ON THE IMPORTANCE OF ONE-TIME KEY PAIRS IN BUYER-SELLER WATERMARKING PROTOCOLS

David M. Williams, Helen Treharne, Anthony T.S. Ho
University of Surrey, Guildford, UK, GU2 7XH
d.m.williams@surrey.ac.uk

Adrian Waller
Thales Research & Technology (UK) Ltd, Reading, UK, RG2 0SB

Keywords: Buyer-Seller Watermarking; protocol; security; unbinding; customers' rights; one-time key pair.

Abstract: In this paper we emphasise the importance of unique certified one-time key pairs in Buyer-Seller Watermarking (BSW) protocols. We distinguish between *reactive unbinding* attacks, in which the seller *reacts* to illicit file sharing by fabricating further evidence of such activity, and *pre-emptive unbinding* attacks, in which the seller gains an advantage by taking action that *pre-empts* the file being shared. We then demonstrate the importance of certified one-time key pairs as a mechanism to avoid pre-emptive unbinding attacks by describing how a recently published BSW protocol is vulnerable to such an attack. We conclude that careful consideration needs to be applied to how best to enforce that the one-time key pairs are truly used just once.

1 INTRODUCTION

Copy deterrence mechanisms discourage illicit duplication and dissemination of copyrighted material by embedding an imperceptible identifying mark within content to trace any illicit file sharing back to the original owner (Barni and Bartolini, 2004).

However, the buyer must be assured that the copyright owner has proof if and only if an illegal act has taken place. The assurance that evidence cannot be fabricated against a buyer is known in the literature as *customers' rights* (Qiao and Nahrstedt, 1998).

Qiao and Nahrstedt observed that previous schemes, in which the seller chose the watermark, failed to protect the copyright. Even if a seller had acted honestly, and is convinced of the buyer's illegal activity, the seller remains unable to prove that it was not they that had leaked the watermarked content.

Memon and Wong proposed a Buyer-Seller Watermarking (BSW) protocol that aimed to resolve this customers' rights issue by withholding knowledge of the watermark or watermarked content from the seller (Memon and Wong, 2001). However, Lei *et al.* later uncovered an issue present in the Memon-Wong (MW) protocol that they coined the *unbinding problem* (Lei et al., 2004). They subsequently proposed their own protocol, their approach to preventing unbinding was twofold: bind a watermark to content using some signed message; and use one-time key pairs to avoid outdated information being used in subsequent transactions.

In this paper we distinguish between two forms of unbinding and emphasise the importance of one-time key pairs in preventing the latter of these. In Section 2 we provide background material on the MW protocol and the unbinding problem presented by Lei *et al.* We distinguish between two forms of unbinding, *reactive* and *pre-emptive* unbinding in Section 3.

The importance of one-time key pairs is demonstrated in Section 4. We omit one-time key pairs from the LYTC protocol leaving it vulnerable to a pre-emptive unbinding attack. We also present a pre-emptive unbinding attack on a recently published protocol in Section 4. Finally we discuss the danger in entrusting buyer's to randomly generate key pairs in Section 6 before drawing conclusions in Section 7.

2 The Unbinding Problem

In this section we describe the protocol proposed in (Memon and Wong, 2001) and the associated unbinding problem presented in (Lei et al., 2004).

The approach taken in (Memon and Wong, 2001) to resolve the customers' rights issue, was to restrict the seller to conduct the watermark embedding in the encrypted domain, using the properties of homomorphic encryption. Thus, knowledge of the watermark and watermarked content are withheld from the seller during embedding and thus the buyer cannot claim that a copy was released by the seller.

During the *watermark generation* phase of the protocol, the buyer receives encrypted watermarks from the trusted third party, signed to certify they are well-formed. The third party need not be involved in transactions between the buyer and seller thus he is said to be offline. Furthermore, the third party is not required to store any data.

