On Designatedly Verified (Non-interactive) Watermarking Schemes

Malapati Raja Sekhar[1], Takeshi Okamoto[2]
and Eiji Okamoto[2]

[1] Applied Statistics Unit, Indian Statistical Institute, Kolkata, India
malapati_r@isical.ac.in
[2] Dept. of Risk Engineering,
Tsukuba University, Japan
{ken,okamoto}@risk.cipher.tsukuba.ac.jp

Abstract. Although many watermarking schemes consider the case of universal verifiability, it is undesirable in some applications. Designated verification is a possible solution for this problem. Watermarking scheme with (non-interactive) designated verification through non-invertible schemes was proposed by Lee *et al* [11] in 2003, to resolve multiple watermarking problem. Yoo *et al* [14] proposed a very similar watermarking scheme. In this paper, we propose a cryptanalytic attack on both of these schemes that allows a dishonest watermarker to send illegal watermarked images and to convince the designated verifier or customer that received watermarked images are valid. We modify the above schemes to overcome the attack. Further, we also propose a new robust watermarking scheme with (non-interactive) designated verification through non-invertible watermarks. Interestingly, our scheme can be extended for joint copyright protection (security of ownership rights for images to be owned by more than one entity).

Keywords Designated verifier watermarks, multiple watermarking, joint copyright protection.

1 Introduction

The rapid (r)evolution of the Internet and the electronic representation have made the transmission of multimedia content such as text, images, audio, and video *easier*. Digital media can be easily accessed or distributed via internet. Due to this, to obtain an illegal copy has become simple with no loss of fidelity, that is, the copy of a digital media object is identical to the original one. Any number of identical copies of digital media objects can be illegally produced. This has increased the potential for possible misuse and theft of such information and significantly increases the problems associated with enforcing copyrights on multimedia information. A straightforward way to protect against above attacks is to completely encrypt the data and thereby require the end user to have the

decryption key for the decoding. However, this process helps only during the transit of object. So after decryption, this process no longer stops illegal distribution. Another way is to protect this data is by applying a digital watermarking methods.

Digital watermarking [2] is the process by which an image is coded with an owner's watermark and this can be done using one of two general approaches. One approach is to transform the host image into its frequency domain representation and embed the watermark data therein. In the other method, the watermark is embedded in the host image directly in the spatial domain. *Although, all the schemes in this paper can be applied to all media objects such as text, video, audio, images etc., we have restricted our discussion to images alone for the sake of brevity.*

Regardless of the embedding method, there are several requirements that the watermarking system must satisfy. First, the watermarked image should retain, as closely as possible, the quality of the original image. This means the presence of watermark should not visually appeared. Second, once the watermark has been recovered from the image, there should be a very high level of certainty whether extracted one is a true watermark or not. Also this method of determination should be simple and accurate. Third, the watermark should be robust to various types of image processing techniques, like compression, which may be applied with the intent to destroy the watermark in the image.

The concept of Undeniable Signature, introduced by Chaum [1], is a kind of digital signature which cannot be verified without interacting with the signer. In some applications, it is important for the signer to decide not only when but also by whom his signatures can be verified due to the blackmailing [6] [8] and mafia [5] attacks. Consider the case: voting center presents a proof to convince a certain voter that his vote was counted while without letting him to convince others (e.g., a coercer) of his vote, which is important to design a receipt-free electronic voting scheme preventing vote buying and coercion. This is the motivation to the concept of "designated verifier proofs" [9]. Designated verifier proofs address trade-off between confidentiality and authenticity. These allow a signer to construct a proof that will convince only a designated verifier. The designated verifier cannot present the proof to convince any third party because he is fully capable of generating the same proof by himself.

A protocol is 'oblivious' when it does not leak any information whether given valid or invalid inputs i.e., in our case, behavior of the prover is identical in both the cases of valid or invalid inputs. Oblivious decision proof is oblivious and computationally minimum-knowledge meta-proof for deciding valid exponentiation. A meta-proof as described in [10], consists of two portions: (1) a "blinded" proof that the "exponentiation is correctly performed" and (2) a proof that first proof is correctly performed. Moreover, oblivious decision proof given in [10] is provably secure.

