Collusion Secure Fingerprinting

CMLab
Outline

• Introduction
• Collusion attack
• Collusion-resilient fingerprinting
  ▫ Combinatorial designed code
  ▫ Error correction code
  ▫ Orthogonal signals
• Research issues
Introduction

- Two types of fingerprinting
  - Type I: Embed fingerprints to contents
  - Type II: Extract fingerprints from contents
Introduction - Type I Fingerprinting

- **What**
  - The process of assigning a unique key for each user's content
- **Why**
  - To identify the person who acquired a particular copy
  - To prevent users from destroying the identification information
  - To trace the traitors
- **How**
  - Watermarking the contents with the unique keys
Introduction - Type II Fingerprinting

- **What**
  - Extract unique and robust image descriptors from contents
- **Why**
  - For the detection of replicas
  - For content-based image retrieval
- **How**
  - Select the features that best describe a given content
Introduction

• Talk focus
  ▫ Type I fingerprinting
• Some related terms
  ▫ Collusion attack
  ▫ Traitor tracing
Introduction-
The Future of Multimedia Security

• What will be the major security problem at the dawn of the 22nd century?

**Ans:** Trust (E. J. Delp, “Multimedia security: the 22nd century approach”, 2005)

  ▫ Who and what I do trust?
  ▫ Is the sensor valid?
  ▫ Is the data valid?
  ▫ Has the data been modified?
Introduction - Fingerprinting

- **Successful application**
  - Hollywood film industry
  - 2004 Oscars
  - Successfully captured a few pirates (in the voting group) leaking the movie to the market

- **Challenging problem**
  - Collusion attack
Introduction

- Collusion attack
  - A coalition of traitors collude to destroy the identification information and leak the content (keys)
Introduction

• **Traitor tracing**
  - Find at least one of the colluders
    \{\text{\text{avatar}}, \text{\text{avatar}}, \text{\text{avatar}}, \ldots, \text{\text{avatar}}\}
  - Can not frame innocent users
    \{\text{\text{avatar}}, \ldots\}
  - Related researches:
    From Broadcast Encryption to Fingerprinting
Collusion attack

- Broadcast encryption
  - e.g. \( u_1: \{k_1, k_3, k_7\} \) \( u_2: \{k_3, k_5, k_7\} \rightarrow \) pirate decoder box: \( \{k_1, k_5, k_7\} \)

- Digital fingerprinting
  - Linear & non-linear collusion attacks

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

- $P_d$ of the $T_N$ statistics under different attacks
  - # of users: 100 code length: 10000 $P_{fp}$: 1%

\[ T_N = y^T \frac{s}{\sqrt{\sigma_d^2 \cdot \|s\|^2}} \]

\[ y = s + d \]

![Graph showing $P_d$ vs. Number of Colluders for different attacks]
Collusion attack

- Analyze the statistical features to improve the detection performance
  - e.g. Observe the histograms of sample means of extracted fingerprints under the average, minimum and randomized negative attacks, respectively $\sigma_w^2 = \frac{1}{9}$, $M = 10000$, $K = 45$
Collusion-resilient fingerprinting

• Boneh-Shaw scheme
  ▫ D. Boneh & J. Shaw, “Collusion-secure fingerprinting for digital data”, CRYPTO’95
  ▫ Watermarking the digital data (e.g. software, documents, music, and videos) with fingerprints
  ▫ Based on the Marking Assumption
  ▫ Too long (too many keys)
    • B. Chor et al. claimed: “It is much less efficient than our schemes”
Collusion-resilient fingerprinting

Marking Assumption

- Any coalition of $c$ users is only capable of creating an object whose fingerprint lies in the feasible set of the coalition
- Feasible set $F(C; \Gamma)$

$$F(C; \Gamma) = \{ w \in (\Sigma \cup \{?\})^l \text{ s.t. } w|_R = w^{(u)}|_R, u \in C \}$$

($\Gamma$ is the code, $C$ is the collusion coalition, $R$ is the set of undetectable positions, and $w$ is the pirate code)

- e.g. A: 3 2 3 1 2
  B: 1 2 2 1 2
  F = $\Gamma'2 \Gamma'12$
Collusion-resilient fingerprinting

- **Definition**: Totally $c$-secure code

A code $\Gamma$ is totally $c$-secure if there exists a tracing algorithm $A$ satisfying the following condition: if a coalition $C$ of at most $c$ users generates a word $x$ then $A(x) \in C$. 