In the *watermark insertion* phase of the protocol, the buyer initiates a transaction sending to the seller some encrypted watermark $S_{sk_t}(E_{pk_b}(w))$ generated and signed by a trusted third party during the watermark generation phase. The seller must also receive an indication of what cover material the buyer wishes to purchase, $arg(c)$, and certification of the buyer's public key. The encrypted watermark signed by the certification authority is also sent to the seller.

As the seller is now in possession of the encrypted watermark $E_{pk_b}(w)$ and can calculate the encrypted content $E_{pk_b}(c)$, they construct the encrypted digital content $E_{pk_b}(W_{wk_s}(c,w))$ by performing the embedding in the encrypted domain[1]. The seller produces $E_{pk_b}(W_{wk_s}(c,w))$ without ever knowing the watermark or the watermarked content in the clear.

It should be possible, once an illicitly shared file is intercepted by the seller, for the original owner to be traced and this proven to an arbitrator. The protocol relies on the buyer participating in the arbitration process, however if a buyer refuses to do so it is considered admission of guilt.

The protocol was shown to be flawed in (Lei et al., 2004). Lei *et al.* presented an *unbinding problem* apparent once the user has illicitly shared content. Should a single file be shared by a buyer, the seller may *react* by embedding this watermark into any other content in order to fabricate evidence of further illicit file sharing against the buyer.

Upon completion of a transaction in which they purchase the content C_1 with watermark W, the buyer B receives the encrypted watermarked content $E_{pk_B}(W_{wk_S}(C_1,W))$. Should B upload the decrypted content $W_{wk_S}(C_1,W)$ onto some file sharing network the seller S may later download the content and extract W in order to trace the piracy back to B. However, once S has extracted W they may embed it within some other content C_2 to produce $W_{wk_S}(C_2,W)$. Thus the evidence of illicit file sharing of C_2 (i.e., $W_{wk_S}(C_2,W)$ and $S_{sk_T}(E_{pk_B}(W))$) can be obtained by S at a time when C_2 has not been shared.

[1] An indexing watermark $\mathbf{v}$ is first embedded to avoid an exhaustive search being performed,. A permutation function σ is then applied whilst embedding the watermark c in the encrypted domain such that the buyer cannot know the signal embedded.

3 REACTIVE AND PRE-EMPTIVE UNBINDING

Lei *et al.* also describe another form of unbinding attack in which the seller gains an advantage by taking action that *pre-empts* the file being shared.

Upon completing a transaction in which they purchase the content C_1 with watermark W_1, the buyer B receives the encrypted watermarked content $E_{pk_B}(W_{wk_S}(C_1,W_1))$. However, during a second transaction, in which B wishes to purchase the content C_2 with watermark W_2, the seller S may choose to distribute the encrypted watermarked content $E_{pk_B}(W_{wk_S}(C_2,W_1))$ to B. Should B ever share the latter content then S may extract the watermark W_1 and embed it within the content C_1. Thus the evidence of illicit file sharing of C_1 (i.e., $W_{wk_S}(C_1,W_1)$ and $S_{sk_T}(E_{pk_B}(W_1))$) can be obtained by S at a time when C_1 has not been shared.

In the first attack scenario, the malicious seller *reacts* to the file sharing maliciously by subsequently extracting the watermark from the shared file and embedding it within another. This is only possible after the file sharing event has occurred. In this paper we shall refer to such an unbinding attack as *Reactive Unbinding*. This is as opposed to what we shall refer to as *Pre-emptive Unbinding* in which the seller gains an advantage by taking action that *pre-empts* the file being shared. The two attacks are only subtly different in the MW protocol, but we shall see that Lei *et al.* adopt different mechanisms to prevent each of the two forms of unbinding.

4 THE IMPORTANCE OF ONE-TIME KEY PAIRS

The approach taken in (Lei et al., 2004) to prevent unbinding, as illustrated in Figure 1, was twofold: bind a watermark to content using some signed message; and use one-time key pairs to avoid outdated information being used in subsequent transactions.