Many watermarking schemes consider the case of universal verifiability. However, it is undesirable in some applications such as watermark based commercial digital contents distribution system. Multiple watermarking problem (the

attacker watermarks an already watermarked image and later claims of ownership) was proposed by Craver *et al* [4]. In 2003, Lee *et al* [11], for the first time, proposed a watermarking scheme with designated verification through noninvertible schemes to resolve multiple watermarking problem. Yoo *et al.* [14] proposed a very similar scheme. These watermarking schemes are based on designated verifier signature schemes [9].

Our Contribution: In this paper, we propose an cryptanalytic attack that allows a dishonest watermarker to send illegal watermarked images and to convince the designated verifier or customer that received watermarked images are valid. Our attack is applicable on both Lee *et al* [11] and Yoo *et al* [14] watermarking schemes. We also show how to modify these schemes to overcome the attack. Then, we propose a new watermarking scheme with (non-interactive) designated verification through non-invertible schemes to resolve the multiple watermarking problem. Our scheme is based on oblivious decision proof [10] mentioned above. We consider the following scenario. An image selling vendor might embed undeniable signature into his images and only allow the paying customers to verify the authentication of the images. If the vendor watermarked a image, he must provide some proofs to convince the customer of the fact. Also, these proofs must be non-transferable, i.e., once a verifier (customer) is convinced that the vendor watermarked the image, then he cannot transfer these proofs to convince any third party. Moreover, as our proposed scheme is based on oblivious decision proofs, joint ownership rights problem (image to be owned more than one person) can be solved by using oblivious multi-party decision proofs.

In the next section, we briefly explain designated verification in watermarking schemes through noninvertible watermarks. In section 3, we briefly describe Yoo *et al* scheme. In section 4, we propose an attack and modification to overcome the proposed attack. In section 5, we propose a new watermarking scheme based on oblivious decision proofs. In section 6, some concluding remarks are presented.

2 Designatedly verified (non-interactive) watermarking schemes

We consider the watermarking systems in an open environment where contents can be easily distributed, for instance, through WWW servers.

Formally, a Watermarking System consists of three probabilistic polynomial time algorithms (GEN_{KEY}, E, D). GEN_{KEY} is the key generation algorithm on input 1^{n_w} (where n_w is a security parameter), outputs the key k. Let I be the image and $W \in \{0,1\}^{n_w}$ be the watermark. The encoding algorithm E receives (I, W, k) as input and produces an output of watermarked image I'. And this marked image I' should be perceptually similar to the original image I. On input (I', I, W, k, Aux), the decoding algorithm D, extracts the watermark W', checks whether $W = W'$ and outputs TRUE or FALSE (Aux could be a cryptographic key). And we require that $D(E(I, W, k), I, W, k, Aux) =$ TRUE for all images I, watermarks W and keys k.

Multiple Watermarking: We consider the attacker's scenario as the attacker watermarks an already watermarked image and later claims ownership. Given a watermarked image, Alice (true watermarker) can extract his mark and likewise Oscar(the attacker). Then, how can we decide who owns the image? Craver *et al* [4] suggested a solution to this problem by defining Non-invertible Watermarks. But he did not suggest the protocol in concrete and there are some unsolved problems [11]. Yoo *et al* [14] considered that undeniable verification can be a possible solution.
Designated verification: Many Digital watermarks are verified as authentic by anyone using the verification process. However, this self-dissemination property is unsuitable for many applications such as watermark based commercial digital contents distribution system. The validity or invalidity of an undeniable watermark can be ascertained by verifier (Bob) in cooperation with original watermarker (Alice). If a confirmation process is needed, Alice gives exponentially-high certainty to the verifier that the watermark does correspond to the legal one. We can use undeniable scheme in the watermarking process.
Verification Mechanism: Multimedia contents company could embed the watermark which was signed using an undeniable scheme. Only someone who had directly purchased the contents from that company could verify the watermark and be certain that the contents were right. Undeniable watermark verification could also be useful in any situation in which an individual wishes to sign a data anonymously.