![Collusion-resilient fingerprinting diagram](image)
Collusion-resilient fingerprinting

Lemma: If $\Gamma$ is a totally $c$-secure code then $C_1 \cap \ldots \cap C_r = \emptyset \Rightarrow F(C_1) \cap \ldots \cap F(C_r) = \emptyset$ for all coalitions $C_1, \ldots, C_r$ of at most $c$ users each.

Boneh & Shaw: “There are no totally $c$-secure codes”

Clearly, it is enough to show that there are no totally 2-secure codes.

Let $\Gamma$ be an arbitrary $(l,n)$-code.

Let $w^{(1)}, w^{(2)}, w^{(3)}$ be three distinct codewords assigned to users $u_1, u_2, u_3$, respectively.

Define the majority word $M = \text{MAJ}(w^{(1)}, w^{(2)}, w^{(3)})$ by

$$Mi = \begin{cases} w_i^{(1)}, & \text{if } w_i^{(1)} = w_i^{(2)} \text{ or } w_i^{(1)} = w_i^{(3)} \\ w_i^{(2)}, & \text{if } w_i^{(2)} = w_i^{(3)} \\ ?, & \text{otherwise.} \end{cases}$$

One can readily verify that the majority word $M$ is feasible for all three coalitions $\{u_1, u_2\}, \{u_1, u_3\}, \{u_2, u_3\}$.

However, the intersection of the coalitions is empty.

$\because$ by the last lemma, the code $\Gamma$ is not totally 2-secure.
Collusion-resilient fingerprinting

- **n-secure code with \( \varepsilon \)-error**
  - \( \Gamma_0(n,d): \ l = d(n-1) = O(n^3 \log(n/\varepsilon)) \) → too long!
  - e.g. \( \Gamma_0(4,3) \)

<table>
<thead>
<tr>
<th></th>
<th>A: 111 111 111</th>
<th>B: 000 111 111</th>
<th>C: 000 000 111</th>
<th>D: 000 000 000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Random</td>
<td>permutation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>111 111 111</td>
<td>011 101 101</td>
<td>000 001 101</td>
<td>000 000 000</td>
</tr>
</tbody>
</table>

Given \( x \in \{0,1\}^l \), find a subset of the coalition that produced \( x \).

1) if \( \text{weight}(x|_{B_1}) > 0 \) then output "user 1 is guilty"

2) if \( \text{weight}(x|_{B_{n-1}}) < d \) then output "user n is guilty"

3) for all \( s = 2 \) to \( n-1 \) do: let \( k = \text{weight}(x|_{B_s}) \).

\[
\text{if } \text{weight}(x|_{B_{s-1}}) < \frac{k}{2} - \frac{k}{2} \sqrt{2 \log \frac{2n}{\varepsilon}} \text{ then output "user } s \text{ is guilty"}
\]
Collusion-resilient fingerprinting

- **c-secure code with ε-error of log length**
  - $\Gamma'(L,N,n,d)$
  - Compose $\Gamma_0(n,d)$ with an $(L,N)$-code $C$
    - $x=(\gamma(1) || \gamma(2) || ... || \gamma(L)) \gamma(i) \in \Gamma_0 \ |\Gamma_0|=\text{alphabet size of } C$
  - $l=O(c^4\log(N/\varepsilon)\log(1/\varepsilon))$

Given $x \in \{0,1\}^l$, find a member of the coalition that produced $x$.

1) Apply algorithm 1 to each of the $L$ components of $x$.
   - for each component $i = 1,...,L$ arbitrarily choose one of the outputs of algorithm $1$. Set $y_i$ to be this chosen output. Note that $y_i$ is a number between 1 and $n$. Next, form the word $y = y_1...y_L$.

2) Find the word $w \in C$ which matches $y$ in the most number of positions (ties are broken arbitrarily).