The one-time key pairs were proposed as a mechanism to prevent pre-emptive unbinding, although

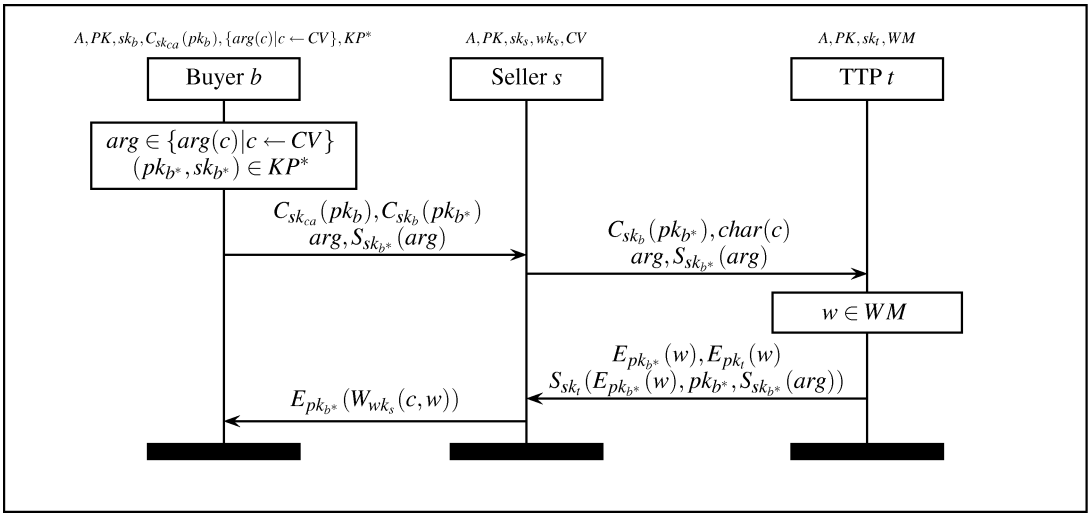

Figure 1: LYTC Protocol

other papers refer to them as anonymous keys such as (Deng and Preneel, 2008), (Ju et al., 2003) and (Shao, 2007). It is out of the scope of this paper to consider anonymity; instead we will demonstrate the importance of one-time key pairs for the overall security of the scheme.

4.1 Lei-Yu-Tsai-Chan (LYTC) Protocol

Figure 1 illustrates the LYTC protocol. The trusted third party is online during each run of the protocol, generating a unique watermark for each transaction. The buyer initiates the protocol by sending $S_{sk_{b^*}}(arg(c))$ to the seller, indicating the content they wish to receive. An anonymous certificate, signed by a certification authority, is sent from the buyer to the seller to certify the the buyer's public key, $E_{sk_{ca}}(pk_b)$, although a standard digital certificate may be used if privacy is not a concern.

The buyer constructs a one-time public-secret key pair (pk_{b^*}, sk_{b^*}) as the unique key pair used during the current transaction. This ensures that outdated information cannot be used in subsequent transactions and thus prevents pre-emptive unbinding. The buyer certifies the public key by constructing a second digital certificate $C_{sk_b}(pk_{b^*})$. The key pk_{b^*} is immediately used to verify the signed agreement sent by the buyer.

Upon verification of the signed agreement, the seller forwards $C_{sk_b}(b^*, pk_{b^*})$ and $E_{sk_{b^*}}(arg(c))$ to the third party. In order to ensure the watermark is suitably robust the seller also sends some characteristics $char(c)$ of the cover material[2].

[2]Lei *et al.* state that the seller may send c if unconcerned about doing so. Evidence of illicit file sharing can be fabricated by any agent in possession of the cover material and watermark. Hence, the third party must not know the cover material as they choose the watermark.