The original watermarker can embed his own secret (true watermark) in cover data and generate watermarked data. And it is distributed to the buyer by using network protocol. And then the authorized user can buy or get this data. However, the attacker can also add his own fabricated watermark to the distributed image. In this case, illegal image can be generated by the attacker. So, if this image will be used by the unauthorized user, we must verify and prove illegality of the image. In our scheme, buyers and the registration center can be designated verifiers. In real-life electronic market places, buyers want to directly ensure that they are purchasing digital items from the real copyright holder. *Therefore, we explain the following protocols for proof of ownership where participation of the registration center is required as another designated verifier.*

3 Yoo *et al* Scheme [14]

Let (GEN_{KEY}, GEN_W, E, D) is a designated verifier proof protocol between prover Alice and verifier Bob. Yoo *et al* consider non-interactive designated verifier proof. Such a scheme can diminish the difference between public watermark and undeniable watermark so that it limits one who can verify it without the provers help, but it does not necessitate interaction.
1. Generating and preprocessing the Keys/Data.

- Public parameters g, p, q are chosen similar to digital signature standard [7].
- Generate prover Alice's secret/public key pair : x_A and $y_A \equiv g^{x_A} \pmod{p}$.

– Generate verifier Bob's secret/public key pair : x_B and $y_B \equiv g^{x_B} \pmod p$. (In this step, we use key generation function GEN_{KEY}).
– Using hash function $h(\cdot)$, generate $m = h(I)$ on cover data I.
– Alice selects $\omega, r, t \in Z_q$.

2. Generating and embedding Watermark.

– The verifier sends the hashed value m to the prover.
– The prover generates watermark W using his own secret key x_A : $W \equiv m^{x_A}(modp)$. (In this step, we use watermark generation function GEN_W and we can tell the prover as a *watermarker* or a *signer*).
– The prover embeds this watermark W into cover data I. Then the watermarked data I_W will be generated. (In this step, we use embedding function E).

3. Constructing a Proof.

– The prover calculates $c \equiv g^{\omega} y_B^r (\text{mod } p), G \equiv g^t (\text{mod } p), M \equiv m^t (\text{mod } p)$, and $d \equiv t + (x_A \cdot \omega) (\text{mod } q)$.
– And the prover sends commitment (ω, r, G, M, d) to the verifier.

4. Verifying the Proof.

– The verifier extracts the embedded pseudo-watermark $W^{'}$ from the $I^{'}$. (In this step, we use detection function D).
– Using the commitment, the verifier calculates $c \equiv g^{\omega} y_B^r (\text{mod } p)$ and $G \cdot y_A^{\omega} \equiv g^d \pmod p$.
– On the extracted pseudo-watermark, the (designated) verifier calculates $M \cdot W^{\omega} \equiv m^d (\text{mod } p)$.

Proof of Ownership: When the ownership is challenged, Alice and Oscar are asked to participate in the proof. Then if anyone *refuses to participate* or *unable to prove*, he/she cannot claim for copyright. If Alice and Oscar both could prove their proofs then who registers the image prior is the true watermarker.
Practicality of Scheme: Yoo *et al* [14] have showed that their scheme is practical. It is based on Cox *et al* [3].

4 Proposed attack and a modification

To cheat a designated verifier Bob, Alice (dishonest watermarker) selects four random numbers $\omega, r, d \in_R Z_q$ and $\overline{W} \in_R Z_p$, and then computes the proof $P = (\omega, r, \overline{G}, \overline{M}, d)$ for an invalid watermark $\overline{W}$ for the hashed image m simultaneously as follows:

$$\begin{cases} c \equiv g^{\omega} y_B^r (\text{mod } p) \\ \overline{G} \equiv g^d . (y_A^{\omega})^{-1} (\text{mod } p) \\ \overline{M} \equiv m^d . (\overline{W}^{\omega})^{-1} (\text{mod } p) \end{cases}$$

