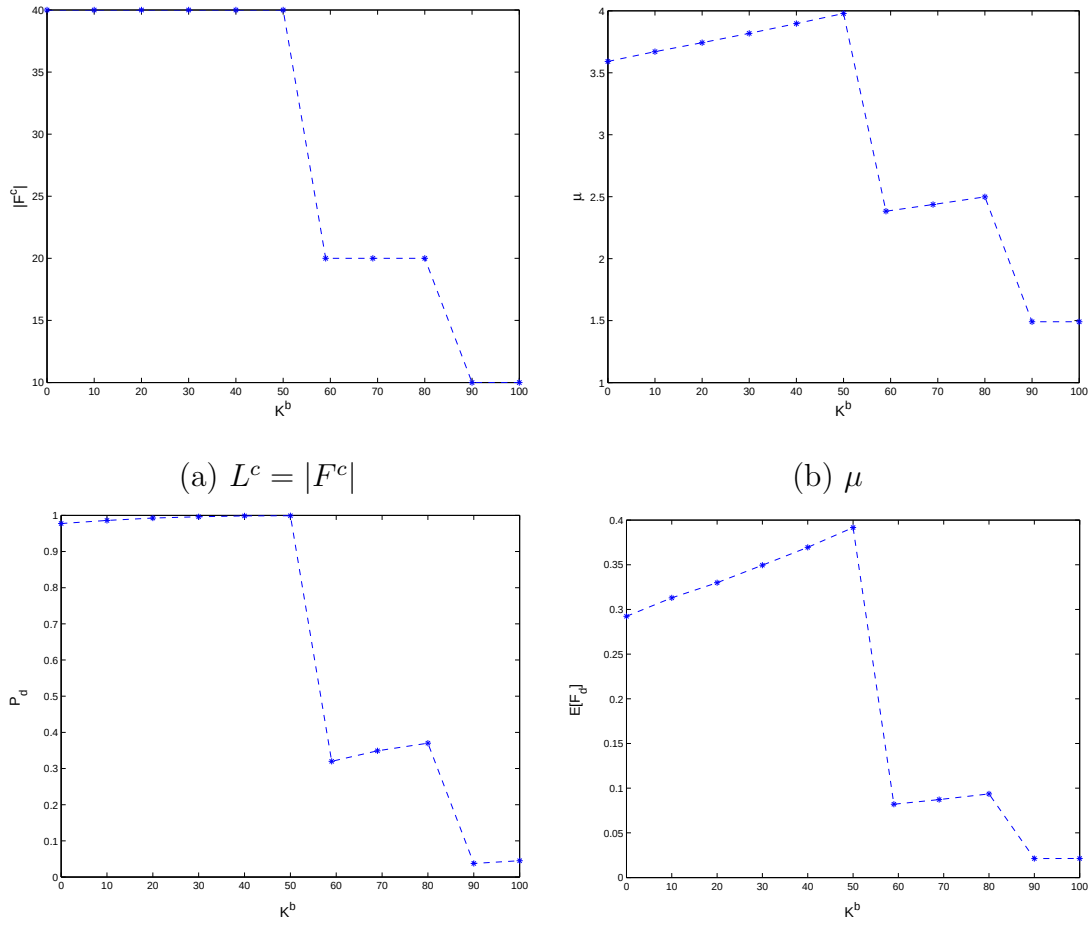
In Figures 4.4, we fix the total number of colluders $K = 150$, and compare the effectiveness of the collusion attacks when the number of colluders in each subgroup $(K^b, K^{b,e1}, K^{all})$ changes and when the temporal resolution of the colluded copy L^c changes. We assume that the colluders generate a colluded copy of the best possible quality under the fairness constraints. In Figure 4.4, $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and



(a) $L^c = |F^c|$ (b) μ

(c) P_d (d) $E[F_d]$

Figure 4.4: Effectiveness of the collusion attacks on scalable fingerprinting systems. Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. $K = 150$ and $(K^b, K^{b,e1}, K^{all})$ are on Line (4.37). $\sigma_n^2/\sigma_W^2 = 2$. $0 \leq K^b, K^{e1}, K^{e2} \leq 150$. $P_{fp} = 10^{-3}$ in (c), and $E[F_{fp}] = 10^{-3}$ in (d).



(a) $L^c = |F^c|$ (b) μ

(c) P_d (d) $E[F_d]$

Figure 4.5: Effectiveness of the collusion attacks on scalable fingerprinting systems. Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. $K = 150$ and $(K^b, K^{b,e1}, K^{all})$ are on Line (4.38). $\sigma_n^2/\sigma_W^2 = 2$. $0 \leq K^b, K^{e1}, K^{e2} \leq 150$. $P_{fp} = 10^{-3}$ in (c), and $E[F_{fp}] = 10^{-3}$ in (d).

they are on the boundary of the fairness constraints (4.21), where

$$\frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}=\frac{N_b}{N_b+N_{e1}+N_{e2}}. \quad (4.37)$$

Figure 4.4 (a) shows the number of frames in the colluded copy L^c for different $(K^b, K^{b,e1}, K^{all})$, and Figure 4.4 (b) shows the corresponding means of the detection statistics. In Figure 4.4 (c), $P_{fp} = 10^{-3}$ and we compare P_d of the collusion attacks with different $(K^b, K^{b,e1}, K^{all})$. In Figure 4.4 (d), $E[F_{fp}] = 10^{-3}$ and we compare $E[F_d]$ of the collusion attacks when $(K^b, K^{b,e1}, K^{all})$ varies.

Similar to Figure 4.4, in Figure 4.5, the total number of colluders is fixed as $K = 150$, and we assume that the colluders generate a colluded copy of the best possible quality under the fairness constraints. In Figure 4.5, $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and they are on another boundary of the fairness constraints (4.25), where

$$\frac{K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}{K^b\sqrt{N_b}+K^{b,e1}\sqrt{N_b+N_{e1}}+K^{all}\sqrt{N_b+N_{e1}+N_{e2}}}=\frac{N_{e2}}{N_b+N_{e1}+N_{e2}}. \quad (4.38)$$

From Figures 4.4 and 4.5, under the fairness constraints, the colluders can generate a colluded copy of higher quality when more colluders have received the enhancement layers from the content owner. Furthermore, when the colluded copy has higher temporal resolution and better quality, the colluders have a larger probability to be captured and the collusion attack is less effective. This is due to the fact that the extracted fingerprint is longer, and it is in agreement with our statistical analysis in Section 4.3.1.

4.4 Resistance of the Scalable Fingerprinting Systems to Collusion Attacks

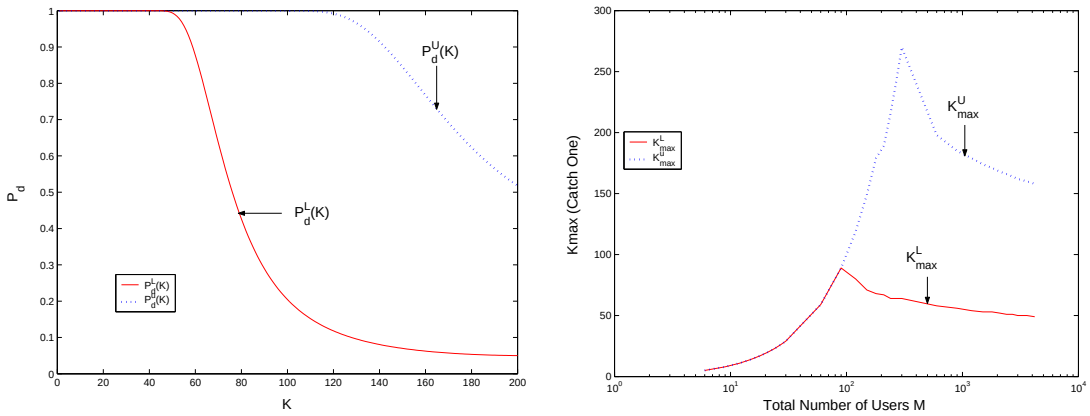
Analysis of the collusion attacks helps to evaluate the traitor tracing capacity of digital fingerprinting systems, and provide guidance to the digital rights enforcers on the design of collusion resistant fingerprinting systems. In this section, we consider the scalable fingerprinting systems in Section 4.1.2, analyze their collusion resistance, and quantify their traitor tracing capacity by studying how many colluders are required for colluders to cause the failure of the fingerprinting systems.

Following the work in [64], we consider three different fingerprinting applications that have different goals of design and different performance criteria: *catch one*, *catch more* and *catch all*; and we will study the collusion resistance of the scalable fingerprinting systems for these three scenarios. In particular, we analyze K_{max}, the maximum number of colluders that the fingerprinting systems can successfully resist under the system requirements.

4.4.1 Catch One

In the *catch one* applications, the goal is to maximize the chance to capture one colluder while minimizing the probability of falsely accusing an innocent users. In such applications, the performance criteria are the probability of capturing at least one colluder P_d and the probability of accusing at least one innocent user P_{fp}. The system requirements are

$$P_d \geq \gamma_d, \quad \text{and} \quad P_{fp} \leq \gamma_{fp}. \tag{4.39}$$



(a) (b)

Figure 4.6: The collusion resistance of the catch one applications. $|\mathbf{U}^b| : |\mathbf{U}^{b,e1}| : |\mathbf{U}^{all}| = 1 : 1 : 1$ and $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. $\sigma_n^2/\sigma_W^2 = 2$. $\gamma_d = 0.8$ and $\gamma_{fp} = 10^{-3}$. In (a), there are a total of 300 users in the system, and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 100$. 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.

Upper and Lower Bounds of K_{max}

From (4.34) and (4.35), if we fix the probability of accusing at least one innocent user $P_{fp} = \gamma_{fp}$, the performance of the detector in Section 4.1.2 depends on many parameters: $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{all}|)$, the lengths of the embedded fingerprints in each layer (N_b, N_{e1}, N_{e2}), the number of colluders in each subgroup $(K^b, K^{b,e1}, K^{all})$, and the length of the extracted fingerprint (or the temporal resolution of the colluded copy L^c, equivalently). Given the system parameters $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{all}|)$ and (N_b, N_{e1}, N_{e2}), for a fixed total number of colluders K, we define

$$
\begin{aligned}
& P_d^U(K) \stackrel{\triangle}{=} \max_{L^c, (K^b, K^{b,e1}, K^{all})} P_d, \\
s.t. \quad & K^b + K^{b,e1} + K^{all} = K, \\
& 0 \le K^b \le |\mathbf{U}^b|,\ 0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|,\ 0 \le K^{all} \le |\mathbf{U}^{all}|,
\end{aligned}
$$

fairness constraints in Table 4.1 are satisfied; (4.40)

$$\text{and} \quad P_d^L(K) \stackrel{\triangle}{=} \min_{L^c,(K^b,K^{b,e1},K^{all})} P_d,$$
$$s.t. \quad K^b + K^{b,e1} + K^{all} = K,$$
$$0 \le K^b \le |\mathbf{U}^b|,\ 0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|,\ 0 \le K^{all} \le |\mathbf{U}^{all}|,$$
fairness constraints in Table 4.1 are satisfied. (4.41)

$P_d^U(K)$ and $P_d^L(K)$ provide the upper and lower bounds of P_d, respectively, for a fixed total number of colluders K. Figure 4.6 (a) shows an example of $P_d^U(K)$ and $P_d^L(K)$ when there are a total of $M = 300$ users with $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 100$ and $\gamma_{fp} = 10^{-3}$.

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

$$K_{max}^U \stackrel{\triangle}{=} \arg_K\{P_d^U(K) \ge \gamma_d,\ P_d^U(K+1) < \gamma_d\}$$
$$\text{and} \quad K_{max}^L \stackrel{\triangle}{=} \arg_K\{P_d^L(K) \ge \gamma_d,\ P_d^L(K+1) < \gamma_d\}. \quad (4.42)$$

Figure 4.6 (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 $\gamma_d = 0.8$. From Figure 4.6 (b), the fingerprinting system can withstand collusion attacks with up to a few dozen colluders.

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

Calculation of the Upper and Lower Bounds of K_{max}

In this section, given $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{all}|)$ and (N_b, N_{e1}, N_{e2}), we analyze how to find K^U_{max} and K^L_{max}.

Given $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{all}|)$ and a fixed total number of colluders K, we define the feasible region of the triplet $(K^b, K^{b,e1}, K^{all})$ as

$$\begin{aligned}\mathbb{FR} \triangleq \big\{&(K^b, K^{b,e1}, K^{all}) : K^b + K^{b,e1} + K^{all} = K, 0 \le K^b \le |\mathbf{U}^b|, \\ &0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|, 0 \le K^{all} \le |\mathbf{U}^{all}|\big\}. \end{aligned} \tag{4.43}$$

To calculate the upper and lower bounds of K_{max}, we have to first calculate $P^U_d(K)$ and $P^L_d(K)$. From the analysis in Section 4.3.1, the detector has the worst performance when the colluded copy contains frames in the base layer only and the extracted fingerprint is of length N_b. Therefore, $P^L_d(K)$ is achieved when $F^c = F_b$. There are no constraints on $(K^b, K^{b,e1}, K^{all})$ and (N_b, N_{e1}, N_{e2}) except $(K^b, K^{b,e1}, K^{all}) \in \mathbb{FR}$.

From Section 4.3.1, for a given K, P_d is maximized when the colluded copy has the highest possible temporal resolution under the fairness constraints. Given (N_b, N_{e1}, N_{e2}) and the total number of colluders K, we define

$$\begin{aligned}\mathbb{RC}^3 \triangleq \Big\{&(K^b, K^{b,e1}, K^{all}) : \\ &\frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} \\ &\le \frac{N_b}{N_b + N_{e1} + N_{e2}}, \\ &\frac{K^{all}\sqrt{N_b + N_{e1} + N_{e2}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} \\ &\ge \frac{N_{e2}}{N_b + N_{e1} + N_{e2}}\Big\}, \end{aligned} \tag{4.44}$$

$$\text{and} \quad \begin{aligned}\mathbb{RC}^2 \triangleq \Big\{&(K^b, K^{b,e1}, K^{all}) : \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + (K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}} \\ &\le \frac{N_b}{N_b + N_{e1}}\Big\}. \end{aligned} \tag{4.45}$$

If $(K^b, K^{b,e1}, K^{all}) \in \mathbb{RC}^3$, the colluders can generate a colluded copy with the highest temporal resolution $F^c = F_b \cup F_{e1} \cup F_{e2}$; and for $(K^b, K^{b,e1}, K^{all}) \in \mathbb{RC}^2$, the colluders can generate a colluded copy with $F^c = F_b \cup F_{e1}$.

If $\mathbb{FR} \cap \mathbb{RC}^3 \neq \emptyset$, there exist at least one $(K^{b*}, K^{b,e1*}, K^{all*}) \in \mathbb{FR}$ such that the colluders can generate a colluded copy of the highest resolution $F^c = F_b \cup F_{e1} \cup F_{e2}$ under the fairness constraints, and

$$
\begin{aligned}
& P_d^U(K) = \max_{F^c = F_b \cup F_{e1} \cup F_{e2}, (K^b, K^{b,e1}, K^{all})} P_d, \qquad (4.46)\\
s.t. \quad & K^b + K^{b,e1} + K^{all} = K,\\
& 0 \leq K^b \leq |\mathbf{U}^b|,\ 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|,\ 0 \leq K^{all} \leq |\mathbf{U}^{all}|,\\
& \frac{K^b \sqrt{N_b}}{K^b \sqrt{N_b} + K^{b,e1} \sqrt{N_b + N_{e1}} + K^{all} \sqrt{N_b + N_{e1} + N_{e2}}} \leq \frac{N_b}{N_b + N_{e1} + N_{e2}},\\
& \frac{K^{all} \sqrt{N_b + N_{e1} + N_{e2}}}{K^b \sqrt{N_b} + K^{b,e1} \sqrt{N_b + N_{e1}} + K^{all} \sqrt{N_b + N_{e1} + N_{e2}}} \geq \frac{N_{e2}}{N_b + N_{e1} + N_{e2}}.
\end{aligned}
$$

From (4.35), maximize P_d when $F^c = F_b \cup F_{e1} \cup F_{e2}$ is equivalent to maximize the corresponding mean of the detection statistics

$$
\mu = \frac{N_b + N_{e1} + N_{e2}}{K^b \sqrt{N_b} + K^{b,e1} \sqrt{N_b + N_{e1}} + K^{all} \sqrt{N_b + N_{e1} + N_{e2}}}, \qquad (4.47)
$$

and it is further equivalent to minimize the denominator of μ. Consequently, the optimization problem of (4.47) can be simplified to

$$
\begin{aligned}
& \min_{(K^b, K^{b,e1}, K^{all})} K^b \sqrt{N_b} + K^{b,e1} \sqrt{N_b + N_{e1}} + K^{all} \sqrt{N_b + N_{e1} + N_{e2}}, \qquad (4.48)\\
s.t. \quad & K^b + K^{b,e1} + K^{all} = K,\\
& 0 \leq K^b \leq |\mathbf{U}^b|,\ 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|,\ 0 \leq K^{all} \leq |\mathbf{U}^{all}|,\\
& \frac{K^b \sqrt{N_b}}{K^b \sqrt{N_b} + K^{b,e1} \sqrt{N_b + N_{e1}} + K^{all} \sqrt{N_b + N_{e1} + N_{e2}}} \leq \frac{N_b}{N_b + N_{e1} + N_{e2}},\\
& \frac{K^{all} \sqrt{N_b + N_{e1} + N_{e2}}}{K^b \sqrt{N_b} + K^{b,e1} \sqrt{N_b + N_{e1}} + K^{all} \sqrt{N_b + N_{e1} + N_{e2}}} \geq \frac{N_{e2}}{N_b + N_{e1} + N_{e2}}.
\end{aligned}
$$

We use linear programming [17] to solve the optimization problem of (4.49), and then calculate the corresponding μ and $P_d^U(K)$.

If $\mathbb{FR} \cap \mathbb{RC}^3 = \emptyset$ and $\mathbb{FR} \cap \mathbb{RC}^2 \neq \emptyset$, the colluders cannot generate a colluded copy with $F^c = F_b \cup F_{e1} \cup F_{e2}$ under the fairness constraints, but they can generate a copy with $F^c = F_b \cup F_{e1}$. In this scenario, $P_d^U(K) = \max\{P_d^{U,1}(K), P_d^{U,2}(K)\}$ where

$$
\begin{aligned}
& P_d^{U,1}(K) = \max_{F^c = F_b \cup F_{e1}, (K^b, K^{b,e1}, K^{all})} P_d, && (4.49)\\
s.t. \quad & K^b + K^{b,e1} + K^{all} = K, \\
& 0 \leq K^b \leq |\mathbf{U}^b|,\ 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|,\ 0 \leq K^{all} \leq |\mathbf{U}^{all}|, \\
& \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + (K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}} \leq \frac{N_b}{N_b + N_{e1}}, \\
& \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} \geq \frac{N_b}{N_b + N_{e1} + N_{e2}}, \\
\text{and} \quad & P_d^{U,2}(K) = \max_{F^c = F_b \cup F_{e1}, (K^b, K^{b,e1}, K^{all})} P_d, && (4.50)\\
s.t. \quad & K^b + K^{b,e1} + K^{all} = K, \\
& 0 \leq K^b \leq |\mathbf{U}^b|,\ 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|,\ 0 \leq K^{all} \leq |\mathbf{U}^{all}|, \\
& \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + (K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}} \leq \frac{N_b}{N_b + N_{e1}}, \\
& \frac{K^{all}\sqrt{N_b + N_{e1} + N_{e2}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} \leq \frac{N_{e2}}{N_b + N_{e1} + N_{e2}}.
\end{aligned}
$$

The optimization problem in (4.49) is equivalent to

$$
\begin{aligned}
& \min_{(K^b, K^{b,e1}, K^{all})} K^b\sqrt{N_b} + (K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}, && (4.51)\\
s.t. \quad & K^b + K^{b,e1} + K^{all} = K, \\
& 0 \leq K^b \leq |\mathbf{U}^b|,\ 0 \leq K^{b,e1} \leq |\mathbf{U}^{b,e1}|,\ 0 \leq K^{all} \leq |\mathbf{U}^{all}|, \\
& \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + (K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}} \leq \frac{N_b}{N_b + N_{e1}}, \\
& \frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} \geq \frac{N_b}{N_b + N_{e1} + N_{e2}},
\end{aligned}
$$

and the optimization problem in (4.50) is equivalent to

$$
\begin{aligned}
& \min_{(K^b, K^{b,e1}, K^{all})} K^b\sqrt{N_b} + (K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}, && (4.52)\\
s.t. \quad & K^b + K^{b,e1} + K^{all} = K,
\end{aligned}
$$

$$0 \le K^b \le |\mathbf{U}^b|,\ 0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|,\ 0 \le K^{all} \le |\mathbf{U}^{all}|,$$
$$\frac{K^b\sqrt{N_b}}{K^b\sqrt{N_b} + (K^{b,e1} + K^{all})\sqrt{N_b + N_{e1}}} \le \frac{N_b}{N_b + N_{e1}},$$
$$\frac{K^{all}\sqrt{N_b + N_{e1} + N_{e2}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} \le \frac{N_{e2}}{N_b + N_{e1} + N_{e2}}.$$

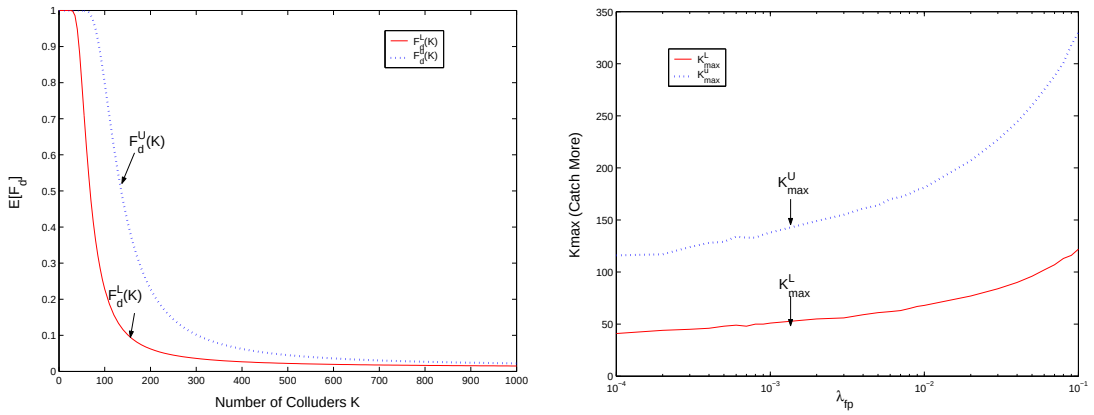
If $\mathbb{FR} \cap \mathbb{RC}^3 = \emptyset$ and $\mathbb{FR} \cap \mathbb{RC}^2 = \emptyset$, then the colluders can only generate a colluded copy of the lowest resolution where $F^c = F_b$, and $P_d^U(K) = P_d^L(K)$ in this scenario.

Given $P_d^U(K)$ and $P_d^L(K)$, the analysis of K_{max}^U and K_{max}^L is the same as in [64] and is omitted here.

Physical Meanings of K_{max}^U and K_{max}^L

From the colluders' point of view, if colluders can collect no more than K_{max}^L independent copies, no matter how they collude, they can never succeed in passing the detector without being captured. 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 and best quality. In the scenario where the colluders collect more than K_{max}^L but fewer than K_{max}^U copies, they can still successfully remove all trace of the fingerprints by generating a colluded copy of the lowest resolution and worst quality. If the colluders wish to generate a colluded copy of better quality, they must take the risk of being captured.

From the content owner's point of view, if he can ensure that potential colluders cannot collect more than K_{max}^L independent copies, the fingerprinting system is essentially collusion resistant.



(a) (b)

Figure 4.7: The collusion resistance of the catch more applications. $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 300$ and $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. $\sigma_n^2/\sigma_W^2 = 2$. In (a), $\lambda_{fp} = 0.01$, and 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}.

4.4.2 Catch More

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

$$E[F_d] \geq \lambda_d, \quad \text{and} \quad E[F_{fp}] \leq \lambda_{fp}. \tag{4.53}$$

Similar to the catch one applications, given $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{all}|)$ and (N_b, N_{e1}, N_{e2}), for fixed $E[F_{fp}] = \lambda_{fp}$ and a fixed total number of colluders K, we define

$$\begin{aligned} & F_d^U(K) \stackrel{\triangle}{=} \max_{L^c, (K^b, K^{b,e1}, K^{all})} E[F_d], \\ s.t. \quad & K^b + K^{b,e1} + K^{all} = K, \end{aligned}$$

$$0 \le K^b \le |\mathbf{U}^b|,\ 0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|,\ 0 \le K^{all} \le |\mathbf{U}^{all}|,$$
fairness constraints in Table 4.1 are satisfied; (4.54)

$$\text{and} \quad F_d^L(K) \stackrel{\triangle}{=} \min_{L^c,(K^b,K^{b,e1},K^{all})} E[F_d],$$
$$s.t. \quad K^b + K^{b,e1} + K^{all} = K,$$
$$0 \le K^b \le |\mathbf{U}^b|,\ 0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|,\ 0 \le K^{all} \le |\mathbf{U}^{all}|,$$
fairness constraints in Table 4.1 are satisfied. (4.55)

Given the total number of colluders K, $F_d^U(K)$ and $F_d^L(K)$ are the upper and lower bounds of $E[F_d]$,respectively. Figure 4.7 (a) shows an example of $F_d^U(K)$ and $F_d^L(K)$ when $\lambda_{fp} = 0.01$.

For the catch more applications, we define

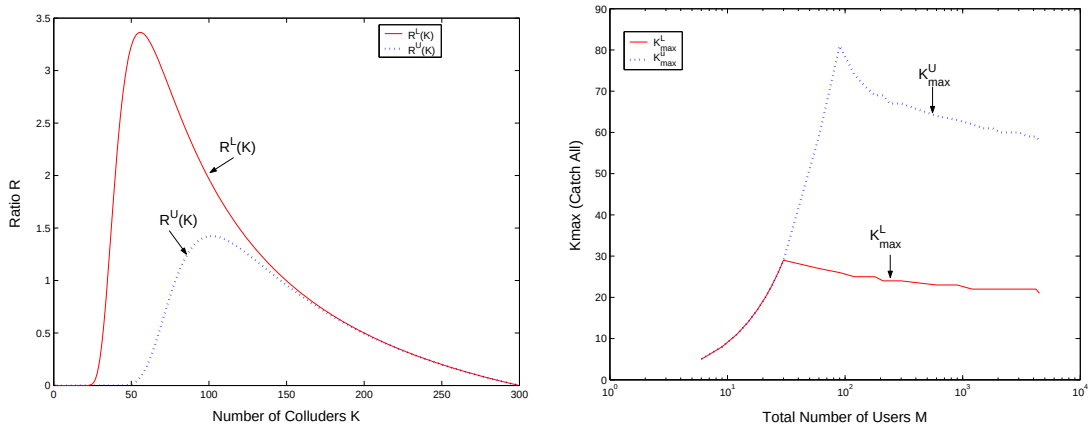
$$\begin{aligned} K_{max}^U &\stackrel{\triangle}{=} \arg_K\{F_d^U(K) \ge \lambda_d,\ F_d^U(K+1) < \lambda_d\} \\ \text{and} \quad K_{max}^L &\stackrel{\triangle}{=} \arg_K\{F_d^L(K) \ge \lambda_d,\ F_d^L(K+1) < \lambda_d\}, \end{aligned} \tag{4.56}$$

which are the upper and lower bounds of K^{max} under the system requirements, respectively. Figure 4.7 (b) shows K_{max}^U and K_{max}^L under different system requirements of λ_{fp} when $\lambda_d = 0.5$. From Figure 4.7 (b), for the catch more applications, the fingerprinting system can resist collusion attacks with a few dozen or even around one hundred colluders, depending on the system requirements.

Given λ_d and λ_{fp}, the analysis of $(F_d^U(K), F_d^L(K))$ and (K_{max}^U, K_{max}^U) in the catch more applications is similar to that in the catch one applications and will be not be repeated.

4.4.3 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



(a) (b)

Figure 4.8: The collusion resistance of the catch all applications. $|\mathbf{U}^b| : |\mathbf{U}^{b,e1}| : |\mathbf{U}^{all}| = 1 : 1 : 1$ and $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. $\sigma_n^2/\sigma_W^2 = 2$. $\theta_d = 0.99$ and $\theta_r = 0.01$. In (a), $M = 300$ and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 100$. We plot R^U and R^L versus the total number of colluders. (b) shows K^U_{max} and K^L_{max} versus the total number of users M.

falsely accused. The set of performance criteria for these applications consists of measuring the efficiency rate

$$R = \frac{(M-K)\cdot E[F_{fp}]}{K \cdot E[F_d]} \tag{4.57}$$

that describes the number of innocents accused per colluder, and the probability of capturing all colluders

$$P_{d,all} = P\left[\min_{i\in SC} T_N^{(i)} > h\right]. \tag{4.58}$$

The system requirements for these applications are

$$R \leq \theta_r, \quad \text{and} \quad P_{d,all} \geq \theta_d. \tag{4.59}$$

Similar to the catch one applications, given $(|\mathbf{U}^b|, |\mathbf{U}^{b,e1}|, |\mathbf{U}^{all}|)$ and (N_b, N_{e1}, N_{e2}),

for a fixed total number of colluders K and fixed $P_{d,all} = \theta_d$, define

$$\begin{aligned} & R^U(K) \stackrel{\triangle}{=} \max_{L^c,(K^b,K^{b,e1},K^{all})} R, \\ s.t. \quad & K^b + K^{b,e1} + K^{all} = K, \\ & 0 \le K^b \le |\mathbf{U}^b|,\ 0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|,\ 0 \le K^{all} \le |\mathbf{U}^{all}|, \\ & \text{fairness constraints in Table 4.1 are satisfied;} \end{aligned} \tag{4.60}$$

$$\begin{aligned} \text{and} \quad & R^L(K) \stackrel{\triangle}{=} \min_{L^c,(K^b,K^{b,e1},K^{all})} R, \\ s.t. \quad & K^b + K^{b,e1} + K^{all} = K, \\ & 0 \le K^b \le |\mathbf{U}^b|,\ 0 \le K^{b,e1} \le |\mathbf{U}^{b,e1}|,\ 0 \le K^{all} \le |\mathbf{U}^{all}|, \\ & \text{fairness constraints in Table 4.1 are satisfied.} \end{aligned} \tag{4.61}$$

Given the total number of colluders K, $R^U(K)$ and $R^L(K)$ are the upper and lower bounds of R, respectively. Figure 4.8 (a) shows an example of $R^U(K)$ and $R^L(K)$ when there are a total of $M = 300$ users with $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 100$ and $\theta_d = 0.99$.