The trusted third party then constructs a robust watermark that is unique to this transaction. They use the public key received in the digital certificate $C_{sk_b}(pk_{b^*})$ to encrypt the watermark ready for use by the seller. It is sent along with the public key used to encrypt it and bound to the signed agreement by the trusted third party by signing a message constructed of all three components. It is this message that prevents *reactive unbinding*. They also encrypt the watermark under their own secret key in case they need to verify the watermark in the arbitration process.

Finally, the seller constructs the watermarked content in the encrypted domain. Once the buyer has received the encrypted, watermarked content they are able to decrypt in order to gain the useful watermarked content that they wished to purchase.

4.2 Omitting One-Time Key Pairs

By binding the watermark to the cover material via the signed message $S_{sk_t}(E_{pk_{b^*}}(w), pk_{b^*}, S_{sk_{b^*}}(arg))$, Lei *et al.* prevent the malicious seller from performing a *reactive unbinding* attack. However, this message alone does not protect against *pre-emptive unbinding*.

Let us suppose that the key pair need not be unique, then a buyer B may use the same key pair (pk_{B^*}, sk_{B^*}) in multiple transactions. Consider the first piece of content C^+ purchased as more expensive than a second piece of content C^-. In the first transaction the seller S receives the signed message $S_{sk_T}(E_{pk_{B^*}}(W), pk_{B^*}, S_{sk_{B^*}}(arg(C^+)))$ be-

fore distributing the encrypted watermarked content $E_{pk_{B^*}}(W_{wk_S}(C^+,W))$ to B.

Subsequently in a second run of the protocol B purchases C^- using the same key pk_{B^*}. As such, S omits the communication with WCA and instead performs the embedding with the same encrypted watermark $E_{pk_{B^*}}(W)$ as in the first transaction.

Should S ever intercept an illicitly shared copy of the less expensive watermarked content $W_{wk_S}(C^-,W)$ then W can be extracted and embedded into the higher priced content to produce $W_{wk_S}(C^+,W)$. This, along with the signed message received signed message from the first transaction $S_{sk_T}(E_{pk_{B^*}}(W), pk_{B^*}, S_{sk_{B^*}}(arg(C^+)))$, is then considered sufficient evidence of illicit file sharing of the more expensive content, when in fact the less expensive watermarked content was illicitly shared.

Should the encryption key pk_{B*} be unique to each transaction then it is not possible for the seller to perform the watermark embedding using an outdated encrypted watermark associated with a previous transaction. Hence, the uniqueness of the one-time key pairs must be assured for the LYTC protocol is not vulnerable to pre-emptive unbinding.

Before discussing how this impacts other BSW protocols in Section 6, we first present a pre-emptive unbinding attack on a recently published protocol.

5 A PRE-EMPTIVE UNBINDING ATTACK ON THE HU-ZHANG (HZ) PROTOCOL

Lei *et al.* included one-time key pairs to prevent pre-emptive unbinding. In this section we shall identify a pre-emptive unbinding attack on the Hu-Zhang (HZ) protocol, illustrated in Figure 2, due to the omission of one-time key pairs.

5.1 Hu-Zhang (HZ) Protocol

In (Hu and Zhang, 2009) a protocol was proposed aiming to increase the efficiency of multiple transactions. The trusted third party is not required to be online during a transaction between the buyer and the seller. As such the HZ protocol is subject to two phases, similar to the MW protocol: the watermark generation phase; and the watermark insertion phase.

In the watermark generation phase, Hu and Zhang propose the novel idea of enabling the buyer to request multiple signed encrypted well-formed watermarks at once. Upon receipt of the buyers digital certificate $C_{sk_{ca}}(b, pk_b)$ and the quantity n of watermarks required, the trusted third party randomly generates n unique watermarks $w_1, w_2, \ldots w_n \in WM$. Each is encrypted using the public key pk_b of the buyer and signed, along with the same public key. Thus for each watermark w_i a message $E_{pk_b}(w_i), S_{sk_t}(pk_b, E_{pk_b}(w_i))$ is sent from the trusted third party to the buyer along with certification of the buyer's public key.

