Anonymous Fingerprinting Based on Committed Oblivious Transfer*

Josep Domingo-Ferrer

Universitat Rovira i Virgili Department of Computer Science, Autovia de Salou s/n, E-43006 Tarragona, Catalonia, Spain jdomingo@etse.urv.es

Abstract. Thwarting unlawful redistribution of information sold electronically is a major problem of information-based electronic commerce. Anonymous fingerprinting has appeared as a technique for copyright protection which is compatible with buyer anonymity in electronic transactions. However, the complexity of known algorithms for anonymous fingerprinting deters their practical implementation, since they rely either on secure multiparty computation or on general zero-knowledge proofs. A scheme for anonymous fingerprinting based on committed oblivious transfer is presented in this paper where all computations can be performed efficiently.

Keywords: Secure electronic commerce, Intellectual property protection, Anonymous fingerprinting, Committed oblivious transfer.

1 Introduction

In information-based electronic commerce, copyright protection of the information being sold is a key problem to be solved, together with secure payment. Fingerprinting is a technique which allows to track redistributors of electronic information. Given an original item of information, an l-uple of marks is probabilistically selected. A mark is a piece of the information item of which two slightly different versions exist. At the moment of selling a copy of the item, the merchant selects one of the two versions for each mark; in other words, she hides an l-bit word in the information, where the i-th bit indicates which version of the data is being used for the i-th mark. Usually, it is assumed that two or more dishonest buyers can only locate and delete marks by comparing their copies (Marking Assumption, [Bone95]).

Classical fingerprinting schemes [Blak86] [Bone95] are symmetrical in the sense that both the merchant and the buyer know the fingerprinted copy. Even if the merchant succeeds in identifying a dishonest buyer, her previous knowledge of the fingerprinted copies prevents her from using them as a proof of redistribution in front of third parties. In [Pfit96], asymmetric fingerprinting was proposed,

 $^{^{\}star}$ This work is partly supported by the Spanish CICYT under grant no. TEL98-0699-C02-02.

H. Imai and Y. Zheng (Eds.): PKC'99, LNCS 1560, pp. 43-52, 1999.

[©] Springer-Verlag Berlin Heidelberg 1999

whereby only the buyer knows the fingerprinted copy; the drawback of this solution is that the merchant knows the buyer's identity even if the buyer is honest. Later ([Pfit97]) the concept of anonymous fingerprinting was introduced; the idea is that the merchant does not know the fingerprinted copy nor the buyer's identity. Upon finding a fingerprinted copy, the merchant needs the help of a registration authority to identify a redistributor. In [Domi98a], a scheme for anonymous fingerprinting is presented where redistributors can be identified by the merchant without help from the authority. The problem with the constructions [Pfit97][Domi98a] is that, being based on secure multiparty computation ([Chau88a]), their complexity is much too high to be implementable in practice. In [Domi98b], an anonymous fingerprinting algorithm is proposed which avoids secure multi-party computation and is based on Rabin's one-out-of-two oblivious transfer; however, this approach also relies on a (unspecified) general zero-knowledge proof whereby the buyer Bob shows to the merchant Mary that a hash value was correctly computed by the buyer.

1.1 Our Result

We present in this paper a scheme for anonymous fingerprinting which is efficiently and completely specified from a computational point of view. The basic primitive used is committed oblivious transfer (see [Crép95]).

Section 2 contains some background on committed oblivious transfer. Section 3 describes the new construction. Section 4 contains a complexity evaluation. Section 5 is a security analysis. Section 6 is a conclusion.

2 Background

In Subsection 2.1, bit commitment with XOR is recalled. This special kind of bit commitment has been shown to be useful for efficient implementation of committed oblivious transfer, which is a concept reviewed in Subsection 2.2.

2.1 Bit Commitment with XOR

In a bit commitment (BC), Mary sends a committed bit a to Bob in such a way that she is able to reveal it later in a *unique* way a but Bob is *not able to find* the value a by himself. Mary cannot change her mind and open a as a.

In [Crép95], bit commitment with XOR was introduced. If a special kind of bit commitments (BCX) is used, then it is possible to prove that some commitments satisfy an XOR-relation, without giving away any other information about the contents of the commitments. In particular, it is possible to prove that two BCXs \boxed{a} and \boxed{b} are equal simply by proving $\boxed{a} \oplus \boxed{b} = 0$; the verifier learns nothing about the bits contained in the commitments, except that they are equal or different.