In the catch all applications, define

$$\begin{aligned} K^U_{max} &\stackrel{\triangle}{=} \arg_K\{R^U(K) \le \theta_r,\ R^U(K+1) > \theta_r\} \\ \text{and} \quad K^L_{max} &\stackrel{\triangle}{=} \arg_K\{R^L(K) \le \theta_r,\ R^L(K+1) > \theta_r\}, \end{aligned} \tag{4.62}$$

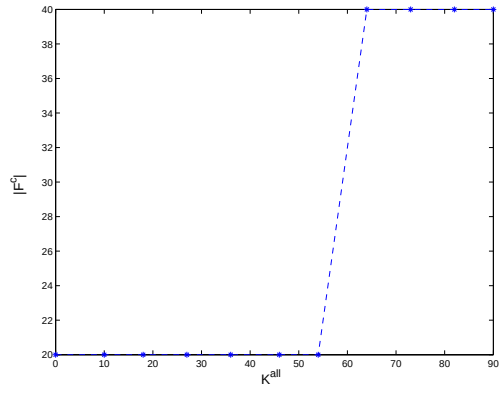
which are the upper and lower bounds of K_{max}. Figure 4.8 (b) shows K^U_{max} and K^L_{max} as functions of the total number of users M under the system requirements of $\theta_d = 0.99$ and $\theta_r = 0.01$. For the catch all applications, from Figure 4.8 (b), the fingerprinting systems are robust against collusion attacks with a few dozen colluders.

Given θ_r and θ_d, the analysis of $(R^U(K), R^L(K))$ and (K^U_{max}, K^U_{max}) in the catch all applications is similar to that in the catch one applications and will be not be repeated.

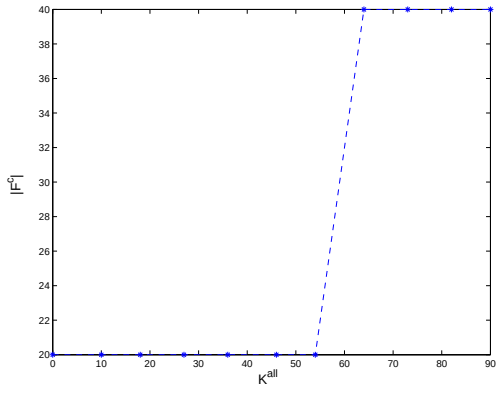
4.5 Simulation Results on Real Video Sequences

To verify the correctness of our analysis on real videos, we choose a typical video sequence "carphone" and use the first 40 frames as an example. Similar to the simulation setup in Section 4.3.2, we choose $F_b = \{j : j = 4k+1, k = 0, \cdots, 9\}$, $F_{e1} = \{j : j = 4k+3, k = 0, \cdots, 9\}$ and $F_{e2} = \{j : j = 2k, k = 1, \cdots, 20\}$ as an example of the temporal scalability, and the lengths of the embedded fingerprints 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 $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. We adopt the human visual model based spread spectrum embedding in [47], and embed the fingerprints in the DCT domain. The fingerprints follow Gaussian distribution $\mathcal{N}(0, 1/9)$, and the fingerprints assigned to different users are generated independently. In one fingerprinted copy, similar to that in [55], the fingerprints embedded in different frames are correlated with each other, depending on the similarity between the host frames.

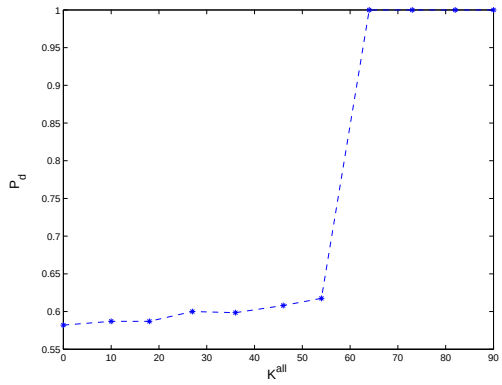
During collusion, we assume that there are a fixed total number of $K = 150$ colluders and the collusion attack is also in the DCT domain. In our simulations, we assume that the colluders use the approximation $\hat{N}_b : \hat{N}_{e1} : \hat{N}_{e2} \approx |F_b| : |F_{e1}| : |F_{e2}| = 1 : 1 : 2$, and apply the intra-group collusion attacks followed by the inter-group attacks as in Section 4.1.2. They further introduce an additive noise $\mathbf{n}_j$ to each frame j in the colluded copy. To be consistent with the simulation setup in Section 4.3.2, we adjust the power of the additive noise such that $\frac{||\mathbf{n}_j/JND_j||^2}{||\mathbf{W}_j^{(i)}||^2} = 2$ for every frame $j \in F^c$ in the colluded copy. JND_j here is the jth frame's just-noticeable-difference from human visual models [47]. In our simulations, we assume that the colluders generate a colluded copy of the best possible quality under the fairness constraints, the same as in Section 4.3.2.



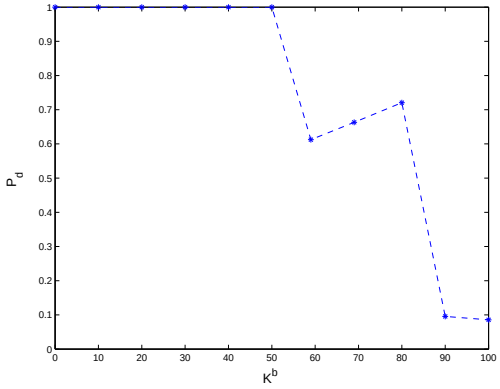
(a) $L^c = |F^c|$ of line (4.37)



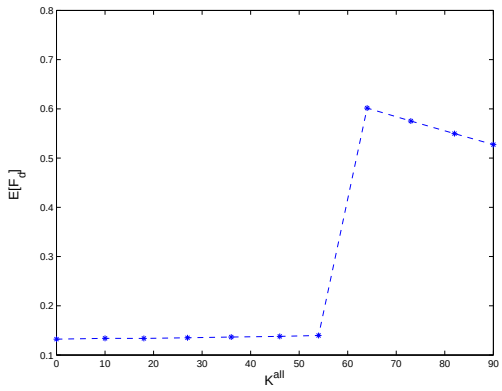
(b) $L^c = |F^c|$ of line (4.38)



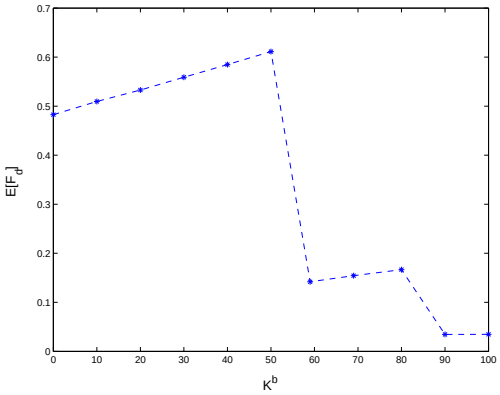
(c) P_d of line (4.37)



(d) P_d of line (4.38)



(e) $E[F_d]$ of line (4.37)



(f) $E[F_d]$ of line (4.38)

Figure 4.9: Simulation results of the collusion attacks on the first 40 frames of "carphone". $(|F_b|, |F_{e1}|, |F_{e2}|) = (10, 10, 20)$. $M = 450$, $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$ and $K = 150$. In (a), (c) and (e), $(K^b, K^{b,e1}, K^{all})$ are on Line (4.37), and in (b), (d) and (f), $(K^b, K^{b,e1}, K^{all})$ are on Line (4.38). $P_{fp} = 10^{-3}$ in (c) and (d), and $E[F_{fp}] = 10^{-3}$ in (e) and (f).

At the detector's side, we consider a non-blind detection scenario where the host signal is removed from the colluded copy before fingerprint detection and colluder identification process. The detector follows the detection process in Section 4.1.2 and estimates the indices of the colluders $\widehat{SC}$.

Figure 4.9 shows the simulation results. In Figure 4.9 (a), (c) and (e), $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and they are on Line (4.37); and in Figure 4.9 (b), (d) and (f), $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and they are on Line (4.38). In Figure 4.9 (c) and (d), we fix $P_{fp} = 10^{-3}$ and compare P_d when $(K^b, K^{b,e1}, K^{all})$ changes. In Figure 4.9 (e) and (f), $E[F_{fp}]$ is fixed as 10^{-3}, and we compare $E[F_d]$ of the collusion attacks with different $(K^b, K^{b,e1}, K^{all})$. From Figure 4.9, the effectiveness of the collusion attacks depends on the perceptual quality of the colluded copy: if the colluded copy has higher resolution and better quality, the extracted fingerprint is longer, and therefore, the colluders have larger probability to be captured. Also, the simulation results on real video sequences are comparable with that in Section 4.3.2.

4.6 Chapter Summary

In this chapter, we have studied the performance of scalable fingerprinting systems where different users received fingerprinted copies of different quality. We have analyzed the fairness constraints on the collusion attacks and provided statistical analysis on the effectiveness of the collusion attacks. We have also investigated the collusion resistance of the scalable fingerprinting systems and studied the maximum number of colluders that the fingerprinting systems can withstand.

We first studied the fairness constraints on the collusion attacks when colluders received fingerprinted copies of different quality. We found that higher resolution

and better quality of the colluded copy puts more severe constraints on the number of colluders in each subgroup. We then analyzed the effectiveness of the collusion attacks. Both our analytical and simulation results have shown that the colluders are more likely to be captured when the colluded copy has higher resolution and better quality. The colluders have to take into consideration the tradeoff between the probability of detection and the perceptual quality of the colluded copy during collusion.

We also studied the collusion resistance of the scalable fingerprinting systems for three different applications with different system requirements, and provided the lower and upper bounds of the maximum number of colluder that the fingerprinting systems can resist. From the colluders' point of view, the upper bound tells the colluders how many independent copies are required to guarantee the success even if the colluded copy has the highest quality. 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 copies more than the lower bound.

Chapter 5

Traitors within Traitors: Strategy and Performance Analysis

Most prior work assumed that during collusion, all colluders would like to share the risk, and they adjust the collusion attack to guarantee that all of them have the same probability of detection. However, there exist selfish colluders who want to minimize their own probability of detection while still profiting from collusion. In order to achieve this goal, they hide from other colluders information of the fingerprinted copies that they received, and process their fingerprinted copies before multiuser collusion.

In this chapter, we investigate this "traitors within traitors" problem in multimedia fingerprinting. We examine the possible pre-collusion processing techniques by the selfish colluders, analyze their effectiveness, and find the optimal pre-collusion processing strategy to minimize their probability of detection under the quality constraints. We also investigate the possible countermeasures by other colluders to protect their own interest and prevent the selfish colluders from processing the copies before collusion.

This chapter is organized as follows. We begin, in Section 5.1, with the system model of traitors within traitors. In Section 5.2, we consider a simple scenario where all colluders receive copies of the same quality, and investigate the possible pre-collusion processing strategy by the selfish colluders. In Section 5.3, we study the pre-collusion processing technique in scalable fingerprinting systems, where different users receive copies of different resolution. Section 5.4 investigates the preliminary countermeasures against pre-collusion processing by other colluders.

5.1 System Model

5.1.1 General Framework of Digital Fingerprinting Systems for Multimedia Forensics

We consider a digital fingerprinting system that consists of three parts: fingerprint embedding, multiuser collusion attacks, and fingerprint detection and colluder identification.

Fingerprint Embedding

Spread spectrum embedding has been widely used in multimedia fingerprinting systems due to its robustness against many attacks [13, 47]. In additive spread spectrum embedding for video applications, for the jth frame in the video sequence represented by a vector $\mathbf{S}_j$ of length N_j, the content owner generates a unique fingerprint $\mathbf{W}_j^{(i)}$ of length N_j for each user $\mathbf{u}^{(i)}$ in the system. The fingerprinted copy that is 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 [47], and it is used to control the energy and achieve the imperceptibility of the embedded fingerprints. Finally, the content owner transmits to each user $\mathbf{u}^{(i)}$ the fingerprinted frames $\left\{\mathbf{X}_j^{(i)}\right\}$.

Previous work has shown that Gaussian distributed fingerprints are more robust against nonlinear collusion attacks [52] and are resilient to the statistical/histogram attacks [15]. Therefore, in this chapter, we consider Gaussian fingerprints and assume that $\left\{\mathbf{W}_j^{(i)}\right\}$ follow normal distribution with zero mean and variance σ_W^2. Furthermore, we apply orthogonal fingerprint modulation [58] and generate the fingerprints for different users independently. In this chapter, to be resistant to intra-content collusion attacks on video watermarking [55,56], in each fingerprinted copy $\{\mathbf{X}_j^{(i)}\}$, the fingerprints $\mathbf{W}_{j_1}^{(i)}$ and $\mathbf{W}_{j_2}^{(i)}$ that are embedded in adjacent frames $\mathbf{S}_{j_1}$ and $\mathbf{S}_{j_2}$, respectively, are correlated with each other. The correlation between $\mathbf{W}_{j_1}^{(i)}$ and $\mathbf{W}_{j_2}^{(i)}$ depends on the similarity between the two host frames $\mathbf{S}_{j_1}$ and $\mathbf{S}_{j_2}$. This is similar to the work in [55].

Multiuser Collusion Attacks

Assume that there are a total of K colluders and SC is the set containing their indices. The colluders first collect a total of K copies of the same content but embedded with different fingerprints, and then apply a multiuser collusion attack to reduce the energy of each of the original fingerprints. In a recently investigation, it has been shown that a nonlinear collusion attack can be modeled as the averaging collusion attack followed by an additive noise [64]. Under the constraints that the colluded copies under different collusion attacks have the same perceptual quality, all collusion attacks have approximately the same performance. Therefore, in this chapter, we consider the averaging based collusion attacks for the simplicity of

analysis.

Fingerprint Detection and Colluder Identification

Once the content owner discovers the existence of the illegal copy in the market, he applies a fingerprint detection and colluder identification process on the suspicious copy. For each frame $\mathbf{V}_j$ in the colluded copy, the detector first extracts the fingerprint $\mathbf{Y}_j$ from $\mathbf{V}_j$. Then, he calculates the similarity between the extracted fingerprint $\{\mathbf{Y}_j\}$ and each of the original fingerprints $\{\mathbf{W}_j^{(i)}\}$, compares with a pre-determined threshold h, and outputs the identities of the estimated colluders $\widehat{SC}$.

To measure the similarity between the extracted fingerprint and the original fingerprint, for each user $\mathbf{u}^{(i)}$, the detector calculates the correlation based detection statistics

$$T_N^{(i)} = \sum_j \langle \mathbf{Y}_j, \mathbf{W}_j^{(i)} \rangle / \sqrt{\sum_j ||\mathbf{W}_j^{(i)}||^2}, \tag{5.1}$$

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

5.1.2 Traitors within Traitors

Before collusion, the colluders have to exchange information of the received copies with each other. The correctness of this information is critical to guarantee the fairness of the collusion and ensure that all colluders have equal probability of detection. In most prior work on multimedia fingerprinting and collusion attacks, it was assumed that all colluders would like to share the risk and have the same probability of detection, and they tell each other the correct information of their received fingerprinted copies. In practice, there are selfish colluders who want to

minimize their own risk of being captured, while they still wish to participate in the collusion in order to profit from the unauthorized redistribution. To achieve this goal, the selfish colluders hide from other colluders information of the fingerprinted copies that they received from the content owner.

Assume that $\mathbf{X}^{(i)}$ is the fingerprinted copy that is received by colluder $\mathbf{u}^{(i)}$. In the scenario where all colluders are willing to share the risk and have equal probability of detection, the multiuser collusion attack is applied to $\{\mathbf{X}^{(k)}\}_{i \in SC}$, and the colluded copy is generated as

$$\mathbf{V} = g\left(\{\mathbf{X}^{(i)}\}_{i \in SC}\right) + \mathbf{n}, \tag{5.2}$$

where $g(\cdot)$ is the collusion function and $\mathbf{n}$ is an additive noise introduced by the colluders to further hinder the detection. Figure 5.1 (a) shows an example of the collusion attack in this scenario.

When there are selfish colluders who wish to minimize their risk, those selfish colluders process their fingerprinted copies before multiuser collusion. During collusion, if other colluders do not discover this pre-collusion processing behavior, and if they use these processed copies instead of the originally received copies, the pre-collusion processing could help the selfish colluders to further reduce their own risk of being captured.

Shown in Figure 5.1 (b) is an example of this scenario. Without loss of generality, assume that colluder $\mathbf{u}^{(i_1)}$ is the selfish colluder who wants to minimize his own risk, and $\mathbf{X}^{(i_1)}$ is the fingerprinted copy that he received from the content owner. Based on $\mathbf{X}^{(i_1)}$, $\mathbf{u}^{(i_1)}$ generates another 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 the other colluders fail to discover $\mathbf{u}^{(i_1)}$'s pre-collusion processing behavior, the colluded copy equals to

$$\mathbf{V}' = g\left(\widetilde{\mathbf{X}}^{(i_1)}, \{\mathbf{X}^{(i)}\}_{i \in SC, i \neq i_1}\right) + \mathbf{n}, \tag{5.3}$$

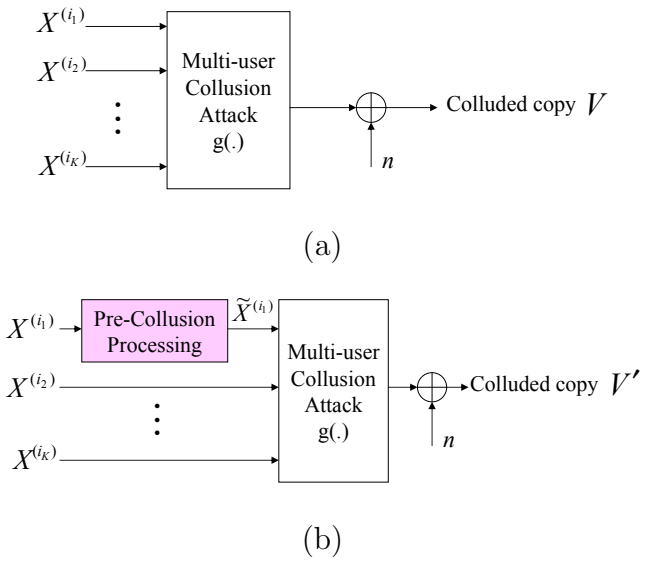


Figure 5.1: (a) The collusion attack when all colluders are willing to share the same risk of being captured. (b) The collusion attack when some selfish colluders want to further reduce their own probability of detection.

where $g(\cdot)$ is the collusion function and $\mathbf{n}$ is an additive noise.

5.1.3 Performance Criteria

To measure the effectiveness of pre-collusion processing in reducing the selfish colluders' probability of detection, we use the following criteria:

- $P_d^{(i)}$: the probability that a colluder $\mathbf{u}^{(i)}$ is successfully captured; and
- P_{fa}: the probability that an innocent user is falsely accused.

For a fixed P_{fa}, we compare a selfish colluder's probability of detection in two scenarios: when the selfish colluder does not apply pre-collusion processing (i.e., he is willing to share the risk with other colluders), and when the selfish colluder processes his fingerprinted copy before collusion. From the selfish colluder's point

of view, the pre-collusion processing technique is more effective when the difference between these two probabilities is larger.

In the example shown in Figure 5.1 (b), in order to cover up the fact that he processes his fingerprinted copy before multiuser collusion, the selfish colluder $\mathbf{u}^{(i_1)}$ has to guarantee that the newly generated copy $\widetilde{\mathbf{X}}^{(i_1)}$ has high quality. We use the mean square error (MSE) between $\widetilde{\mathbf{X}}^{(i_1)}$ and $\mathbf{X}^{(i_1)}$, or equivalently the PSNR for image and video applications, to measure the effect of pre-collusion processing on the perceptual quality of the fingerprinted copies.

5.2 Energy Attenuation of the Embedded Fingerprints During Pre-collusion Processing

For a selfish colluder, to further reduce his own probability of detection, one possible solution is to apply pre-collusion processing to attenuate the energy of the embedded fingerprint. An example is to replace each segment of the fingerprinted signal with another, seemingly similar segment from different spatial or temporal regions of the content, e.g., averaging or swapping consecutive frames of similar content [56].

In this section, we take frame averaging as an example, and analyze its effects on the probability of detection as well as the perceptual quality of the fingerprinted copies. We consider a simple scenario where all users in the system receive fingerprinted copies of the same quality. Our analysis can be extended to scalable fingerprinting systems, where different users receive copies of different quality.

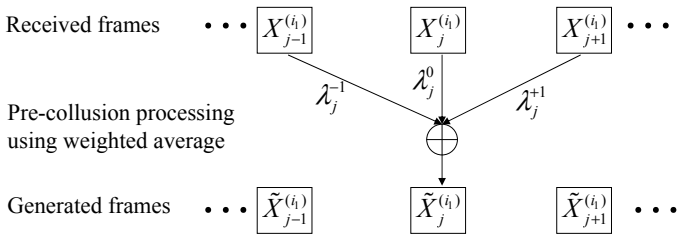


Figure 5.2: Applying weighted average during pre-collusion processing.

5.2.1 Pre-collusion Processing Using Weighted Average

In this chapter, we assume that the selfish colluder uses a simple linear interpolation based average during pre-collusion processing.[1] For a selfish colluder $\mathbf{u}^{(i_1)}$, assume that $\{\mathbf{X}_j^{(i_1)}\}_{j=1,2,\cdots}$ are the fingerprinted frames that he received from the content owner, and $\mathbf{X}_{j-1}^{(i_1)}$, $\mathbf{X}_j^{(i_1)}$ and $\mathbf{X}_{j+1}^{(i_1)}$ are three consecutive frames. As shown in Figure 5.2, for each frame j, $\mathbf{u}^{(i_1)}$ applies weighted average to the three adjacent frames, $\mathbf{X}_{j-1}^{(i_1)}$, $\mathbf{X}_j^{(i_1)}$ and $\mathbf{X}_{j+1}^{(i_1)}$, and generates a new frame $\widetilde{\mathbf{X}}_j^{(i_1)}$ by

$$\widetilde{\mathbf{X}}_j^{(i_1)} = \lambda_j^{-1} \cdot \mathbf{X}_{j-1}^{(i_1)} + \lambda_j^{0} \cdot \mathbf{X}_{j-1}^{(i_1)} + \lambda_j^{+1} \cdot \mathbf{X}_{j+1}^{(i_1)}, \tag{5.4}$$

where $0 \leq \lambda_j^{-1}, \lambda_j^{0}, \lambda_j^{+1} \leq 1$ and $\lambda_j^{-1} + \lambda_j^{0} + \lambda_j^{+1} = 1$. For simplicity, we let $\lambda_j^{-1} = \lambda_j^{+1} = (1 - \lambda_j^{0})/2$, and give equal weights to the neighboring frames $\mathbf{X}_{j-1}^{(i_1)}$ and $\mathbf{X}_{j+1}^{(i_1)}$. The selfish colluder $\mathbf{u}^{(i_1)}$ repeats this process for every frame in the video sequence and generates $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}_{j=1,2,\cdots}$.

For simplicity, we assume that there is only one selfish colluder $\mathbf{u}^{(i_1)}$ in this section.[2] If the other colluders do not discover $\mathbf{u}^{(i_1)}$'s pre-collusion processing

[1] A selfish colluder can also apply more complicated motion based interpolation [2,8], and the analysis will be similar.

[2] When there are multiple selfish colluders using weighted average during pre-collusion processing, the analysis is similar and not repeated here.

actions, then the jth frame in the colluded copy equals to

$$\mathbf{V}'_j = \frac{\sum_{i \in SC, i \neq i_1} \mathbf{X}_j^{(i)}}{K} + \frac{\lambda_j^{-1} \cdot \mathbf{X}_{j-1}^{(i_1)} + \lambda_j^0 \cdot \mathbf{X}_{j-1}^{(i_1)} + \lambda_j^{+1} \cdot \mathbf{X}_{j+1}^{(i_1)}}{K} + \mathbf{n}_j, \tag{5.5}$$

where $\mathbf{n}_j$ is an additive noise.

5.2.2 Performance Analysis and Selection of the Optimal Weight Vector

By processing the fingerprinted frames before collusion, the selfish colluder wishes to minimize his own probability of detection while maintaining the perceptual quality of $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$. In this section, we will first analyze the quality of the newly generated frames $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$ and the selfish colluder's probability of detection, and then study the selection of the optimal weight vector $[\lambda_1^0, \lambda_2^0, \cdots]$.

Analysis of Perceptual Quality

If $\widetilde{\mathbf{X}}_j^{(i_1)}$ is generated as in (5.4), then the MSE between $\widetilde{\mathbf{X}}_j^{(i_1)}$ and $\mathbf{X}_j^{(i_1)}$ is

$$\begin{aligned} MSE_j &= ||\widetilde{\mathbf{X}}_j^{(i_1)} - \mathbf{X}_j^{(i_1)}||^2 = \left(\frac{1-\lambda_j^0}{2}\right)^2 \cdot \phi_j, \\ \text{where} \quad \phi_j &= 4||\mathbf{X}_j^{(i_1)}||^2 + ||\mathbf{X}_{j-1}^{(i_1)}||^2 + ||\mathbf{X}_{j+1}^{(i_1)}||^2 \\ &\quad -4\langle \mathbf{X}_{j-1}^{(i_1)}, \mathbf{X}_j^{(i_1)} \rangle - 4\langle \mathbf{X}_j^{(i_1)}, \mathbf{X}_{j+1}^{(i_1)} \rangle + 2\langle \mathbf{X}_{j-1}^{(i_1)}, \mathbf{X}_{j+1}^{(i_1)} \rangle. \end{aligned} \tag{5.6}$$

In (5.6), $||\mathbf{X}_j^{(i_1)}||$ is the Euclidean norm of $\mathbf{X}_j^{(i_1)}$, and $\langle \mathbf{X}_{j-1}^{(i_1)}, \mathbf{X}_j^{(i_1)} \rangle$ is the correlation between $\mathbf{X}_{j-1}^{(i_1)}$ and $\mathbf{X}_j^{(i_1)}$. From (5.6), a larger λ_j^0 implies a smaller MSE_j. Consequently, from the perceptual quality's point of view, $\mathbf{u}^{(i_1)}$ should choose a larger λ_j^0. When $\lambda_j^0 = 1$ and colluder $\mathbf{u}^{(i_1)}$ does not apply pre-collusion processing, $\widetilde{\mathbf{X}}_j^{(i_1)} = \mathbf{X}_j^{(i_1)}$ and it has the best possible quality.

Analysis of Probability of Detection

Given the colluded copy as in (5.5), the fingerprint that is extracted from frame j in the colluded copy $\mathbf{V}'_j$ is

$$\mathbf{Y}_j = \frac{\sum_{i\in SC, i\neq i_1} \mathbf{W}_j^{(i)}}{K} + \frac{\lambda_j^{-1}\cdot \mathbf{W}_{j-1}^{(i)} + \lambda_j^{0}\cdot \mathbf{W}_{j-1}^{(i)} + \lambda_j^{+1}\cdot \mathbf{W}_{j+1}^{(i)}}{K} + \mathbf{d}_j, \qquad (5.7)$$

where $\mathbf{d}_j$ contains terms that are independent of the embedded fingerprints $\{\mathbf{W}_j^{(i)}\}$. For simplicity, we assume that $\mathbf{d}_j$ are i.i.d. and follow Gaussian distribution $\mathcal{N}(0,\ \sigma_n^2)$.

It is straightforward to show that given the colluder set SC and the index of the selfish colluder i_1, the detection statistics follow Gaussian distribution with mean $\mu^{(i)}$ and variance σ_n^2, i.e.,

$$p\left(T_N^{(i)}|SC, i_1\right) \sim \mathcal{N}\left(\mu^{(i)}, \sigma_n^2\right). \qquad (5.8)$$

The detection statistics have a zero mean for an innocent user and a positive mean for a guilty colluder. Consequently, the probability of accusing an innocent user and the probability of capturing a guilty colluder $\mathbf{u}^{(i\in SC)}$ are

$$P_{fa} \approx Q\left(\frac{h}{\sigma_n}\right) \quad \text{and} \quad P_d^{(i)} \approx Q\left(\frac{h-\mu^{(i)}}{\sigma_n}\right), \qquad (5.9)$$

respectively, where $Q(\cdot)$ is the Gaussian tail function. Therefore, for fixed σ_n^2 and P_{fa}, a colluder $\mathbf{u}^{(i)}$ has a smaller probability of detection when $\mu^{(i)}$ is smaller, and minimizing the probability of detection is equivalent to minimizing the mean of the detection statistics.

For the selfish colluder $\mathbf{u}^{(i_1)}$,

$$\mu^{(i_1)} \ = \ \sum_j \mu_j^{(i_1)},$$

$$\text{where} \quad \mu_j^{(i_1)} = \frac{1-\lambda_j^0}{2} \cdot \frac{\langle \mathbf{W}_{j-1}^{(i_1)}, \mathbf{W}_j^{(i_1)} \rangle}{K\sqrt{\sum_l ||\mathbf{W}_l^{(i_1)}||^2}} + \lambda_j^0 \frac{||\mathbf{W}_j^{(i_1)}||^2}{K\sqrt{\sum_l ||\mathbf{W}_l^{(i_1)}||^2}} + \frac{1-\lambda_j^0}{2} \cdot \frac{\langle \mathbf{W}_j^{(i_1)}, \mathbf{W}_{j+1}^{(i_1)} \rangle}{K\sqrt{\sum_l ||\mathbf{W}_l^{(i_1)}||^2}}. \tag{5.10}$$

In (5.10), $\langle \mathbf{W}_{j-1}^{(i)}, \mathbf{W}_j^{(i)} \rangle$ is the correlation between $\mathbf{W}_{j-1}^{(i)}$ and $\mathbf{W}_j^{(i)}$, and $\langle \mathbf{W}_j^{(i)}, \mathbf{W}_{j+1}^{(i)} \rangle$ is the correlation between $\mathbf{W}_j^{(i)}$ and $\mathbf{W}_{j+1}^{(i)}$. From the fingerprint design in Section 5.1.1,

$$\langle \mathbf{W}_{j-1}^{(i)}, \mathbf{W}_j^{(i)} \rangle \leq ||\mathbf{W}_j^{(i)}||^2, \quad \text{and} \quad \langle \mathbf{W}_j^{(i)}, \mathbf{W}_{j+1}^{(i)} \rangle \leq ||\mathbf{W}_j^{(i)}||^2. \tag{5.11}$$

From (5.10) and (5.11), if $\lambda_1^0, \cdots, \lambda_{j-1}^0, \lambda_{j+1}^0, \cdots$ are fixed, $\mu^{(i_1)}$ is a non-decreasing function of λ_j^0 and is minimized when $\lambda_j^0 = 0$. Consequently, from the detection probability's point of view, $\mathbf{u}^{(i_1)}$ should choose a smaller λ_j^0 to reduce his own probability of detection.

