International Journal of Cryptology Research 1 (1): 55-64 (2009)

Design Flaws of a Secure Watermarking Scheme for Buyer-Seller Identification and Copyright Protection

Geong Sen Poh and Keith M. Martin
Information Security Group, Royal Holloway, University of London,
Egham, Surrey, TW20 0EX, United Kingdom
Email: {g.s.poh, keith.martin}@rhul.ac.uk

ABSTRACT

A buyer-seller protocol for content rights protection deters dishonest buyers from illegally distributing bought content. This can be achieved by giving the seller the capability to trace and identify these buyers, while also allowing the seller to prove illegal acts to a third party. Many protocols have been proposed, one of the most recent being the protocol of Ahmed et al. in the EURASIP Journal on Applied Signal Processing in 2006. This protocol is interesting in that it uses a new design method, but we show that its claims of copyright infringement protection and buyer-seller identification can be defeated. We remark that the main reason for the security flaws is due to the misinterpretation of the design principles of a buyer-seller protocol, in which we give a brief definition, and further suggest a way forward in designing protocols of this nature.

INTRODUCTION

The ease of copying digital content has raised security concerns about illegal distribution of copyright materials. One of the many methods proposed to address this concern is copy deterrence through fingerprinting schemes [2][16]. These schemes were proposed as a mechanism to allow a user (e.g. a seller) to embed a unique watermark into content such as digital images and movies. The seller can then trace and identify a dishonest buyer if an illegal copy is found. However, fingerprinting schemes do not give the seller the ability to prove to a third party the fact that a dishonest buyer has illegally distributed content. This is because a buyer can claim that it is the seller who distributed it since the seller owns the watermark, which can be the case when a scrupulous seller is trying to frame a buyer. Hence asymmetric fingerprinting schemes [3][12][13] and buyer-seller watermarking protocols [4][6][8][9][10] were proposed to address these issues. One of the most recent protocols was proposed by Ahmed et al. [1]. This protocol is of interest because it uses a different design approach, notably it does not require any special cryptographic primitive such as homomorphic encryption scheme [11]. It claims to provide:

- *buyer-seller identification*, in which the buyer can reveal the identity of the seller and his or her own identity when the marked content is received;

- *copyright infringement protection*, in which the seller can trace and identify a buyer from an illegal copy, and further prove this fact to a third party;
- *ownership verification*, in which the seller can claim ownership of content when multiple ownership claims occur.

We show how the provisions of *buyer-seller identification* and *copyright infringement protection* are flawed, unless both the seller and the buyer are *fully trusted*, which in this case defeat the very purpose of constructing the protocol, where the seller and the buyer are assumed to be mutually distrustful parties.

In the rest of the paper we briefly discuss the properties of buyer-seller watermarking protocols, followed by an explanation of notation and building blocks. We then describe Ahmed et al.'s protocol and define our attacks.

Finally we analyse Ahmed et al.'s protocol based on our attacks before concluding with some design recommendations.

PROPERTIES OF BUYER-SELLER WATERMARKING PROTOCOLS

In this section we briefly define the main properties of a buyer-seller protocol. By defining these properties we compare the difference between the standard definitions that can be found in existing protocols [4][6][8][9][12] with Ahmed et al.'s claimed properties.

- *Traceability.* A legitimate, but dishonest, buyer who illegally distributes purchased content can be traced by the seller.
- *Framing Resistance.* An honest buyer cannot be falsely accused of illegal distribution by other parties.
- *Non-repudiation of redistribution.* A dishonest buyer found to have illegally redistributed purchased content cannot deny this fact by claiming that these copies were created and distributed by the seller. In other words, the seller obtains proof of illegal activity of the buyer.

We observe that the proposal by Ahmed et al. aims to provide *traceability* and *non-repudiation of redistribution* under the tag of copyright infringement protection. However *framing resistance* was not considered in their proposal (although they mention that it has been addressed in [10]). We will demonstrate that not considering framing resistance is in fact a critical failure which makes this protocol unable to provide copyright infringement protection. We note that our observations are independent of (and thus we

do not further consider) the novel ownership verification property of Ahmed et al.'s protocol.