In the watermark insertion phase the buyer chooses which watermark from the generation phase to use for the current transaction. The buyer sends to the seller a common agreement, along with a signature used to *bind* the watermark to the cover material, in the message $arg, S_{sk_b}(E_{pk_b}(w), arg)$. This is sent with messages m_w and m_b, received in the watermark generation phase. The seller verifies the signatures and embeds the watermark in the encrypted domain, sending the result $E_{pk_b}(W_{wk_s}(c,w))$ to the buyer.

5.2 A Pre-emptive Unbinding Attack

One-time key pairs are not used in (Hu and Zhang, 2009) and no alternative mechanism for preventing pre-emptive unbinding is provided, which leads to the following attack:-

Upon completing a transaction in which they purchase the content C^+ with watermark W, the buyer B receives the encrypted watermarked content $E_{pk_B}(W_{wk_S}(C^+,W))$. During a second transaction, in which the B purchases less expensive content C^- the seller S ignores the watermark received but instead embeds W received in the first transaction. Finally, the seller distributes the encrypted watermarked content $E_{pk_B}(W_{wk_S}(C^-,W))$.

Should the buyer share the less expensive content, S may extract W and embed it within the more expensive content C^+. Thus evidence of illicit file sharing of C^+ (i.e., $W_{wk_S}(C^+,W)$, $S_{sk_B}(E_{pk_B}(W), arg(C^+))$ and $S_{sk_T}(pk_B, E_{pk_B}(W))$) can be obtained by the seller at a time when C^+ has not been shared.

This attack closely follows the pre-emptive unbinding attack on the MW protocol described in Section 3. It differs only in what constitutes sufficient evidence of file sharing. It demonstrates that signing a message to bind the watermark to the cover material does not alone prevent unbinding as a mechanism also needs to be adopted to avoid outdated information being used in subsequent transactions.

Any BSW protocol that fails to adopt a mechanism for avoiding pre-emptive unbinding is vulnerable to attack. It has been demonstrated in (Poh and Martin, 2008) and (Deng and Preneel, 2008) that the protocol proposed in (Ibrahim et al., 2007) is flawed, however it is also vulnerable to the pre-emptive unbinding attack described in this section.