Selection of the Optimal Weight Vector

During pre-collusion processing, the selfish colluders wish to minimize their own probability of detection while maintaining the quality of the fingerprinted copies. Consequently, for a selfish colluder $\mathbf{u}^{(i_1)}$, the selection of the weight vector $[\lambda_1^0, \lambda_2^0, \cdots]$ can be modeled as

$$\begin{aligned} &\min_{\{\lambda_j^0\}} \left\{ \mu^{(i_1)} = \sum_j \mu_j^{(i_1)} \right\} \\ s.t. \quad &MSE_j \leq \varepsilon, \ 0 \leq \lambda_j^0 \leq 1, \ j = 1, 2, \cdots, \end{aligned} \tag{5.12}$$

where ε is the threshold on the perceptual quality. In our model of weighted average, $\{\lambda_j^0\}$ for different frames are selected independently. Thus, minimizing $\mu^{(i_1)}$ over the entire video sequence is equivalent to minimizing $\mu_j^{(i_1)}$ in (5.10) for each

frame j independently. Therefore, the optimization problem in (5.12) is equivalent to: for each frame j,

$$\begin{aligned} & \min_{\lambda_j^0} \ \mu_j^{i_1} \\ s.t. \quad & MSE_j \leq \varepsilon, \ 0 \leq \lambda_j^0 \leq 1, \end{aligned} \tag{5.13}$$

Given ϕ_j as defined in (5.6), we can show that the solution to (5.13) is

$$\lambda_j^* \ = \ \max\left\{0, \ 1 - 2 \cdot \sqrt{\varepsilon/\phi_j}\right\}. \tag{5.14}$$

5.2.3 Simulation Results

In our simulations, we choose sequence "carphone", and use the first 40 frames as an example. At the content owner's side, we adopt the human visual model based spread spectrum embedding [47], and embed fingerprints in the DCT domain. The fingerprints follow Gaussian distribution $\mathcal{N}(0, \sigma_W^2)$ with $\sigma_W^2 = 1/9$, and fingerprints for different users are generated independently. In each fingerprinted copy, similar to the work in [55], fingerprints embedded in adjacent frames are correlated with each other, and the correlation depends on the similarity between the two host frames.

At the colluder's side, we assume that there are a total of 150 colluders. For simplicity, we assume that there is only one selfish colluder and he applies weighted average as in (5.4) during pre-collusion processing. In addition, we assume that after the multiuser collusion attack, the colluders add an additive noise to further hinder the detection. In this chapter, we let the noise term $\mathbf{d}_j$ in (5.7) have variance $\sigma_n^2 = 2\sigma_W^2$, and other values of σ_n^2 will give the same trend.

At the detector's side, we consider a non-blind detection scenario. The detector first registers the test signal with respect to the host signal, then removes the host

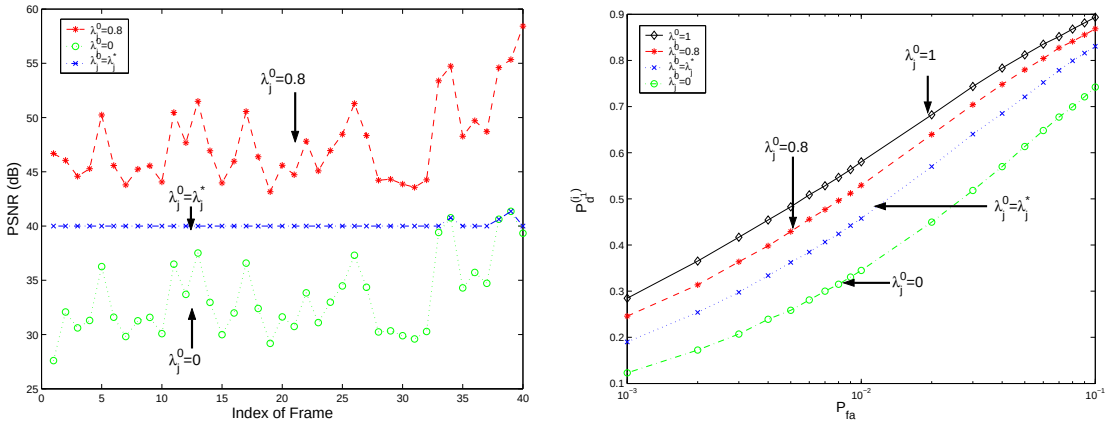


Figure 5.3: Simulation results of the weighted average on sequence "carphone". Assume that there are a total of $K = 150$ colluders and there is only one selfish colluder $\mathbf{u}^{(i_1)}$. $\{\lambda_j^*\}$ are the solution of (5.14) where ε is chosen to satisfy $PSNR_j \geq 40dB$ for all frame j. (Left): PSNR of the newly generated copy $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$ compared with the originally received fingerprinted frames $\{\mathbf{X}_j^{(i_1)}\}$. (Right): the selfish colluder's probability of detection $P_d^{(i_1)}$.

signal from the test copy, and finally applies the fingerprint detection process in Section 5.1.1.

Figure 5.3 shows the simulation results of weighted average on sequence "carphone". For each frame j, $PSNR_j$ is defined as PSNR of $\widetilde{\mathbf{X}}_j^{(i_1)}$ compared with $\mathbf{X}_j^{(i_1)}$. In Figure 5.3, $\{\lambda_j^*\}$ are the solution of (5.14) where ε is chosen to satisfy $PSNR_j \geq 40dB$ for all frames. In our simulations, we consider four different scenarios where $\lambda_j^0 = 1$, $\lambda_j^0 = 0.8$, $\lambda_j^0 = \lambda_j^*$, and $\lambda_j^0 = 0$, respectively[3]. Figure 5.3 (a) and (b) compare the perceptual quality of $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$ and the selfish colluder $\mathbf{u}^{(i_1)}$'s probability of detection, respectively, when $\{\lambda_j^0\}$ take different values.

[3] $\lambda_j^0 = 1$ corresponds to the scenario where the selfish colluder $\mathbf{u}^{(i_1)}$ does not apply pre-collusion processing before multiuser collusion.

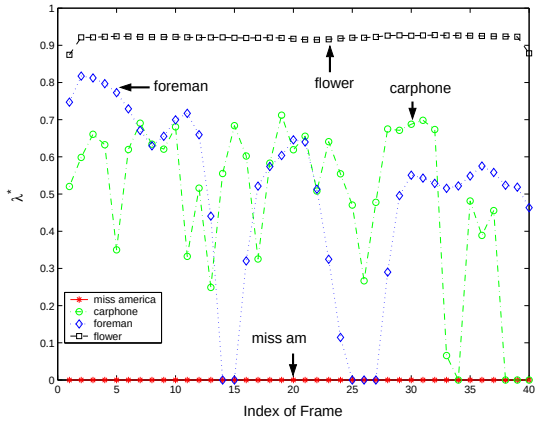


Figure 5.4: λ_j^* of (5.14) for different sequences where ε is chosen to satisfy $PSNR_j \geq 40dB$ for all frames in $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$.

From Figure 5.3, a selfish colluder can reduce his own probability of detection by applying the weighted average before multiuser collusion. By choosing $\{\lambda_j^0\}$ of smaller values, the selfish colluder has a smaller probability of detection while sacrificing the quality of the newly generated copy. Therefore, during pre-collusion processing, the selfish colluder has to consider the tradeoff between the probability of detection and the perceptual quality.

We then compare the solution of $\{\lambda_j^0\}$ in (5.14) for different sequences. We choose four representative video sequences: "miss america" that has large smooth region and slow motion, "carphone" and "foreman" that are moderately complicated, and "flower" where the high frequency band has large energy and the camera moves quickly. We choose the threshold ε in (5.14) such that $PSNR_j \geq 40dB$ for all frames in $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$. Figure 5.4 shows the solutions of (5.14) for various sequences. From Figure 5.4, for sequences that have slow motion ("miss america"), a selfish colluder can choose $\{\lambda_j^0\}$ with small values, e.g., around 0, without significant quality degradation; for sequences that have moderate motion ("carphone" and

"foreman"), λ_j^* is around 0.5; while for sequences with fast movement ("flower"), a selfish colluder has to choose large $\{\lambda_j^0\}$, e.g., larger than 0.9, to ensure the perceptual quality.

5.3 Modifying Resolution of Received Copies During Pre-collusion Processing

In the previous section, we have shown how a selfish colluder can modify the content of the received frames to minimize his probability of detection under the quality constraints, and it can be applied to all video fingerprinting systems. In this section, we consider scalable fingerprinting systems in which users receive copies of different quality, and study how a selfish colluder can modify the resolution of his fingerprinted copy to minimize his risk. For simplicity, in this section, we assume that the selfish colluders only change the resolution of their received copies and do not further apply weighted average during pre-collusion processing.

5.3.1 Changing the Resolution of the Fingerprinted Copies Before Collusion

Assume that $F^{(i_1)}$ contains the indices of the frames that a selfish colluder $\mathbf{u}^{(i_1)}$ subscribes to, and $\{\mathbf{X}_j^{(i)}\}$ are the fingerprinted frames that he receives from the content owner. Before collusion, the selfish colluder $\mathbf{u}^{(i_1)}$ processes his received copy and generates another copy $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$, whose temporal resolution is different from that of $\{\mathbf{X}_j^{(i_1)}\}$. Assume that $\widetilde{F}^{(i_1)}$ contains the indices of the frames in $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$ and $\widetilde{F}^{(i_1)} \neq F^{(i_1)}$. During collusion, $\mathbf{u}^{(i_1)}$ uses the newly generated copy $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}_{j\in\widetilde{F}^{(i_1)}}$, instead of $\{\mathbf{X}_j^{(i_1)}\}_{j\in F^{(i_1)}}$.

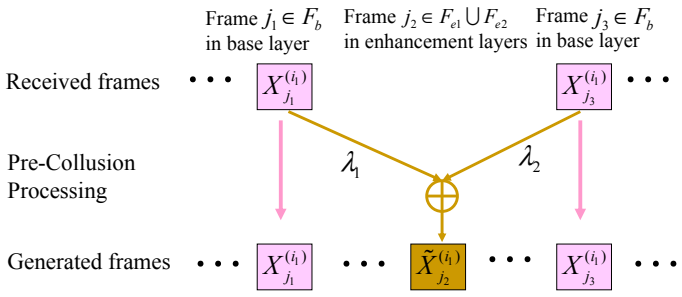


Figure 5.5: An example of cheat upward where $F^{(i_1)} = F_b$ and $\widetilde{F}^{(i_1)} = F_b \cup F_{e1} \cup F_{e2}$.

We consider a simple scenario where $\widetilde{F}^{(i_1)} \in \{F_b,\ F_b \cup F_{e1},\ F_b \cup F_{e1} \cup F_{e2}\}$. For simplicity, we assume that there is only one selfish colluder $\mathbf{u}^{(i_1)}$ who changes the resolution of his copy before multiuser collusion, and no colluders apply weighted average. Our analysis can be extended to complicated scenarios where there are multiple selfish colluders.

For a selfish colluder $\mathbf{u}^{(i_1)}$ who changes the resolution of his received copy during pre-collusion processing, we define the processing parameter as $CP^{(i_1)} \triangleq \left(F^{(i_1)}, \widetilde{F}^{(i_1)}\right)$. If $\widetilde{F}^{(i_1)} \supset F^{(i_1)}$, i.e., a selfish colluder $\mathbf{u}^{(i_1)}$ subscribes to the low quality version and he tells other colluders that he has a copy of higher frame rate, we call it "cheat upward". If $\widetilde{F}^{(i_1)} \subset F^{(i_1)}$, i.e., $\mathbf{u}^{(i_1)}$ subscribes to the high quality version and he tells others that he only has a copy of lower resolution, we call it "cheat downward".

Cheat Upward

In this type of pre-collusion processing, a selfish colluder $\mathbf{u}^{(i_1)}$ subscribes to a copy of low frame rate while telling other colluders that he received a copy of higher resolution. As an example, we consider a scenario where the processing parameter is $CP^{(i_1)} = \left(F^{(i_1)} = F_b, \widetilde{F}^{(i_1)} = F_b \cup F_{e1} \cup F_{e2}\right)$. In the example shown in Figure 5.5, a selfish colluder receives the fingerprinted base layer only, and applies cheat

upward to generate a copy $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}$ that contains frames in all three layers. He then tells the other colluders that $\{\widetilde{\mathbf{X}}_j^{(i_1)}\}_{j\in F_b\cup F_{e1}\cup F_{e2}}$ is the copy that he received.

Pre-collusion Processing

We assume that for every frame $j \in F^{(i_1)} = F_b$ that $\mathbf{u}^{(i_1)}$ received, the selfish colluder simply copies $\mathbf{X}_j^{(i_1)}$ and let $\widetilde{\mathbf{X}}_j^{(i_1)} = \mathbf{X}_j^{(i_1)}$. During pre-collusion processing, $\mathbf{u}^{(i_1)}$ needs to forge $\widetilde{\mathbf{X}}_j^{(i_1)}$ for frame $j \in F_{e1} \cup F_{e2}$ in the enhancement layers that he 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 are received by $\mathbf{u}^{(i_1)}$. 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 $j_1 < j_2 < j_3$, in this chapter, we consider a simple interpolation based method and let

$$\begin{aligned} \widetilde{\mathbf{X}}_{j_2}^{(i_1)} &= \lambda_1 \cdot \mathbf{X}_{j_1}^{(i_1)} + \lambda_2 \cdot \mathbf{X}_{j_3}^{(i_1)}, \\ \text{where} \quad \lambda_1 &= \frac{j_3 - j_2}{j_3 - j_1} \quad \text{and} \quad \lambda_2 = \frac{j_2 - j_1}{j_3 - j_1}. \end{aligned} \tag{5.15}$$

Other complicated algorithms, e.g., motion based interpolation [2, 8], can also be used to improve the quality of the forged frames, and the analysis will be similar.

Perceptual Quality

When the selfish colluder applies cheat upward, he has to forge frames in the enhancement layers that he did not receive from the content owner. To cover up the fact he processed his fingerprinted copy before collusion and make other colluders believe him, the selfish colluder must ensure that the fake enhancement layers generated by himself have high quality.

In this section, we examine the perceptual quality of the forged enhancement layers and study the quality constraints on cheat upward. As an example, we consider a scenario where the processing parameter is $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$, and use the simple interpolation based method in (5.15).

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

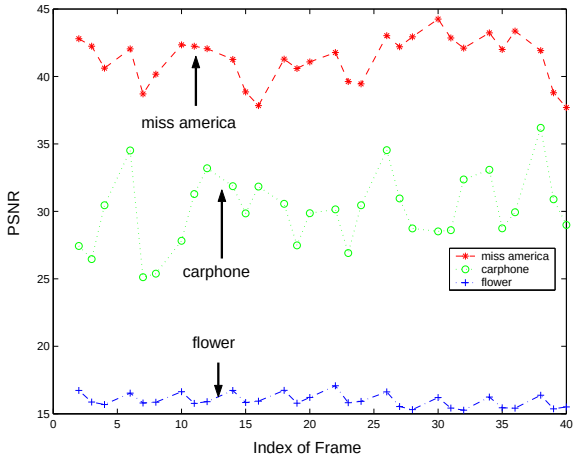


Figure 5.6: The quality of the enhancement layers that is forged by the selfish colluder during pre-collusion processing. The processing parameter is $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$ and the interpolation based method in (5.15) is used. $F_b = \{1, 5, 9, \cdots\}$, $F_{e1} = \{3, 7, 11, \cdots\}$ and $F_{e2} = \{2, 4, 6, 8, \cdots\}$.

enhancement layers, assume that $\mathbf{X}_j^{(i_1)}$ is the fingerprinted frame that he would have received if he had subscribed to frame j. In our simulations, we choose $\mathbf{X}_j^{(i_1)}$ as the ground truth and calculate the PSNR of $\widetilde{\mathbf{X}}_j^{(i_1)}$ compared with $\mathbf{X}_j^{(i_1)}$.[4]

Figure 5.6 shows the results on the first 40 frames of sequence "miss america", "carphone" and "flower". From Figure 5.6, for sequence "miss america" with flat regions and slow motion, the selfish colluder can forge enhancement layers of high quality (around $40dB$ in PSNR); for sequence "flower" that has fast movement, the selfish colluder can only generate low-quality enhancement layers (only $15dB$ in PSNR), and therefore, might not be able to apply cheat upward due to the

[4]In practice, the selfish colluder $\mathbf{u}^{(i_1)}$ does not have $\mathbf{X}_j^{(i_1)}$ and cannot use objective criteria to measure the quality of $\widetilde{\mathbf{X}}_j^{(i_1)}$. He can only subjectively judge the quality himself. The results shown here is only for the purpose of performance comparison.

quality constraints.[5]

Analysis of the Probability of Detection

For a selfish colluder, to analyze the effectiveness of cheat upward on reducing his risk of being captured, we compare his probability of detection when the selfish colluder applies cheat upward with that when he does not process his fingerprinted copy before collusion. We use the example in Figure 5.5 where the processing parameter is $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$.

We first consider the scenario where $\mathbf{u}^{(i_1)}$ does not apply pre-collusion processing, and we assume that

- $SC^b = \{i \in SC : F^{(i)} = F_b\}$ contains the indices of the colluders who subscribes to copies of lowest resolution and only receive the base layer from the content owner;
- $SC^{b,e1} = \{i \in SC : F^{(i)} = F_b \cup F_{e1}\}$ contains the indices of the colluders who receive both the base layer and the enhancement layer 1 from content owner;
- and $SC^{all} = \{i \in SC : F^{(i)} = F_b \cup F_{e1} \cup F_{e2}\}$ 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.

Given $(K^b, K^{b,e1}, K^{all})$ and (N_b, N_{e1}, N_{e2}), the colluders first decide the resolution of the colluded copy F^c, and then choose the collusion parameters $\{\beta_k\}_{k=1,2,3}$

[5]Motion based interpolation [2, 8] can be used to improve the quality. However, for some sequences with fast movement and complex scene composition, e.g., "football" and "flower", even with motion based interpolation, the selfish colluder may still not be able to forge enhancement layers of good enough quality to use. Therefore, the selfish colluders may not be able to apply cheat upward on those complicated sequences.

and $\{\alpha_l\}_{l=1,2}$ according to Table 4.1. In this scenario, for each frame $j \in F_b$ in the base layer, the extracted fingerprint is

$$\mathbf{Y}_j = \frac{\beta_1 \cdot \mathbf{W}_j^{(i_1)}}{K^b} + \sum_{i \in SC^b, i \neq i_1} \frac{\beta_1 \cdot \mathbf{W}_j^{(i)}}{K^b} + \sum_{i \in SC^{b,e1}} \frac{\beta_2 \cdot \mathbf{W}_j^{(i)}}{K^{b,e1}} + \sum_{i \in SC^{all}} \frac{\beta_3 \cdot \mathbf{W}_j^{(i)}}{K^{all}} + \mathbf{n}_j, \tag{5.16}$$

where $\mathbf{n}_j$ is the additive noise. We assume that $\mathbf{n}_j$ follows Gaussian distribution $\mathcal{N}(0, \sigma_n^2)$ for simplicity.

At the detector's side, following the detection procedure in Section 5.1.1, $\breve{F}^{(i)} = F^{(i)} \cap F^c = F_b$, and the detector calculates the detection statistics

$$T_N^{(i_1)} = \frac{\sum_{j \in F_b} \langle \mathbf{Y}_j, \mathbf{W}_j^{(i_1)} \rangle}{\sqrt{\sum_{j \in F_b} ||\mathbf{W}_j^{(i_1)}||^2}}, \tag{5.17}$$

compares it with the threshold h and decides if $\mathbf{u}^{(i_1)}$ is a possible colluder. It is straightforward to show that given the colluder set SC and the extracted fingerprint as in (5.16), the detection statistics in (5.17) follows distribution

$$\begin{aligned} p\left(T_N^{(i_1)} | SC\right) &\sim \mathcal{N}(\mu^{(i_1)}, \sigma_n^2), \\ \text{where} \quad \mu^{(i_1)} &= \frac{\beta_1}{K^b} \sqrt{\sum_{j \in F_b} ||\mathbf{W}_j^{(i_1)}||^2}. \end{aligned} \tag{5.18}$$

We then consider the scenario where $\mathbf{u}^{(i_1)}$ applies cheat upward during pre-collusion processing, and assume that

- $\widetilde{SC}^b = \{i \in SC : \widetilde{F}^{(i)}\}$ contains the indices of the colluders who *claim* that they received the base layer only;
- $\widetilde{SC}^{b,e1} = \{i \in SC : \widetilde{F}^{(i)} = F_b \cup F_{e1}\}$ is the set containing the indices of the colluders who *claim* that they have received both the base layer and enhancement layer 1;

- and $\widetilde{SC}^{all} = \{i \in SC : \widetilde{F}^{(i)} = F_b \cup F_{e1} \cup F_{e2}\}$ is the set containing the indices of the colluders who *claim* that they have received all three layers.

Define $\widetilde{K}^b, \widetilde{K}^{b,e1}$ and $\widetilde{K}^{all}$ as the number of colluders in $\widetilde{SC}^b, \widetilde{SC}^{b,e1}$ and $\widetilde{SC}^{all}$, respectively.

If $\mathbf{u}^{(i_1)}$ is the only selfish colluder and the processing parameter is $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$, then we have $\widetilde{SC}^b = SC^b \setminus \{i_1\}$[6], $\widetilde{SC}^{b,e1} = SC^{b,e1}$ and $\widetilde{SC}^{all} = SC^{all} \cup \{i_1\}$. Consequently, $\widetilde{K}^b = K^b - 1$, $\widetilde{K}^{b,e1} = K^{b,e1}$ and $\widetilde{K}^{all} = K^{all} + 1$. If other colluders do not discover $\mathbf{u}^{(i_1)}$'s processing before collusion, given $(\widetilde{K}^b, \widetilde{K}^{b,e1}, \widetilde{K}^{all})$ and (N_b, N_{e1}, N_{e2}), following Table 4.1, the colluders first decide on the indices of the frames in the colluded copy $\widetilde{F}^c$, and then choose the parameters $\{\widetilde{\beta}_k\}_{k=1,2,3}$ and $\{\widetilde{\alpha}_l\}_{l=1,2}$ accordingly. For fair comparison, if $(\widetilde{K}^b, \widetilde{K}^{b,e1}, \widetilde{K}^{all})$ and (N_b, N_{e1}, N_{e2}) satisfy the fairness constraints listed in Table 4.1, we choose $\widetilde{F}^c = F^c$.

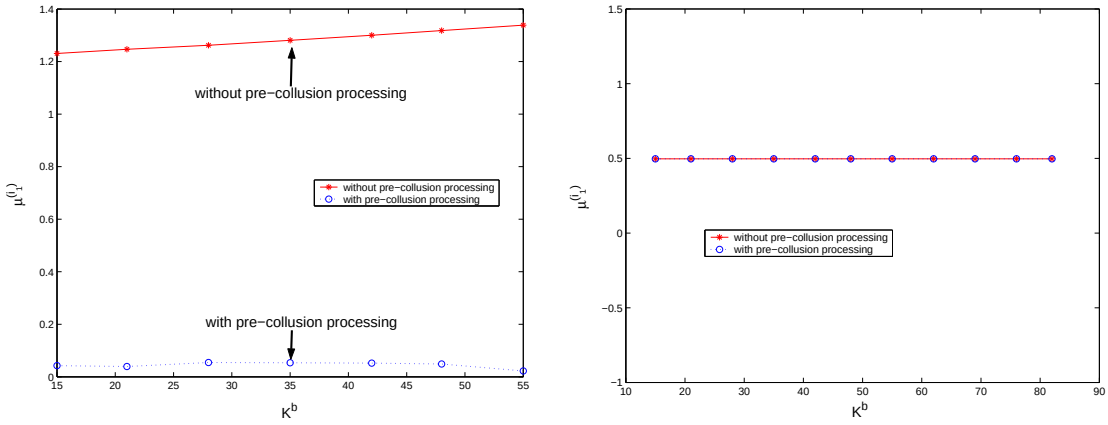
At the detector's side, in this scenario, for frame $j \in F_b$ in the base layer, the extracted fingerprint is

$$\mathbf{Y}_j = \frac{\widetilde{\beta}_3 \cdot \mathbf{W}_j^{(i_1)}}{\widetilde{K}^{all}} + \sum_{i \in SC^b, i \neq i_1} \frac{\widetilde{\beta}_1 \cdot \mathbf{W}_j^{(i)}}{\widetilde{K}^b} + \sum_{i \in SC^{b,e1}} \frac{\widetilde{\beta}_2 \cdot \mathbf{W}_j^{(i)}}{\widetilde{K}^{b,e1}} + \sum_{i \in SC^{all}} \frac{\widetilde{\beta}_3 \cdot \mathbf{W}_j^{(i)}}{\widetilde{K}^{all}} + \mathbf{n}_j. \tag{5.19}$$

Under the assumption that $\mathbf{n}_j$ follows the same Gaussian distribution $\mathcal{N}(0, \sigma_n^2)$ as in (5.16), given the colluder set SC, the identity of the selfish colluder i_1, and the pre-collusion processing parameter $CP^{(i_1)}$, the detection statistics in (5.17) follow the distribution

$$\begin{aligned} p\left(T_N^{(i_1)} | SC, i_1, CP^{(i_1)}\right) &\sim \mathcal{N}(\widetilde{\mu}^{(i_1)}, \sigma_n^2), \\ \text{where} \quad \widetilde{\mu}^{(i_1)} &= \frac{\widetilde{\beta}_3}{\widetilde{K}^{all}} \sqrt{\sum_{j \in F_b} ||\mathbf{W}_j^{(i_1)}||^2}. \end{aligned} \tag{5.20}$$

[6]For two sets A and B where $B \subseteq A$, $A \setminus B \stackrel{\triangle}{=} \{i : i \in A, i \notin B\}$.



(a) $F^c = F_b \cup F_{e1} \cup F_{e2}$ (b) $F^c = F_b$

Figure 5.7: Means of the selfish colluder's detection statistics when he applies cheat upward. $(F_b, F_{e1}, F_{e2}) = (10, 10, 20)$, and $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. There are a total of $M = 450$ users in the system and a total of $K = 150$ colluders. $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and $(K^b, K^{b,e1}, K^{all})$ are on the line (5.21). $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$.

From (5.18) and (5.20), if the same threshold h is used at the detector's side, comparing $P_d^{(i_1)}$ of these two scenarios is equivalent to comparing $\mu^{(i_1)}$ in (5.18) with $\widetilde{\mu}^{(i_1)}$ in (5.20).

To compare the values of the two means, we consider the following scalable fingerprinting systems. We observe that for video sequences like "miss america", "carphone" and "foreman", each frame has approximately 3000 $\sim$ 7000 embeddable coefficients, depending on the characteristics of the video sequence. As an example, we assume that the length of the embedded fingerprints in each frame is 5000, and we test on a total of 40 frames. We choose $F_b = \{j : j = 4k + 1, k = 0, \cdots, 9\}$, $F_{e1} = \{j : j = 4k + 3, k = 0, \cdots, 9\}$ and $F_{e2} = \{j : j = 2k, k = 1, \cdots, 20\}$ as an example of the temporal scalability, and the lengths of the fin-

gerprints embedded in the base layer, enhancement layer 1 and enhancement layer 2 are $N_b = 50000$, $N_{e1} = 50000$ and $N_{e2} = 100000$, respectively. We assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. Each user is assigned a unique fingerprint following Guassian distribution $\mathcal{N}(0, \sigma_W^2)$ with $\sigma_W^2 = 1/9$.

We assume that there are a total of $K = 150$ colluders, $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and $(K^b, K^{b,e1}, K^{all})$ are on the line

$$\frac{K^{all}\sqrt{N_b + N_{e1} + N_{e2}}}{K^b\sqrt{N_b} + K^{b,e1}\sqrt{N_b + N_{e1}} + K^{all}\sqrt{N_b + N_{e1} + N_{e2}}} = \frac{N_{e2}}{N_b + N_{e1} + N_{e2}}, \quad (5.21)$$

which is the boundary of the fairness constraints for $F^c = F_b \cup F_{e1} \cup F_{e2}$ in Table 4.1. Other values of $(K^b, K^{b,e1}, K^{all})$ and (N_b, N_{e1}, N_{e2}) give the same trend.

Given the above scalable fingerprinting system, Figure 5.7 compares $\mu^{(i_1)}$ in (5.18) with $\widetilde{\mu}^{(i_1)}$ in (5.20). In Figure 5.7 (a), $F^c = F_b \cup F_{e1} \cup F_{e2}$ and the colluded copy has the highest resolution; and in Figure 5.7 (b), $F^c = F_b$ and the colluded copy contains frames in the base layer only. From Figure 5.7, cheat upward can help the selfish colluder to further reduce his probability of detection when the colluded copy is of high quality; while it can not lower the selfish colluder's risk when the colluders decide to generate a copy of low resolution. This is because when $F^c = F_b$, no matter how many frames that $\mathbf{u}^{(i_1)}$ claims that he has received, only those in the base layer are used during collusion. Therefore, in such a scenario, cheat upward cannot help the selfish colluders to further reduce his risk. To generalize, cheat upward is effective in reducing a selfish colluder $\mathbf{u}^{(i_1)}$'s probability of detection only if $F^c \supset F^{(i_1)}$.

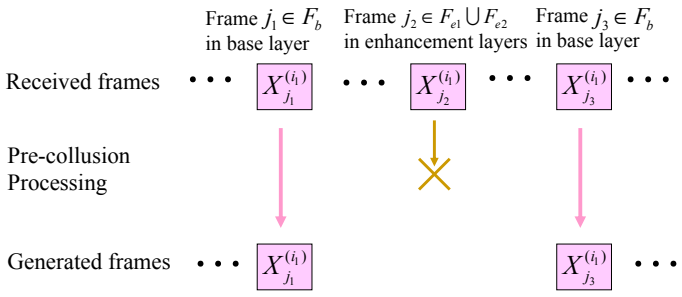


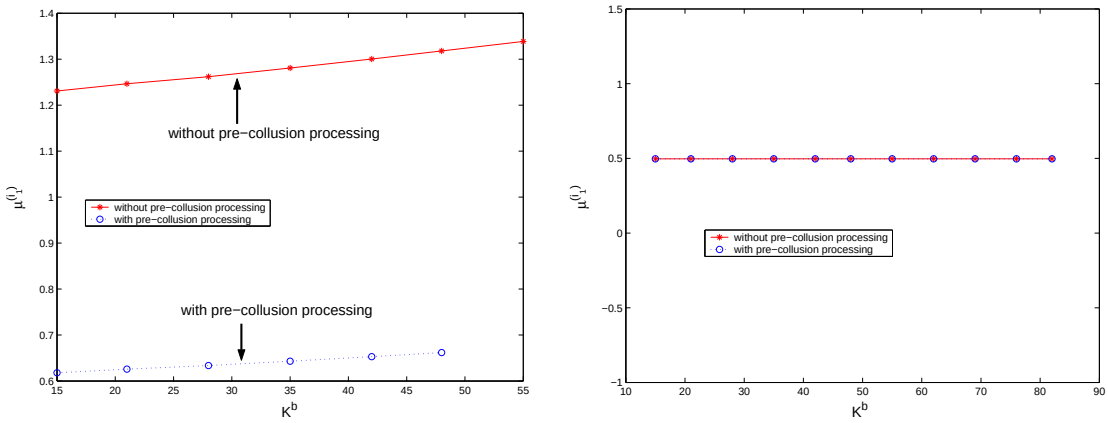
Figure 5.8: An example of cheat downward where $F^{(i_1)} = F_b \cup F_{e1} \cup F_{e2}$ and $\widetilde{F}^{(i_1)} = F_b$.