NOTATION AND BUILDING BLOCKS

In this section we explain the notation required and provide a brief description of the building blocks needed to illustrate Ahmed et al.'s protocol and the subsequent analysis. Details can be found in [1].

Notation. In Table 1 we list the main entities involved and also the objects exchanged between them.

TABLE 1 : Notation

Entities/Objects	***Descriptions***
B	*Buyer*
S	*Seller*
RC	*Registration Center who issues key pairs for the buyer and the seller*
A	*An arbiter settles dispute of illegal distribution between the buyer and the seller*
X	*An original content*
V	*Seller's copyright ownership watermark to uniquely identify a buyer*
W	*Buyer's watermark*
X'	*Intermediate content where V is embedded into X*
X''	*Marked content where W is embedded into X'*
X^*	*Illegal copy of a marked content*
Y^*	*Seller's watermark extracted from X^**
W^*	*Buyer's watermark extracted from X^**

Building Blocks. The protocol requires a standard digital signature scheme with message recovery such as RSA [14]. We denote the signature generation algorithm with signing key skY as $Sig_{skY}(.)$ and the signature verification algorithm with verification key vkY as $Ver_{vkY}(.)$. The protocol also requires a standard cryptographic hash function, such as SHA-2 [7]. We denote this hash function as $H(.)$.

We describe briefly the watermark embedding and extraction algorithms proposed in the protocol. Note that we only describe the technique used, which is important for the analysis we present later. Details of these algorithms can be obtained from the original paper [1].

Two embedding algorithms were proposed, but here we generalize them into one algorithm, $\mathrm{Emb}(.)$, since both of them use the spread spectrum watermarking technique of Cox et al. [5]. Briefly, Emb(.) takes as input a watermark $V = (v_1, v_2, ...v_n)$ and content $X = (x_1, x_2, ..., x'_n)$, both sequences of real numbers, and outputs the marked content $X' = (x'_1, x'_2, ..., x'_n)$ and optionally a watermark key, in which the embedding process is defined by:

$$x'_i = x_i(1 + \alpha v_i) \qquad or \qquad x'_i = x_i + \alpha v_i \qquad for \ 1 \le i \le n, \tag{1}$$

where α is a real number determined by how much degradation of X is allowed when V is embedded.

Similarly, there are two detection algorithms, which again we simplify into one algorithm, $Det(.)$. On input of a seller's watermark and possibly marked content, $Det(.)$ outputs a watermark. We will not discuss how detection works as it does not affect our analysis. It suffices to note that one of the proposed detection algorithms uses a special primitive known as Independent Component Analysis (ICA) [15], a blind source separation technique to *separate* the content and the watermark during detection.

AHMED ET AL.'S COPYRIGHT PROTECTION PROTOCOLS

In this section we describe, at a conceptual level only, the three protocols proposed by Ahmed et al. [1]. These are the watermarked image generation and distribution protocol, the buyer-seller identification protocol and the copyright infringement protocol.

Also, it is assumed that there is an initialisation phase before the execution of the protocol, where each party Y is issued with a key pair (skY, vkY). The verification key vkY allows other parties to authenticate Y.

Watermarked Image Generation and Distribution Protocol

Figure 1 illustrates the protocol flow. Buyer B initiates the protocol by sending a purchase request to seller S. Next the seller S contacts the registration centre RC to request a seller's watermark V. After verifying the request, RC provides V and a signature consisting of the time of the request and other important information, to S. After receiving this message, S embeds V into content X, resulting in the marked content X'. In addition,

S computes $B's$ watermark, $W = H(vkS, vkB)$ and generates the marked content and a watermark public key, $(X'', K_S) = Emb(W, X')$. The marked content X'' the key K_S and a hash value $h_V = H(V)$ are sent to B together with a signature to ensure the origin and integrity of these messages. When B receives these messages, he obtains a watermark $W^* = Det(K_S, X'')$, and verifies W^* by comparing it with the hash value resulted from $H(vkS, vkB)$. If both are identical, then B can be sure that W^* reflects $B's$ and $S's$ identities. Next B acknowledges receiving content by generating a signature $S_B = Sig_{skB}(H(X'', h_v, vkS))$ and sends it to S. Finally, S checks the signature S_B.

