Security Flaws in Kuribayashi-Tanaka Fingerprinting Protocol

Yongdong Wu
System and Security department
Institute for Infocomm Research,Singapore
wydong@i2r.a-star.edu.sg

Abstract—The buyer-seller watermarking protocol enables a seller to successfully identify at least one traitor from a pirated copy, while prevent the seller from framing an innocent buyer. Kuribayashi and Tanaka proposed a fingerprinting protocol whose objectives are: (1) a buyer can obtain a watermarked copy anonymously, (2) a seller can trace the illegal distributor effectively. However, this paper points out that Kuribayashi-Tanaka fingerprinting scheme has flaws in achieving the above objectives. Furthermore, we improve Kuribayashi-Tanaka fingerprinting scheme so as to patch the security flaws.

I. INTRODUCTION

In a content distribution application, a seller encapsulates and disseminates digital content which is disseminated to the buyers. This kind of business model opens the door for circulating pirated copies since digital content can be reproduced easily. Thus, it is becoming an important concern to protect ownership of digital audio, image and video materials. Since it is hard to prevent illegal copy of digital content, traitor tracing is proposed to find the illegal distributor who is a legitimate buyer but subsequently distributes his copy illegally. Presently, two kinds of traitor tracing methods have been proposed: Key-based scheme and watermark-based scheme.

In the key-based tracing method (e.g., [1]-[6]), the original content is encrypted first, and then the encrypted content is distributed over a broadcast public channel, but the key for decrypting the protected content is embedded into a device or decoder. Once a key is found from a confiscated device or decoder, the original buyer who is assigned with the key is identified as a traitor.

Instead of hiding a decryption key into a device, watermark-based method (e.g., [7]-[15]) embeds a unique watermark for identifying the buyer into the content, and distributes the protected content in a unicast channel. Once a watermark is retrieved from a suspected copy, the buyer with the watermark is accused of being guilty. More precisely, a watermarking scheme for traitor tracing involves three steps: Firstly, an owner embeds into a cover image a watermark that identifies the buyer. Secondly, if a suspicious image is found, the owner will detect the watermark in the image. Once the watermark of a specific buyer is identified, the owner will take the case to a court. Finally, the authority will independently detect the watermark again in the image in question. If the watermark is really found, the traitor is confirmed.

Generally, key-based method is suitable for distribution on a large scale, but lost control after the buyer decrypts the content; on the contrary, watermark-based scheme is able to control the content even the buyer has the content, but is applicable to disseminate valuable content on a small scale. Albeit both schemes can protect the copyright of the seller, they can not protect the buyer’s copyright since the seller knows the private information of the buyer. Concretely, the accusation against the charged traitor, who was the buyer in some earlier transaction, could be repudiable because the original seller also has access to the decoder/copy and, hence, is able to release such a decoder/replica on his own to frame the buyer [16] [17]. Up to now, it is still open to protect the buyer’s copyright in the key-based scheme. But there are some interactive watermark-based buyer-seller protocols [18]-[20] for invisible watermarking to solve this buyer’s copyright problem. In the protocol, the seller does not know the watermark of the buyer, so the seller cannot create copies of the protected content containing the buyer’s watermark. Therefore, if the seller finds an unauthorized copy and identifies the buyer from the unauthorized copy, the buyer can not deny since the unauthorized copy must be originated from the buyer herself.

Kuribayashi and Tanaka [21] proposed a new buyer-seller scheme (called as KT scheme or fingerprinting for short). KT scheme encrypts the buyer’s identification and embeds the encrypted identification into the image. It enables the buyer to obtain a watermarked copy anonymously, while the seller can trace the illegal distributor yet. However, this paper shows that KT scheme fails to not only protect the anonymity of buyer identification, but also prevent buyer cheating. To overcome the flaws, we improve the KT scheme by selecting appropriate parameters and modifying protocol.

The remainder of this paper is as follows. Section II overviews KT fingerprinting scheme. Section III analyzes the security flaws of KT fingerprinting scheme and presents countermeasures. Section IV discusses the attack possibility and experimental results. Finally, Section V concludes the paper.