Cheat Downward

In cheat downward, a selfish colluder receives a copy of high resolution and tells other colluders that he only has a copy of low quality. Shown in Figure 5.8 is an example of cheat downward where $\mathbf{u}^{(i_1)}$ subscribes to all three layers while claiming that he only has the fingerprinted base layer. In this example, during pre-collusion processing, $\mathbf{u}^{(i_1)}$ simply keeps frames in the base layer and drops those in the enhancement layers.

In cheat downward, the selfish colluder does not need to forge any frames, and therefore, the quality constraints are satisfied. For cheat downward, the analysis of the selfish colluder's probability of detection is similar to that for cheat upward, and is not repeated here.

Figure 5.9 compares the means of the selfish colluder's detection statistics when he uses cheat downward with that when he does not apply pre-collusion processing. The setup of the scalable fingerprinting system in Figure 5.9 is the same as that in Figure 5.7. In Figure 5.9, the processing parameter is $CP^{(i_1)} = (F_b \cup F_{e1} \cup F_{e2}, F_b)$, and $F^c = F_b \cup F_{e1} \cup F_{e2}$ and $F^c = F_b$ in Figure 5.9 (a) and (b), respectively. Similar to cheat upward, when the colluded copy has high resolution, the selfish colluder



(a) $F^c = F_b \cup F_{e1} \cup F_{e2}$ (b) $F^c = F_b$

Figure 5.9: Means of the selfish colluder's detection statistics when he applies cheat downward. $(F_b, F_{e1}, F_{e2}) = (10, 10, 20)$ and $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. There are a total of $M = 450$ users in the system and a total of $K = 150$ colluders. $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and $(K^b, K^{b,e1}, K^{all})$ are on the line (5.21). $CP^{(i_1)} = (F_b \cup F_{e1} \cup F_{e2}, F_b)$.

can reduce his own probability of detection by applying cheat downward before multiuser collusion; while when the resolution of the colluded copy is low, cheat downward cannot further lower the selfish colluder's risk. Cheat downward can reduce the selfish colluder's probability of detection only when the colluded copy has high resolution and $F^c \supset \widetilde{F}^{(i_1)}$.

5.3.2 Performance Comparison of Different Strategy

For each selfish colluder, if he wants to modify the resolution of his fingerprinted copy during pre-collusion processing, he has two choices. For example, for a selfish colluder $\mathbf{u}^{(i_1 \in SC^b)}$ who receives the base layer only, he can apply cheat upward with two different processing parameters: $CP_1^{(i_1)} = (F_b, F_b \cup F_{e1})$ and $CP_2^{(i_1)} =$

$(F_b, F_b \cup F_{e1} \cup F_{e2})$. We assume that the fake enhancement layers generated in cheat upward satisfy the quality constraints, and other colluders do not discover the selfish colluder's pre-collusion processing behavior. In this section, we compare the effectiveness of different processing parameters in reducing the selfish colluder's risk of being captured.

From the analysis in the previous section, neither cheat upward nor cheat downward can further reduce the selfish colluder's probability of detection when $F^c = F_b$ and the colluded copy contains the base layer only. Therefore, in this section, we only consider the scenario where the colluded copy contains the enhancement layers and F^c equals to either $F_b \cup F_{e1}$ or $F_b \cup F_{e1} \cup F_{e2}$.

Our simulation setup is similar to that in Section 5.3.1. We assume each frame has 5000 embeddable coefficients and we test on a total of 40 frames. We consider a temporally scalable video coding system where $F_b = \{j : j = 4k + 1, k = 0, \cdots, 9\}$, $F_{e1} = \{j : j = 4k + 3, k = 0, \cdots, 9\}$ and $F_{e2} = \{j : j = 2k, k = 1, \cdots, 20\}$, and the lengths of the fingerprints embedded in the base layer, enhancement layer 1 and enhancement layer 2 are $N_b = 50000$, $N_{e1} = 50000$ and $N_{e2} = 100000$, respectively. We further assume that there are a total of $M = 450$ users in the system, and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. For each user, we generate a unique fingerprint following Gaussian distribution $\mathcal{N}(0, \sigma_W^2)$ where $\sigma_W^2 = 1/9$, and in each fingerprinted copy, we embed correlated fingerprints in adjacent frames. In addition, fingerprints for different users are independent of each other.

During collusion, we assume that there are a total of $K = 150$ colluders. $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and $(K^b, K^{b,e1}, K^{all})$ are on the line (5.21). We further assume that the additive noise $\mathbf{n}_j$ follows distribution $\mathcal{N}(0, \sigma_n^2)$ where $\sigma_n^2 = 2\sigma_W^2$. In our simulations, we assume that there is only one selfish colluder $\mathbf{u}^{(i_1)}$ and other

colluders do not discover the pre-collusion processing by the selfish colluder.

At the detector's side, we assume that there is a registration module that registers each frame in the colluded copy with respect to the corresponding frame in the host signal, and we consider a non-blind detection scenario where the host signal is first removed from the test copy before detection.

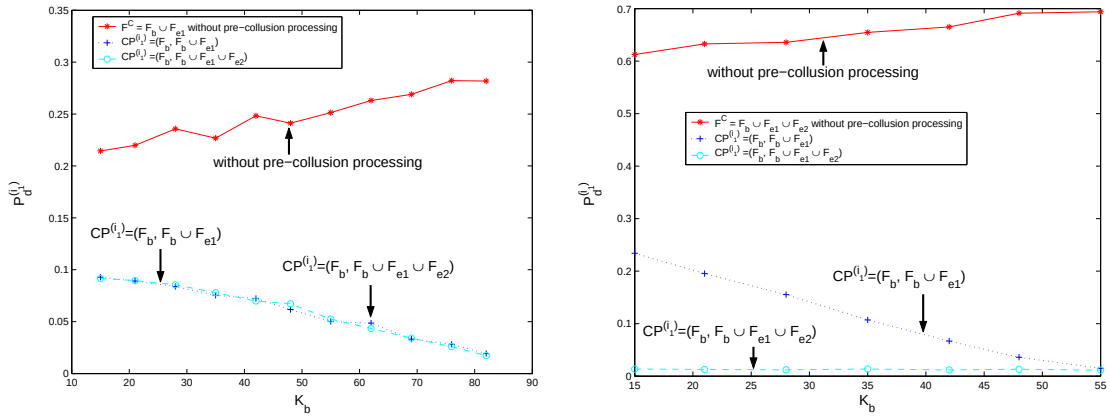
For Selfish Colluders in Subgroup SC^b

For a selfish colluder $\mathbf{u}^{(i_1)}$ where $i_1 \in SC_b$ and $F^{(i_1)} = F_b$, he can use cheat upward with two different processing parameters: $CP_1^{(i_1)} = (F_b, F_b \cup F_{e1})$ and $CP_2^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$. In this section, we compare the effectiveness of these two parameters in reducing $\mathbf{u}^{(i_1)}$'s probability of detection $P_d^{(i_1)}$.

Figure 5.10 shows our simulation results. In Figure 5.10, we fix the probability of accusing an innocent user P_{fa} as 0.01, and compare $P_d^{(i_1)}$ of different processing parameters. $F^c = F_b \cup F_{e1}$ and $F^c = F_b \cup F_{e1} \cup F_{e2}$ in Figure 5.10 (a) and (b), respectively. From the selfish colluder's point of view, when $F^c = F_b \cup F_{e1}$, the two parameters have almost identical performance. If $F^c = F_b \cup F_{e1} \cup F_{e2}$, $CP_2^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$ gives the selfish colluder a smaller probability of detection than $CP_2^{(i_1)} = (F_b, F_b \cup F_{e1})$. Therefore, under the quality constraints, a selfish colluder in SC^b should choose cheat upward with processing parameter $CP_2^{(i_1)}$ to minimize his own risk of being detected.

For Selfish Colluders in Subgroup $SC^{b,e1}$

For a selfish colluder $\mathbf{u}^{(i_1 \in SC^{b,e1})}$ who receives the base layer and the enhancement layer 1 from the content owner, he can apply cheat downward with processing parameter $CP_1^{(i_1)} = (F_b \cup F_{e1}, F_b)$ during pre-collusion processing, or use cheat



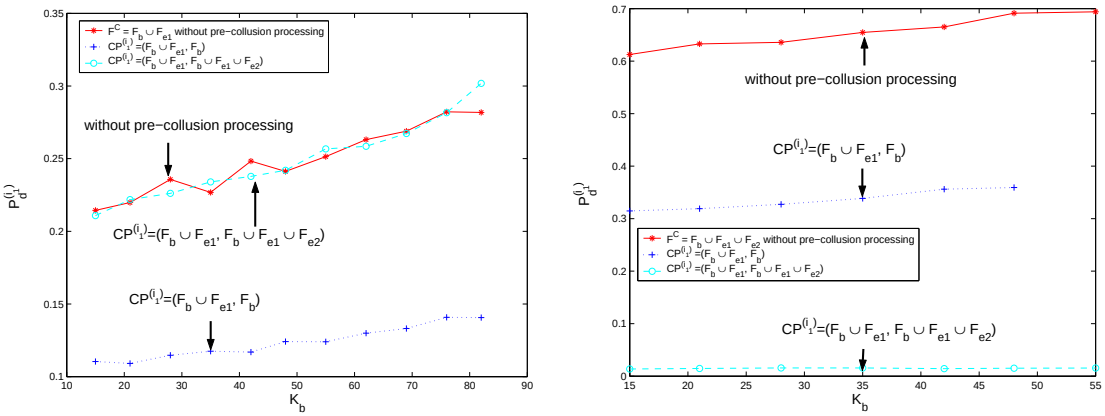
(a) $F^c = F_b \cup F_{e1}$ (b) $F^c = F_b \cup F_{e1} \cup F_{e2}$

Figure 5.10: Performance comparison of different processing parameters for selfish colluders in SC^b. $(F_b, F_{e1}, F_{e2}) = (10, 10, 20)$ and $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. The total number of colluders is $K = 150$. $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and $(K^b, K^{b,e1}, K^{all})$ are on the line (5.21). $P_{fa} = 0.01$.

upward with parameter $CP_2^{(i_1)} = (F_b \cup F_{e1}, F_b \cup F_{e1} \cup F_{e2})$.

Figure 5.11 shows the simulation results. When $F^c = F_b \cup F_{e1}$, $CP_1^{(i_1)}$ reduces $\mathbf{u}^{(i_1)}$'s probability of detection while $CP_2^{(i_1)}$ cannot lower the selfish colluder's risk. This is because when $F^c = F_b \cup F_{e1}$, $F^c \supset \widetilde{F}^{(i_1)}$ for cheat downward with parameter $CP_1^{(i_1)}$, while $F^c \not\supset F^{(i_1)}$ for cheat upward with parameter $CP_2^{(i_1)}$. The simulation results are in agreement with our analysis in the previous section. When $F^c = F_b \cup F_{e1} \cup F_{e2}$, both $CP_1^{(i_1)}$ and $CP_2^{(i_1)}$ reduce $\mathbf{u}^{(i_1)}$'s probability of detection, while $P_d^{(i_1)}$ of $CP_2^{(i_1)}$ is smaller than that of $CP_1^{(i_1)}$.

Consequently, in order for a selfish colluder in subgroup $SC^{b,e1}$ to minimize his own risk, when the colluded copy has medium resolution, he should use cheat downward with processing parameter $CP_1^{(i_1)} = (F_b \cup F_{e1}, F_b)$; and when the col-



(a) $F^c = F_b \cup F_{e1}$ (b) $F^c = F_b \cup F_{e1} \cup F_{e2}$

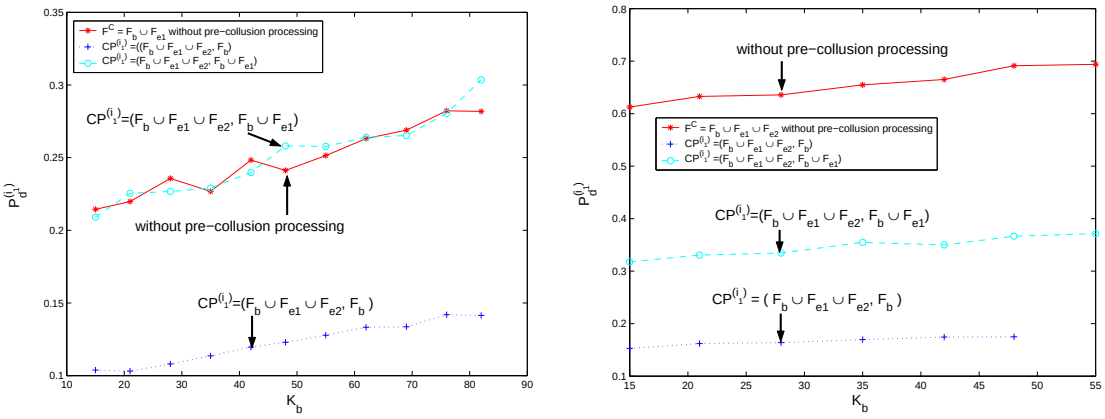
Figure 5.11: Performance comparison of different processing parameters for selfish colluders in $SC^{b,e1}$. $(F_b, F_{e1}, F_{e2}) = (10, 10, 20)$ and $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. The total number of colluders is $K = 150$. $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and $(K^b, K^{b,e1}, K^{all})$ are on the line (5.21). $P_{fa} = 0.01$.

luded copy has high resolution, he should apply cheat upward with $CP_2^{(i_1)} = (F_b \cup F_{e1}, F_b \cup F_{e1} \cup F_{e2})$.

For Selfish Colluders in SC^{all}

For a selfish colluder $\mathbf{u}^{(i_1)}$ in subgroup SC^{all} who receives all three layers, during pre-collusion processing, he can use cheat downward with two different parameters: $CP_1^{(i_1)} = (F_b \cup F_{e1} \cup F_{e2}, F_b)$ and $CP_2^{(i_1)} = (F_b \cup F_{e1} \cup F_{e2}, F_b \cup F_{e1})$. Figure 5.12 shows the simulation results.

From Figure 5.12, when $F^c = F_b \cup F_{e1}$, $CP_1^{(i_1)}$ reduces $\mathbf{u}^{(i_1)}$'s probability of detection, while $CP_2^{(i_1)}$ does not change the selfish colluder's risk of being captured. When $F^c = F_b \cup F_{e1} \cup F_{e2}$, both processing parameters reduce $\mathbf{u}^{(i_1)}$'s probability of



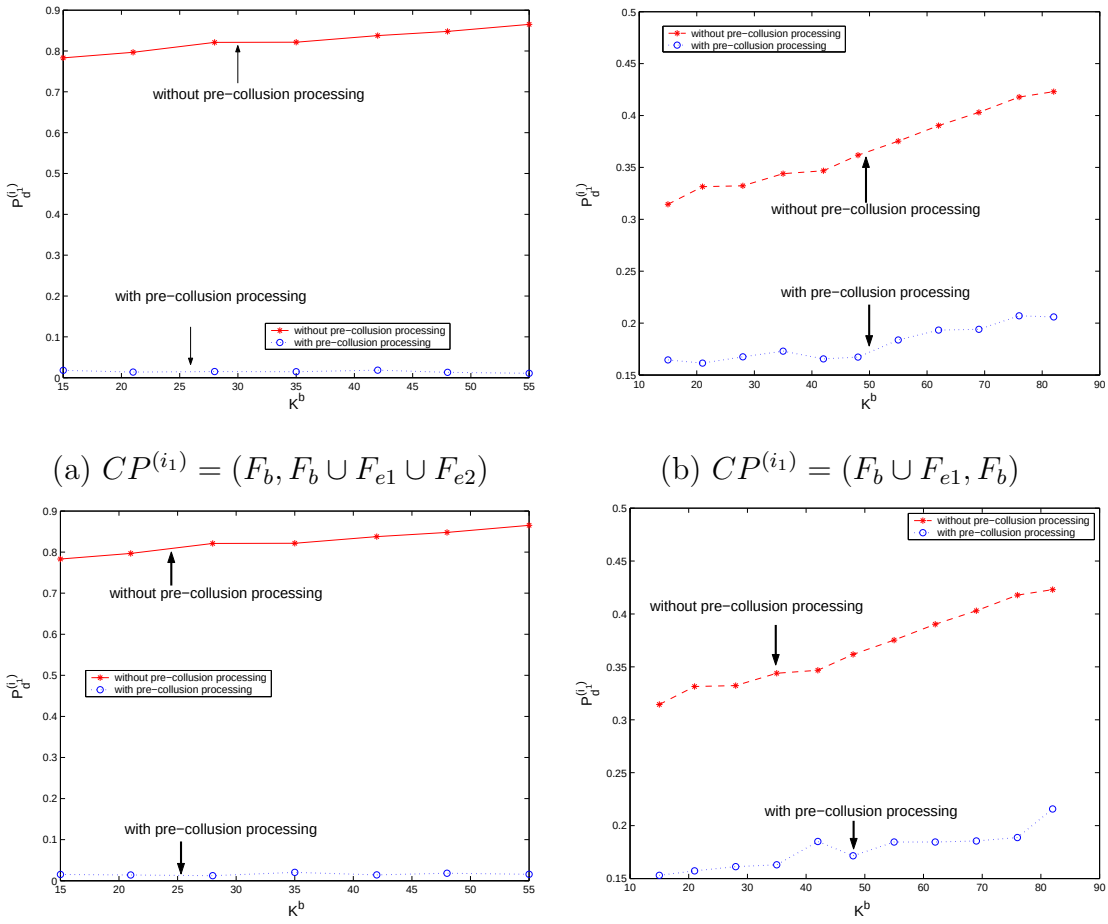
(a) $F^c = F_b \cup F_{e1}$ (b) $F^c = F_b \cup F_{e1} \cup F_{e2}$

Figure 5.12: Performance comparison of different processing parameters for selfish colluders in SC^{all}. $(F_b, F_{e1}, F_{e2}) = (10, 10, 20)$ and $(N_b, N_{e1}, N_{e2}) = (50000, 50000, 100000)$. Assume that there are a total of $M = 450$ users and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. The total number of colluders is $K = 150$. $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and $(K^b, K^{b,e1}, K^{all})$ are on the line (5.21). $P_{fa} = 0.01$.

detection, while $P_d^{(i_1)}$ of $CP_1^{(i_1)}$ is smaller than that of $CP_2^{(i_1)}$. Consequently, from the selfish colluder's point of view, for colluder $\mathbf{u}^{(i_1)}$ in subgroup SC^{all}, he should always choose cheat downward with parameter $CP_1^{(i_1)} = (F_b \cup F_{e1} \cup F_{e2}, F_b)$ during pre-collusion processing to minimize his own probability of detection.

5.3.3 Simulation Results on Real Video

We verify the correctness of our analysis on real videos, and choose the first 40 frames of the sequence "carphone" as an example. Similar to that in Section 5.3.1, we consider a temporally scalable video coding system where $F_b = \{j : j = 4k + 1, k = 0, \cdots, 9\}$, $F_{e1} = \{j : j = 4k + 3, k = 0, \cdots, 9\}$ and $F_{e2} = \{j : j = 2k, k = 1, \cdots, 20\}$. The length of the embedded fingerprints in the base



(a) $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$

(b) $CP^{(i_1)} = (F_b \cup F_{e1}, F_b)$

(c) $CP^{(i_1)} = (F_b \cup F_{e1}, F_b \cup F_{e1} \cup F_{e2})$

(d) $CP^{(i_1)} = (F_b \cup F_{e1} \cup F_{e2}, F_b)$

Figure 5.13: Simulation results of changing the resolution of the received copies during pre-collusion processing on the first 40 frames of sequence carphone. $(F_b, F_{e1}, F_{e2}) = (10, 10, 20)$. The total number of users is $M = 450$ and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. There are a total number of $K = 150$ colluders, $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and $(K^b, K^{b,e1}, K^{all})$ are on the line (5.21). P_{fa} is fixed as 10^{-2}. In (a) and (c), $F^c = F_b \cup F_{e1} \cup F_{e2}$. In (b) and (d), $F^c = F_b \cup F_{e1}$.

layer, enhancement layer 1 and enhancement layer 2 are $N_b = 72222$, $N_{e1} = 71926$ and $N_{e2} = 143820$, respectively. We assume that the total number of users is $M = 450$ and $|\mathbf{U}^b| = |\mathbf{U}^{b,e1}| = |\mathbf{U}^{all}| = 150$. We adopt the human visual model based spread spectrum embedding in [47], and embed the fingerprints in the DCT domain. The fingerprints follow Gaussian distribution $\mathcal{N}(0, \sigma_W^2)$ with $\sigma_W^2 = 1/9$, and the fingerprints assigned to different users are generated independently. 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 assume that there are a total of $K = 150$ colluders. $0 \leq K^b, K^{b,e1}, K^{all} \leq 150$ and they are on the line (5.21). We consider a simple scenario where there is only one selfish colluder who changes the resolution of his received copy before collusion, and no colluders apply weighted average. Furthermore, we assume that no colluders discover the pre-collusion processing by the selfish colluder. In our simulations, we adjust the power of the additive noise $\mathbf{n}_j$ such that $||\mathbf{n}_j||^2/||\mathbf{W}_j^{(i)}||^2 = 2$ for every frame $j \in F^c$ in the colluded copy.

We simulate the non-blind detection scenario where the detector first subtracts the host signal from the test copy before fingerprint detection, and then follows the procedure in Section 5.1.1.

Figure 5.13 shows the simulation results. In Figure 5.13 (a), the selfish colluder $\mathbf{u}^{(i_1)}$ is in subgroup SC^b and the processing parameter is $CP^{(i_1)} = (F_b, F_b \cup F_{e1} \cup F_{e2})$. In Figure 5.13 (b) and (c), the selfish colluder is in subgroup $SC^{b,e1}$ and the processing parameters are $CP^{(i_1)} = (F_b \cup F_{e1}, F_b)$ and $CP^{(i_1)} = (F_b \cup F_{e1}, F_b \cup F_{e1} \cup F_{e2})$, respectively. In Figure 5.13 (d), the selfish colluder is in subgroup SC^{all}, and the processing parameter is $CP^{(i_1)} = (F_b \cup F_{e1} \cup F_{e2}, F_b)$. In Figure 5.13 (a) and (c), $F^c = F_b \cup F_{e1} \cup F_{e2}$ and the colluded copy has high resolution, and in Figure

5.13 (b) and (d), $F^c = F_b \cup F_{e1}$ and the colluded copy has medium quality.

From Figure 5.13, under the quality constraints, if the colluded copy contains the enhancement layers and if the processing parameter is chosen properly, changing the resolution of the received copies can significantly reduce the selfish colluder's risk of being detected. In addition, the simulation results on real video agree with our analysis in Section 5.3, and are comparable with the simulation results in Section 5.3.2.

5.4 Countermeasures against Pre-collusion Processing

From the previous sections, by processing his fingerprinted copy before multiuser collusion, a selfish colluder can reduce his probability of detection, especially when the colluded copy has high quality. To prevent such pre-collusion processing and protect his own interest, before collusion, each colluder has to check the integrity of all the fingerprinted frames from the other colluders. In this section, we discuss some preliminary countermeasures against pre-collusion processing by selfish colluders.

In this chapter, we assume that when transmitting the fingerprinted copies through networks, the the service provider enables the users to authenticate each fingerprinted frame that is distributed to them. Possible authentication methods include [36,37,50,53,66,67,74]. If the content owner or the service provider provides such verification tools for the receivers, then before collusion, for each fingerprinted frame $\widetilde{\mathbf{X}}_j^{(i)}$ from colluder $\mathbf{u}^{(i)}$, all other colluders should verify whether it has been tampered using the same verification tools. This integrity check can help to detect

pre-collusion processing using weighted average. Furthermore, it will also assist the other colluders to detect cheat upward by a selfish colluder, who has to forge the enhancement layers that he did not subscribe to. Consequently, verification of each fingerprinted frame before multiuser collusion will help to detect and prevent a selfish colluder from applying both weighted average and cheat upward before multiuser collusion.

To detect cheat downward, if the content owner also provide tools to verify the indices of the frames that user $\mathbf{u}^{(i)}$ received during transmission, then the other colluders can use the same tools to verify $\widetilde{F}^{(i)}$ before multiuser collusion. If the content owner does not provide methods to verify $\widetilde{F}^{(i)}$, the colluders have to find other ways to guarantee the fairness of the collusion attacks. Note that for a selfish colluder $\mathbf{u}^{(i_1)}$, cheat downward will not further reduce his probability of detection if $F^c \subset \widetilde{F}^{(i_1)}$. Consequently, after the colluders verify each fingerprinted frame from all colluders, if they cannot verify $\widetilde{F}^{(i)}$, they have to choose $F^c = \bigcap_{i \in SC} \widetilde{F}^{(i)}$ to guarantee the absolute fairness of collusion. This implies that it is very likely that the colluders can only generate a colluded copy of low quality.

5.5 Chapter Summary

In this chapter, we have studied the traitors within traitors problem in multimedia fingerprinting, where some selfish colluders process their fingerprinted copies before multiuser collusion to minimize their own probability of detection under the quality constraints. We have studied the possible pre-collusion processing strategy by the selfish colluders, and analyzed their effects on both the selfish colluders' probability of detection and the perceptual quality of the fingerprinted copies. We have also studied possible countermeasures by other colluders to detect and prevent such

pre-collusion processing in order to protect their own interest.

We first proposed to use weighted average to attenuate the energy of the originally embedded fingerprints, and analyzed its performance. From both our analysis and simulation results, this pre-collusion processing technique reduces the selfish colluder's probability of detection at the cost of quality degradation. We then studied the selection of the optimal weight vector to minimize the selfish colluder's risk under the quality constraints. This type of pre-collusion processing can be applied to all video fingerprinting systems.

We also studied a specific pre-collusion processing technique on scalable fingerprinting systems, in which different colluders receive copies of different quality. In such fingerprinting systems, a selfish colluder can also modify the resolution of his received copy during pre-collusion processing. From our analytical and simulation results, with careful selection of the processing parameter, this type of pre-collusion processing can decrease the selfish colluder's risk of being captured when the colluded copy is of medium or high resolution. For this type of pre-collusion processing, we also analyzed the optimal processing parameter for selfish colluders in different subgroups to minimize their probability of detection under the quality constraints.

Finally, we studied some preliminary countermeasures for other colluders to detect and prevent pre-collusion processing. To detect weighted average and cheat upward, each colluder should check the integrity and verify the authenticity of each fingerprinted frame from other colluders. To prevent cheat downward, if the content owner does not provide tools for users to verify the indices of the received frames, the colluders have to generate a colluded copy of low quality in order to guarantee the absolute fairness of collusion.

Chapter 6

Secure Fingerprint Multicast for Video Streaming

In video streaming applications, a large amount of data has to be transmitted to a large number of users with limited bandwidth available under stringent latency constraints. This requires the service provider to minimize the communication cost in transmitting each copy in order to support as many users as possible. The uniqueness of each distributed copy in digital fingerprinting makes it even more critical to have secure fingerprint multicast schemes, which reduce the bandwidth requirement while protecting the secrecy of the multimedia content as well as each embedded fingerprint.

All the prior work on secure fingerprint multicast considered applications where the goal of the fingerprinting system is to be resistant to collusion attacks with a few colluders, e.g., seven or ten traitors, and designed the efficient distribution schemes accordingly. In video streaming applications, there are usually a large number of users (e.g., several thousand users), and therefore, potentially a large number of colluders (e.g., a few dozen or maybe even a hundred colluders). Some prior work

[58, 61, 62] has shown that with proper design and embedding of the fingerprints, the fingerprinting systems can resist collusion attacks with dozens of colluders, e.g., up to 60 colluders. In this chapter, we consider video applications where the fingerprinting system aims to survive collusion attacks with dozens of or even a hundred colluders, adopt the fingerprinting systems with strong traitor tracing capability [58, 62], and study the secure and efficient distribution of fingerprinted copies in such applications.

In this chapther, we consider the group-oriented fingerprint design in [62] as an example, and study the secure fingerprint multicast schemes for tree based fingerprinting systems. Section 6.2 introduces the tree based fingerprint design and Section 6.1 analyzes the security requirement in video streaming applications. In Section 6.3, we discuss the simple solution of pure unicast scheme, where each fingerprinted copy is unicasted to the corresponding user. In Section 6.4, we propose a general fingerprint multicast scheme that can be used with most spread spectrum embedding based fingerprinting systems, and in Section 6.5, we propose a joint fingerprint design and distribution scheme which utilizes the special structure of the fingerprint design to further improve the bandwidth efficiency.

6.1 Secure Video Streaming

In video streaming applications, to protect the welfare and interests of the content owner, it is critical to ensure the proper distribution and authorized usage of multimedia content. To be specific, the desired security requirements in video streaming applications are[1]:

[1]Note that depending on the applications, there might be other security requirements except these listed in this chapter, e.g., sender authentication and data integrity verification [45]. It

1. *Secrecy of the video content*: Only legitimate users who have registered with the content owner/service provider can have access to the content of video sequences. Proper encryption should be applied to prevent outsiders who do not subscribe to the service from estimating the video's content.

2. *Traitor tracing*: After the data are distributed to the legitimate users, the content owner has to protect multimedia from unauthorized redistribution. Digital fingerprinting is one possible solution to trace traitors and thwart the illegal information leakage.

3. *Robustness of the embedded fingerprints*: If digital fingerprinting is used for traitor tracing, it is required that the embedded fingerprints can survive common signal processing (e.g., compression), attacks on a single copy [14, 28], as well as multiuser collusion attacks [13, 52].

4. *Anti-framing*: The clear text of a fingerprinted copy is known only by the corresponding legitimate user whose fingerprint is embedded in that copy, and no other users of the service can access that copy in clear text and frame an innocent user.

In particular, we will explain the anti-framing requirement in detail. In digital fingerprinting applications, different fingerprinted copies do not differ significantly from each other. If the content owner or the service provider does not protect the transmitted bit streams appropriately, it is very easy for an attacker, who subscribes to the video streaming service, to impersonate an innocent user of the service.

---

is out of the scope of this chapter and we assume that the distribution systems have already included the corresponding security modules if required.