Buyer-Seller Identification Protocol

This protocol is used by B (or other authorised parties) to verify that B is the legitimate buyer of the marked content X''. The buyer B computes $W^* = Det(K_S, X'')$ and verifies the watermark W^* by comparing it with the hash value $H(vkS, vkB)$.

Copyright Infringement Protocol

The seller S nitiates this protocol when an illegal copy of content X^* is found. It consists of two stages:

Stage I. A judge follows the *Buyer-Seller Identification Protocol*. If W^* extracted from X^* matches $H(vkS, vkB)$ then B is guilty. However, B is capable of removing W from the original marked content since he knows W. So if W^* cannot be found from W^*, the judge moves to **Stage II**.

Stage II. The seller S supplies the judge X, X'', X^*, V and S_B. The judge extracts V^* from X^* based on X'' and retrieves $h^* = H(X'', h_V, vkS)$ from S_B. Next the judge computes $h = H(X'', h_v, vkS)$ based on X'' and V supplied by S. If $h^* = h$ then the judge is sure that B bought content X'' from S and B is considered guilty when V^* matches V and X^* is similar to X''.

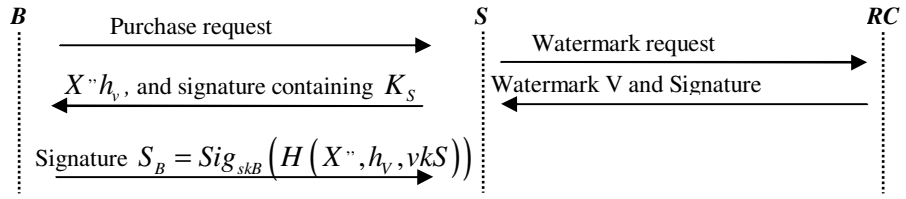

Figure 1: Watermarked Image Generation and Distribution

ATTACKS

We now define our attacks. The first two attacks are common to buyer-seller watermarking protocols [4][6][8][9], but not explicitly defined in these previous protocols. The third attack is a common attack on the underlying embedding algorithms [5], while the fourth attack is a new attack that we have defined based on the design weakness of Ahmed et al's protocol.

Framing Attack. This is an attack where the seller has complete knowledge of all the watermarks required in a protocol execution. Given knowledge of these watermarks, the attack is successful if the seller S is able to prove to the arbiter A that illegal copies of marked content belong to a particular buyer B even though B has not bought this content, or has bought this content but did not distribute copies of it illegally.

Buyer-denial Attack. This is an attack where a buyer B who illegally distributed copies of content attempt to deny doing so when confronted by the seller S and the arbiter A. The attack is successful if the buyer B is capable of showing that S has the capability of launching a framing attack against him.

Watermark-removal Attack. This is an attack where a party attempts to remove or replace a watermark from content and hence make the watermark untraceable, or trace to a different identity. The attack is successful if this attempt is successful.

Buyer-runaway Attack. This is an attack where a buyer B "runs away" with the received message and halts the protocol. The attack is successful if the buyer B is able to halt the protocol with marked content safely received without sending back an acknowledgement signature to S as evidence that B bought this content.

ANALYSIS

In this section we analyse Ahmed et al's protocol against the four attacks defined previously.