II. KURIBAYASHI-TANAKA FINGERPRINTING PROTOCOL

To make it easy for cross-reference, the same notations as [21] are used in this paper. Kuribayashi-Tanaka scheme [21] includes 2 major protocols: watermarking protocol including watermark embedding and watermark detecting, and binary proof protocol. In the watermarking protocol,
it employs additive-homomorphic algorithm (e.g. Okamoto-Uchiyama Encryption Scheme [22]) so as to hide the buyer’s identification.

A. Okamoto-Uchiyama Encryption Scheme

Let \( p \) and \( q \) be two large primes of \( k \) bits, \( N = p^2q \).
\[
\gcd(p, q - 1) = 1 \quad \text{and} \quad \gcd(p - 1, q) = 1.
\]
Randomly choose \( g \in \mathbb{Z}_N = \{1, 2, \ldots, N-1\} \) such that the order of \( g_p = g^{p-1} \mod p^2 \) is \( p \). Denote \( h = g^N \mod N \), the public key is \((N, g, h, k)\) and the secret key is \((p, q)\).

To encrypt a message \( m \in \{0, 2^{k-1}\} \subset \mathbb{Z}_p \), select a random number \( r \in \mathbb{Z}_N \), generate the ciphertext as
\[
c = E(m, r) = g^m h^r \mod N
\]
where \( E(\cdot) \) is an encryption function.

To decrypt a ciphertext \( c \), let \( C_p = c^{p-1} \mod p^2 \),
\[
m = D(c) = \frac{C_p - 1}{g_p - 1} \mod p.
\]

Okamoto-Uchiyama Encryption Scheme has additive homomorphic property. That is to say, \( \forall m_1, m_2 \in \mathbb{Z}_p \), such that \( m_1 + m_2 < 2^{k-1} \),
\[
E(m_1, r_1) \cdot E(m_2, r_2) = E(m_1 + m_2, r_1 + r_2)
\]
where \( r_1 \) and \( r_2 \) are two random numbers.

B. KT watermarking protocol

In the KT protocol, there are two participants: Buyer \( B \) and Seller \( M \). The buyer \( B \) has an authenticated identification \( id \) whose binary representation is
\[
id = w_0 + w_1 \cdot 2^1 + \cdots + w_{l-1} \cdot 2^{l-1},
\]
\( w_j \in \{0, 1\} \). The seller \( M \) knows \( W = g^{id} \mod N \) in advance. To embed the \( id \) into the image, both seller and buyer perform the watermarking scheme as shown in Fig.1.

1. After receiving the purchase request from the buyer \( B \), the seller \( M \) selects a random \( a \in \mathbb{Z}_p \) and send \( a \) to the buyer.

2. The buyer constructs a random sequence \( \{a_j \in \mathbb{Z}_N, j = 0, 1, \ldots, l-1\} \) such that
\[
a = a_0 + a_1 \cdot 2^1 + \cdots + a_{l-1} \cdot 2^{l-1},
\]
calculates
\[
com_j = E(w_j, a_j) = g^{w_j} h^{a_j} \mod N,
\]
and sends \( com_j \) \( (j = 0, 1, \ldots, l-1) \) to the seller.

3. The seller verifies the validity of \( com_j \) by checking
\[
\prod com_j^2 \equiv h^a W,
\]
\[\text{That is to say}, \ (g_p)^p = 1 \mod p^2, \ \text{but} \ \forall x \in (0, p), \ (g_p)^x \neq 1 \mod p^2\]

Afterwards, the seller performs DCT (Discrete Cosine Transform) on the image so as to obtain the DCT coefficients. For each DCT coefficient \( I_i \) to be manipulated, select \( b_i \in \mathbb{Z}_N \),
\[
Y_i = g^{I_i} h^{b_i} \mod N = E(I_i + Tw_j, b_i + T a_j).
\]
where \( T \) is a predefined parameter to control the watermark invisibility. Send \( Y_i \) to the buyer as well as all the un-manipulated DCT coefficients.

4. The buyer decrypts \( Y_i \) and obtains the watermarked DCT coefficient \( I_i = I_i + Tw_j \).

As a result, the buyer anonymously obtains the same watermarked copy as that generated with Cox’s [11] additive watermarking. Hence, the watermark detecting method is the same as that in Cox’s method [11]. Since the detection method is well known in watermarking domain, we omit the detail for simplicity. Usually, a block of DCT coefficients is used to embed the same bit \( w_j \) so as to increase detection robustness.