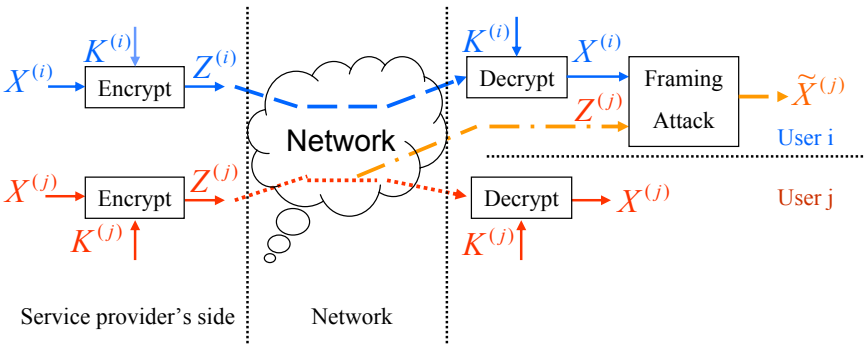


Figure 6.1: An example of framing attack on fingerprinting systems.

Figure 6.1 shows an example of the framing attack. Assume that $K^{(i)}$ and $K^{(j)}$ are the secret keys of user $\mathbf{u}^{(i)}$ and $\mathbf{u}^{(j)}$, respectively; $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$ are the clear text versions of two fingerprinted copies for $\mathbf{u}^{(i)}$ and $\mathbf{u}^{(j)}$, respectively; and $\mathbf{Z}^{(i)}$ and $\mathbf{Z}^{(j)}$ are the cipher text versions of $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$ encrypted with $K^{(i)}$ and $K^{(j)}$, respectively. $\mathbf{u}^{(i)}$ first decrypts $\mathbf{Z}^{(i)}$ that is transmitted to him and reconstructs $\mathbf{X}^{(i)}$. Assume that he also intercepts $\mathbf{Z}^{(j)}$ that is transmitted to $\mathbf{u}^{(j)}$. Without appropriate protection by the content owner or the service provider, $\mathbf{u}^{(i)}$ can compare $\mathbf{Z}^{(j)}$ with $\mathbf{X}^{(i)}$, estimates $\mathbf{X}^{(j)}$ without knowledge of $K^{(j)}$, and generates $\widetilde{\mathbf{X}}^{(j)}$ of good quality, which is an estimated version of $\mathbf{X}^{(j)}$. $\mathbf{u}^{(i)}$ can then redistribute $\widetilde{\mathbf{X}}^{(j)}$ or use $\widetilde{\mathbf{X}}^{(j)}$ during collusion. This framing puts innocent user $\mathbf{u}^{(j)}$ under suspicion and disables the content owner from capturing attacker $\mathbf{u}^{(i)}$. The content owner must prohibit such framing attacks.

In summary, before transmission, the content owner should embed unique and robust fingerprints in each distributed copy, and apply proper encryption to the bit streams to protect both the content of the video as well as each fingerprinted coefficient in every fingerprinted copy.

6.2 Tree Based Fingerprint Design

From Section 6.1, traitor tracing capability is a fundamental requirement for content protection and digital rights enforcement in networked video applications. This section introduces the tree based fingerprint design [62], which can resist collusion attacks by a few dozen colluders.

It was observed in [62] that a subgroup of users are more likely to collude with each other than others due to geographical or social reasons. A tree based fingerprint design was proposed in [61], [62] to explore the hierarchical relationship among users. In their fingerprint design, users that are more likely to collude with each other are assigned correlated fingerprints to improve the robustness against collusion attacks.

For simplicity, a symmetric tree structure is used where the depth of each leaf node is L and each node at level $l-1$ ($l = 1, \cdots, L$) has the same number of children nodes D_l. In a simple example of the tree structure shown in Figure 6.2, it is assumed that

- the users in the subgroup $\mathbf{U}^{1,1}$ are equally likely to collude with each other with probability p_3;
- each user in the subgroup $\mathbf{U}^{1,1}$ is equally likely to collude with the users in the subgroup $\mathbf{U}^{1,2}$ with probability $p_2 < p_3$;
- and each user in the subgroup $\mathbf{U}^{1,1} \cup \mathbf{U}^{1,2}$ is equally likely to collude with the users in other subgroups with probability $p_1 < p_2 < p_3$.

A unique basis fingerprint $\mathbf{a}^{(i_1,\cdots,i_l)}$ following Gaussian distribution $\mathcal{N}(0, \sigma_W^2)$ is generated for each node $[i_1, \cdots, i_l]$ in the tree except the root node and all the basis fingerprints $\{\mathbf{a}\}$ are independent of each other. For each user, all the

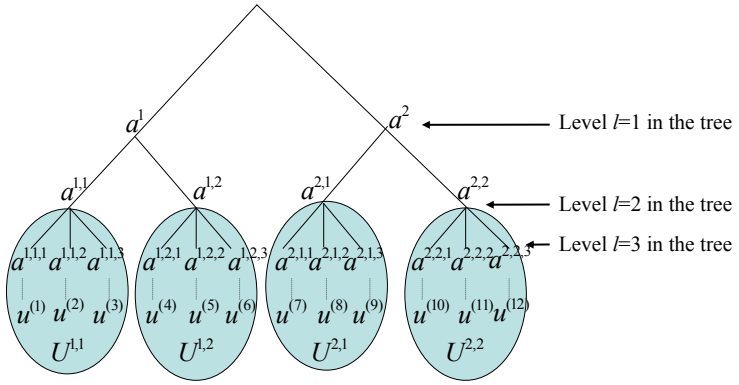


Figure 6.2: A tree-structure based fingerprinting scheme with $L = 3$, $D_1 = D_2 = 2$ and $D_3 = 3$.

fingerprints that are on the path from its corresponding leaf node to the root node are assigned to him. For example, in Figure 6.2, the fingerprints $\mathbf{a}^1$, $\mathbf{a}^{1,1}$ and $\mathbf{a}^{1,1,1}$ are embedded in the fingerprinted copy $\mathbf{X}^{(1)}$ that is distributed to user $\mathbf{u}^{(1)}$.

Assume that there are a total of K colluders and SC is the set containing the indices of the colluders. Given K different copies $\{\mathbf{X}^{(i)}\}_{i\in SC}$, the colluders generate the colluded copy $\mathbf{V} = g\left(\{\mathbf{X}^{(i)}\}_{i\in SC}\right)$ where $g(\cdot)$ is the collusion function.

In the detection process, the detector first extracts the fingerprint $\mathbf{Y}$ from $\mathbf{V}$. In [61], [62], a *multi-stage* colluder identification scheme was proposed and is as follows.

Detection at the first level of the tree: The detector correlates the extracted fingerprint $\mathbf{Y}$ with each of the D_1 fingerprints $\{\mathbf{a}^{i_1}\}_{i_1=1,\cdots,D_1}$ at level 1 and calculates the detection statistics

$$T^{i_1} =< \mathbf{Y}, \mathbf{a}^{i_1} > /||\mathbf{a}^{i_1}||,\ i_1 = 1, \cdots, D_1, \tag{6.1}$$

where $||\mathbf{a}||$ is the Euclidean norm of $\mathbf{a}$. The estimated guilty regions at level 1 are

$$GR(1) = \{[i_1] : T^{i_1} > h_1\}, \tag{6.2}$$

where h_1 is a predetermined threshold for fingerprint detection at the first level in the tree.

Detection at level $2 \leq l \leq L$ in the tree: Given the previously estimated guilty regions $GR(l-1)$, for each $[i_1, \cdots, i_{l-1}] \in GR(l-1)$, the detector calculates the detection statistics

$$T^{i_1,\cdots,i_{l-1},i_l} =< \mathbf{Y}, \mathbf{a}^{i_1,\cdots,i_{l-1},i_l} > /||\mathbf{a}^{i_1,\cdots,i_{l-1},i_l}||,\ i_l = 1, \cdots, D_l, \tag{6.3}$$

and narrows down the guilty regions to

$$GR(l) = \left\{[i_1, \cdots, i_l] : [i_1, \cdots, i_{l-1}] \in GR(l-1),\ T^{i_1,\cdots,i_l} \geq h_l\right\} \tag{6.4}$$

where h_l is a predetermined threshold for fingerprint detection at level l in the tree. Finally, the detector outputs the estimated colluder set

$$\widehat{SC} = \left\{\mathbf{u}^{(i)} : i = [i_1, \cdots, i_L] \in GR(L)\right\}. \tag{6.5}$$

6.3 The Pure Unicast Distribution Scheme

The most straightforward way to distribute the fingerprinted copies is the pure unicast scheme, where each fingerprinted copy is encoded independently, encrypted with the corresponding user's secret key and unicasted to him. It is simple and has limited requirement on the receivers' computation capability. However, from the bandwidth's point of view, it is inefficient because the required bandwidth is proportional to the number of users while the difference between different copies is small.

In this chapter, in the pure unicast scheme, to prevent outside attackers from estimating the video content, the generalized index mapping [21], [70] is used to encrypt portions of the compressed bit streams that carry the most important

information of the video content: the DC coefficients in the Intra blocks and the motion vectors in the Inter blocks. Applying the generalized index mapping to the fingerprinted AC coefficients can prevent the attackers from framing an innocent user at the cost of inroducing significant bit rate overhead.[2] In this chapter, to protect the fingerprinted coefficients without significant bit rate overhead, similar to the encryption scheme in [49], we apply the stream cipher [39] from traditional cryptography to the compressed bit streams of the AC coefficients.[3] It has no impact on the compression efficiency. In addition, the bit stuffing scheme [70] is used to prevent the encrypted data from becoming identical to some headers/markers.

6.4 The General Fingerprint Multicast Distribution Scheme

In this section, we propose a general fingerprint multicast distribution scheme that can be used with most multimedia fingerprinting systems where the spread spectrum embedding is adopted. We consider a video fingerprinting and distribution system that uses MPEG-2 encoding standard. For simplicity, we assume that all the distributed copies are encoded at the same bit rate and have approximately the same perceptual quality. To reduce the computation cost at the sender's side, fingerprints are embedded in the DCT domain. The block based human visual models [47] are used to guarantee the imperceptibility and control the energy of the embedded fingerprints.

[2]From [70], the bit rate is increased by more than 5.9% if two nonzero AC coefficients in each Intra block are encrypted.

[3]We only encrypt the content-carrying fields and the headers/markers are transmitted in clear text.

From human visual models [47], not all DCT coefficients are embeddable due to the imperceptibility constraints on the embedded fingerprints, and a non-embeddable coefficient has the same value in all copies. To reduce the required bandwidth in transmitting the non-embeddable coefficients, we propose a general fingerprint multicast scheme: the non-embeddable coefficients are multicasted to all users, and the coefficients left are embedded with unique fingerprints and unicasted to each user.[4]

In the general fingerprint multicast scheme, the transmitted video sequences are encrypted in the same way as in the pure unicast scheme. To guarantee that no outsiders can access the video content, a key that is shared by all users is used to encrypt the multicasted bit stream by applying the generalized index mapping to the DC coefficients in the Intra blocks and the motion vectors in the Inter blocks. To protect the fingerprinted coefficients, each unicasted bit stream is encrypted with the corresponding user's secret key. The stream cipher [39] is applied to the unicasted bit streams with headers/markers intact. Finally, the bit stuff scheme [70] is used to ensure that the cipher text does not duplicate MPEG headers/markers.

Figure 6.3 shows the MPEG-2 based general fingerprint multicast scheme for video on demand applications where the video is stored in compressed format. Assume that K^m is a key that is shared by all users, and $K^{(i)}$ is user $\mathbf{u}^{(i)}$'s secret key. The key steps in the fingerprint embedding and distribution at the server's side are as follows.

1. A unique fingerprint is generated for each user.

[4]We assume that each receiver has moderate computation capability and can listen to at least 2 channels simultaneously to reconstruct one video sequence. We also assume that the receivers have large enough buffers to smooth out the jittering of delays among different channels.

2. The compressed bit stream is split into two parts: the first one includes motion vectors, quantization factors and other side information and is not altered, and the second one contains the coded DCT coefficients and is variable length decoded.

3. Motion vectors, quantization factors and other side information are left intact, and only the values of the DCT coefficients are changed. For each DCT coefficient, if it is not embeddable, it is variable length coded with other non-embeddable coefficients. Otherwise, first, it is inversely quantized. Then for each user, the corresponding fingerprint component is embedded using spread spectrum embedding, and the resulting fingerprinted coefficient is quantized and variable length coded with other fingerprinted coefficients.

4. The non-embeddable DCT coefficients are encrypted with K^m and multicasted to all users, together with the positions of the embeddable coefficients in the 8×8 DCT blocks, motion vectors and other shared information; the coded fingerprinted DCT coefficients are encrypted with each user's secret key $\{K^{(i)}\}$ and unicasted to them.

For live applications where the video is compressed and transmitted at the same time, the fingerprint embedding and distribution process is similar to that for video on demand applications.

The decoder at user $\mathbf{u}^{(i)}$'s side is the same for both types of applications and is similar to a standard MPEG-2 decoder. After decrypting, variable length decoding and inversely quantizing both the unicasted bit stream to user $\mathbf{u}^{(i)}$ and the multicasted bit stream to all users, the decoder puts each reconstructed DCT coefficient in its original position in the 8×8 DCT block. Then, it applies inverse DCT and motion compensation to reconstruct each frame.

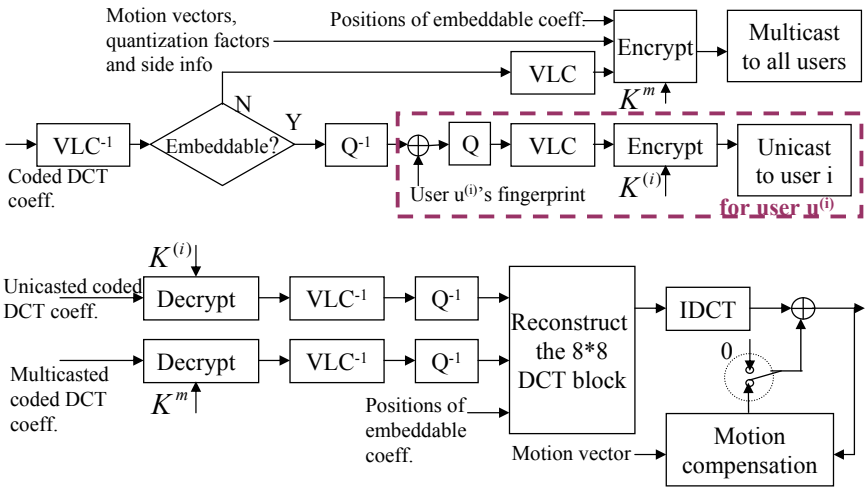


Figure 6.3: The MPEG-2 based general fingerprint multicast scheme for video on demand applications. Top: the fingerprint embedding and distribution process at the server's side, bottom: the decoding process at the user's side.

6.5 The Tree Based Joint Fingerprint Design and Distribution Scheme

The general fingerprint multicast scheme proposed in the previous section is design for the general fingerprinting applications that use spread spectrum embedding. In this section, to further improve the bandwidth efficiency, we utilize the tree structure of the embedded fingerprints and propose a joint fingerprint design and distribution scheme.

In this section, we first compare two fingerprint modulation schemes commonly used in the literature: the CDMA based and the TDMA based fingerprint modulation. We compare their bandwidth efficiency and their robustness against collusion attacks in the tree based fingerprinting systems. Then in Section 6.5.2, we propose a joint fingerprint design and distribution scheme that achieves both the robust-

ness against collusion attacks and the bandwidth efficiency of the distribution scheme. In Section 6.5.3, we take the computation constraints into consideration, and adjust the joint fingerprint design and distribution scheme to minimize the communication cost under the computation constraints.

6.5.1 The CDMA Based and The TDMA Based Fingerprint Modulation

In the tree based fingerprint design, a unique basis fingerprint $\mathbf{a}^{i_1,\cdots,i_l}$ following Gaussian distribution $\mathcal{N}(0,\sigma_W^2)$ is generated for each node $[i_1,\cdots,i_l]$ in the tree, and the basis fingerprints $\{\mathbf{a}\}$ are independent of each other. For user $\mathbf{u}^{(i)}$ whose index is $i=[i_1,\cdots,i_L]$, a total of L fingerprints $\mathbf{a}^{i_1},\mathbf{a}^{i_1,i_2},\cdots,\mathbf{a}^{i_1,\cdots,i_L}$ are embedded in the fingerprinted copy $\mathbf{X}^{(i)}$ that is distributed to him. Assume that the host signal $\mathbf{S}$ has a total of N embeddable coefficients. There are two different methods to embed the L fingerprints into the host signal $\mathbf{S}$: the CDMA based and the TDMA based fingerprint modulation.

The CDMA Based Fingerprint Modulation

In the CDMA based fingerprint modulation, the basis fingerprints $\{\mathbf{a}\}$ are of the same length N and equal energy. User $\mathbf{u}^{(i)}$'s fingerprint $\mathbf{W}^{(i)}$ is generated by

$$\mathbf{W}^{(i)}=\sqrt{\rho_1}\,\mathbf{a}^{i_1}+\sqrt{\rho_2}\,\mathbf{a}^{i_1,i_2}+\cdots+\sqrt{\rho_L}\,\mathbf{a}^{i_1,i_2,\cdots,i_L}, \tag{6.6}$$

and the fingerprinted copy distributed to $\mathbf{u}^{(i)}$ is $\mathbf{X}^{(i)}=\mathbf{S}+\mathbf{W}^{(i)}$ where $\mathbf{S}$ is the host signal. In (6.6), $\{0\le\rho_l\le 1\}_{l=1}^{L}$ with $\sum_{j=1}^{L}\rho_j=1$ are determined by the probabilities of users under different tree branches to collude with each other. They are used to control the energy of the embedded fingerprints at each level and the correlation between fingerprints assigned to different users.

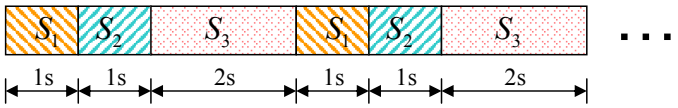


Figure 6.4: An example of the partitioning of the host signal for a tree with $L = 3$ and $[\rho_1,\ \rho_2,\ \rho_3] = [1/4,\ 1/4,\ 1/2]$.

The TDMA Based Fingerprint Modulation

In the TDMA based fingerprint modulation, the host signal $\mathbf{S}$ is divided into L non-overlapping parts $\mathbf{S}_1, \cdots, \mathbf{S}_L$, such that the number of embeddable coefficients in $\mathbf{S}_l$ is $N_l = \rho_l N$ with $\sum_{l=1}^{L} N_l = N$. An example of the partitioning of the host signal is shown in Figure 6.4 for a tree with $L = 3$, $[\rho_1,\ \rho_2,\ \rho_3] = [1/4,\ 1/4,\ 1/2]$ and $[N_1,\ N_2,\ N_3] = N[1/4,\ 1/4,\ 1/2]$. For every 4 seconds, all the frames in the 1st second belong to $\mathbf{S}_1$, all the frames in the 2nd second are in $\mathbf{S}_2$ and all the frames in the last two seconds are in $\mathbf{S}_3$. If the video sequence is long enough, the number of embeddable coefficients in $\mathbf{S}_l$ is approximately N_l.

In the TDMA based fingerprint modulation, the basis fingerprints $\{\mathbf{a}^{i_1,\cdots,i_l}\}$ at level l are of length N_l. In the fingerprinted copy $\mathbf{X}^{(i)}$ that is distributed to user $\mathbf{u}^{(i)}$, the basis fingerprint $\mathbf{a}^{i_1,\cdots,i_l}$ at level l is embedded in the lth part of the host signal $\mathbf{S}_l$, and the lth part of the fingerprinted copy $\mathbf{X}^{(i)}$ is $\mathbf{X}_l^{(i)} = \mathbf{S}_l + \mathbf{a}^{i_1,\cdots,i_l}$.

Comparison of the Performance of the CDMA Based and the TDMA Based Fingerprint Modulation

To compare the CDMA based and the TDMA based fingerprint modulation schemes in the tree based fingerprinting systems, we measure the energy of the fingerprints that are embedded in different parts of the fingerprinted copies. Assume that the host signal $\mathbf{S}$ is partitioned into L non-overlapping parts $\{\mathbf{S}_l\}_{l=1,\cdots,L}$ where there

are N_l embeddable coefficients in $\mathbf{S}_l$, the same as in the TDMA based modulation. We also assume that for user $\mathbf{u}^{(i)}$, $\mathbf{W}_l^{(i)}$ is the fingerprint that is embedded in $\mathbf{S}_l$, and $\mathbf{X}_l^{(i)} = \mathbf{S}_l + \mathbf{W}_l^{(i)}$ is the lth part of the fingerprinted copy that is distributed to $\mathbf{u}^{(i)}$. Define $E_{k,l}$ as the energy of the basis fingerprint $\mathbf{a}^{i_1,\cdots,i_k}$ at level k that is embedded in $\mathbf{X}_l^{(i)}$, and $E_l \stackrel{\triangle}{=} \sum_{k=1}^{L} E_{k,l}$ is the energy of $\mathbf{W}_l^{(i)}$. We further define a matrix $\mathbf{P}$ whose element at row k and column l is $p_{k,l} \stackrel{\triangle}{=} E_{k,l}/E_l$, and it is the ratio of the energy of the kth level fingerprint $\mathbf{a}^{i_1,\cdots,i_k}$ embedded in $\mathbf{X}_l^{(i)}$ over the energy of $\mathbf{W}_l^{(i)}$. The $\mathbf{P}$ matrices for the CDMA based and the TDMA based fingerprint modulation schemes are

$$\mathbf{P}^{CDMA} = \begin{pmatrix} \rho_1 & \rho_1 & \cdots & \rho_1 \\ \rho_2 & \rho_2 & \cdots & \rho_2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_L & \rho_L & \cdots & \rho_L \end{pmatrix}_{L\times L} \quad \text{and} \quad \mathbf{P}^{TDMA} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{L\times L} \tag{6.7}$$

respectively. In addition, in the TDMA based fingerprint modulation scheme,

$$\mathbf{P}^{TDMA} \begin{bmatrix} N_1 & N_2 & \cdots & N_L \end{bmatrix}^T = N \begin{bmatrix} \rho_1 & \rho_2 & \cdots & \rho_L \end{bmatrix}^T \tag{6.8}$$

and $\sum_{l=1}^{L} N_l = N$, where N is the total number of embeddable coefficients in $\mathbf{S}$.

Comparison of Bandwidth Efficiency

First, in the TDMA based modulation scheme, $p_{k,l} = 0$ for $k > l$, and therefore, the lth part of the fingerprinted copy $\mathbf{X}_l^{(i)}$ is only embedded with the basis fingerprints at level $k \le l$ in the tree. Note that the basis fingerprints $\{\mathbf{a}^{i_1,\cdots,i_k}\}_{k\le l}$ are shared by all the users in the subgroup $\mathbf{U}^{i_1,\cdots,i_l} \stackrel{\triangle}{=} \{\mathbf{u}^{(j)}, j = [j_1,\cdots,j_l,\cdots,j_L] : j_1 = i_1,\cdots,j_l = i_l\}$, so is $\mathbf{X}_l^{(i)}$. Consequently, in the TDMA based fingerprint modulation, the distribution system can not only multicast the non-embeddable coefficients to all users, it can also multicast part of the fingerprinted coefficients

that are shared by a subgroup of users to them. In the CDMA based fingerprint modulation, $p_{k,l} > 0$ for $k > l$ and the distribution system can only multicast the non-embeddable coefficients. Therefore, from the bandwidth efficiency's point of view, the TDMA based modulation is more efficient than the CDMA based fingerprint modulation.

Comparison of Resistance to Collusion

Second, in the TDMA based modulation scheme, $p_{k,l} = 0$ for $k \neq l$ and the basis fingerprint $\mathbf{a}^{i_1,\cdots,i_l}$ at level l are only embedded in the lth part of the fingerprinted copy $\mathbf{X}_l^{(i_1,\cdots,i_l)}$. With the TDMA based modulation scheme, by comparing all the fingerprinted copies that they have, the colluders can distinguish different parts of the fingerprinted copies that are embedded with fingerprints at different levels in the tree. They can also figure out the structure of the fingerprint tree and the positions of all colluders in the tree. Based on the information they collect, they can apply a specific attack against the TDMA based fingerprint modulation, *the interleaving based collusion attack*.

Assume that SC is the set containing the indices of all colluders, and $\{\mathbf{X}^{(k)}\}_{k\in SC}$ are the fingerprinted copies that they received. In the interleaving based collusion attacks, the colluders divide themselves into L subgroups $\{SC_l \subseteq SC\}_{l=1,\cdots,L}$, and there exists at least one $1 \leq l < L$ such that the lth subgroup SC_l and the $(l+1)$th subgroup SC_{l+1} are under different branches in the tree and are non-overlapping, i.e., $SC_l \cap SC_{l+1} = \emptyset$. The colluded copy $\mathbf{V}$ contains L non-overlapping parts $\{\mathbf{V}_l\}_{l=1,\cdots,L}$, and the colluders in the subgroup SC_l generate the lth part of the colluded copy by $\mathbf{V}_l = g\left(\{\mathbf{X}_l^{(i)}\}_{i\in SC_l}\right)$ where $g(\cdot)$ is the collusion function. Figure 6.5 shows an example of the interleaving based collusion attack on the tree based fingerprint design of Figure 6.2. Assume that $SC =$

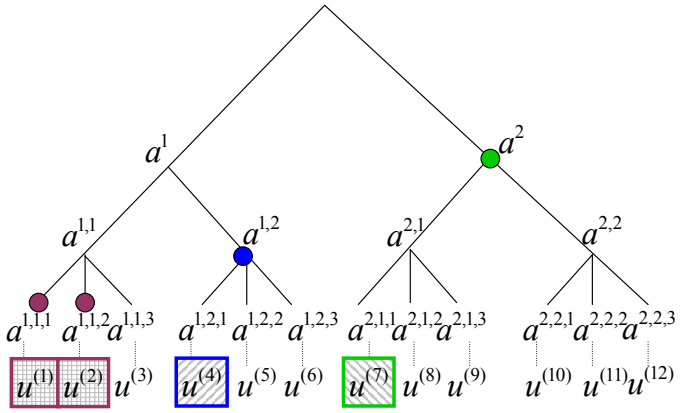


Figure 6.5: An example of the interleaving based collusion attack on the tree based fingerprinting system shown in Figure 6.2 with the TDMA based fingerprint modulation.

$\{1 = [1,1,1], 2 = [1,1,2], 4 = [1,2,1], 7 = [2,1,1]\}$ is the set containing the indices of the colluders. The colluders choose $SC_1 = \{7\}, SC_2 = \{4\}$ and $SC_3 = \{1,2\}$, and generate the colluded copy $\mathbf{V}$ where

$$\begin{aligned} \mathbf{V}_1 &= \mathbf{X}_1^{(7)} = \mathbf{S}_1 + \mathbf{a}^2, \\ \mathbf{V}_2 &= \mathbf{X}_2^{(4)} = \mathbf{S}_2 + \mathbf{a}^{1,2}, \\ \text{and} \quad \mathbf{V}_3 &= (\mathbf{X}_3^{(1)} + \mathbf{X}_3^{(2)})/2 = \mathbf{S}_3 + \left(\mathbf{a}^{1,1,1} + \mathbf{a}^{1,1,2}\right)/2. \end{aligned} \tag{6.9}$$

In the detection process, at the first level in the tree, although both $\mathbf{a}^1$ and $\mathbf{a}^2$ are guilty, the detector can only detect the existence of $\mathbf{a}^2$ because $\mathbf{a}^1$ is not in any part of the colluded copy $\mathbf{V}$. The detector outputs the estimated guilty region $GR(1) = [2]$. At the second level, the detector tries to detect whether $[2,1]$ and $[2,2]$ are the guilty sub-regions, and finds out neither of these two are guilty since $\mathbf{a}^{2,1}$ and $\mathbf{a}^{2,2}$ are not in $\mathbf{V}$. To continue the detection process, the detectors has to check the existence of each of the four fingerprints $\{\mathbf{a}^{i_1,i_2}\}$ in $\mathbf{V}$. This detection process is equivalent to the detection of independent fingerprints. The performance

of the detection process in the TDMA based fingerprint modulation is worse than that of the CDMA based fingerprint modulation [62], and it is due to the special structure of the fingerprint design and the unique "multi-stage" detection process in the tree based fingerprinting systems.

To summarize, in the tree based fingerprinting systems, the TDMA based fingerprint modulation improves the bandwidth efficiency of the distribution system at the cost of the robustness against collusion attacks.

6.5.2 The Joint Fingerprint Design and Distribution Scheme

In the joint fingerprint design and distribution scheme, the content owner first applies the tree based fingerprint design and generates the fingerprint tree, the same as in [61], [62]. Then, he embeds the fingerprints using the joint TDMA and CDMA fingerprint modulation scheme, which improves the bandwidth efficiency without sacrificing the robustness. Finally, the content owner distributes the fingerprinted copies to users using the proposed distribution scheme.

Design of the Joint TDMA and CDMA Fingerprint Modulation

To achieve both the robustness against collusion attacks and the bandwidth efficiency of the distribution scheme, we propose a *joint TDMA and CDMA fingerprint modulation scheme*, whose $\mathbf{P}$ matrix is an upper triangular matrix. In $\mathbf{P}^{Joint}$, we let $p_{k,l} = 0$ for $k > l$ to achieve the bandwidth efficiency. For $k \leq l$, we choose $0 < p_{k,l} \leq 1$ to achieve the robustness. Take the interleaving based collusion attack shown in Figure 6.5 as an example, in the joint TDMA and CDMA fingerprint modulation, although $\mathbf{a}^1$ is not in $\mathbf{V}_1$, it can still be detected from $\mathbf{V}_2$ and $\mathbf{V}_3$. Consequently, the detector can apply the "multi-stage" detection and narrow

down the guilty-region step by step, the same as in the CDMA based fingerprint modulation.