Call BCX operations the following: creation of a BCX, opening of a BCX and proof that a constant number of BCXs satisfy a given linear relation. Then, if

m is the security parameter, it is argued in [Crép95] that each BCX operation can be implemented using O(m) BC operations, where BC denotes plain bit commitment. Unless otherwise specified, all commitments mentioned in the rest of this paper are BCX commitments.

2.2 Committed Oblivious Transfer

Oblivious transfer was originally invented by Rabin [Rabi81]. Mary has one secret; the protocol allows Bob to learn the secret with probability 1/2; whatever they do, Mary and Bob cannot modify the probability of Bob learning the secret; moreover, Mary cannot infer from the protocol whether Bob learned the secret or not. A slight variation of the above yields Rabin's one-out-of-two oblivious transfer, whereby Mary has two secrets and the protocol allows Bob to learn one of them; the probability of Bob learning either secret is 1/2; whatever they do, Mary and Bob cannot modify that probability; moreover, Mary cannot infer from the protocol which was the secret learned by Bob. A provably secure protocol for implementing Rabin's oblivious transfers can be found in [Berg85].

In a one-out-of-two oblivious transfer (OT), Bob has to choose between learning bit a_0 or a_1 prepared by Mary but she does not learn his choice b. If m is the security parameter, it is well known ([Crép88]) that OT can be constructed using O(m) of Rabin's oblivious transfers.

Now let us turn to committed oblivious transfer (COT). Suppose that Mary is committed to bits a_0 , a_1 and Bob is committed to bit b. After running COT(a_0 , a_1)(b) Bob knows a_b and is committed to a_b . Mary, whatever she does, cannot use the protocol to learn information on b and Bob, whatever he does, cannot use the protocol to learn information on $a_{\bar{b}}$.

COT was introduced in [Crép90] under the label "Verifiable Oblivious Transfer"; unfortunately, that first protocol used $O(m^3)$ OTs. In [Gold91], a more efficient protocol for COT was presented as "Preprocess-Oblivious-Transfer". In the best case, such a proposal requires $O(m^2)$ OTs. In [Crép95], a protocol for COT was proposed that used O(m) OTs and O(m) BCX operations (the latter can be replaced by $O(m^2)$ BC operations).

3 Anonymous Fingerprinting Based on Committed Oblivious Transfer

In this section, a fingerprinting scheme is presented which provides anonymity and has the advantage of being efficient and completely specified from a computational point of view. This was not the case for previous asymmetric and anonymous fingerprinting schemes.

3.1 Merchandise Initialization

Let the information *item* to be fingerprinted be n bits long. For i = 1 to n, the merchant Mary creates two versions $item_i^0$ and $item_i^1$ of the i-th bit $item_i$.

Both versions differ only for bit positions containing one mark (in the sense of Section 1).

Now, for i = 1 to n, Mary commits, using BCXs, to $item_i^0$ and to $item_i^1$ to get $item_i^0$ and $item_i^1$. The 2n BCXs are stored by Mary for later use.

Mary sends to the registration authority Ron a signed and time-stamped message containing a short description of item (but not the full item) as well as a list of the l < n bit positions in item containing a mark.

Note 1. The only reason to use BCXs for $item_i^0$ and $item_i^1$ instead of plain BCs is that $item_i^0$ and $item_i^1$ are used as inputs to a COT in the fingerprinting protocol of Subsection 3.3. As mentioned above and justified in [Crép95], using BCXs allows an efficient construction for COT (the XOR property is used only inside the COT construction).

3.2 Buyer Registration

Let p be a large prime such that q=(p-1)/2 is also prime. Let G be a group of order p, and let g be a generator of G such that computing discrete logarithms to the base g is difficult. Assume that both the buyer Bob and the registration authority Ron have ElGamal-like public-key pairs ([ElGa85]). Bob's secret key is x_B and his public key is $y_B=g^{x_B}$. The registration authority Ron uses his secret key to issue certificates which can be verified using Ron's public key. The public keys of Ron and all buyers are assumed to be known and certified.