According to the watermark detecting method, the output watermark \( w_j' \) is binary. That is to say, the detector always reports a binary string \( w_0' w_1' \ldots w_{l-1}' \) as the traitor’s identification. If \( w_j \) is not binary, the seller can not trace the traitor correctly. Thus, it is critical to prove \( w_j \) is binary in KT protocol.

\text{Buyer \( B \)} \hspace{1cm} \text{Seller \( M \)}

\begin{align*}
\text{id} &= \sum w_j \cdot 2^j & \text{W} &= g^{id} \mod N \\
\text{a}_j &\in \mathbb{Z}_N & \text{a} &\in (2^l, N) \\
\text{a} &= \sum \text{a}_j \cdot 2^j & \text{com}_j &= g^{w_j} h^{a_j} \\
\text{com}_j &= \prod \text{com}_j^2 \equiv h^a W \\
\text{I}_i &= \text{D}(Y_i) & \text{Y}_i &= g^{I_i} h^{b_i} \text{com}_j^T \\
&= \text{I}_i + Tw_j & & \text{I}_i = \text{I}_i + Tw_j
\end{align*}

Fig. 1. KT watermarking protocol in a buyer-seller transaction. \( \text{D}(\cdot) \) is a decryption function whose decryption key is known to the buyer only. The original image is \( I = \{I_i\} \) and the watermarked image is \( I = \{I_i\} \). To be robust, one bit will be embedded into a block of DCT coefficients.

C. KT Binary-proof Protocol

In order to prove that the embedding message is binary without disclosing buyer’s \( id \), both the seller and the buyer must perform a Zero-knowledge proof for the binary \( w_j \) as Fig.2.

2. After receiving \( \text{com}_j \), the seller selects two random numbers \( t_j, c_j \in \mathbb{Z}_N \) such that \( t_j + c_j < 2^{k-1} \).

(2.1) Seller computes
\[
\text{COM}_j = \text{com}_j^T g^{c_j} = E(w_j t_j + c_j, a_j t_j),
\]
sends \( \text{COM}_j \) to the buyer.
Indeed, it is not guaranteed in KT fingerprinting scheme. The identification to preserve the anonymity/privacy of the buyer. That is to say, required to be randomly selected from group $Z_N$. Scheme can not meet both the above requirements. The second one is more strict since it demands $A. Attack 1: Exposing the Buyer

Usually, anonymity [23] has two properties: One is identification hiding, and another is un-linkability. The former requires that the identification such as buyer’s name be unknown to the seller. The second one is more strict since it demands that any one but the buyer can not tell whether two transactions are executed by the same buyer or not. Unfortunately, KT scheme can not meet both the above requirements.

1) Disclosing Identification: One goal of KT protocol is to preserve the anonymity/privacy of the buyer. That is to say, the identification $id$ should be unknown to the malicious seller. Indeed, it is not guaranteed in KT fingerprinting scheme.

According to $com_j = g^{w_j}h^{a_j}$, $j = 0, 1, \ldots, l - 1$, if the seller $M$ can obtain $a_j$, he is able to decide whether $w_j = 0$ or $w_j = 1$.

In the fingerprinting protocol of Subsection II-B, $a_i$ is required to be randomly selected from group $Z_N$. However, it is impossible to meet this requirement due to the limitation of Eq.(3):

- The possible interval for selecting $a_j$ is much smaller than $Z_N$.
- $a_j$s are not independent. If one is selected from a large size set, others must be selected from a small set.

Indeed, this limitation has shown in the process of selecting the sequence $\{a_j\}$ in the original paper [21]. Since the only requirement for selecting $a_i$ is that $a_i \in (2^l, N)$, the seller may maliciously choose the smallest one, i.e., $a = 2^l + 1$ such that

$$a = 2^l + 1 = a_0 + a_1 \cdot 2^1 + \cdots + a_{l-1} \cdot 2^{l-1}$$  \hspace{1cm} (3)$$

many of $\{a_j\}$ may be small. For example, $a_{l-1} \in \{0, 1, 2\}$ in Eq.(3). As a result, if the set for selecting each $a_j$ is of small size, for instance, of size $S = 2^{10}$, the seller can deduce $a_j$ and $w_j$ with the commitment $com_j$. Specifically, to obtain buyer $id$, the seller will

1) build a table $T$ which includes three columns as shown Table III-A.1. Column 1 is the index for $a_j$, column 2 is $g^{a_j}h^{a_j}$, and Column 3 is $g^{w_j}h^{a_j}$, where $a_j \in [0, S)$.