Framing and Buyer-denial attacks

S knows both the watermarks V and W and can launch a framing attack on the buyer. As can be observed from the protocols previously discussed, when an illegal copy X^* is found, S identifies B by extracting V^* and/or W^* from this copy. After doing so, S searches his database for the acknowledgment signature S_B to prove that B indeed bought the content that is similar to the illegal copy. We argue that B can be framed by a scrupulous seller S since S knows both V and W and can embed these watermarks into any content. Even worse, S has S_B and can prove that B bought the content when instead copies of it may have been illegally distributed by S. Conversely, *B can launch the buyer-denial attack* and claim that it is S who distributed illegal copies of content. This creates a deadlock scenario and can only be solved if we assume *either S or B is honest*. Hence as long as the scenario of framing or denial exists, the seller can not prove to a third party that the buyer is guilty, and as Ahmed et al. did not consider the possibility of framing attack, we would argue that proving B guilty of illegal content distribution based on the watermarks V, W and B's acknowledgment signature in their protocol is flawed.

Buyer-runaway attack

A buyer B can launch the buyer-runaway attack by choosing not to send back the acknowledgment signature S_B since B has already obtained the content. Without S_B, there is no way for S to prove to a third party that an illegal copy of bought content belongs to B. While S can always request B to send S_B, a dishonest buyer B can either refuse (or pretend) that he has sent it to S. The only resolution in the above scenario is for S to blacklist B from the client database. However, note that if S's main purpose is to trace and blacklist malicious buyers, a fingerprinting scheme [16] where S generates, embeds and traces the watermark into content is sufficient. There is no need for an interactive protocol like the one proposed. So in this case *B must be honest* for S to be able to prove B's dishonesty, which is a contradiction.

Watermark-removal attack

A buyer B can launch a watermark-removal attack by computing the watermark $W = H(vkS, vkB)$ and trivially subtract W from the marked content (which is also mentioned as possible by the Ahmed et al. in their paper). What we stress here is that B can also *replace* the watermark since W is the hash value of two public verification keys. In this case B can hash any other buyer's verification key together with the seller's verification key, resulting in a new watermark $W^U = H(vkS, vkU)$, and can then embed this watermark into content, which totally dismisses the credibility of the identification protocol and Stage I of the copyright infringement protocol. The watermark can be replaced as below (following formula (1)):

$$x^U = x_i(1+\alpha w_i)/(1+\alpha w_i).(1+\alpha w^U) \; or \; x^U{}_i = x_i + \alpha w_i - \alpha w_i + \alpha w^U{}_i$$

$$for \; 1 \le i \le n \qquad (2)$$

We further note that in fact anyone authorized to check watermark W can launch a watermark-removal attack since vkS and vkB are in the public domain. So watermark W seems to not serve any purpose at all. Identification of the buyers still requires the seller S to detect watermark V from any content, rendering the process of embedding and detecting W redundant, where a conventional fingerprinting scheme is sufficient based on embedding and detecting V.

CONCLUSION

We have analysed a buyer-seller protocol for content rights protection and showed that it contains serious flaws by demonstrating four attacks that can be run against it. One of the reasons why these flaws occurred appears to be due to a misinterpretation of the basic properties required by a buyer-seller watermarking scheme. We have shown that by underestimating the need for framing resistance, the protocol of Ahmed et al was left unable to function as a secure buyer-seller watermarking protocol. It is thus important to identify the correct properties required by such protocols prior to any design attempt. A second possible reason for the weaknesses in this protocol is the lack of a sound design framework. In order to minimize the risk of design flaws, we suggest that any secure buyer-seller watermarking protocol requires a clear statement of:

- the trust assumptions concerning the parties involved;
- who has potential knowledge of the embedded watermarks;

- the attack scenario (capability of the adversaries);
- the properties required.

REFERENCES

[1]. Ahmed, F., Sattar, F., Siyal, M. Y. and Yu, D. 2006. A Secure Watermarking Scheme for Buyer-Seller Identification and Copyright Protection. *EURASIP Journal on Applied Signal Processing*, 2006 (Article ID 56904), 15 pages.