3) Let $u$ be the user whose codeword is derived from $w \in C$
   - output "user $u$ is guilty"
Collusion-resilient fingerprinting

- Three main trends of collusion-resilient fingerprint codes
  - Combinatorial designed code
  - Error correction code
  - Orthogonal signals
Collusion-resilient fingerprinting

- Combinatorial designs
  - D. R. Stinson & R. Wei, “Combinatorial properties and constructions of traceability schemes and frameproof codes”, 1997
  - Give combinational descriptions of the following two objects in terms of set systems
    - Traceability schemes for broadcast encryption
    - Frameproof codes for digital fingerprinting
  - <Definition> $t-(v, k, \lambda)$ design

A $t-(v, k, \lambda)$ design is a set system $(\mathcal{X}, \mathcal{B})$, where $|\mathcal{X}| = v$, $|B| = k$ for every $B \in \mathcal{B}$, and every $t$-subset of $\mathcal{X}$ occurs in exactly $\lambda$ blocks in $\mathcal{B}$. 
Collusion-resilient fingerprinting

- **t-design (t-(v,k,\lambda))**
  - 2-design: BIBD (Balanced incomplete block design)
  - (9,3,1)-BIBD
    \{0,1,6\}, \{0,2,5\}, \{0,3,4\}, \{1,2,4\}, \{3,5,6\}, \{1,5,7\}, \{5,4,8\}, \{4,6,7\}, \{6,2,8\}, \{2,3,7\}, \{3,1,8\}, \{0,7,8\}
Collusion-resilient fingerprinting

**<Definition> c-Traceability Scheme \((c-TS(k,n,l))\)**

Suppose any exposed user \(U\) is a member of the coalition \(C\) whenever a pirate decoder \(F\) is produced by \(C\) and \(|C| \leq c\). Then the scheme is called a \(c\)-traceability scheme and it is denoted by \(c-TS(k,n,l)\).

**<Definition> c-Frame Proof Codes \((c-FPC(l,n))\)**

A \((l,n)\) code \(\Gamma\) is called a \(c\)-frameproof code if, for every \(C \subseteq \Gamma\) such that \(|C| \leq c\), we have \(F(C) \cap \Gamma = C\).

We say that \(\Gamma\) is a \(c-FPC(l,n)\).
Collusion-resilient fingerprinting

• <Theorem>

\[ \exists \text{ a } c\text{-FPC}(v,b) \iff \exists \text{ a set system } (X, \mathcal{B}) \text{ such that } |X| = v, \]
\[ |\mathcal{B}| = b \text{ and for any subset of } d \leq c \text{ blocks } B_1, B_2, \ldots, B_d \in \mathcal{B}, \]
\[ \text{there does not exist a block } B \in \mathcal{B}\setminus\{B_1, B_2, \ldots, B_d\} \text{ such that } \]
\[ \bigcap_{i=1}^{d} B_i \subseteq B \subseteq \bigcup_{i=1}^{d} B_i \]
Collusion-resilient fingerprinting

• <Theorem>

\[ \exists a \ c-TS(k, b, v) \iff \exists \ a \ set \ system \ (X, B) \ such \ that \ |X| = v, \]
\[ |B| = b \ and \ |B| = k \ for \ every \ B \in B, \ with \ the \ property \ that \]
for every choice of \( d \leq c \) blocks \( B_1, B_2, \ldots, B_d \in B \) and for any
\( k \)-subset \( F \subseteq \bigcup_{j=1}^{d} B_j \), there does not exist a block
\( B \in B\{B_1, B_2, \ldots, B_d \} \) such that \( |F \cap B_j| \leq |F \cap B| \) for \( 1 \leq j \leq d \)
Collusion-resilient fingerprinting

- <Theorem> If there exists a $c$-$TS(k, b, v)$, then there exists a $c$-$FPC(v, b)$.

<proof>

Let $(\mathcal{X}, \mathcal{B})$ be the set system corresponding to a $c$-$TS(k, b, v)$. We prove that $(\mathcal{X}, \mathcal{B})$ is a $c$-$FPC(v, b)$.

Suppose no; then there exist $d \leq c$ blocks, $(B_1, B_2, \ldots, B_d) \in \mathcal{B}$, and a block $B \in \mathcal{B} \setminus \{B_1, B_2, \ldots, B_d\}$ such that $B \subseteq \bigcup_{i=1}^{d} B_i$.