At level 1, $p_{1,1} = 1$. At level $2 \le l \le L$, given $p_{l,l}$, we seek $\{p_{k,l}\}_{k<l}$ to satisfy

$$E_{1,l} : E_{2,l} : \cdots : E_{l-1,l} = \rho_1 : \rho_2 : \cdots, \rho_{l-1}. \tag{6.10}$$

We can show that

$$\begin{aligned} p_{k,l} &= \frac{\rho_k}{\rho_1 + \cdots + \rho_{l-1}} (1 - p_{l,l}) \quad \text{for } 1 \le k < l \le L, \\ \text{and} \quad \mathbf{P}^{Joint} &= \begin{pmatrix} 1 & 1 - p_{2,2} & \cdots & (1 - p_{L,L}) \frac{\rho_1}{1-\rho_L} \\ 0 & p_{2,2} & \cdots & (1 - p_{L,L}) \frac{\rho_2}{1-\rho_L} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{L,L} \end{pmatrix}_{L \times L}. \end{aligned} \tag{6.11}$$

Given $\{p_{l,l}\}_{l=1,\cdots,L}$ and $\mathbf{P}_{Joint}$ as in (6.11), we seek $N_1, N_2, \cdots, N_L$ to satisfy

$$\begin{aligned} \mathbf{P}^{Joint} \begin{bmatrix} N_1 & N_2 & \cdots N_L \end{bmatrix}^T &= N \begin{bmatrix} \rho_1 & \rho_2 & \cdots \rho_L \end{bmatrix}^T \\ \text{s. t.} \quad \sum_{l=1}^{L} N_l &= N, \quad 0 \le N_l \le N. \end{aligned} \tag{6.12}$$

From (6.11), when $p_{L,L} = \rho_L$, it is the CDMA based fingerprint modulation. Therefore, we only consider the case where $p_{L,L} > \rho_L$. Define

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} 1 & 1 - p_{2,2} & \cdots & (1 - p_{L-1,L-1}) \frac{\rho_1}{\sum_{k=1}^{L-2} \rho_k} \\ 0 & p_{2,2} & \cdots & (1 - p_{L-1,L-1}) \frac{\rho_2}{\sum_{k=1}^{L-2} \rho_k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{L-1,L-1} \end{pmatrix}_{(L-1)\times(L-1)} \\ \text{and} \quad \mathbf{B} &= \frac{1 - p_{L,L}}{1 - \rho_L} \begin{pmatrix} \rho_1 & \rho_1 & \cdots & \rho_1 \\ \rho_2 & \rho_2 & \cdots & \rho_2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{L-1} & \rho_{L-1} & \cdots & \rho_{L-1} \end{pmatrix}_{(L-1)\times(L-1)}. \end{aligned} \tag{6.13}$$

We can show that (6.12) can be rewritten as:

$$\begin{pmatrix} & \mathbf{A}-\mathbf{B} & \\ p_{L,L} & p_{L,L} \cdots & p_{L,L} \end{pmatrix}_{L\times(L-1)} \begin{bmatrix} N_1 \\ \vdots \\ N_{L-1} \end{bmatrix} = (p_{L,L}-\rho_L)\, N \begin{bmatrix} \frac{\rho_1}{1-\rho_L} \\ \vdots \\ \frac{\rho_{L-1}}{1-\rho_L} \\ 1 \end{bmatrix},$$

$$\text{and} \quad N_L = N - \sum_{l=1}^{L-1} N_l. \tag{6.14}$$

$$\text{Define} \quad \mathbf{Q} \triangleq \begin{pmatrix} & \mathbf{A}-\mathbf{B} & \\ p_{L,L} & p_{L,L} \cdots & p_{L,L} \end{pmatrix}_{L\times(L-1)}$$

$$\text{and} \quad \underline{c} \triangleq (p_{L,L}-\rho_L)\, N \begin{bmatrix} \frac{\rho_1}{1-\rho_L} & \frac{\rho_2}{1-\rho_L} & \cdots & \frac{\rho_{L-1}}{1-\rho_L} & 1 \end{bmatrix}^T. \tag{6.15}$$

Given $\{p_{l,l}\}_{l=1,\cdots,L}$, if $\mathbf{Q}$ is of full rank, then the least square solution to (6.14) is

$$\begin{bmatrix} N_1 & N_2 & \cdots & N_{L-1} \end{bmatrix}^T = \mathbf{Q}^\dagger \underline{c} \quad \text{and} \quad N_L = N - \sum_{l=1}^{L-1} N_l, \tag{6.16}$$

where $\mathbf{Q}^\dagger = \left(\mathbf{Q}^T\mathbf{Q}\right)^{-1}\mathbf{Q}$ is the pseudo inverse of $\mathbf{Q}$. Finally, we need to verify the feasibility of the solution (6.16), i.e., if $0 \le N_l \le N$ for all $1 \le l \le L$. If not, another set of $\{p_{l,l}\}_{l=1,\cdots,L}$ has to be used.

Fingerprint Embedding and Detection in the Joint TDMA and CDMA Modulation

In the joint TDMA and CDMA fingerprint modulation scheme, given $\mathbf{P}^{Joint}$ as in (6.11) and $\{N_l\}_{l=1,\cdots,L}$ as in (6.16), for each basis fingerprint $\mathbf{a}^{i_1,\cdots,i_l}$ at level l in the tree,

$$\mathbf{a}^{i_1,\cdots,i_l} = \mathbf{a}_l^{i_1,\cdots,i_l} \Cup \mathbf{a}_{l+1}^{i_1,\cdots,i_l} \Cup \cdots \Cup \mathbf{a}_L^{i_1,\cdots,i_l}, \tag{6.17}$$

where $\{\mathbf{a}_k^{i_1,\cdots,i_l}\}_{k=l,\cdots,L}$ follow Gaussian distribution $\mathcal{N}(0,\sigma_W^2)$ and are independent of each other. $\mathbf{a}_k^{i_1,\cdots,i_l}$ for $k \geq l$ is of length N_k, and is embedded in $\mathbf{S}_k$. "$\Cup$" is the concatenation operator. For user $\mathbf{u}^{(i=[i_1,\cdots,i_L])}$, the lth part of the fingerprinted copy that $\mathbf{u}^{(i)}$ receives is

$$\mathbf{X}_l^{(i_1,\cdots,i_l)} = \mathbf{S}_l + \mathbf{W}_l^{(i_1,\cdots,i_l)}, \tag{6.18}$$

where

$$\mathbf{W}_l^{(i_1,\cdots,i_l)} = \sqrt{p_{1,l}}\mathbf{a}_l^{i_1} + \sqrt{p_{2,l}}\mathbf{a}_l^{i_1,i_2} + \cdots + \sqrt{p_{l,l}}\mathbf{a}_l^{i_1,\cdots,i_l}. \tag{6.19}$$

During collusion, assume that there are a total of K colluders and SC is the set containing the indices of all colluders. Assume that the colluders divide them into L subgroups $\{SC_l \subseteq SC\}_{l=1,\cdots,L}$. For each $1 \leq l \leq L$, given the K copies $\{\mathbf{X}_l^{(k)}\}_{k\in SC}$, the colluders in SC_l generate the lth part of the colluded copy by $\mathbf{V}_l = g\left(\{\mathbf{X}_l^{(k)}\}_{k\in SC_l}\right)$. Assume that $\mathbf{V} = \mathbf{V}_1 \Cup \cdots \Cup \mathbf{V}_L$ is the colluded copy that is generated by the colluders.

At the detector's side, given the colluded copy $\mathbf{V}$, for each $1 \leq l \leq L$, the detector first extracts the fingerprint $\mathbf{Y}_l$ from $\mathbf{V}_l$, and the detection process is similar to that in Section 6.2.

Detection at the first level of the tree: The detector correlates the extracted fingerprint $\{\mathbf{Y}_l\}_{l=1,\cdots,L}$ with each of the D_1 fingerprints $\{\mathbf{a}^{i_1}\}_{i_1=1,\cdots,D_1}$ at level 1 and calculates the detection statistics

$$T^{i_1} = \sum_{k=1}^{L} < \mathbf{Y}_k, \mathbf{a}_k^{i_1} > / \sqrt{\sum_{k=1}^{L} ||\mathbf{a}^{i_1}||^2}, \ i_1 = 1, \cdots, D_1. \tag{6.20}$$

The estimated guilty regions at level 1 are

$$GR(1) = \{[i_1] : T^{i_1} > h_1\}, \tag{6.21}$$

where h_1 is a predetermined threshold for fingerprint detection at the first level in the tree.

Detection at level $2 \leq l \leq L$ in the tree: Given the previously estimated guilty regions $GR(l-1)$, for each $[i_1, i_2, \cdots, i_{l-1}] \in GR(l-1)$, the detector calculates the detection statistics

$$T^{i_1,\cdots,i_{l-1},i_l} = \sum_{k=l}^{L} < \mathbf{Y}_k, \mathbf{a}_k^{i_1,\cdots,i_{l-1},i_l} > / \sqrt{\sum_{k=l}^{L} ||\mathbf{a}^{i_1,\cdots,i_{l-1},i_l}||^2}, \; i_l = 1, \cdots, D_l, \quad (6.22)$$

and narrows down the guilty regions to

$$GR(l) = \left\{ [i_1, \cdots, i_l] : [i_1, \cdots, i_{l-1}] \in GR(l-1), T^{i_1,\cdots,i_l} \geq h_l \right\}, \quad (6.23)$$

where h_l is a predetermined threshold for fingerprint detection at level l in the tree. Finally, the detector outputs the estimated colluder set

$$\widehat{SC} = \left\{ \mathbf{u}^{(i)} : i = [i_1, \cdots, i_L] \in GR(L) \right\}. \quad (6.24)$$

Fingerprint Distribution in the Joint Fingerprint Design and Distribution Scheme

In the joint fingerprint design and distribution scheme, given the fingerprinted copy $\{\mathbf{X}^{(i)}\}$ as in (6.18), the MPEG-2 based fingerprint distribution scheme for video on demand applications is shown in Figure 6.6. Assume that K^m is a key that is shared by all users, $K^{(i_1,\cdots,i_l)}$ is a key shared by a subgroup of users $\mathbf{U}^{(i_1,\cdots,i_l)}$, and $K^{(i)}$ is user $\mathbf{u}^{(i)}$'s secret key. The encryption method in the joint fingerprint design and distribution scheme is the same as that in the general fingerprint multicast and is not repeated. The key steps in the fingerprint embedding and distribution process at the server's side are as follows.

- For each user $\mathbf{u}^{(i)}$, the fingerprint $\mathbf{W}^{(i)}$ is generated as in (6.19).
- The compressed bit stream is split into two parts: the first one includes motion vectors, quantization factors and other side information and is not

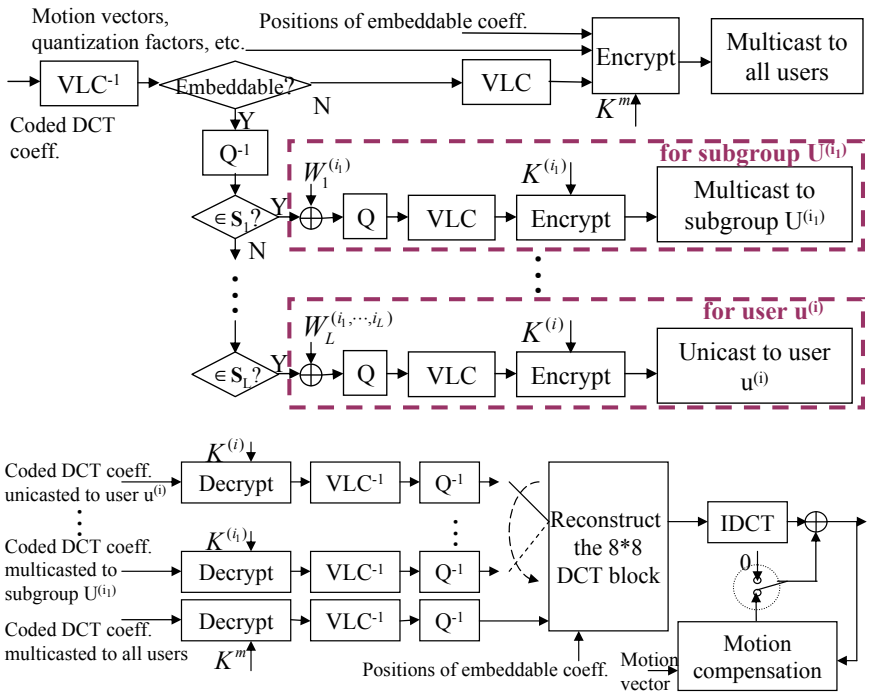


Figure 6.6: The MPEG-2 based joint fingerprint design and distribution scheme for video on demand applications. Top: the fingerprint embedding and distribution process at the server's side, bottom: the decoding process at the user's side.

altered, and the second one contains the coded DCT coefficients and is variable length decoded.

- Only the values of the DCT coefficients are modified, and the first part of the compressed bit stream is left unchanged. For each DCT coefficient, if it is not embeddable, it is variable length coded with other non-embeddable DCT coefficients. If it is embeddable, first, it is inversely quantized. If it belongs to $\mathbf{S}_l$, for each subgroup $\mathbf{U}^{i_1,\cdots,i_l}$, the corresponding fingerprint component in $\mathbf{W}_l^{(i_1,\cdots,i_l)}$ is embedded using spread spectrum embedding, and the resulting fingerprinted coefficients is quantized and variable length coded with other fingerprinted coefficients in $\mathbf{X}_l^{(i_1,\cdots,i_l)}$.

- The coded non-embeddable DCT coefficients are encrypted with key K^m and multicasted to all users, together with the positions of the embeddable coefficients in the 8×8 DCT blocks, motion vectors and other shared information. For $1 \leq l < L$, the coded fingerprinted coefficients in $\mathbf{X}_l^{(i_1,\cdots,i_l)}$ are encrypted with key $K^{(i_1,\cdots,i_l)}$ and multicasted to the users in the subgroup $\mathbf{U}^{i_1,\cdots,i_l}$. The coded fingerprinted coefficient in $\mathbf{X}_L^{(i)}$ are encrypted with user $\mathbf{u}^{(i)}$'s secret key and unicasted to him.

The decoder at user $\mathbf{u}^{(i)}$'s side is similar to that in the general fingerprint multicast scheme. The difference is that the decoder has to listen to $L+1$ bit streams in the joint fingerprint design and distribution scheme instead of 2 in the general fingerprint multicast scheme.

6.5.3 Joint Fingerprint Design and Distribution under Computation Constraints

Compared with the general fingerprint multicast scheme, the joint fingerprint design and distribution scheme further reduces the communication cost by multicasting some of the fingerprinted coefficients that are shared by a subgroup of users to them. However, it increases the total number of multicast groups that the sender needs to manage and the number of channels that each receiver downloads data from.

In the general fingerprint multicast scheme shown in Figure 6.3, the sender sets up and manages 1 multicast group, and each user listens to 2 bit streams simultaneously to reconstruct the fingerprinted video sequence. In the joint fingerprint design and distribution scheme, the sender has to set up a multicast group for every subgroup of users represented by a node in the upper $L-1$ levels in the tree. For

a tree structure with $L = 4$ and $[D_1, D_2, D_3, D_4] = [4, 5, 5, 100]$, the total number of multicast groups needed is 125. Also, each user has to listen to $L = 4$ different multicast groups and 1 unicast channel. In practice, the underlying network might not be able to support so many multicast groups simultaneously, and it could be beyond the sender's capability to manage this huge number of multicast groups at one time. It is also possible that the receivers can only receive data from a small number of channels simultaneously due to computation and buffer constraints.

To address this computation constraints, we adjust the joint fingerprint design and distribution scheme to minimize the overall communication cost under the computation constraints.

For a fingerprint tree of level L and degrees $[D_1, \cdots, D_L]$, if the sender sets up a multicast group for each subgroup of users represented by a node in the upper l levels in the tree, then the total number of multicast groups is $MG(l) \triangleq 1 + D_1 + \cdots + \prod_{m=1}^{l} D_m$. Also, each user listens to $RB(l) \triangleq l + 2$ channels. Assume that $\overline{MG}$ is the maximum number of multicast groups that the network can support and the sender can manage at once. We further assume that each receiver can only listen to no more than $\overline{RB}$ channels. Define the computation constraint parameter as $\left(\overline{MG}, \overline{RB}\right)$, and define

$$L' \triangleq \max\left\{l : MG(l) \leq \overline{MG}, RB(l) \leq \overline{RB}\right\}. \tag{6.25}$$

To satisfy the computation constraints $\left(\overline{MG}, \overline{RB}\right)$, we adjust the fingerprint distribution scheme in 6.5.2 as follows. Step 1, 2 and 3 in the distribution scheme in 6.5.2 are not changed, and Step 4 is modified to:

- The coded non-embeddable DCT coefficients are encrypted with key K^m and multicasted to all users, together with the positions of the embeddable coefficients in the 8×8 DCT blocks, motion vectors and other shared information.

- For each subgroup of users $\mathbf{U}^{i_1,\cdots,i_l}$ corresponding to a node at level $l \leq L'$ in the tree, a multicast group is set up and the fingerprinted coefficients in $\mathbf{X}_l^{(i_1,\cdots,i_l)}$ are encrypted with key $K^{(i_1,\cdots,i_l)}$ and multicasted to users in $\mathbf{U}^{i_1,\cdots,i_l}$.

- For $L' < m \leq L-1$, there are two possible methods to distribute the fingerprinted coefficients in $\mathbf{X}_m^{(i_1,\cdots,i_{L'},\cdots,i_m)}$ to a subgroup of users $\mathbf{U}^{i_1,\cdots,i_{L'},\cdots,i_m}$, and the one that has a smaller communication cost is chosen.

  - First, after encrypting the encoded fingerprinted coefficients in $\mathbf{X}_m^{(i_1,\cdots,i_m)}$ with key $K^{(i_1,\cdots,i_m)}$, the encrypted bit stream can be multicasted to the users in the subgroup $\mathbf{U}^{i_1,\cdots,i_{L'}}$. Since $K^{(i_1,\cdots,i_m)}$ is known only to the users in the subgroup $\mathbf{U}^{i_1,\cdots,i_m}$, only they can decrypt the bit stream and reconstruct $\mathbf{X}^{(i_1,\cdots,i_m)}$. This is similar to the distribution scheme in [10].

  - The fingerprinted coefficients in $\mathbf{X}^{(i_1,\cdots,i_m)}$ can also be unicasted to each user in the subgroup $\mathbf{U}^{i_1,\cdots,i_m}$ after encryption, the same as in the general fingerprint multicast scheme.

- The fingerprinted coefficients in $\mathbf{X}_L^{(i_1,\cdots,i_L)}$ are encrypted with user $\mathbf{u}^{(i=[i_1,\cdots,i_L])}$'s secret key $K^{(i)}$ and unicasted to him.

This will minimize the communication cost under the computation constraints $\left(\overline{MG}, \overline{RB}\right)$.

6.6 Chapter Summary

In this chapter, we have studied the secure distribution of fingerprinted copies for video streaming applications where a large amount of data have to be transmitted

to a large number of users in real time. In particular, we consider applications that require strong traitor tracing capability and can survive collusion attacks with up to a few dozen colluders, and we have proposed two secure fingerprint multicast schemes: a general fingerprint multicast scheme and a joint fingerprint design and distribution scheme.

We first observed that not all coefficients are embeddable in spread spectrum embedding due to perceptual constraints, and a non-embeddable coefficient has the same value in all copies. To reduce the communication cost in transmitting these non-embeddable coefficients, we proposed a general fingerprint multicast scheme that can be used with most spread spectrum embedding based fingerprinting systems. In this scheme, the non-embeddable coefficients that are shared by all users are multicasted, while the embeddable coefficients are embedded with each user's unique fingerprint and unicasted to the corresponding user.

We then proposed a joint fingerprint design and distribution scheme that explores the special structure of the fingerprint design to further improve the bandwidth efficiency. We first proposed a joint TDMA and CDMA fingerprint modulation scheme for the tree based fingerprint design. It enables the service provider to further multicast some fingerprinted coefficients that are shared by a subgroup of users to them, while maintaining the robustness of the embedded fingerprints against collusion. Based on the proposed fingerprint modulation scheme, we proposed a fingerprint multicast scheme that minimizes the overall communication cost under the computation constraints.

Chapter 7

Secure Fingerprint Multicast: Performance Analysis and Comparison

In Chapter 6, we have studied the secure distribution of fingerprinted copies in video streaming applications, and we have proposed two secure fingerprint multicast schemes: a general fingerprint multicast scheme that can be used with most spread spectrum embedding based fingerprinting systems, and a joint fingerprint design and distribution scheme that utilizes the special structure of the fingerprint design to further improve the bandwidth efficiency.

In this chapter, we analyze the performance of these two fingerprint multicast schemes, including the bandwidth efficiency, the robustness of the embedded fingerprints and the perceptual quality of the reconstructed fingerprints at the decoder's side. In Section 7.1, we analyze the bandwidth efficiency of the two multicast schemes and compare it with that of the pure unicast scheme, where each fingerprinted copy is unicasted to the corresponding user. In Section 7.2, we

compare the robustness of the embedded fingerprints using the joint TDMA and CDMA fingerprint modulation with that of the fingerprint embedded using the CDMA based fingerprint modulation, and equivalently, the resistance of the embedded fingerprints in the three schemes. In Section 7.3, we analyze the perceptual quality of the reconstructed video sequence at the decoder's side, and propose a fingerprint drift compensation scheme for the two fingerprint multicast schemes.

7.1 Analysis of Bandwidth Efficiency

To analyze the bandwidth efficiency of the secure fingerprint multicast schemes proposed in Chapter 6, we compare their communication costs with that of the pure unicast scheme. In this section, we assume that the fingerprinted copies in all schemes are encoded at the same targeted bit rate R.

To be consistent with general Internet routing where hop-count is the widely used metric for route cost calculation [11], we use the hop-based link usage to measure the communication cost and set the cost of all edges to be the same. To transmit a package of length Len^{unit} to a multicast group of size M, it was shown in [7, 11] that the normalized multicast communication cost can be approximated by $C^{unit}_{multi}(M)/C^{unit}_{uni}(M) = M^{EoS}$, where $C^{unit}_{multi}(M)$ is the communication cost using multicast, $C^{unit}_{uni}(M)$ is the average communication cost per user using unicast and EoS is the economies-of-scale factor. It was shown in [7] that EoS is between 0.66 and 0.7 for realistic networks. In this chapter, we choose $EoS \approx 0.7$.

7.1.1 The "multicast only" scenario

For the purpose of performance comparison, we consider another special scenario where the video streaming applications require the service provider to prevent outsiders from estimating the video's content, but do not require the traitor tracing capability. In this scenario, we apply the general index mapping to encrypt the DC coefficients in the Intra blocks and the motion vectors in Inter block; and the AC coefficients are left unchanged and transmitted in clear text. Since the copies that are distributed to different users are the same, the service provider can use a single multicast channel for the distribution of the encrypted bit stream to all users. We call this particular scenario, which does not require the traitor tracing capability and uses multicast channels only, the *"multicast only"*; and we compare the communication cost of the "multicast only" with that of the proposed secure fingerprint multicast schemes to illustrate the extra communication overhead introduced by the traitor tracing requirement.

For a given video sequence and a targeted bit rate R, we assume that in the pure unicast scheme, the average size of the compressed bit streams that are unicasted to different users is Len^{pu}. Define Len^{mo} as the length of the bit stream that is multicasted to all users in the "multicast only" scenario. Note that in the pure unicast scheme, the streaming cipher that we applied to the AC coefficients in each fingerprinted copy does not increase the bit rate and keep the compression efficiency unchanged. Consequently, we have $Len^{mo} \approx Len^{pu}$.

For a multicast group of size M, we further assume that the communication cost of the pure unicast scheme is C^{pu}, and C^{mo} is the communication cost in the "multicast only". We have

$$C^{pu}(M) \;=\; M \times C_{uni}^{unit}(M) \times Len^{pu}/Len^{unit},$$

$$\text{and} \quad C^{mo}(M) \;=\; C^{unit}_{multi}(M) \times Len^{mo}/Len^{unit}. \tag{7.1}$$

We define the communication cost ratio of the "multicast only" as

$$\gamma^{mo}(M) \triangleq \frac{C^{mo}(M)}{C^{pu}(M)} \approx M^{-0.3}, \tag{7.2}$$

and it depends only on the total number of users M.

7.1.2 The General Fingerprint Multicast Scheme

For a given video sequence and a targeted bit rate R, we assume that in the general fingerprint multicast scheme, the bit stream that is multicasted to all users is of length Len^{fm}_{multi}, and the average size of different bit streams that are unicasted to different users is Len^{fm}_{uni}. For a multicast group of size M, we further assume that the communication cost of the general fingerprint multicast scheme is C^{fm}. We have

$$C^{fm}(M) \;=\; C^{unit}_{multi}(M) \times Len^{fm}_{multi}/Len^{unit} + M \times C^{unit}_{uni}(M) \times Len^{fm}_{uni}/Len^{unit}. \tag{7.3}$$

The *coding parameter* is defined as $CP \triangleq \left(Len^{fm}_{multi} + Len^{fm}_{uni}\right)/Len^{pu}$, and the *unicast ratio* is defined as $UR \triangleq Len^{fm}_{uni}/\left(Len^{fm}_{multi} + Len^{fm}_{uni}\right)$. Then the communication cost ratio of the general fingerprint multicast scheme is

$$\gamma^{fm}(M) \triangleq \frac{C^{fm}(M)}{C^{pu}(M)} \approx CP\left\{UR + (1-UR)M^{-0.3}\right\}. \tag{7.4}$$

The smaller the communication cost ratio γ^{fm}, the more efficient the general fingerprint multicast scheme. Given the multicast group size M, the efficiency of the general fingerprint multicast scheme is determined by the coding parameter and the unicast ratio.

Coding Parameters

Four factors affect the coding parameters.

- For each fingerprinted copy, two different sets of motion vectors and quantization factors are used: the general fingerprint multicast scheme uses those calculated from the original unfingerprinted copy, while the pure unicast scheme uses those calculated from the fingerprinted copy itself. Note that the original unfingerprinted copy and the fingerprinted copy are similar to each other, so are both sets of parameters. Therefore, the difference between these two sets of motion vectors and quantization factors has negligible effect on the coding parameters.

- In the general fingerprint multicast scheme, headers and side information have to be inserted in each unicasted bit stream for synchronization. We follow the MPEG-2 standard and observe that this extra overhead consumes no more than 0.014 bit per pixel (bpp) per copy and is much smaller than the targeted bit rate R. Therefore, its effect on the coding parameters can be ignored.

- In the variable length coding stage, the embeddable and the non-embeddable coefficients are coded together in the pure unicast scheme while they are coded separately in the general fingerprint multicast scheme. Figure 7.1 shows the histograms of the (run length, value) pairs of the "carphone" sequence at $R = 1Mbps(1.3bpp)$ in both schemes. From Figure 7.1, the (run length, value) pairs generated by the two schemes have approximately the same distribution. Thus, encoding the embeddable and the non-embeddable coefficients together or separately does not affect the coding parameters. The

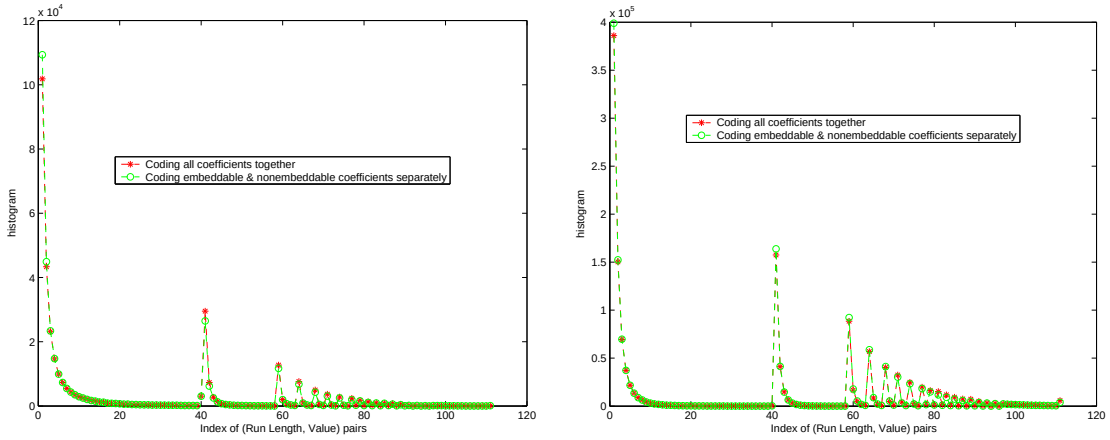


Figure 7.1: Histograms of the (run length, value) pairs of the "carphone" sequence that are variable length coded in the two schemes. $R = 1Mbps$. The indices of the (run length, value) pairs are sorted first in the ascending order of the run length, and then in the ascending order of the value. Left: in the Intra coded blocks, right: in the Inter coded blocks.

same conclusion can be drawn for other sequences and for other bit rates.

- In the general fingerprint multicast scheme, the positions of the embeddable coefficients have to be encoded and transmitted to the decoders. The encoding procedure is as follows.
  - For each 8×8 DCT block, first, an 8×8 mask is generated where a bit '0' is assigned to each non-embeddable coefficient and a bit '1' is assigned to each embeddable coefficient. Since DC coefficients are not embedded with fingerprints [47], the mask bit at the DC coefficient's position is skipped and only the 63 mask bits at the AC coefficients' positions are encoded.
  - Observing that most of the embeddable coefficients are in the low fre-

quencies, the 63 mask bits are zigzag scanned in the same way as in the JPEG baseline compression.

- Run length coding is applied to the zigzag scanned mask bits followed by huffman coding.
- An "End of Block" (EOB) marker is inserted after encoding the last mask bit whose value is 1 in the block.

Communication Cost Ratio

We choose three representative sequences: "miss america" with large smooth regions, "carphone" that is moderately complicated and "flower" that has large high frequency coefficients. Listed in Table 7.1 are the coding parameters, the unicast ratios and the communication cost ratios of these sequences at $R = 1.3bpp$. Figure 7.2 (a) also shows the communication cost ratios of the three sequences.

For M in the range between 1000 and 10000, compared with the pure unicast scheme, the general fingerprint multicast scheme reduces the communication cost by 48% to 84%, depending on the values of M and the characteristics of sequences. Given a sequence and a targeted bit rate R, the performance of the general fin-

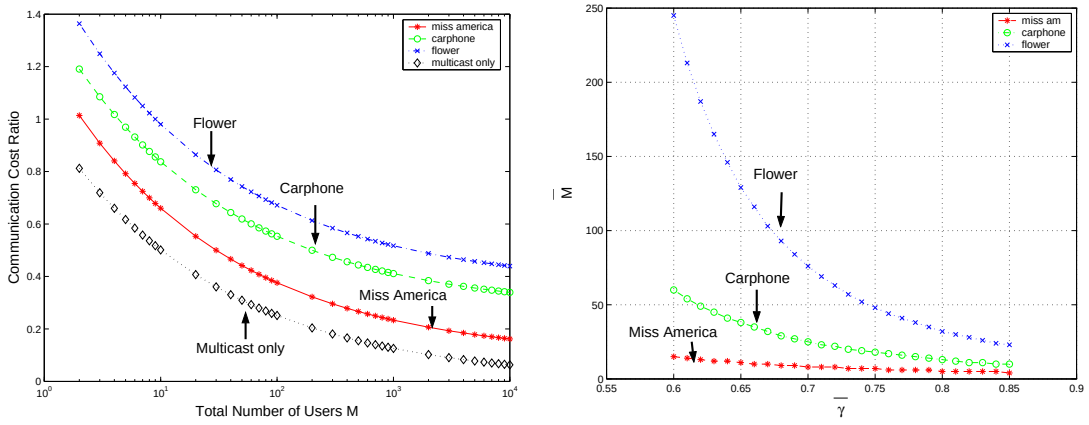
Table 7.1: Performance of the general fingerprint multicast scheme at $R = 1.3bpp$.