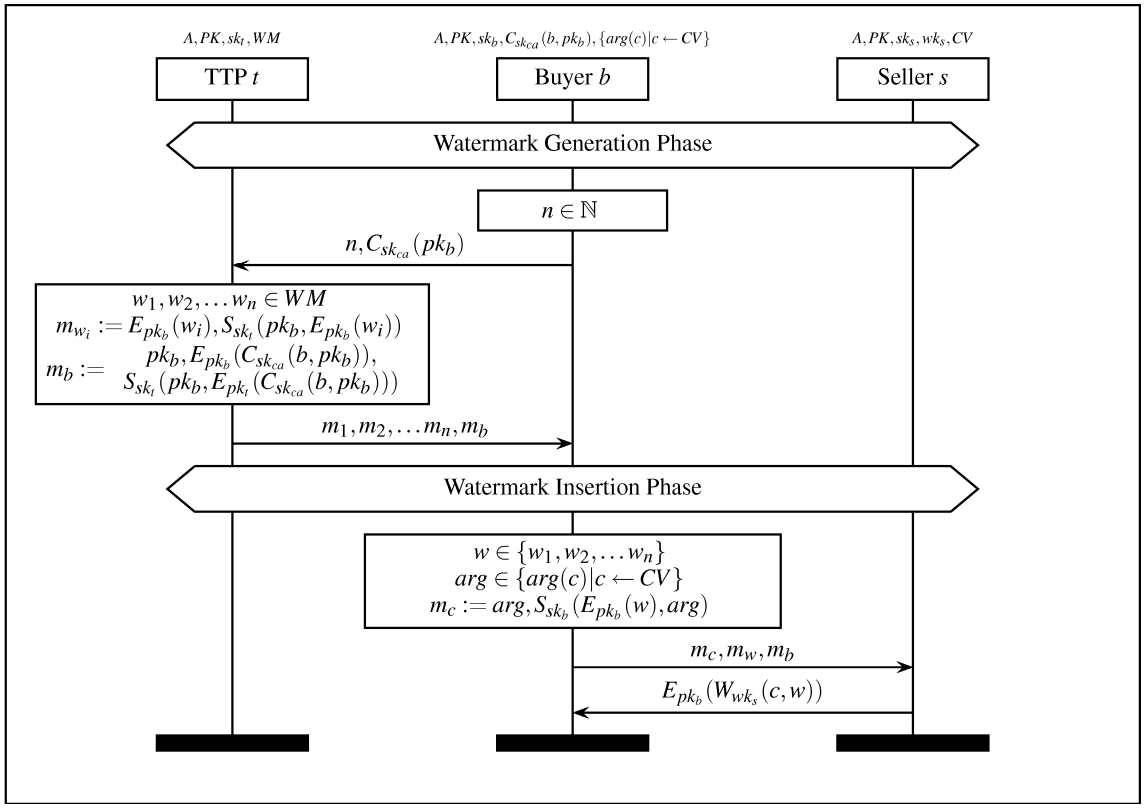

Figure 2: HZ Protocol

6 ASSURING KEY PAIRS ARE USED ONLY ONCE

From the customers' right issue, Qiao and Nahrstedt observed that watermarking schemes in which the seller chose the watermark to be embedded failed to protect the copyright (Qiao and Nahrstedt, 1998). Should the seller fail to address this issue an arbitrator is unable to discern whether it is the buyer or seller that has acted maliciously. Thus, it is in the interest of a dishonest buyer to act in a manner that renders themselves vulnerable to an attack as they may as a consequence construct a plausible denial of the illicit file sharing they wish to perform.

In (Williams et al., 2008a), the protocol proposed in (Ibrahim et al., 2007) was shown to be vulnerable to unbinding if the buyer leaves himself open to attack. The buyer can then share files without the seller being able to prove precisely which files were shared and escape prosecution. Although unbinding is only possible once a file is shared the fabrication of evidence implies a failure to resolve the customers' rights issue.

As it is in the interest of the buyer for the protocol to fail to protect their rights, this implies it is in their interest to leave themselves vulnerable to unbinding. We demonstrated in Section 4.2 that the resolution of the customers' rights issue in (Lei et al., 2004) is dependent upon the uniqueness one-time key pairs. An unbinding attack is possible on the LYTC protocol should the same key pair be used in multiple transactions. A similar vulnerability in (Shao, 2007) was presented in (Williams et al., 2008b) As preventing pre-emptive unbinding is dependent upon the uniqueness of the one-time key pairs it is apparent that entrusting the random generation of one-time key pairs to the buyer puts the protocols security at risk.

A natural choice of whom ensures the uniqueness of key pairs thus becomes the seller. The digital certificate $C_{sk_b}(pk_{b^*})$ must be checked against all other certificates used in previous transactions. Duplicate certificates must be rejected and the seller may later be required to prove this action to the arbiter. These may not be trivial tasks if the protocol is deployed on a large scale e-commerce system in which a great number of certificates must be stored and cross referenced in each transaction. Therefore careful consideration must be applied to how best to assure the uniqueness of the one-time key pairs.

7 CONCLUSION

In this paper we emphasised the difference between reactive and pre-emptive unbinding and demonstrated how individual mechanisms must be used to ensure that BSW protocols are vulnerable to neither form of attack.

One-time key pairs are the mechanism adopted by Lei *et al.* to avoid pre-emptive unbinding attacks in which the seller gains an advantage by taking action that *pre-empts* the file being shared. We have demonstrated that the omission of one-time key pairs from a BSW protocol may leave it vulnerable to a pre-emptive unbinding attack as illustrated by the attack on the HZ protocol in Section 5.