Then $|B \cap B_j| \leq |B \cap B|$ for $1 \leq j \leq d$. 

$\rightarrow$
Collusion-resilient fingerprinting

<Theorem> \( \exists t - (v, k, 1)\) design \( \Rightarrow \exists c - FPC(v, \binom{v}{t} / \binom{k}{t}) \),

where \( c = \left\lfloor \frac{(k - 1)}{(t - 1)} \right\rfloor \)

<Theorem> \( \exists t - (v, k, 1)\) design \( \Rightarrow \exists c - TS(k, \binom{v}{t} / \binom{k}{t}, v) \),

where \( c = \left\lfloor \sqrt{(k - 1)} / (t - 1) \right\rfloor \)

- <Theorem>
  \[ \forall v \equiv 1, 3 \mod 6 \exists 2 - FPC(v, v(v - 1) / 6) \]
  \[ \forall v \equiv 1, 4 \mod 12 \exists 3 - FPC(v, v(v - 1) / 12) \]
  \[ \forall v \equiv 1, 5 \mod 20 \exists 4 - FPC(v, v(v - 1) / 20) \]

- <Theorem>
  \[ \forall \text{prime power } q \exists \left\lfloor \sqrt{q} \right\rfloor - TS(q + 1, q^2 + q + 1, q^2 + q + 1) \]

- It needs tricks to construct a suitable code
Collusion-resilient fingerprinting

- Error correcting code (ECC)
  - Field
    - A field $F$ is a set that has two operations defined on it: addition and multiplication, such that the following axioms are satisfied.
      - The set is an abelian group under addition
      - The field is closed under multiplication, and the set of nonzero elements is an abelian group under multiplication
      - The distribution law $(a+b)c=ac+bc$ holds for all $a$, $b$, $c$ in the field

(An abelian group, also called a commutative group, is a group satisfying the requirement that the product of elements does not depend on their order)
Collusion-resilient fingerprinting

- **Galois field**
  - A field with \( q \) elements
  - Denoted by \( GF(q) \)
- **The smallest field**
  - Must have ‘zero’ and ‘one’

\[
\begin{array}{ccc}
+ & 0 & 1 \\
0 & 0 & 1 \\
1 & 1 & 0 \\
\end{array}
\]

- \( GF(2) \)
Collusion-resilient fingerprinting

- A **linear code** is a subspace of $GF(q)^n$
  - The theory of vector spaces can be used to study these codes
  - Any set of basis vectors for the subspace can be used as rows to form a $k \times n$ matrix $G$ called the generator matrix of the code
  - The row space of $G$ is the linear code $C$, any codeword is a linear combination of the rows of $G$
  - The set of $q^k$ codewords is called an $(n,k)$ linear code
    - The dimension of the whole space $GF(q)^n$ is $n$
    - There are $q^k$ codewords
Collusion-resilient fingerprinting

• Encoding
  ▫ Any one-to-one paring of $k$-tuples and codewords can be used as an encoding procedure
  ▫ $c = iG$
    $c$: codeword
    $i$: information word
    $G$: generator matrix
Collusion-resilient fingerprinting

- **Example**

\[ G = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \]

\[ i = [0 \ 1 \ 1] \]

\[ c = iG = [0 \ 1 \ 1] \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} = [0 \ 1 \ 1 \ 1 \ 0] \]
Collusion-resilient fingerprinting

- parity-check matrix $H$
  - A generator matrix of $C$’s dual code
  - $H^TC = 0$
    - e.g. $H = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$

specifies that for each codeword, digits 1 and 2 should sum to zero and digits 3 and 4 should sum to zero
Collusion-resilient fingerprinting

- $G = [I_k | P]$
- $H = [-P^T | I_{n-k}]$
- $G = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$
- $H = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$

$Hx = 0$

$\forall x \in C$  the syndrome of the received $z = x + e$ is defined to be:

$Hz = H(x + e) = Hx + He = 0 + He = He$
Collusion-resilient fingerprinting

• Syndrome decoding

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Coset Leader</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>0</td>
</tr>
<tr>
<td>001</td>
<td>$e_6$</td>
</tr>
<tr>
<td>010</td>
<td>$e_5$</td>
</tr>
<tr>
<td>011</td>
<td>$e_1$</td>
</tr>
<tr>
<td>100</td>
<td>$e_4$</td>
</tr>
<tr>
<td>101</td>
<td>$e_2$</td>
</tr>
<tr>
<td>110</td>
<td>$e_3$</td>
</tr>
<tr>
<td>111</td>
<td>$e_1 + e_4$, $e_2 + e_5$, $e_3 + e_6$</td>
</tr>
</tbody>
</table>
Collusion-resilient fingerprinting

- Error correcting code (ECC)
  - Has systematic encoding and decoding algorithms
  - Distance-based decoding criterion
    - Resilient to attacks for multimedia data
      → Robust than the $t$-design codes
Collusion-resilient fingerprinting

- **<Definition> c-TA codes (c-traceability codes)**
  A code $C$ is a $c$-TA code if for all coalitions $C_i$ of size at most $c$, if $w \in \text{desc}(C_i)$ then there exists $x \in C_i$ such that $|I(x,w)| > |I(z,w)|$ for all $z \in C - C_i$.

- **<Theorem>**
  Suppose that $C$ is an $(l,n)$-code having minimum Hamming distance $d > l(1 - 1/c^2)$. Then $C$ is $c$-TA code.

- **Larger minimum distances imply higher tracing ability**
  - Unfortunately, it’s **not easy to find** practical ECCs with so large minimum distances!!
Collusion-resilient fingerprinting

- e.g. 2-TA:
  \((15,3,11)_4\)-BCH
  \((120,2,96)_4\)-code

- It’s not easy to find practical ECCs with so large minimum distances

- Reed-Solomon code
  - \((l,k,d)\)-RS over GF\((q^m)\): \(d=l-k+1=q^m-1-k+1=q^m-k\)

- A code for sequential traitor tracing

\[
(Tr(\alpha x_1^s + \beta), Tr(\alpha x_2^s + \beta),..., Tr(\alpha x_q^s + \beta)) \quad (Tr(x) = x + x^q + x^{q^2} + ... + x^{q^k})
\]
\[
\alpha \in GF(q^k)^*, \beta \in \{\beta_1, \beta_2, ..., \beta_q\}
\]
\[
k = 2t, \ s < q^{k/2+1}, \text{ and } \exists r \text{ such that } r|t, q^r = -1 \pmod{s}
\]
Collusion-resilient fingerprinting

- Assign one of the $q$ versions for each movie segment ($q$-ary code)
  - H. Jin, J. Lotspiech, & S. Nusser (IBM Almaden research), 2004

<table>
<thead>
<tr>
<th>Movie Version</th>
<th>#0 segment</th>
<th>#1 segment</th>
<th>#2 segment</th>
<th>...</th>
<th>#14 segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0</td>
<td>v1</td>
<td>v4</td>
<td>v15</td>
<td>...</td>
<td>v9</td>
</tr>
<tr>
<td>M1</td>
<td>v5</td>
<td>v9</td>
<td>v0</td>
<td>...</td>
<td>v13</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>M254</td>
<td>v1</td>
<td>v15</td>
<td>v10</td>
<td>...</td>
<td>v4</td>
</tr>
<tr>
<td>M255</td>
<td>v9</td>
<td>v7</td>
<td>v0</td>
<td>...</td>
<td>v6</td>
</tr>
</tbody>
</table>
Collusion-resilient fingerprinting

• Orthogonal signals
  ▫ Orthogonal fingerprint
    • Orthogonal modulation
      \[ w_j = u_j \]  # of fingerprints = # of ortho. bases
    • Modulate the code based on BIBD by noise-like signals
      \[
      w_j = \sum_{i=1}^{B} b_{ij}u_i
      \]
      # of fingerprints >>> # of orthogonal bases
Collusion-resilient fingerprinting

- \textbf{Theorem} Let \((X,A)\) be a \((v,k,1)\)-BIBD and \(M\) the corresponding incidence matrix. If the codevectors are assigned as the bit complement of the columns of \(M\), then the resulting scheme is a \((k-1)\)-resilient AND-ACC.