[2]. Blakley, G. R., Meadows, C. and Purdy, G. B. 1985. Fingerprinting Long Forgiving Messages. In Williams, H. C. (editor), *Advances in Cryptology – CRYPTO '85*, volume 218 of *Lecture Notes in Computer Science*, 180-189, Springer-Verlag.

[3]. Camenisch, J. 2000. Efficient Anonymous Fingerprinting with Group Signatures. In Okamoto, T. (editor), *Advances in Cryptology – Asiacrypt 2000*, volume 1976 of *Lecture Notes in Computer Science*, 415-428, Springer-Verlag.

[4]. Choi, J.-G., Sakurai, K. and Park, J.-H. 2003. Does It Need Trusted Third Party? Design of Buyer-Seller Watermarking Protocol without Trusted Third Party. In Zhou, J., Yung, M. and Han, Y. (editors), *Applied Cryptography and Network Security - ACNS 2003*, volume 2846 of *Lecture Notes in Computer Science*, 265-279, Springer-Verlag.

[5]. Cox, I. J., Kilian, J., Leighton, F. T. and Shamoon, T. G. 1997. Secure Spread Spectrum Watermarking for Multimedia. *IEEE Trans. On Image Processing*, 6(12), 1673-1687.

[6]. Goi, B.-M., Phan, Raphael C.-W., Yang, Y., Bao, F., Deng, Robert H. and Siddiqi, M. U. 2004. Cryptanalysis of Two Anonymous Buyer-Seller Watermarking Protocols and an Improvement for True Anonymity. In Jakobsson, M., Yung, M. and Zhou, J. (editors), *Applied Cryptography and Network Security - ACNS 2004*, volume 3089 of *Lecture Notes in Computer Science*, 369-382, Springer-Verlag.

[7]. ISO/IEC Standard. 2004. Information technology -- Security techniques -- Hash-functions -- Part 3: Dedicated hash-functions. ISO/*IEC 10118-3:2004*.

[8]. Ju, H. S., Kim, H. J., Lee, D. H. and Lim, J. I. 2002. An Anonymous Buyer-Seller Watermarking Protocol with Anonymity Control. In Lee, P. J. and Lim, C. H. (editors), *Information Security and Cryptology - ICISC 2002*, volume 2587 of *Lecture Notes in Computer Science*, 421-432, Springer-Verlag.

[9]. Lei, C.-L., Yu, P.-L., Tsai, P.-L. and Chan, M.-H. 2004. An Efficient and Anonymous Buyer-Seller Watermarking Protocol. *IEEE Trans. on Image Processing*, 13(12), 1618-1626.

[10]. Memon, N. and Wong, P. W. 2001. A Buyer-Seller Watermarking Protocol. *IEEE Trans. on Image Processing*, 10(4), 643-649.

[11]. Paillier, P. 1999. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In Stern, J. (editor), *Advances in Cryptology - EUROCRYPT '99*, volume 1592 of *Lecture Notes in Computer Science*, 223-238, Springer-Verlag.

[12]. Pfitzmann, B. and Schunter, M. 1996. Asymmetric Fingerprinting. In Maurer, U. M. (editor), *Advances in Cryptology - EUROCRYPT '96*, volume 1070 of *Lecture Notes in Computer Science*, 84-95, Springer-Verlag.

[13]. Pfitzmann, B. and Waidner, M. 1997. Anonymous Fingerprinting. In Fumy, W. (editor), *Advances in Cryptology - EUROCRYPT '97*, volume 1233 of *Lecture Notes in Computer Science*, 88-102, Springer-Verlag.

[14]. Rivest, R. L., Shamir, A. and Adleman, L. 1978. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Commun. of the ACM*, 2(2), 120-126.

[15]. Hyvärinen, A 1999. Survey on Independent Component Analysis. *Neural Computing Surveys*, 2, 94-128.

[16]. Wagner, N. R. 1983. Fingerprinting. *In IEEE Symposium on Security and Privacy*, 18-22.