In Section 6 we demonstrated the advantage gained by the buyer in behaving in a manner that leaves him vulnerable to an attack. In such a scenario an arbitrator is unable to verify which protocol participant had acted improperly and to what extent. It should not therefore be left to the buyer to ensure that the same key pair is used in multiple transactions. As such we conclude that careful consideration needs to be paid to how best enforce that the key pairs are truly used just once.

REFERENCES

Barni, M. and Bartolini, F. (2004). *Watermarking Systems Engineering*. Marcel Dekker, Inc.

Deng, M. and Preneel, B. (2008). Attacks on two buyer-seller watermarking protocols and an improvement for revocable anonymity. *Electronic Commerce and Security, International Symposium*, 0:923–929.

Hu, Y. and Zhang, J. (2009). A secure and efficient buyer-seller watermarking protocol. *Journal of Multimedia*, 4(3):161–168.

Ibrahim, I., El-Din, S. N., and Hegazy, A. (2007). An effective and secure buyer seller watermarking protocol. In *Third International Symposium on Information Assurance and Security*, pages 21–28.

Ju, H., Kim, H., Lee, D., and Lim, J. (2003). An anonymous buyer-seller watermarking protocol with anonymity control. In *International Conference on Information Security and Cryptology*, pages 421–432.

Lei, C., Yu, P., Tsai, P., and Chan, M. (2004). An efficient and anonymous buyer-seller watermarking protocol. *IEEE Transactions on Image Processing*, 13(12):1618–1626.

Mauw, S. and Bos, V. (2001). Drawing Message Sequence Charts with LaTeX. *TUGBoat*, 22(1-2):87–92.

Memon, N. and Wong, P. W. (2001). A buyer seller watermarking protocol. *IEEE Transactions on Image Processing*, 10(4):643–649.

Poh, G. and Martin, K. (2008). On the (in)security of two buyer-seller watermarking protocols. In *SECRYPT 2008 - International Conference on Security and Cryptography*, pages 253–260.

Qiao, L. and Nahrstedt, K. (1998). Watermarking schemes and protocols for protecting rightful ownership and customer's rights. *Journal of Visual Communication and Image Representation*, 9(3):194–210.

Shao, M. (2007). A privacy-preserving buyer-seller watermarking protocol with semi-trust third party. In *Trust, Privacy and Security in Digital Business*, pages 44–53.

Williams, D., Treharne, H., Ho, A., and Culnane, C. (2008a). Using a formal analysis technique to identify an unbinding attack on a buyer-seller watermarking protocol. In *10th ACM Workshop on Multimedia and Security*, pages 205–214.

Williams, D., Treharne, H., Ho, A., and Waller, A. (2008b). Formal analysis of two buyer-seller watermarking protocols. In *7th International Workshop on Digital Watermarking*.

APPENDIX

The following notation is used throughout.

- A set of all agents
- WM, CV sets of all watermarks and cover material respectively
- PK set of all public keys
- b, s, t, ca variables to model agents, respectively buyers, sellers, trusted third parties and certification authorities
- w, c variables to model watermarks and content taken from the set of all watermarks WM and set of all digital content CV, respectively
- (pk_a, sk_a) public-secret key pair belonging to agent a
- wk_s watermarking key belonging to seller s
- $arg(c)$ common agreement identifying content c
- $C_{sk_{ca}}(a, pk_a)$ digital certificate binding an agent to their public key signed under sk_{ca}
- $C_{sk_{ca}}(pk_a)$ anonymous certificate signed under sk_{ca}
- $S_{sk_a}(m)$ message m signed under sk_a
- $E_{pk_a}(m)$ message m encrypted under pk_a
- $W_{wk_s}(c, w)$ watermark w embedded within c using watermarking key wk_s

Lowercase values are considered variable whereas those in uppercase are concrete. Protocols are presented as message sequence charts (Mauw and Bos, 2001) in conjunction this notation.