- \((7,3,1)\)-BIBD
  \[
  \{\{0,1,3\},\{0,2,5\},\{0,4,6\},\{1,2,4\},\{1,5,6\},\{2,3,6\},\{3,4,5\}\}
  \]
Collusion-resilient fingerprinting

- The bit-complement of the incidence matrix for a (7,3,1)-BIBD

\[
C = \begin{pmatrix}
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1
\end{pmatrix}
\]

- Modulate the codewords by 7 orthogonal bases

\[
w_1 = -u_1 + u_2 - u_3 - u_4 + u_5 - u_6 + u_7, \\
w_2 = -u_1 - u_2 + u_3 + u_4 + u_5 - u_6 + u_7, \\
\vdots \\
\vdots \\
\vdots \\
w_7 = +u_1 + u_2 + u_3 - u_4 - u_5 - u_6 + u_7.
\]
Collusion-resilient fingerprinting

- If user 1 and user 2 are colluders
  - Average attack: the pirate code is \((w_1 + w_2)/2\)
  - The coefficient vector is \((-1,0,0,0,1,0,1)\)
  - 1 occurs in the fifth and seventh location uniquely identifies users 1 and 2 as colluders

- Experiments
  - \((16,4,1)\)-BIBD
  - 20 users
  - \(c=3\)
Collusion-resilient fingerprinting

- Tree-structured detection

\[ T_N = y^T s_U / \sqrt{\sigma_d^2 \cdot \|s_U\|^2} \]

Lenna Colluded with Fingerprints \( U_1, U_2, \) and \( U_4 \)

- \( T_N|U_1, U_2 = 21.88 \)
- \( T_N|U_3, U_4 = 10.00 \)
- \( T_N|U_5, U_8 = -0.01 \)
- \( T_N|U_1 = 15.18 \)
- \( T_N|U_2 = 14.55 \)
- \( T_N|U_3 = -0.15 \)
- \( T_N|U_4 = 14.09 \)

\( U_1, U_2, U_3, U_4, U_5, U_6, U_7, U_8 \)
Collusion-resilient fingerprinting

- **Orthogonal signals**
  - Gaussian signals (suitable for multimedia content)
  - Performance is limited if the number of Gaussian signals is larger
Collusion-resilient fingerprinting

- Joint fingerprint embedding and decryption based on coefficient set scrambling
  - D. Kundur & K. Karthik, 2004
Collusion-resilient fingerprinting

PSNR=22dB  PSNR=34dB

Original, encrypted, and fingerprinted images
Collusion-resilient fingerprinting

- Make the colluded copy useless
  - The host signal is warped randomly prior to watermarking
Collusion-resilient fingerprinting
Collusion-resilient fingerprinting

- **Fingerprint multicast in secure video streaming**
  - H. V. Zhao & K. J. R. Liu, 2006
  - Tree-structure-based fingerprint scheme
Collusion-resilient fingerprinting

- MPEG-2-based joint fingerprint design and distribution scheme for video on demand applications

The fingerprint embedding and distribution process at the server’s side
Collusion-resilient fingerprinting

- Hierarchical traceability code
  - Y. T. Lin & J. L. Wu, 2006

The fingerprint of $u_x$ is $w_0w_1w_2w_3w_4...w_1$
Collusion-resilient fingerprinting

- Discriminate different groups
  - Pseudorandomly modulate each group by a different sequence
Collusion-resilient fingerprinting

- **Traitor-within-traitor behavior forensics**
  - H. V. Zhao & K. J. R. Liu, 2006
  - Investigate the precollusion processing strategies for selfish colluders to minimize their risk

\[
\tilde{X}^{(i_1)}_j = \lambda_j(-1) \cdot X^{(i_1)}_{j-1} + \lambda_j(0) \cdot X^{(i_1)}_j + \lambda_j(+1) \cdot X^{(i_1)}_{j+1}.
\]
Collusion-resilient fingerprinting

- Traitor-within-traitor behavior forensics
Collusion-resilient fingerprinting

- Traitor-within-traitor behavior forensics
Collusion-resilient fingerprinting

- **Concatenated collusion-secure code**
  - Y. T. Lin & J. L. Wu, 2007
  - **Outer code:** resisting collusion attacks
  - **Inner code:** satisfying the Strict Marking Assumption