Sequence	Parameters		$\gamma^{fm}(\gamma^{mo})$			$\bar{M}$	
	CP	UR	$M = 1000$	$M = 5000$	$M = 10^4$	$\bar{\gamma} = 0.7$	$\bar{\gamma} = 0.8$
miss america	1.23	0.07	0.23	0.18	0.16	8	5
carphone	1.40	0.19	0.41	0.35	0.34	25	13
flower	1.65	0.23	0.52	0.46	0.44	76	32
multicast only	–	–	0.13	0.08	0.06	–	–

gerprint multicast scheme improves as the multicast group size M increases. For example, for the "carphone" sequence at $R = 1.3bpp$, $\gamma^{fm} = 0.41$ when there are a total of $M = 1000$ users, and it drops to 0.34 when M is increased to 10000. Also, given M, the performance of the general fingerprint multicast scheme depends on the characteristics of video sequences. For sequences with large smooth regions, the embedded fingerprints are shorter. Therefore, fewer bits are needed to encode the positions of the embeddable coefficients, and fewer DCT coefficients are transmitted through unicast channels. So the general fingerprint multicast scheme is more efficient. On the contrary, for sequences where the high frequency band has large energy, more DCT coefficients are embeddable. Thus, the general fingerprint multicast scheme is less efficient since the coding parameter and the unicast ratio are larger. When there are a total of $M = 5000$ users, the communication cost ratio is 0.18 for sequence "miss america" and is 0.46 for sequence "flower".

From Table 7.1 and Figure 7.2, if we compare the communication cost of the general fingerprint multicast with that of the "multicast only" scenario, enabling traitor tracing in video streaming applications introduces an extra communication overhead of 10% to 40%, depending on the characteristics of video sequences. For sequences with fewer embeddable coeffients, e.g, "miss america", the length of the embedded fingerprints is shorter, and applying digital fingerprinting increases the communication cost by a smaller percentage (around 10%). For sequences that have much more embeddable coefficients, e.g., "flower", more DCT coefficients have to be transmitted through unicast channels in the general fingerprint multicast scheme, and it increases the communication cost by a larger percentage (approximately 40%).

In addition, the general fingerprint multicast scheme performs worse than the



(a) $\gamma^{fm}(M)$ and $\gamma^{mo}(M)$ versus M (b) $\bar{M}$ versus $\bar{\gamma}$

Figure 7.2: Performance of the general fingerprint multicast scheme at $R = 1.3bpp$.

pure unicast scheme when M is small. Therefore, given the coding parameter and the unicast ratio, the pure unicast scheme is preferred when the communication cost ratio γ is larger than a threshold $\bar{\gamma}$, i.e., when M is smaller than $\bar{M}$ where

$$\bar{M} = \left\lceil \left(\frac{1 - UR}{\bar{\gamma}/CP - UR} \right)^{10/3} \right\rceil. \tag{7.5}$$

The ceil function $\lceil x \rceil$ returns the minimum integer that is not smaller than x. $\bar{M}$ of different sequences for different $\bar{\gamma}$ are listed in Table 7.1 and shown in Figure 7.2 (b). For $\bar{\gamma} = 0.8$ and $R = 1.3bpp$, $\bar{M}$ is 5 for sequence "miss america", 13 for "carphone" and 32 for "flower".

7.1.3 Joint Fingerprint Design and Distribution Scheme

For a given video sequence and a targeted bit rate R, we assume that in the joint fingerprint design and distribution scheme, the bit stream that is multicasted to all users is of length Len_{multi}^{joint} where $Len_{multi}^{joint} = Len_{multi}^{fm}$. For any two nodes $[i_1, \cdots, i_l] \neq [j_1, \cdots, j_l]$ at level l in the tree, we further assume that the bit streams that are transmitted to the users in the subgroups $\mathbf{U}^{i_1,\cdots,i_l}$ and $\mathbf{U}^{j_1,\cdots,j_l}$ are

approximately of the same length Len_l^{joint}.

In the joint fingerprint design and distribution scheme, all the fingerprinted coefficients inside one frame are variable length coded together. Therefore, the histograms of the (run length, value) pairs in the joint fingerprint design and distribution scheme are the same as that in the general fingerprint multicast scheme. If we ignore the impact of the headers/markers that are inserted in each bit stream, we have

$$\begin{aligned} & Len_1^{joint} + \cdots + Len_L^{joint} \approx Len_{uni}^{fm}, \\ \text{and} \quad & \frac{Len_{multi}^{joint} + \sum_{l=1}^{L} Len_l^{joint}}{Len^{pu}} \approx CP. \end{aligned} \tag{7.6}$$

Furthermore, fingerprints at different levels are embedded into the host signal periodically. In the simple example shown in Figure 6.4, the period is 4 seconds. If this period is small compared with the overall length of the video sequence, we can have the approximation that

$$\begin{aligned} & Len_1^{joint} : \cdots : Len_L^{joint} \approx N_1 : \cdots : N_L, \\ \text{and} \quad & Len_l^{joint} \approx \frac{N_l}{N} \cdot Len_{uni}^{fm}, \ 1 \le l \le L. \end{aligned} \tag{7.7}$$

In the joint fingerprint design and distribution scheme, to multicast the non-embeddable DCT coefficients and other shared side information to all users, the communication cost is

$$C_{multi}^{joint} = C_{multi}^{unit}(M) \times Len_{multi}^{joint} / Len^{unit}, \tag{7.8}$$

where M is the total number of users. For $l \le L'$, to multicast the fingerprinted coefficients in $\mathbf{X}_l^{(i_1,\cdots,i_l)}$ to the users in $\mathbf{U}^{i_1,\cdots,i_l}$, the communication cost is

$$C_l^{joint} = C_{multi}^{unit}(M_l) \times Len_l^{joint} / Len^{unit}, \tag{7.9}$$

where $M_l \triangleq \prod_{m=l+1}^{L} D_m$. There are M/M_l such subgroups. For $L' < l \leq L-1$, to distribute the fingerprinted coefficients in $\mathbf{X}_l^{(i_1,\cdots,i_{L'},\cdots,i_l)}$ to the users in the subgroup $\mathbf{U}^{i_1,\cdots,i_{L'},\cdots,i_l}$, the communication cost is

$$C_l^{joint} = \min\left\{C_{multi}^{unit}(M_{L'}) \times Len_l^{joint}/Len^{unit}\ ,\ M_l \cdot C_{uni}^{unit}(M_l) \times Len_l^{joint}/Len^{unit}\right\}, \tag{7.10}$$

where the first term is the communication cost if they are multicasted to users in the subgroup $\mathbf{U}^{i_1,\cdots,i_{L'}}$, and the second term is the communication cost if they are unicasted to each user in the subgroup $\mathbf{U}^{i_1,\cdots,i_{L'},\cdots,i_l}$. Finally, the communication cost of distributing the fingerprinted coefficients in $\mathbf{X}_L^{(i_1,\cdots,i_L)}$ to user $\mathbf{u}^{(i_1,\cdots,i_L)}$ is

$$C_L^{joint} = M \cdot C_{uni}^{unit}(M) \times Len_L^{joint}/Len^{unit}. \tag{7.11}$$

The overall communication cost of the joint fingerprint design and distribution scheme is

$$C^{joint} = C_{multi}^{joint} + \sum_{l=1}^{L} \frac{M}{M_l} \cdot C_l^{joint}, \tag{7.12}$$

and the communication cost ratio $\gamma^{joint} \triangleq \frac{C^{joint}}{C^{pu}}$ is

$$\begin{aligned}\gamma^{joint} \approx \quad & CP\left\{(1-UR)\cdot M^{-0.3} + UR\cdot\sum_{l=1}^{L'}\frac{N_l}{N}\cdot M_l^{-0.3}\right.\\ & \left. +UR\cdot\sum_{l=L'+1}^{L-1}\frac{N_l}{N}\cdot\min\left(\frac{M_{L'}^{0.7}}{M_l},\ 1\right) + UR\cdot\frac{N_L}{N}\right\}. \end{aligned}\tag{7.13}$$

Listed in Table 7.2 are the communication cost ratios of the joint fingerprint design and distribution scheme under different L' for sequence "miss america", "carphone" and "flower". $L' = 0$ corresponds to the general fingerprint multicast scheme. We consider three scenarios where the numbers of users are 1000, 5000 and 10000 respectively. The tree structures of the three scenarios are listed in Table 7.2. In the three cases considered, compared with the pure unicast scheme, the

Table 7.2: The communication cost ratios of the joint fingerprint design and distribution scheme. $L' = 0$ is the general fingerprint multicast scheme. $R = 1.3bpp$, $p = 0.95$.

	L'	MG	RB	miss america	carphone	flower	multicast only
$M = 1000$, $L = 3$,	0	1	2	0.23	0.41	0.52	
$\underline{D} = [2, 5, 100]$,	1	3	3	0.22	0.34	0.43	0.13
$\underline{\rho} = [1/4, 1/4, 1/2]$	2	13	4	0.20	0.31	0.39	
$M = 5000$, $L = 4$,	0	1	2	0.18	0.35	0.46	
$\underline{D} = [2, 5, 5, 100]$,	1	3	3	0.16	0.30	0.39	0.08
$\underline{\rho} = [1/6, 1/6, 1/6, 1/2]$	2	13	4	0.15	0.27	0.35	
	3	65	5	0.14	0.25	0.32	
$M = 10000$, $L = 4$,	0	1	2	0.16	0.34	0.43	
$\underline{D} = [4, 5, 5, 100]$,	1	5	3	0.14	0.28	0.37	0.06
$\underline{\rho} = [1/6, 1/6, 1/6, 1/2]$	2	25	4	0.13	0.26	0.33	
	3	125	5	0.13	0.23	0.30	

joint fingerprint design and distribution scheme reduces the communication cost by 57% to 87%, depending on the total number of users, network and computation constraints, and the characteristics of video sequences.

Given a sequence, the larger the L', i.e., the larger the $\overline{MG}$ and $\overline{RB}$, the more efficient the joint fingerprint design and distribution scheme. This is because more fingerprinted coefficients can be multicasted. Take the "carphone" sequence with $M = 1000$ users as an example, in the general fingerprint multicast scheme, $\gamma^{fm} = 0.41$. If $L' = 1$, the joint fingerprint design and distribution scheme reduces the communication cost ratio to 0.34, and it is further dropped to 0.31 if $\overline{MG} \geq 13$ and $\overline{RB} \geq 4$.

Also, compared with the general fingerprint multicast scheme, the extra communication cost saved by the joint fingerprint design and distribution scheme varies from sequence to sequence. For sequences that have more embeddable coefficients, the joint fingerprint design and distribution improves the bandwidth efficiency by a much larger percentage. For example, for $M = 5000$ and $L' = 2$, compared with the general fingerprint multicast scheme, the joint fingerprint design and distribution scheme further reduces the communication cost by 10% for sequence "flower", while it only further improves the bandwidth efficiency by 3% for sequence "miss america". However, for sequence "miss america" with $M = 5000$ users, the general fingerprint multicast scheme has already reduced the communication cost by 82%. Therefore, for sequences with fewer embeddable coefficients, the general fingerprint multicast scheme is recommended to reduce the bandwidth requirement at a low computation cost. The joint fingerprint design and distribution scheme is preferred on sequences with much more embeddable coefficients to achieve the bandwidth efficiency under network and computation constraints.

Compared with the "multicast only" scenario, the joint fingerprint design and distribution scheme enables the traitor tracing capability by increasing the communication cost by 6% to 30%, depending on the characteristics of the video sequence as well as the network and computation constraints. Compared with the "multicast only", for sequences with fewer embeddable coefficients, the joint fingerprint design and distribution scheme increases the communication cost by a smaller percentage (around 6% to 10% for sequence "miss america"); while for sequences with much more embeddable coefficients, the extra communication overhead introduced is larger (around 24% to 30% for sequence "flower").

7.2 Robustness of the Embedded Fingerprints

In this section, we take the tree based fingerprint design as an example, and compare the robustness of the embedded fingerprints in different schemes. In the pure unicast scheme and the general fingerprint multicast scheme, we use the CDMA based fingerprint modulation to be robust against interleaving based collusion attacks; and in the joint fingerprint design and distribution scheme, the joint TDMA and CDMA fingerprint modulation scheme proposed in Section 6.5.2 In Chapter 6 is used. In this section, we compare the collusion resistance of the fingerprints embedded using the joint TDMA and CDMA fingerprint modulation scheme with that of the fingerprints embedded using the CDMA based fingerprint modulation.

7.2.1 Digital Fingerprinting System Model

Spread spectrum embedding [14,47] is widely used in digital fingerprinting systems due to its robustness against many single-copy attacks. In spread spectrum em-

bedding, the fingerprint is additively embedded into the host signal, and human visual models are used to control the energy and the imperceptibility of the the embedded fingerprints. In this chapter, we use the the block based human visual models and follow the embedding method in [47].

At the attackers' side, since spread spectrum embedding has been proven to be robust against attacks on a single copy, e.g., compression and lower pass filtering, we focus on the multiuser collusion which is more challenging. Under those single-copy attacks, the performance of the joint TDMA and CDMA fingerprint modulation is similar to that of the watermarking systems in [14, 47] and is not repeated here.

During collusion, we assume that there are a total of K colluders and SC is the set containing the indices of all colluders. In the joint TDMA and CDMA fingerprint modulation, the colluders can apply the interleaving based collusion attacks, where the colluders divide themselves into L subgroups and $\{SC_l \subseteq SC\}_{l=1,\cdots,L}$ contain the indices of the colluders in the L subgroups, respectively. The colluders in subgroup SC_l generate the lth part of the colluded copy by $\mathbf{V}_l = g\left(\{\mathbf{X}_l^{(i)}\}_{i\in SC_l}\right)$ where $g(\cdot)$ is the collusion function. In the CDMA based fingerprint modulation, the colluders cannot distinguish fingerprints at different levels in the tree and cannot apply interleaving based collusion attacks. Consequently, $SC_1 = \cdots = SC_L = SC$ for collusion attacks on the CDMA based fingerprint modulation.

In a recent investigation [63], we have shown that nonlinear collusion attacks can be modeled as the averaging collusion attack followed by an additive noise. Under the constraint that the perceptual quality of the attacked copies under different collusion attacks are the same, different collusion attacks have almost identical performance. Therefore, we only consider the averaging collusion attack.

At the detector's side, we consider a non-blind detection scenario, where the host signal $\mathbf{S}$ is available to the detector and is first removed from the colluded copy $\mathbf{V}$ before fingerprint detection and colluder identification. Note that different from other data hiding applications where the host signal is not available to the detector and 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 third party who can have access to the original host signal. In addition, prior work has shown that the non-blind detection has a better performance than the blind detection [58], [63]. Therefore, in this chapter, we consider non-blind detection to improve the detection performance and the collusion resistance of the fingerprinting systems.

From the other point of view, with spread spectrum embedding, in the blind detection, the host signal serves as an additional noise with very large energy during the detection process, and the blind detection can be regarded as a non-blind scenario with very low watermark to noise ratio (WNR). Thus, the analysis of the detection statistics for the blind scenario will be similar, and we will observe similar trend. Consequently, for the purpose of comparing the robustness of the joint TDMA and CDMA fingerprint modulation with that of the CDMA based modulation, our assumption of the non-blind detection scenario is justified.

7.2.2 Performance Criteria

To measure the robustness of the joint TDMA and CDMA fingerprint modulation scheme against collusion attacks, we adopt the commonly used criteria in the literature [58], [63]: the probability of capturing at least one colluder (P_d), and the probability of accusing at least one innocent user (P_{fp}).

In this chapter, we assume that the colluders collude under the fairness constraint, i.e., all colluders share the same risk and are equally likely to be detected. Assume that A and B are two non-overlapping subgroups of colluders, and SC_A and SC_B are the sets containing the indices of the colluders in A and B, respectively. $SC_A \cap SC_B = \emptyset$, and we define the fairness parameter $FP(SC_A, SC_B)$ as

$$\begin{aligned} FP(SC_A, SC_B) &\triangleq \frac{F_d(SC_A)}{F_d(SC_B)}, \\ \text{where} \quad F_d(SC_A) &= \frac{\sum_{i \in SC_A} I[i \in \widehat{SC}]}{|SC_A|} \\ \text{and} \quad F_d(SC_B) &= \frac{\sum_{i \in SC_B} I[i \in \widehat{SC}]}{|SC_B|}. \end{aligned} \tag{7.14}$$

In (7.14), $I[\cdot]$ is the indication function, $|SC_A|$ and $|SC_B|$ are the number of colluders in SC_A and SC_B, respectively, and $\widehat{SC}$ is the estimated colluder set output by the detector. If $FP(SC_A, SC_B) \approx 1$ for any (SC_A, SC_B) with $SC_A \cap SC_B = \emptyset$, then the collusion attack is fair and each colluder is equally likely to be detected. If $FP(SC_A, SC_B) \gg 1$ or $FP(SC_A, SC_B) \ll 1$ for some pair of (SC_A, SC_B), some colluders are more likely to be detected than others and the collusion attack is not fair.

7.2.3 Statistical Analysis of the Probability of Detection

The work in [62] provided detailed analysis of the probability of detection for the CDMA based fingerprint modulation, and it is not repeated here. In this section, we focus on the analysis of the joint TDMA and CDMA fingerprint modulation.

At the detector's side, given the lth part of the colluded copy $\mathbf{V}_l$, the detector

first extracts the fingerprint

$$\begin{aligned}\mathbf{Y}_l &= \frac{1}{K_l}\sum_{i=[i_1,\cdots,i_L]\in SC_l}\sqrt{p_{1,l}}\cdot\mathbf{a}_l^{i_1}+\sqrt{p_{2,l}}\cdot\mathbf{a}_l^{i_1,i_2}+\cdots+\sqrt{p_{l,l}}\cdot\mathbf{a}_l^{i_1,\cdots,i_l}+\mathbf{n}_l \\ &= \frac{1}{K_l}\sum_{k=1}^{l}\sum_{i_1,\cdots,i_k}K_l^{i_1,\cdots,i_k}\sqrt{p_{k,l}}\cdot\mathbf{a}_l^{i_1,\cdots,i_k}+\mathbf{n}_l. \qquad (7.15)\end{aligned}$$

In (7.15), $K_l^{i_1,\cdots,i_k}\triangleq\sum_{j=[j_1,\cdots,j_L]\in SC_l}I[j_1=i_1,\cdots,j_k=i_k]$ is the number of colluders in SC_l that are in the subregion represented by $[i_1,\cdots,i_k]$, $K_l=\sum_{i_1,\cdots,i_k}K_l^{i_1,\cdots,i_k}$ is the number of colluders in SC_l, and $\mathbf{n}_l$ is the additive noise that the colluders add to the colluded copy $\mathbf{V}_l$ to further hinder the detection performance. In this chapter, for simplicity, we assume that the additive noise $\mathbf{n}_l$ are i.i.d. and follow Gaussian distribution $\mathcal{N}(0,\sigma_n^2)$.

Following the statistical analysis in [62], given $\mathbf{P}^{Joint}$ as in (6.11) and $\{N_l\}_{l=1,\cdots,L}$ as in (6.16), we can show that at level l, $T^{i_1,\cdots,i_l}$ can be approximated by a normal distribution

$$\begin{aligned}T^{i_1,\cdots,i_l} &\sim \mathcal{N}\left(\mu^{i_1,\cdots,i_l},\sigma_n^2\right), \\ \text{where}\quad \mu^{i_1,\cdots,i_l} &= \left\{\sum_{r=l}^{L}\sqrt{p_{l,r}}\left(\frac{K_r^{i_1,\cdots,i_l}}{K_r}\cdot N_r\right)\right\}\sigma_W\Big/\sqrt{\sum_{t=l}^{L}N_t}. \qquad (7.16)\end{aligned}$$

In (7.16), σ_W^2 is the variance of $\{\mathbf{a}\}$, and σ_n^2 is the variance of the additive noise $\mathbf{n}_l$. The analysis of P_d and P_{fp} is similar to that in [62] and is omitted.

We then analyze the robustness of the joint TDMA and CDMA fingerprint modulation scheme under the interleaving based collusion attacks. From (7.16), if σ_n^2 and the thresholds used during detection $\{h_l\}$ are fixed, comparing the probability of detection is equivalent to comparing the means of the detection statistics. Therefore, in this paper, we focus on the analysis of the detection statistics' means.

We first analyze the means of the detection statistics when detecting guilty regions at upper levels in the fingerprint tree. Under the interleaving based col-

lusion attacks, we consider a colluder $\mathbf{u}^{(i)}$ where $i = [i_1, \cdots, i_k, \cdots i_L] \in SC_l$. For a guilty node $[i_1, \cdots, i_k]$ at upper level $k < l$ in the tree, we have $K_l^{i_1,\cdots,i_k} > 0$. Consequently, from (7.16), even if $i \notin SC_k$ and $K_k^{i_1,\cdots,i_k} = 0$, we still have

$$\mu^{i_1,\cdots,i_k} \geq \sqrt{p_{k,l}} \left(\frac{K_l^{i_1,\cdots,i_k}}{K_l} \cdot N_l \right) \sigma_W \Big/ \sqrt{\sum_{t=l}^{l} N_t} > 0. \tag{7.17}$$

Therefore, in the joint TDMA and CDMA fingerprint modulation scheme, the guilty region $[i_1, \cdots, i_k]$ at the upper level of the tree can be detected even under the interleaving based collusion.

We then analyze the means of the detection statistics when detecting guilty regions at lower levels in the fingerprint tree. Assume that the depth of the fingerprint tree is L. First, we consider a Type I interleaving based collusion attacks where colluders in subgroup SC_{L-1} and colluders in subgroup SC_L are under different branches of the tree and $SC_{L-1} \cap SC_L = \emptyset$. In addition, for any $[i_1, \cdots, i_{L-1}, i_L] \in SC_{L-1}$ and $[j_1, \cdots, j_{L-1}, j_L] \in SC_L$, $[i_1, \cdots, i_{L-1}] \neq [j_1, \cdots, j_{L-1}]$. The example shown in Figure 6.5 belongs to this type of collusion attacks.

We consider two colluder $\mathbf{u}^{(i)}$ and $\mathbf{u}^{(j)}$ where $i = [i_1, \cdots, i_L] \notin SC_L$ and $j = [j_1, \cdots, j_L] \in SC_L$. At level L in the tree, for colluder $\mathbf{u}^{(i)}$ who is not in the subgroup SC_L, $K_L^{i_1,\cdots,i_L} = 0$; while $K_L^{j_1,\cdots,j_L} > 0$ for colluder $\mathbf{u}^{(j)}$ who is in the subgroup SC_L. Therefore, from (7.16), at level L in the tree, the means of the detection statistics for user $\mathbf{u}^{(i)}$ and user $\mathbf{u}^{j)}$ are

$$\mu^{i_1,\cdots,i_L} = 0 \quad \text{and} \quad \mu^{j_1,\cdots,j_L} = \frac{\sqrt{\rho_{L,L} N_L \ \sigma_W^2}}{K_L} > 0 \tag{7.18}$$

respectively. Consequently, in the joint TDMA and CDMA fingerprint modulation scheme, under the Type I interleaving based collusion attacks, the colluders in the subgroup SC_L are more likely to be detected than other colluders. So the Type I interleaving based collusion attacks are not fair collusion attacks.

Then, we consider a Type II interleaving based collusion attacks where $SC_L = SC$ but $SC_l \subset SC$ for some $l < L$. As an example, we consider the scenario where $SC_{L-1} \subset SC_L = SC$, and for any $i \in SC_{L-1}$ and $j \in SC_L \setminus SC_{L-1}$, $[i_1, \cdots, i_{L-1}] \neq [j_1, \cdots, j_{L-1}]$.[1] This corresponds to the scenario where all colluders participate in the generation of $\mathbf{V}_L$, but some of the colluders do not participate in the generation of $\mathbf{V}_{L-1}$. Take the fingerprint tree in Figure 6.2 as an example, if user $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \mathbf{u}^{(4)}$ and $\mathbf{u}^{(7)}$ are the colluders, and if the colluders choose $SC_1 = \{7\}, SC_2 = \{4\}$ and $SC_3 = \{1, 2, 4, 7\}$, then this is a Type II interleaving based collusion attack.

We consider two colluders $\mathbf{u}^{(i)}$ and $\mathbf{u}^{(j)}$, where $i \in SC_{L-1},\ i \in SC_L$ and $j \notin SC_{L-1},\ j \in SC_L$. Under the Type II interleaving based collusion, for colluder $\mathbf{u}^{(i)}$

$$K_{L-1}^{i_1,\cdots,i_{L-1}} > 0 \quad \text{and} \quad K_L^{i_1,\cdots,i_{L-1}} > 0, \tag{7.19}$$

and for colluder $\mathbf{u}^{(j)}$,

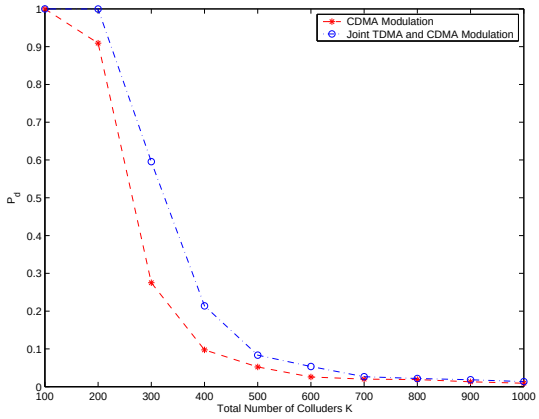
$$K_{L-1}^{i_1,\cdots,i_{L-1}} = 0 \quad \text{and} \quad K_L^{i_1,\cdots,i_{L-1}} > 0. \tag{7.20}$$

Consequently, from (7.16), when detecting guilty regions at level $L-1$ in the tree,

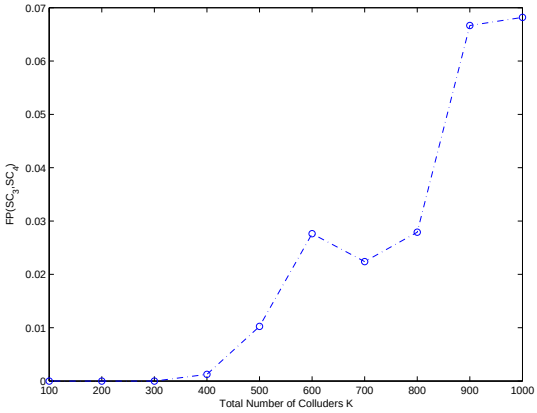
$$\begin{aligned} \mu^{i_1,\cdots,i_{L-1}} &= \frac{K_{L-1}^{i_1,\cdots,i_{L-1}}}{K_{L-1}} \sqrt{\frac{p_{L-1,L-1}}{N_{L-1}+N_L}} \cdot N_{L-1}\sigma_W \\ &\quad + \frac{K_L^{i_1,\cdots,i_{L-1}}}{K_L} \sqrt{\frac{p_{L-1,L}}{N_{L-1}+N_L}} \cdot N_L\sigma_W, \\ \text{and} \quad \mu^{j_1,\cdots,j_{L-1}} &= \frac{K_L^{j_1,\cdots,j_{L-1}}}{K_L} \sqrt{\frac{p_{L-1,L}}{N_{L-1}+N_L}} \cdot N_L\sigma_W. \end{aligned} \tag{7.21}$$

Since $K_{L-1}^{(i)} > 0$, we have $\mu^{j_1,\cdots,j_{L-1}} < \mu^{i_1,\cdots,i_{L-1}}$ in almost all cases. So in the joint TDMA and CDMA fingerprint modulation, under the Type II interleaving based collusion attacks, the colluders in SC_{L-1} are more likely to be detected than other colluders. Consequently, the Type II interleaving based collusion attacks are not fair collusion attacks either.

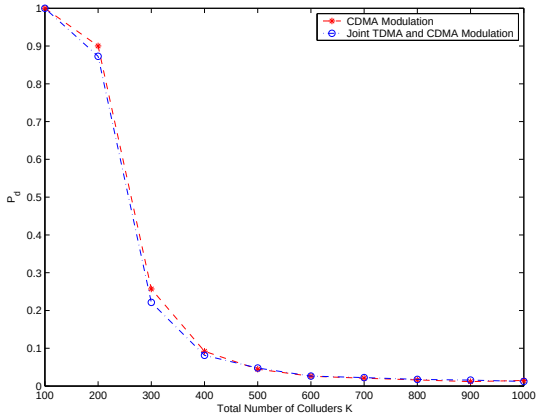
[1]For two sets A and B where $A \supseteq B$, $A \setminus B \triangleq \{i : i \in A, i \notin B\}$.



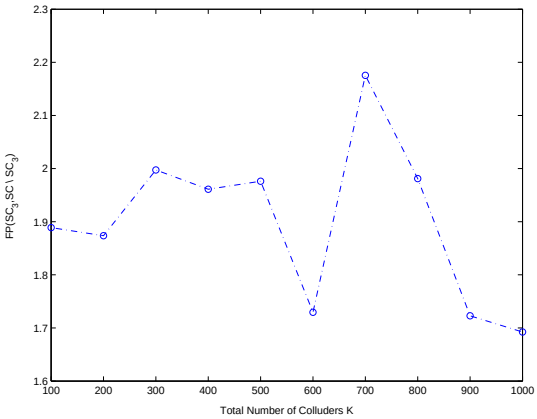
(a) P_d



(b) $FP(SC_{L-1}, SC_L)$



(c) P_d



(d) $FP(SC_{L-1}, SC \setminus SC_{L-1})$

Figure 7.3: Performance of the joint TDMA and CDMA fingerprint modulation scheme under interleaving based collusion attacks. $L = 4$, $[D_1, D_2, D_3, D_4] = [4, 5, 5, 100]$ and $[\rho_1, \rho_2, \rho_3, \rho_4] = [1/6, 1/6, 1/6, 1/2]$. $N = 10^6$, $\sigma_n^2 = 2\sigma_W^2$ and $P_{fp} = 10^{-2}$. $p = 0.95$. Top: under Type I interleaving based collusion attacks, bottom: under Type II interleaving based collusion attacks.