Protocol 1

- 1. Ron chooses a random nonce $x_r \in \mathbf{Z}_p$ and sends $y_r = g^{x_r}$ to B.
- 2. Bob chooses secret random s_1 and s_2 in \mathbb{Z}_p such that $s_1 + s_2 = x_B \pmod{p}$ and sends $S_1 = y_r^{s_1}$ and $S_2 = y_r^{s_2}$ to Ron. Bob convinces Ron in zero-knowledge of possession of s_1 and s_2 . The proof given in [Chau88b] for showing possession of discrete logarithms may be used here. The buyer Bob computes an ElGamal public key $y_1 = g^{s_1} \pmod{p}$ and sends it to Ron.
- 3. Ron checks that $S_1S_2 = y_B^{x_r}$ and $y_1^{x_r} = S_1$. Ron returns to Bob a certificate $Cert(y_1)$. The certificate states the correctness of y_1 .

By going through the registration procedure above several times, Bob can obtain several different certified keys y_1 .

Note 2. If Bob is represented by his smart card, then the private key x_B is the smart card's private key, which is recorded in PROM by the card manufacturer or issuer. Having Bob represented by a tamper-proof smart card has several advantages, as will be discussed in Notes 3, 4 and 6 below.

3.3 Fingerprinting

If we denote by $item^*$ the fingerprinted copy of the original information item being sold, the fingerprinting protocol can be specified as follows:

Protocol 2

- 1. For i=1 to n, the merchant Mary shuffles the pairs $(item_i^0, item_i^1)$ to obtain $(item_i^{(0)}, item_i^{(1)})$. Mary records the result of the shuffling in the purchase record.
- 2. For i=1 to n, Mary and Bob run $COT(item_i^{(0)})$, $item_i^{(1)})$ (b_i), where b_i is a bit value chosen by Bob and b_i is a BCX. In this way, Bob obtains item_i^* and returns to Mary a signed commitment $item_i^*$ on it. This commitment is signed using the private key s_1 corresponding to the public key y_1 registered in Protocol 1.

Note 3 (Collusion-resistance). The information embedded in the fingerprinted copy is formed by bits b_i , for which $item_i^0 \neq item_i^1$. Assume that there are l such bits b_i , with $l \leq n$. If Bob takes part in the fingerprinting process through a tamper-proof device such as a smart card, then assumptions about the structure of the embedded information can be made. A possibility is for Mary to provide Bob's card with information on which are the l bit positions containing a mark; such information should be encrypted under the card's pseudonymous public key y_1 . Then the card could be programmed to choose the l embedded bits as a random codeword of a c-secure code ([Bone95]), which would provide protection against buyer collusions. Bob should not learn the codeword chosen by his card. It is worth mentioning that this way of using smart cards to counter buyer collusion also applies to the scheme described in [Domi98b] if OTs or COTs are used in that scheme instead of Rabin's oblivious transfers (i.e. if the buyer's card is allowed to input its choice to oblivious transfers).

Note 4. A second advantage of Bob taking part in the fingerprinting process through a tamper-proof smart card is that Step 1 of Protocol 2 is not needed. However, if the choice of b_i is known and controlled by Bob personally (instead of a smart card), shuffling is necessary because otherwise Bob could go twice through the fingerprinting protocol (perhaps under different pseudonyms), first with $b_i = 0$ and then with $b_i = 1$, which with probability 1 would allow him to discover whether $item_i$ contains a mark, i.e. whether $item_i^0 \neq item_i^1$. This would be against the marking assumption stated in Section 1.

3.4 Identification

Following [Pfit96], it only makes sense to try to identify a redistributor if the redistributed copy $item^{red}$ is not too different from the original item:

Definition 1. Let sim be an arbitrary relation where $sim(item^{red}, item)$ means that a redistributed illegal copy $item^{red}$ is still so close to item that the merchant Mary wants to identify the original buyer.

If $sim(item^{red}, item)$ holds, then it is reasonable to assume that $item^{red}$ contains a substantial number of bits which are (perhaps modified) copies of $item_1^*, \dots, item_n^*$, for some fingerprinted version $item^*$ of item.

Protocol 3

1. Upon detecting a redistributed item^{red}, Mary determines whether

$$sim(item^{red}, item)$$

holds for some information item on sale. If not, Mary quits the protocol.