The outer code

- codeword 1
- codeword 2
- codeword $i$
- codeword $n_o$

The inner code

- codeword 1
- codeword 2
- codeword $q$

A $q$-ary symbol

Each $q$-ary symbol is encoded by the inner code

One example of the nested codewords
Collusion-resilient fingerprinting

- **Outer code:** encoding group numbers
- **Inner code:** encoding member numbers and symbols

<table>
<thead>
<tr>
<th>Group #</th>
<th>Codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4 2 0 ... 6</td>
</tr>
<tr>
<td>2</td>
<td>7 0 3 ... 2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>g</td>
<td>1 0 4 ... 1</td>
</tr>
</tbody>
</table>

The octal outer code

\[ g = 1 \quad m = 2 \]

The binary inner code

\[ 7 \quad 0 \quad 3 \quad ... \quad 2 \]

\[ 000...1 \quad 000...0 \quad 101...0 \quad 010...0 \]
Collusion-resilient fingerprinting

- Video fingerprinting for large user group
  - S. He & M. Wu, 2007
  - A small $\rho$ (the correlation between fingerprint sequences), i.e. a large minimum distance, leads to higher detection accuracy
Collusion-resilient fingerprinting

- **Two schemes**
  - **ECCFP with building-blocks**
  - **Directly apply ECCFP**

![Diagram](image)
Collusion-resilient fingerprinting

- Collusion resistance under interleaving collusion (copy-and-paste attack)
  - Outer code: $\text{RS (62,2,61)}_{64}$ for $64^2$ users
  - Inner code: $\text{RS (6,2,5)}_8$ for 64 symbols
Collusion-resistant video fingerprinting based on temporal oscillation


- Protect the traceable code
- Make the colluded version of the video useless
Collusion-resilient fingerprinting

Original video

\[ \text{Shot change detection} \]

Curve fingerprints producing

Motion analysis

Spatial warping

Temporal fingerprint

\[ \oplus \]

Traceability fingerprint

\[ 1101 \ldots 101 \]

Traceability code

Fingerprinted video
Collusion-resilient fingerprinting

- Oscillation
  - User1: 0.0161
  - User2: 0.0156
  - User3: 0.0176
  - User4: 0.0200

<table>
<thead>
<tr>
<th>Number of colluders</th>
<th>Average attack (PSNR)</th>
<th>Copy-and-paste attack (Oscillation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>26.4624 dB</td>
<td>0.2409</td>
</tr>
<tr>
<td>3</td>
<td>24.0246 dB</td>
<td>0.3846</td>
</tr>
<tr>
<td>4</td>
<td>23.8947 dB</td>
<td>0.4775</td>
</tr>
</tbody>
</table>

\[
\text{Oscillation} = \frac{\sum_{i}^{\|f(i) - f(i-1)\|^2}}{\sum_{i}^{\|f(i)\|^2}}
\]

, where \(f(i)\) is the embedding curve function
Collusion-resilient fingerprinting

- Fingerprinting while demodulating
  - S. Lian & Z. Wang, 2008
  - The multimedia content is modulated by pseudorandom sequences at the server side
  - The encrypted content is then demodulated under the control of the fingerprint code at the customer side
Collusion-resilient fingerprinting

- Fingerprinting while demodulating
Research issues

- **Collusion-resilient fingerprint scheme**
  - Robust against collusion attacks
    - Fingerprint - Hard to be removed by collusion attacks
    - Content - Useless after being attacked
  - Traceability code
    - Short $l$
    - Large $n$
    - Large $c$
  - Tracing algorithm & detection strategy
    - High hit ratio
    - Low false alarm rate ✔
    - Fast
    - Low computational overhead
- **Applications**
  - Broadcast environment
  - Streaming data
Collusion-resilient fingerprinting

• Dan Boneh

• Associate Professor, [Computer Science](http://crypto.stanford.edu/~dabo/) and [Electrical Engineering](http://crypto.stanford.edu/~dabo/), Stanford University.
Collusion attack

- K. J. Ray Liu
- Z. J. Wang
- Min Wu
- H. V. Zhou