2) select a small $a$, e.g, $a = 2^l + 1$, and send $a$ to the buyer in step 1 of Fig.1. Let $j = l - 1$.

3) receive the commitment $com_j$ in step 2 of Fig.1. Search $T$ against $com_j$ sequentially. To search promptly, only check the entries whose $a_j \cdot 2^l \leq a$.

4) If $com_j$ is not an entry of $T$, goto step (5). But if $com_j$ is an entry of column 2 of table $T$, $w_j = 0$; otherwise, $w_j = 1$. $a \leftarrow a - a_j \cdot 2^j$.

5) $j \leftarrow j - 1$. If $j \geq 0$, goto (3).

With the above attack, the seller obtains a portion of $id$ bits.

### Table I

<table>
<thead>
<tr>
<th>$a$</th>
<th>$com_j(w_j = 0)$</th>
<th>$com_j(w_j = 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$1$</td>
<td>$g$</td>
</tr>
<tr>
<td>1</td>
<td>$h$</td>
<td>$gh$</td>
</tr>
<tr>
<td>2</td>
<td>$h^2$</td>
<td>$gh^2$</td>
</tr>
<tr>
<td></td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$a_j$</td>
<td>$h^{a_j}$</td>
<td>$gh^{a_j}$</td>
</tr>
<tr>
<td></td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$S - 1$</td>
<td>$h^{S-1}$</td>
<td>$gh^{S-1}$</td>
</tr>
</tbody>
</table>

2) Linking transaction: According to the fingerprinting protocol in Fig.1, the seller has to extract $W$ from his database. Hence the seller can decide whether two transactions are performed by the same buyer or not by checking $W$. In other words, the fingerprinting protocol enables to decide whether two transactions are performed by the same buyer or not. Although unlinkability is not one of the designed goal in KT scheme and may be not serious in the transactions, it results in the following collison attack.

---

(2.2) Buyer decrypts $COM_j$ as $D(COM_j) = w_jt_j+c_j$, selects a random $r_j \in R Z_N$, and calculates a commitment $\overline{com}_j$ from $D(COM_j)$. The buyer sends $\overline{com}_j$ to the seller.

(2.3) Seller sends the random number $t_j$ and $c_j$ to the buyer.

(2.4) Buyer verifies $COM_j$ according to $t_j$ and $c_j$. If the result is positive, she discloses the random number $r_j$. Only if

$$\overline{com}_j = \overline{com}_j^{t_j+c_j} h^{r_j},$$  \hspace{1cm} (2)$$

the seller will regard $w_j$ as binary.

With the binary proof, the seller is convinced that each $com_j$ is a commitment to a binary bit $w_j$. However, the following Section will show that the binary proof is vulnerable.
3) Collusion attack: Since KT does not satisfy unlinkability property, the seller can start a collusion attack with transaction history. Technically, if the buyer performs the watermarking protocol (e.g., buy different images) several times with the same identification \( id \), the seller merges the obtained bits of \( id \), hence discloses \( id \) completely. Specifically,

(1) Let \( id = \{ w_{0}, w_{1}, \ldots, w_{l-1} \} \), where \( w_{j} \notin \{ 0,1 \} \).
(2) If \( w_{j} \notin \{ 0,1 \} \), start the attack in Subsection III-A.1 so as to update \( id \).
(3) Repeat the second step with another new transaction using the same \( id \) for some rounds as necessary. Therefore, the buyer’s identification is disclosed gradually.

4) Countermeasure: In the above attack, the seller is able to obtain \( w_{j} \) by brute force since the set for selecting \( a_{j} \) is small in the original work [21]. Naturally, if each \( a_{j} \) is selected from a large set, the above attack will be impractical. Hence, there are two countermeasures:

\( C_{1} \): Let \( a \in R (2^{l} + k, N) \) instead of \( a \in R (2^{l}, N) \). \( C_{2} \): Let \( a_{j} \in R (-N, N) \) instead of \( a_{j} \in R Z_{N} \).