7.2.4 Simulation Results

Resistance to Interleaving Based Collusion Attacks

Figure 7.3 shows the simulation results of the joint TDMA and CDMA fingerprint modulation scheme under both types of interleaving based collusion attacks. Our simulation is set up as follows. For the tested video sequences, the number of embeddable coefficients is in the order of 10^6 per second. So we choose $N = 10^6$ and assume that there are a total of $M = 10^4$ users. Following the tree based fingerprint design in [61], [62], we consider a symmetric tree structure with $L = 4$ levels, $[D_1, D_2, D_3, D_4] = [4, 5, 5, 100]$ and $[\rho_1, \rho_2, \rho_3, \rho_4] = [1/6, 1/6, 1/6, 1/2]$. In our simulations, the basis fingerprints $\{\mathbf{a}\}$ in the fingerprint tree follow Gaussian distribution $\mathcal{N}(0, \sigma_W^2)$ with $\sigma_W^2 = 1/9$. In the joint TDMA and CDMA fingerprint modulation scheme, for simplicity, we let $p_{2,2} = \cdots = p_{L,L} = p$ for the matrix $\mathbf{P}^{Joint}$ in (6.11) and choose $p = 0.95$ for the above fingerprint tree structure. A smaller value of p should be used if L is larger or the total number of nodes at the upper $L - 1$ levels in the tree is larger.

At the attackers' side, we consider the most effective collusion pattern on the tree based fingerprint design, where colluders are from all the 100 subgroups at level 3. We assume that each of the 100 subgroups has the same number of colluders. In the joint TDMA and CDMA fingerprint modulation scheme, for both types of interleaving based collusion attacks, we choose different subgroups of colluders as $SC_1 = \{i = [i_1, i_2, i_3, i_4] \in SC : i_1 = 1\}$, $SC_2 = \{i = [i_1, i_2, i_3, i_4] \in SC : i_1 = 2\}$ and $SC_3 = \{i = [i_1, i_2, i_3, i_4] \in SC : i_1 = 3\}$. In the Type I interleaving based collusion attacks, we choose $SC_4 = SC \setminus SC_3$. In the Type II interleaving based collusion attacks, $SC_4 = SC$. In the CDMA based fingerprint modulation scheme, similarly, we assume that colluders are from all the 100 subgroups at level 3 in the

tree, and each subgroup at level 3 in the tree has equal number of colluders. In the CDMA based fingerprint modulation, the colluders cannot distinguish fingerprints at different levels, and they apply the *pure averaging collusion attack* where $SC_1 = \cdots = SC_L = SC$. Also, we choose $\sigma_n^2 = 2\sigma_W^2$ for all collusion attacks in (7.16). Other values of σ_n^2 give the same trend and are not shown here.

Figure 7.3 (a) and (b) show the simulation results of the Type I interleaving base collusion, while Figure 7.3 (c) and (d) show the simulation results of the Type II interleaving based collusion.

In Figure 7.3 (a) and (c), given the total number of colluders K, we compare P_d of the joint TDMA and CDMA fingerprint modulation under the interleaving based collusion attacks with that of the CDMA based fingerprint modulation scheme under the pure averaging collusion attacks. As an example, we fix P_{fp} as 10^{-2}. From Figure 7.3 (a) and (c), the performance of the joint TDMA and CDMA fingerprint modulation under the interleaving based collusion is approximately the same or even better than that of the CDMA based fingerprint modulation under the pure averaging collusion attacks.

Figure 7.3 (b) and (d) show the fairness parameters of the two types of interleaving based collusion attacks in the joint TDMA and CDMA fingerprint modulation. From Figure 7.3 (b), under the Type I interleaving based collusion attacks, $FP(SC_{L-1}, SC_L) \ll 1$. Therefore, the colluders in the subgroup SC_L are much more likely to be detected than those in SC_{L-1}, which is in agreement with our analysis. From 7.3 (d), under the Type II interleaving based collusion attacks, $FP(SC_{L-1}, SC \setminus SC_{L-1}) \approx 1.9$, and the colluders in the subgroup SC_{L-1} are more likely to be detected than other colluders, which is also consistent with the analysis.

Therefore, the performance of the joint TDMA and CDMA fingerprint modu-

lation scheme under the interleaving based collusion attacks is approximately the same as, and may be even better than, that of the CDMA fingerprint modulation scheme under the pure averaging collusion attacks. Furthermore, we have shown that neither of the two types of interleaving based collusion attacks are fair in the joint TDMA and CDMA fingerprint modulation scheme, and some colluders are more likely to be captured than others. Consequently, to guarantee the fairness of the collusion attacks, the colluders cannot use the interleaving based collusion attacks in the joint TDMA and CDMA fingerprint modulation.

Resistance to the Pure Averaging Collusion Attacks

In this section, we study the detection performance of the joint TDMA and CDMA fingerprint modulation under the pure averaging collusion attacks where $SC_1 = SC_2 = \cdots = SC_L = SC$. We compare the detection performance of the Joint TDMA and CDMA fingerprint modulation with that of the CDMA fingerprint modulation. In both fingerprint modulation schemes, all colluders have equal probability of detection under this type of collusion, and the pure averaging attacks are fair collusion attacks. The simulation setup is the same as in the previous section and Figure 7.4 shows the simulation results. We consider two possible collusion patterns. In the first one, we assume that one region at level 1 is guilty and it has two guilty sub-regions at level 2. For each of the two guilty regions at level 2, we assume that all its five children at level 3 are guilty and colluders are present in 10 out of 100 subgroups at level 3. This collusion pattern corresponds to the case where the fingerprint tree matches the hierarchical relationship among users. In the second one, we assume that all the 100 subgroups at level 3 are guilty, and this collusion pattern happens when the fingerprint tree does not reflect the

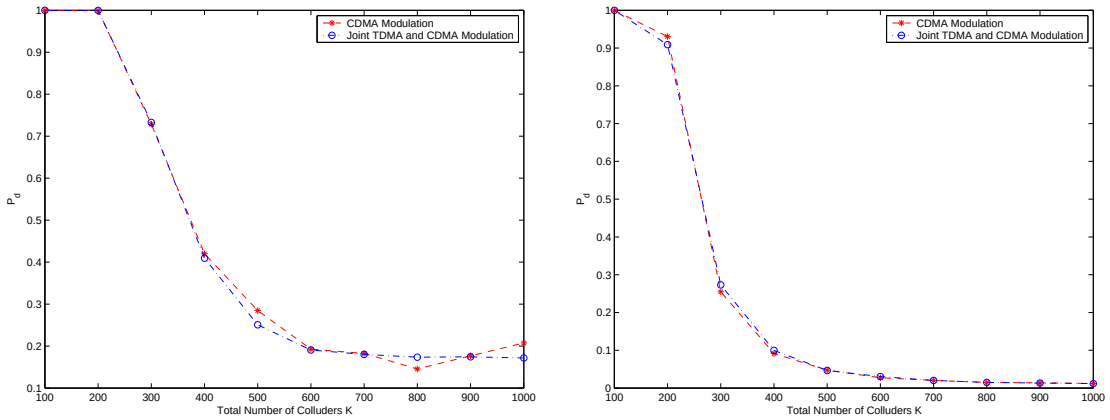


Figure 7.4: P_d of the joint TDMA and CDMA fingerprint modulation scheme under the pure average attacks. $L = 4$, $[D_1, D_2, D_3, D_4] = [4, 5, 5, 100]$ and $[\rho_1, \rho_2, \rho_3, \rho_4] = [1/6, 1/6, 1/6, 1/2]$. $N = 10^6$, $\sigma_n^2 = 2\sigma_W^2$ and $P_{fp} = 10^{-2}$. $p = 0.95$. Left: colluders are from 10 subgroups at level 3 in the tree, right: colluders are from all the 100 subgroups at level 3 in the tree.

real hierarchical relationship among users. We further assume that each guilty subgroup at level 3 has the same number of colluders in both collusion patterns.

From Figure 7.4, the joint TDMA and CDMA fingerprint modulation scheme has approximately the same performance as the CDMA based fingerprint modulation scheme under the pure averaging collusion attacks. Both fingerprint modulation schemes perform better when the fingerprint tree design matches the hierarchical relationship among users and the colluders are present in fewer subgroups in the tree.

To summarize, under the constraint that all colluders share the same risk and have equal probability of detection, the joint TDMA and CDMA fingerprint modulation has approximately identical performance as the CDMA based fingerprint modulation, and the embedded fingerprints in the three secure fingerprint distribution schemes have the same collusion resistance.

7.3 Fingerprint Drift Compensation

In both the general fingerprint multicast scheme and the joint fingerprint design and distribution scheme, the video encoder and the decoder use the reconstructed *unfingerprinted* and *fingerprinted* copies, respectively, as references for motion compensation. The difference, which is the embedded fingerprint, will propagate to the next frame. Fingerprints from different frames will accumulate and cause the perceptual quality degradation of the reconstructed frames at the decoder's side. A drift compensation signal, which is the embedded fingerprint in the reference frame(s) with motion, has to be transmitted to each user. It contains confidential information of the embedded fingerprint in the reference frame(s) and is unique to each user. Therefore, it has to be transmitted seamlessly with the host signal to the decoder through unicast channels. Since the embedded fingerprint propagates not only to the embeddable coefficients but also to the non-embeddable ones, fully compensating the drifted fingerprint will significantly increase the communication cost.

To reduce the communication overhead introduced by full drift compensation, we propose to compensate the drifted fingerprint that propagates to the embeddable coefficients only and ignore the rest. Shown in Figure 7.5 is the fingerprint drift compensation scheme in the general fingerprint multicast scheme for video on demand applications. The one in the joint fingerprint design and distribution scheme is similar and omitted. The calculation of the drift compensation signal is similar to that in [26]. Step 3 in the fingerprint embedding and distribution process is modified as follows. For each DCT coefficient, if it is not embeddable, it is variable length coded with other non-embeddable coefficients. Otherwise, first, it is inversely quantized. *Then for each user, the corresponding fingerprint compo-*

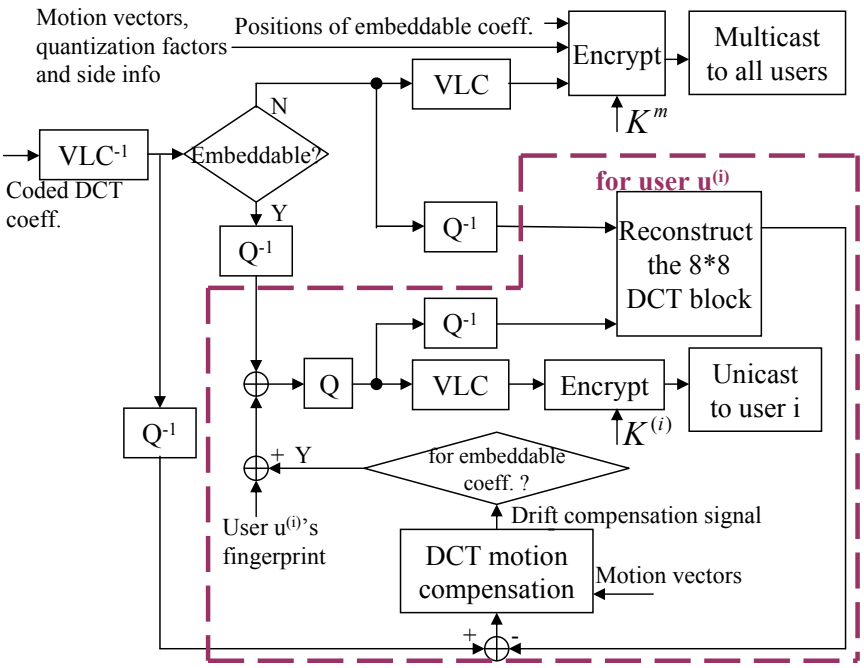


Figure 7.5: The proposed fingerprint drift compensation scheme in the general fingerprint multicast for VoD applications.

nent is embedded, the corresponding drift compensation component is added, and the resulting fingerprinted and compensated coefficient is quantized and variable length coded with other fingerprinted and compensated coefficients.

In Table 7.3, we compare the quality of the reconstructed sequences at the decoder's side in three scenarios: $PSNR_f$ is the average PSNR of the reconstructed frames with full drift compensation; $PSNR_n$ is the average PSNR of the reconstructed frames without drift compensation; and $PSNR_p$ is the average PSNR of the reconstructed frames in the proposed drift compensation scheme. Compared with the reconstructed frames with full drift compensation, the reconstructed frames without drift compensation have an average of $1.5 \sim 2dB$ loss in PSNR, and those using the proposed drift compensation have an average of $0.5dB$ loss. Therefore, the proposed drift compensation scheme improves the quality of the

Table 7.3: Perceptual quality of the reconstructed frames at the decoder's side at bit rate $R = 1.3bpp$.

Sequence	$PSNR_f(dB)$	$PSNR_n(dB)$	$PSNR_p(dB)$
miss america	44.89	42.73	44.31
carphone	40.45	38.05	39.88
flower	31.53	30.01	30.92

reconstructed frames at the decoder's side without extra communication overhead.

7.4 Chapter Summary

In this chapter, we have analyzed the performance of the two fingerprint multicast schemes proposed in Chapter 6, including the bandwidth efficiency, the robustness of the embedded fingerprints, and the perceptual quality of the reconstructed sequence at the decoder's side.

We first analyzed the bandwidth efficiency of the two fingerprint multicast schemes. Compared with the pure unicast scheme, the general fingerprint multicast scheme reduces the communication cost by 48% to 84%, depending on the total number of users and the characteristics of sequences; and the joint fingerprint design and distribution scheme reduces the bandwidth requirement by 57% to 87%, depending on the number of users, the characteristics of sequences, and network and computation constraints.

If we compare the three distribution schemes: the pure unicast scheme, the general fingerprint multicast scheme, and the joint fingerprint design and distribution scheme, the pure unicast scheme is preferred when there are only a few users

in the system, e.g., around ten or twenty users; and the other two schemes should be used when there are a large number of users, e.g., thousands of users. Compared with the general fingerprint multicast scheme, the joint fingerprint design and distribution scheme further improves the bandwidth efficiency by increasing the complexity of the underlying network and that of the receivers. Therefore, for sequences that have fewer embeddable coefficients, e.g., sequence "miss america", the general fingerprint multicast scheme is preferred to achieve the bandwidth efficiency at a low computation cost. For sequences with much more embeddable coefficients, e.g., sequence "flower", the joint fingerprint design and distribution scheme is recommended to reduce the communication cost under network and computation constraints.

We then analyzed the collusion resistance of the embedded fingerprints in different schemes. We have shown that with the joint TDMA and CDMA fingerprint modulation, although the colluders can still apply the interleaving based collusion, some colluders have larger probability to be detected than the others. To guarantee that all colluders share the same risk and have equal probability of being captured, the colluders can only apply the pure averaging collusion, under which the proposed joint TDMA and CDMA fingerprint modulation has approximately identical performance as the CDMA based fingerprint modulation scheme.

Finally, we analzyed the perceptual quality of the reconstructed sequences at the receiver's side. We have shown that the proposed fingerprint drift compensation scheme improves PSNR of the reconstructed frames by an average of $1 \sim 1.5\ dB$ without increasing the communication cost.

Chapter 8

Conclusions and Future Research

In this thesis, we have examined and explored various aspects of multimedia fingerprinting, including the analysis of collusion resistance as well as secure fingerprint multicast for video streaming.

We first investigated order statistics based nonlinear collusion attacks and analyzed their effectiveness in defeating multimedia fingerprinting systems. We also analyzed the detection performance of several commonly used detectors in the literature and compared their performance under nonlinear collusion attacks. To improve the performance of the detection statistics under collusion attacks, during fingerprint detection, we utilized the statistical features of the extracted fingerprints and proposed a preprocessing technique specifically for collusion scenario. We showed that the preprocessing techniques improve the collusion resistance of multimedia fingerprinting systems.

We then investigated collusion attacks on scalable fingerprinting systems, where users received copies of different quality due to bandwidth and computation constraints. We first analyzed the fairness constraints on the collusion attacks, which requires all colluders share the same risk and have equal probability of detection.

We then investigated the tradeoff between the probability of detection and the perceptual quality during collusion. Finally, we analyzed the collusion resistance of the scalable fingerprinting systems for different applications with different requirements, and provided the lower and upper bounds of K_{max}.

We also investigated the traitors within traitors problem in multimedia fingerprinting, where selfish colluders process their received copies before multiuser collusion to further reduce their own probability of detection. We explored the possible strategy by those selfish colluders, analyzed their performance, and investigated the optimal pre-collusion processing technique for selfish colluders to minimize their risk of being captured under the quality constraints. For other colluders who wish to protect their own interest, we also provided preliminary countermeasures to detect and prevent such pre-collusion processing.

In this thesis, we also address the secure distribution of uniquely fingerprinted copies for video streaming applications with stringent latency constraints. We proposed two secure fingerprint multicast schemes: the general fingerprint multicast scheme that can be used with most spread spectrum embedding based fingerprinting systems, and the joint fingerprint design and distribution scheme that explores the special structure of fingerprint design to further reduce the communication cost. We compared their performance, including the communication cost and the robustness against collusion attacks, and analyzed the tradeoff between the bandwidth efficiency and computation complexity. We also proposed a fingerprint drift compensation scheme to improve the quality of the reconstructed sequences at the decoder's side without extra communication overhead.

Digital fingerprinting and traitor tracing in multimedia forensics is at its young age, and there are many more interesting research directions that need to be further

investigated.

First, our current work on secure fingerprint multicast considered a simple scenario where all users receive copies of the same quality. As we have pointed out in Chapter 4, scalability is usually required for video coding and transmission to address the heterogeneity of networks as well as the variation of users' computation capability. Consequently, it is important to investigate secure fingerprint multicast for scalable video fingerprinting and coding, which is more realistic and practical. In addition, observing that both Internet and wireless networks change dynamically over time, the service provider is obliged to adjust the distribution schemes according to the bandwidth fluctuations, and it is crucial to have flexible secure fingerprint multicast schemes that can address both the heterogeneity and the dynamically changing nature of networks.

In addition, in our work, we assumed that the networks are error-free for simplicity and considered simple scenarios where users receive bit streams correctly. In practice, data transmitted through networks suffer from bit errors and packet losses, especially for wireless networks. For video applications, the extensive use of predictive and variable-length coding in video compression techniques renders the compressed bit streams especially vulnerable to transmission errors, and the sender has to undergo a channel encoding stage to protect compressed video from transmission errors. Therefore, it is important to investigate the error control and error concealment mechanisms in secure fingerprint multicast, and examine the Quality of Service (QoS) management for secure distribution of fingerprinted copies in video streaming applications.

Finally, in digital rights management systems, traditional cryptography and multimedia forensics are tightly connected with each other, and neither can stand

alone. It will be fruitful to investigate the combination of multimedia forensics and traditional cryptography, and examine the benefit of this combination in order to complement each other. This investigation will lead to the general framework of digital rights management for multimedia, and provide a basis for the design of multimedia security systems.

BIBLIOGRAPHY

[1] J. Apostolopoulos, W. Tan, and S. Wee. Video streaming: Concepts, algorithms, and systems. Technical Report HPL-2002-260, HP Labs, 2002.

[2] S. Baker, R. Gross, I. Matthews, and T. Ishikawa. Lucas-kanade 20 years on: A unifying framework. *To Appear in the International Journal of Computer Vision*, 2004.

[3] D. Boneh and J. Shaw. Collusion-secure fingerprinting for digital data. *IEEE Trans. Information Theory*, 44(5):1897–1905, Sept. 1998.

[4] I. Brown, C. Perkins, and J. Crowcroft. Watercasting: Distributed watermarking of multicast media. *Network Group Commuincation, Pisa, Italy*, pages 286–300, Nov 1999.

[5] L. Camp. First principles of copyright for DRM design. *IEEE Internet Computing*, pages 59–65, May/June 2003.

[6] G. Caronni and C. Schuba. Enabling hierarchical and bulk-distribution for watermarked content. *The* 17^{th} *Annual computer Security Applications Conference, New Orleans, LA*, Dec. 2001.

[7] R. Chalmers and K. Almeroth. Modeling the branching characteristics and efficiency gains in global multicast trees. *IEEE InfoCom 2001*, 1:449–458, April 2001.

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

[9] B. Chor, A. Fiat, and M. Manor. Tracing traitors. *IEEE Trans. Information Theory*, 46(3):893–910, May 2000.

[10] H. Chu, L. Qiao, and K. Nahrstedt. A secure multicast protocol with copyright protection. *ACM SIGCOMM Computer Communications Review*, 32(2):42–60, April 2002.

[11] J. Chuang and M. Sirbu. Pricing multicast communication: A cost-based approach. *Telecommunication Systems*, 17(3):281–297, 2001.

[12] I. Cox, J. Bloom, and M. Miller. *Digital Watermarking: Principles and Practice*. Morgan Kaufmann, 2001.

[13] I. Cox, J. Killian, F. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Trans. on Image Processing*, 6(12):1673–1687, Dec. 1997.

[14] I. Cox and J.P Linnartz. Some general methods for tampering with watermaking. *IEEE J. Select. Areas Commun.*, 16(4):587–593, May 1998.

[15] S. Craver, B. Liu, and W. Wolf. Histo-cepstral analysis for reverse-engineering watermarks. 38*th Conference on Information Sciences and Systems*, pages 824–826, March 2004.

[16] S. Craver, N. Memon, B. Yeo, and M. Yeung. Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications. *IEEE J. Select. Areas Commun.*, 16(4):572–586, May 1998.

[17] G. Dantzig. *Linear Programming and Extensions*. Princeton University Press, 1963.

[18] H. A. David. *Order Statistics*. New York: John Wiley and Son, 2nd edition, 1981.

[19] J. Dittmann, P. Schmitt, E. Saar, J. Schwenk, and J. Ueberberg. Combining digital watermarks and collusion secure fingerprints for digital images. *SPIE Journal of Electronic Imaging*, 9(4):456–467, Oct. 2000.

[20] 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*, 1592:140–149, 2001.

[21] J. Wen et al. A format-compliant configurable encryption framework for access control of video. *IEEE Trans. on Circuits & Systems for Video Technology*, 12(6):545–557, June 2002.

[22] A. Fiat and T. Tassa. Dynamic tracing traitors. *Advances in Cryptology – Crypto '99, Lecture Notes in Computer Science*, 1666:354–371, 1999.

[23] W. Gander and W. Gautschi. Adaptive quadrature - revised. *BIT 40(1)*, pages 84–101, March 2000.

[24] H. Gou and M. Wu. Data hiding in curves for collusion resistant digital fingerprinting. *IEEE Int. Conf. on Image Processing*, Oct. 2004.

[25] H. Gou and M. Wu. Fingerprinting curves. *Int. Workshop on Digital Watermarking*, Oct. 2004.

[26] F. Hartung and B. Girod. Watermarking of uncompressed and compressed video. *Signal Processing*, 66(3):283–301, 1998.

[27] F. Hartung and F. Ramme. Digital rights management and watermarking of multimedia content for M-commerce applications. *IEEE Communications Magazine*, pages 78–84, Nov. 2000.

[28] F. Hartung, J. Su, and B. Girod. Spread spectrum watermarking: Malicious attacks and counterattacks. *Proc. SPIE, Security and Watermarking of Multimedia Contents, Electronic Imaging*, pages 147–158, Jan. 1999.

[29] S. He and M. Wu. Performance study of ecc-based collusion resistant multimedia fingerprinting. *Proceedings of the 38th CISS*, pages 827–832, March 2004.

[30] P. Judge and M. Ammar. Whim: Watermarking multicast video with a hierarchy of intermediaries. *Proc. NOSSDAC, Chapel Hill, NC*, June 2000.

[31] P. Judge and M. Ammar. Security issues and solutions in multicast content distribution: A survey. *IEEE Network*, pages 30–36, Jan./Feb. 2003.

[32] J. Killian, T. Leighton, L. R. Matheson, T. G. Shamoon, R. Tajan, and F. Zane. Resistance of digital watermarks to collusive attacks. Technical Report TR-585-98, Department of Computer Science, Princeton Univ., 1998.

[33] D. Konstantas and D. Thanos. Commercial dissemination of video over open networks: issues and approaches. Technical report, Object Systems Group, Center Universitaire d'Informatique of Univ. of Geneva, 2000.

[34] D. Kundur and K. Karthik. Video fingerprinting and encryption principles for digital rights management. *Proceedings of the IEEE*, 92(6):918– 932, June 2004.

[35] G. C. Langelaar, I. Setyawan, and R. Lagendijk. Watermarking digital image and video data: A state-of-the-art overview. *IEEE Signal Processing Mag.*, 17(5):20–46, Sept. 2000.

[36] C. Lin and S. Chang. Image watermarking for tamper detection. *IEEE Int. Conf. on Image Processing*, 1998.

[37] C. Lin and S. Chang. Semi-fragilewatermarking for authenticating jpeg visual content. *SPIE International Conf. on Security and Watermarking of Multimedia Contents II (EI00)*, 3971, 2000.

[38] J. Lubin, J. Bloom, and H. Cheng. Robust, content-dependent, high-fidelity watermark for tracking in digital cinema. *Security and Watermarking of Multimedia Contents V, Proc. SPIE*, 5020, 2003.

[39] A. Menezes, P. Oorschot, and S. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.

[40] R. Parviainen and R. Parnes. Enabling hierarchical and bulk-distribution for watermarked content. *Proceedings of the IFIP TC6/TC11 International Conference on Communications and Multimedia Security Issues*, 192, May 2001.

[41] S. Paul. *Multicast on the Internet and its application*. Kluwer Academic Publisher, 1998.

[42] F. Petitcolas, R. Anderson, and M. Kuhn. Attacks on copyright marking systems. 2*nd Workshop on Info. Hiding, Lecture Notes in Computer Science*, pages 218–238, April 1998.

[43] F. Petitcolas, R. Anderson, and M. Kuhn. Information hiding - a survey. *Proc. IEEE*, 87:1062–1078, July 1999.

[44] B. Pfitzmann and M. Waidner. Asymmetric fingerprinting for larger collusions. 4*th ACM Conference on Computer and Communication Security*, pages 151–160, 1997.

[45] C. Pfleeger. *Security in Computing*. Prentice Hall PTR, 1996.

[46] A. Piva, F. Bartolini, and M. Barni. Managing copyright in open networks. *IEEE Internet Computing*, pages 18–26, May/June 2002.

[47] C. Podilchuk and W. Zeng. Image adaptive watermarking using visual models. *IEEE J. Select. Areas in Commun.*, 16(4):525–540, May 1998.

[48] H. V. Poor. *An Introducton to Signal Detection and Estimation*. Springer Verlag, 2nd edition, 1999.

[49] L. Qiao and K. Nahrstedt. A new algorithm for mpeg video encryption. *Proc. Int. Conf. on Imaging Science, Systems and Technology*, pages 21–29, June 1997.

[50] M. Schneider and S. Chang. A robust content based dig. signature for image authentication. *IEEE Int. Conf. on Image Processing*, 1996.

[51] S. Siwek. Copyright industries in the u.s. economy, the 2004 report. *International Intellectural Property Alliance* (IIPA), Oct. 2004.

[52] H. Stone. Analysis of attacks on image watermarks with randomized coefficients. Technical Report 96-045, NEC Research Instistute, 1996.

[53] D. Storck. A new approach to integrity of digital images. *IFIP Conf. on Mobile Communication*, 1996.

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

[55] K. Su, D. Kundur, and D. Hatzinakos. Statistical invisibility for collusion-resistant digital video watermarking. *to appear in IEEE Tran. on Multimedia*, 2004.

[56] M. Swason, B. Zhu, and A. Tewfik. Multiresolution scene-based video watermarking using perceptual models. *IEEE Journal on Sel. Area in Comm.*, 16(4):540–550, May 1998.

[57] L. Tang. Methods for encrypting and decrypting MPEG video data efficiently. *Proceedings of 4th* ACM *Inter. Conf. on Multimedia,*, pages 219–229, 1996.

[58] W. Trappe, M. Wu, Z. J. Wang, and K. J. R. Liu. Anti-collusion figerprinting for multimedia. *IEEE Trans. on Signal Processing*, 51(4):1069–1087, April 2003.

[59] U. Varshney. Multicast over wireless networks. *Communications of the ACM*, 45:31–37, Dec. 2002.

[60] Y. Wang, J. Ostermann, and Y. Zhang. *Video Processing and Communications*. Prentice Hall, 1st edition, 2002.

[61] Z. J. Wang, M. Wu, W. Trappe, and K. J. R. Liu. Anti-collusion of group-oriented fingerprinting. *Proc. IEEE Int. Conf. on Multimedia & Expo*, 2003.

[62] Z. J. Wang, M. Wu, W. Trappe, and K. J. R. Liu. Group-oriented fingerprinting for multimedia forensics. *to appear in EURASIP Journal on Applied Signal Processing*, 2004.

[63] 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*, IV:724–727, April 2003.

[64] Z. J. Wang, M. Wu, H. Zhao, W. Trappe, and K. J. R. Liu. Collusion resistance of multimedia fingerprinting using orthogonal modulation. *to apprea in IEEE Trans. on Image Processing*, 2005.

[65] D. Wu, Y. Hou, W. Zhu, Y. Zhang, and J. Peha. Streaming video over the interent: Approaches and directions. *Ieee Transactions on Circuits and Systems for Video Technology*, 11(3):282–300, March 2001.

[66] M. Wu. Multimedia data hiding. *Ph. D Dissertation, Princeton University*, 2001.

[67] M. Wu and B. Liu. Watermarking for image authentication. *IEEE Int. Conf. on Image Processing*, 1998.

[68] M. Wu and B. Liu. Data hiding in image and video: Part-I fundamental issues and solutions. *IEEE Tran. on Image Processing*, 12(6):685–695, June 2003.

[69] M. Wu and B. Liu. *Multimedia Data Hiding*. Springer Verlag, Jan. 2003.

[70] M. Wu and Y. Mao. Communication-friendly encryption of multimedia. *Proc. IEEE Multimedia Signal Processing Workshop*, Dec. 2002.

[71] M. Wu, W. Trappe, Z. J. Wang, and K. J. R. Liu. Collusion resistant fingerprint for multimedia. *IEEE Signal Processing Magazine*, 21(2):15–27, March 2004.

[72] T. Wu and S. Wu. Selecive encryption and watermarking of mpeg video. *Proc. Int. Conf. on Imaging Science, Systems, and Technology*, June 1997.

[73] Y. Yacobi. Improved Boneh-Shaw content fingerprinting. *Topics in Cryptology - CT-RSA 2001, The Cryptographer's Track at RSA Conf. 2001, Proc., Lecture Notes in Computer Science*, 2020:378–391, 2001.

[74] M. Yeung and F. Mintzer. An invisible watermarking technique for image verification. *IEEE Int. Conf. on Image Processing*, 1997.

[75] F. Zane. Efficient watermark detection and collusion security. *Proc. of Financial Cryptography, Lecture of Notes in Computer Science*, 1962:21–32, Feb. 2000.

[76] W. Zeng and B. Liu. A statistical watermark detection technique without using original images for resolving right ownerships of digital images. *IEEE Trans. on Image Processing*, 8(11):1534–1548, Nov. 1999.