After that, Alice embeds $\overline{W}$ into image and sends the watermarked image to Bob, and stores the proof $P = (\omega, r, \overline{G}, \overline{M}, d)$ securely. When Bob needs to check the validity of watermark $\overline{W}$, Alice sends the proof P to Bob. It is easy to see that Bob will believe that proof P for the watermark $\overline{W}$ is Alice's valid proof for watermarked object $I^{'}$, since the proof satisfies verification equations, that is, $G \cdot y_A^{\omega} \equiv g^d \pmod{p}$ and $M \cdot W^{\omega} \equiv m^d \pmod{p}$. However, in later disputes, Alice can convince a third party (e.g. a judge) that $\overline{W}$ is indeed invalid by using a denial protocol (Chaum's zero-knowledge protocol [1]). Attack on Lee *et al* scheme [11] is very similar to the above and so we omit the details here.
The above attack "seems to be" rectified by using a hash function $h = (c, M, G)$ in the proof while computing d, *i.e.*, $d \equiv t + x_A(\omega + h) \pmod{q}$. Correspondingly, the verification equations change to $G \cdot y_A^{\omega+h} \equiv g^d \pmod{p}$ and
$M \cdot W^{\omega+h} \equiv m^d \pmod{p}$.
But there is still an attack, on this modified scheme, which is similar to the Wang's attack [13]. Dishonest watermarker selects four random numbers
$\omega, r, t, \bar{t} \in_R Z_q$, and then computes the proof $P = (\omega, r, G, \overline{M}, d)$ for an invalid watermark $\overline{W}$ for hashed image m simultaneously as follows:

$$\begin{cases} c \equiv g^{\omega} y_B^r \pmod{p} \\ G \equiv g^t \pmod{p} \\ \overline{M} \equiv m^{\bar{t}} \pmod{p} \\ d \equiv t + x_A(h + \omega) \pmod{q} \\ \overline{W} \equiv m^{x_A} \cdot m^{(t-\bar{t})(h+\omega)^{-1}} \pmod{p} \end{cases}$$

It is clear that $P = (\omega, r, G, \overline{M}, d)$ is a valid proof for the invalid watermark $\overline{W}$. However, this new attack can be rectified by letting the prover to prove $log_m M = log_g G$ using equality of discrete logarithms protocol given in Jakobsson *et al* [10] in addition to modified verification equations.
Analysis of Lee et al scheme [11] is similar to the above.

5 Proposed Scheme

Here we propose a new scheme which is based on oblivious decision proof given in [10]. We follow the similar notation of the above scheme. We call the prover as a *watermarker* or a *signer*.
1. Generating the Keys (Key generation algorithm : GEN_{KEY})

- Let p and q be two primes where $q|p-1$ and g is a generator of the subgroup G_q of Z_p^* of order q.
- Prover Alice generates secret key $x_A \in_R Z_q$ and her public key is $y_A \equiv g^{x_A} \pmod{p}$.
- Verifier Bob generates secret key $x_B \in_R Z_q$ and his public key is $y_B \equiv g^{x_B} \pmod{p}$.
- Using perceptual hash function [12] $h(\cdot)$, generate $m = h(I)$ on cover data I.

2. Generating and embedding watermark (Watermark embedding algorithm E)

– The prover generates watermark W using his secret key x_A :
$W \equiv m^{x_A} (\bmod p)$.
– The prover embeds this watermark W into cover data I. Then the watermarked data I_W will be generated.

3. Constructing the Proof.

– *Setup*: The prover selects $t \in_R Z_q$.
– *First-order Proof*: The prover generates and computes what corresponds to a first order proof, i.e., the triple $(\overline{W}, \overline{\sigma}, \overline{m}) = (W^t, m^{tx_A}, m^t)$. And he sends it to the verifier.
– *Second-order proof*: The prover proves that $log_m \overline{m} = log_W \overline{W}$ and $log_{\overline{m}} \overline{\sigma} = log_g y_A$. The proofs of equality of discrete logs are based on non-interactive oblivious decision proofs [10].