Method \( C_{1} \) enables to select \( a_{j} \) from the set of size at least \( 2^{k} \). Since \( C_{2} \) enables \( a_{j} \) to be negative, \( a_{j} (j = 1, 2, \ldots, l-1) \) can be selected from a large set in a roughly independent way. In either \( C_{1} \) or \( C_{2} \) countermeasures, the buyer should check \( a \) against the interval after receiving \( a \) in step 2 of Fig.1.

B. Attack 2: Cheating the Seller

In the binary proof protocol, the buyer selects \( r_{j} \) so as to calculate \( \widehat{com}_{j} \) as a commitment. After obtaining \( t_{j} \) and \( c_{j} \), the buyer sends \( r_{j} \) back to the seller for verification. However, if a crafty buyer can find another \( r'_{j} \) such that

\[
\widehat{com}_{j} = \widehat{com}_{j}^{t_{j}+c_{j}r'_{j}} \mod N
\]

holds, \( w_{j} \) may be not binary and hence the seller is fooled. To accomplish it, the buyer will make sure that

\[
\begin{align*}
\widehat{com}_{j} & = g^{w_{j}h_{j}^{t_{j}+c_{j}r'_{j}}} = g^{w_{j}N_{a_{j}}} = g^{w_{j}+N_{a_{j}}} \\
g^{t_{j}+c_{j}r'_{j}} & = (g^{w_{j}+N_{a_{j}}})^{(t_{j}+c_{j})} g^{N_{r_{j}}} \\
 & = g^{(w_{j}+N_{a_{j}})(t_{j}+c_{j})+N_{r_{j}}} \mod N
\end{align*}
\]

Rewrite Eq.(4) with Eq.(5) and Eq.(6), we have

\[
g^{(w_{j}+N_{a_{j}})(t_{j}+c_{j})+N_{r_{j}}} \mod N = (w_{j}+N_{a_{j}})(t_{j}+c_{j}) + N_{r_{j}} \mod \phi(N)
\]

\[
N(r'_{j} - r_{j}) = (w_{j}+N_{a_{j}})(w_{j}-1)t_{j} \mod \phi(N)
\]

i.e.,

\[
p | w_{j} \quad \text{or} \quad p | (w_{j}-1) \quad \text{(8)}
\]

Denote \( c = (w_{j}+N_{a_{j}})(w_{j}-1)/p \), rewrite Eq.(7) as

\[
P^{2}q(r'_{j} - r_{j}) = p \times c \times t_{j} \mod p(p-1)(q-1)
\]

\[
r'_{j} = r_{j} + c \times t_{j}/p \mod \phi(pq) + k \phi(pq)
\]

\[
\phi(pq) = (p-1)(q-1), k = 0, 1, \ldots, p-1
\]

Thus, if the buyer selects \( w_{j} \) according to Eq.(8), and then sends \( r'_{j} \) to the seller in step (2.4) of Fig.2, the seller will be convinced that Eq.(4) holds, and wrongly believes that any \( w_{j} \) is binary. Consequently, the seller can not trace the traitor correctly since the embedded identification is not identical to the traitor’s identification. In order to mount this attack, \( id \) should be large due to restriction of Eq.(8). Please notice the difference between the collusion attack in Subsection III-A.3 and the protocol attack. The collusion attack exploits the weakness of small \( id \) by the seller, while the attack to binary protocol uses the weakness of large \( id \) by the buyer. Hence, KT scheme should be upgraded other than a simple parameter selection.

1) Bug removal: Besides the above attack, KT scheme has a minor bug. In Fig.2, \( t_{j}, c_{j} \in Z_{N} \), thus the buyer will obtain

\[
D(COM_{j}) = w_{j}t_{j} + c_{j} \mod p \neq w_{j}t_{j} + c_{j}
\]

The buyer will use \( (w_{j}t_{j} + c_{j} \mod p) \) to compute \( \widehat{com}_{j} \), but the seller will use \( t_{j} + c_{j} \) for verification in Eq.(2). If \( w_{j}t_{j} + c_{j} > p \), the verification will fail. That is to say, a honest buyer can not always pass the verification in the binary-proof protocol. To remove this bug, we modify the protocol as \( t_{j}, c_{j} \in (0, 2^{k-2}) \).

2) Countermeasure and modification: There are two ways to foil the attack on binary-proof protocol. One is to strictly register the \( id \). Another is to modify the protocol itself.

Strict registration: In order to invalidate the binary proof protocol, the buyer has to meet the following requirements.

\[
w_{j} \equiv 0 \mod p \quad \text{or} \quad w_{j} = 1 \mod p
\]

\[
id \equiv w_{0} + w_{1} \times 2^{l} + \ldots + w_{l} \times 2^{l} \mod \phi(N)
\]

To this end, the buyer has to select \( p \) and \( q \) first, and then, he selects \( w_{j} \) according Eq.(10). Finally, his identification \( id \) will be available due to Eq.(11). Clearly, it is necessary to prevent the buyer from selecting \( id \) randomly, the registrant should restrict \( id \) in an uncontrollable way, e.g., the hash value of the name of the buyer.

Protocol improvement: Although a strict registration process is able to defeat the attack, it results in inconvenience for the application. It is preferable to improve the protocol so as to defend against the above attack. Technically, the commitment to \( r_{j} \) should be known to the seller before the seller sends out \( t_{j} \). In the countermeasure shown in Fig.3, if the buyer sends \( r'_{j} \) yielded from Eq.(9), the seller can expose the cheating behavior. Fig.3 is the upgraded binary-proof protocol.
B. Collision Possibility

In step (2), $\text{rand}(\cdot)$ is used to generate a random $a_i$ which is equal to $2^l$, rewrite Eq.(12) as

$$p^2q(a_i - a_j) \equiv 0 \mod \phi(N)$$

Since $-2^{l+1} < a_i - a_j < 2^{l+1}$, Eq.(14) has a unique solution $a_i = a_j$. Thus, $(w_i, a_i) = (w_j, a_j)$. It contradicts our assumption. Thus, the search result in Subsection III-A.1 is always correct.

C. Attack with Single Transaction

Based on the attack in Subsection III-A.1, if $a_j \in [0, S)$, we can search table $T$, so as to obtain $w_j$. Fig.4 describes the probability of recovering 50% of $\{w_0, w_1, \ldots, w_{l-1}\}$ given that $a$ is in $(2^{64}, 2^{128})$. Thus, we can see that the original KT scheme is vulnerable to our attack. Meanwhile, the low recovery probability in case of large $a$ demonstrates that the protection methods in Subsection III-A is effective.

D. Attack with Multiple Transactions

In this experiment, the same $id$ is used repeatedly, but the sequence $\{a_j\}$ will be changed in every transaction whatever $a$ is. Fig.5 illustrates the results of collusion attack. From the experiments, we observe

- the collusion attack is viable when multiple transactions are performed. Especially, collusion attack is effective to recover $id$ bits in the first 10 transactions, but it will not very useful for more than 20 transactions.
- the bigger $a$ is, the less $id$ is exposed.
- LSBs (Least Significant Bits) are harder to be disclosed than MSBs (Most Significant Bits).

All the observations support that the security strength is increased with larger interval for selecting $a$. The observations are in consistent with our intuition and analysis. Meanwhile, the observations also support the countermeasures.
V. CONCLUSION AND FUTURE WORK

Buyer-seller protocol enables the buyer to embed her identification in a stealthy way. KT buyer-seller scheme enables the seller to embed an identification ciphertext but extract the identification plaintext. Unfortunately, KT is vulnerable to the present attacks due to collusion attack and protocol flaw. The collusion attack exploits the weakness of small id such that the seller can deduce the identification of the buyer, while the latter uses the weakness of large id such that the buyer may cheat the seller with a forged identification. This paper proposes countermeasures to above attacks by choosing appropriate parameters and upgraded protocol.

Although KT scheme can be improved to defeat the attacks, its computation and communication cost are of high. In fact, the other watermark-based buyer-seller schemes consume much computational time and heavy bandwidth too. On the other hand, few solutions provide key-based buyer-seller scheme. Hence, it is still a challenge to provide a satisfactory buyer-seller scheme.

REFERENCES