4. Validating the Proof (Watermark extraction algorithm D)

– The verifier extracts the embedded pseudo-watermark $W^{'}$ from the $I^{'}$. While participating in the equality of discrete log proofs, he considers the extracted watermark.
– A verifier accepts the proof iff $\overline{W} = \overline{\sigma}$ and if he is convinced with the *second order proof.*

Multiple Ownership: Clearly the above protocol can also extended in the case of multiple ownership problem [4], that is, the case of image to be owned by more than one entity. The proof is very similar to the case of oblivious multiparty protocol given in [10]. Note that in the literature, *multiple ownership* is also known as *joint copyright protection.*
Security of the scheme: Our scheme is based on oblivious decision proofs, which are provably secure under random oracle model as showed in [10].
Practicality aspects: Since the watermark embedding and extraction procedure is similar to Yoo *et al* [14], so their implementation also work here too.

6 Conclusion

The importance of (non-interactive) designatedly verifiable non-invertible watermarking schemes is inevitable. Firstly, they solve Craver *et al*'s [4] multiple watermarking problem and secondly they restrict, by whom verification to be done. In this paper, we first proposed cryptanalytic attack on Yoo *et al* scheme [14] (and on Lee *et al* scheme [11] is similar). Then a modification for the resistance against the attack and we also proposed a new scheme which is robust (as our proof for validity of watermark is based on a provable oblivious decision proof). Importantly, our scheme can be used for solving "joint copyright protection".

References

1. D Chaum, *Undeniable signatures*, CRYPTO'89, LNCS 435, Springer-Verlag, pp. 212-216, 1990.

2. I Cox, J Bloom and M Miller, *Digital Watermarking, Principles & Practice*, Morgan Kaufmann, 2001.
3. I Cox, J Kilian, T Leighton and T Shamoon, *A secure, robust watermark for multimedia*, Information Hiding, LNCS 1174, Springer-Verlag, pp. 185-206, 1996.
4. S Craver, N Memon, B Yeo and M Yeung, *Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks, and Implications*, IEEE Journal on Selected Areas in Communications 16(4), pp. 573-586, 1998.
5. Y Desmedt, C Goutier, and S Bengio, *Special uses and abuses of the Fiat- Shamir passport protocol*, CRYPTO'87, LNCS 293, Springer-Verlag, pp. 21-39, 1988.
6. Y Desmedt and M Yung, *Weaknesses of undenaiable signature schemes*, EUROCRYPT'91, LNCS 547, Springer-Verlag, pp.205-220, 1992.
7. FIPS PUB 186, *Digital Signature Standard*, February 1, 1993.
8. M Jakobsson, *Blackmailing using undeniable signatures*, EUROCRYPT'94, LNCS 950, Springer-Verlag, pp.425-427, 1994.
9. M Jakobsson, K Sako, and R Impagliazzo. *Designated Verifier Proofs and Their Applications*, EUROCRYPT'96, LNCS 1070, Springer-Verlag, pp. 143-154, 1996.
10. M Jakobsson and C Schnorr, *Efficient Oblivious Proofs of Correct Exponentiation*, Communications and Multimedia Security (CMS), Kluwer Academic Publishers, pp. 71-84, 1999.
11. H Lee and I Lee, *Designated Verification of Digital Watermark for Network Based Image Distribution*, ICCS'03, LNCS 2660, Springer-Verlag, pp.1069-1078, 2003.
12. V Monga and B L Evans, *Robust perceptual image hashing using feature points*, IEEE Conference on Image Processing, Vol. 3, pp.677-680, 2004.
13. G Wang, *An attack on not-interactive designated verifier proofs for undeniable signatures*, e-print archive, 2003.
14. H Yoo, H Lee, S Lee, and J Lim, *Designated Verification of Non-invertible Watermark*, ISC'03, LNCS 2851, Springer-Verlag, pp.338-351, 2003.