- 2. Mary looks in her purchase record for all entries corresponding to sales of item. Each entry contains the buyer-signed BCX bit commitments for the fingerprinted copy item₁*, ..., item_n*.
- 3. Mary sends a signed and time-stamped copy of item^{red} to the authority Ron and all (pseudonymous) buyers having bought a copy of item. Requiring Mary to give away item^{red} for free to the suspect buyers is meant to thwart her from systematically and unjustly accusing all buyers of false redistributions. In other words, item^{red} represents no gift only for those buyers having purchased something similar to item^{red}.
- 4. Take all suspect pseudonymous buyers in turn and do the following until a redistributor is found or all buyers have been examined:
 - (a) Using a coin-flipping protocol, Mary and the pseudonymous buyer agree on $l_1 \leq l < n$ bit positions. If the resulting positions contain less than $l_2 \leq l_1$ marks, then Mary requests the buyer to start again the coin flipping protocol to agree on a new set of positions. The procedure is repeated until the resulting positions contain l_3 marks, with $l_2 \leq l_3 \leq l_1$.
 - (b) The pseudonymous buyer opens his BCX bit commitments corresponding to the l_1 agreed bit positions.
 - (c) If all l₃ opened commitments agree with the corresponding bit values in item^{red}, then Mary takes this as a proof of redistribution (see Note 6 below). Otherwise the suspect pseudonymous buyer is declared innocent and will be given by Mary a new fingerprinted copy (this is necessary because the buyer has been forced to reveal l₃ out of the l commitments in his fingerprinted copy; an honest buyer who is declared suspect several times might end up with virtually all his commitments opened).
- 5. Mary presents the opened signed commitments to the authority Ron asking for identification of the dishonest buyer. The opened commitments constitute a proof of redistribution, together with the signed item^{red} sent to Ron at Step 2 and the list of mark positions in item sent to Ron during merchandise initialization.

Note 5. If Ron refuses to collaborate in Protocol 3, his role can be performed by an arbiter except buyer identification and mark recognition. Replace "identify buyer" by "declare Ron guilty". If a suspect pseudonymous buyer refuses to collaborate, then the transcript of the protocol is sent to Ron, asking for identification. If the parameter l_2 is tuned properly, the risk of unjustly accusing a buyer is sufficiently low not to deter suspect buyers from proving their innocence (see Section 5).

Note 6. A third advantage of having buyers use tamper-proof smart cards during fingerprinting is that the embedded information (i.e. the set of marks) can be assumed to be a codeword of an error-correcting code with minimal distance d > 1. In this case, finding $l_3 - d + 1$ matches in Substep 4c of Protocol 3 suffices to declare a buyer guilty.

4 Complexity Analysis

The complexity of the construction of Section 3 is next assessed.

Merchandise initialization involves a digital signature and 2n BCX commitments, where n is the bitlength of the information item to be fingerprinted. If m is the security parameter, each BCX commitment requires O(m) plain bit commitments BC, as mentioned in Subsection 2.1. Thus merchandise initialization requires O(nm) BCs. However, notice that merchandise initialization is an off-line procedure that is only run once for each information item on sale.

The buyer registration protocol requires five exponentiations and a zero-knowledge proof for showing possession of discrete logarithms (an efficient protocol for such a proof can be found in [Chau88b]).

The fingerprinting protocol basically involves n committed oblivious transfers and n signatures (on the commitments resulting from the COTs). From Section 2.2, the n COTs are equivalent to O(nm) plain oblivious transfers OT and $O(nm^2)$ plain bit commitments BC.

The identification protocol requires opening O(n) BCX commitments. This is equivalent to opening O(nm) plain BCs. In addition, one instance of the fingerprinting protocol should be run for each suspect buyer who cannot be found guilty.

Note 7. Previously proposed anonymous fingerprinting protocols rely on computationally unspecified black boxes: secure multiparty computation in the case of [Pfit97] and [Domi98a] or a generic zero-knowledge proof in the case of [Domi98b]. Therefore, implementation of such protocols is far from obvious. The construction in this paper does not suffer from this problem, because it is based on well-known primitives.

5 Security Analysis

We analyze in this section the security of the construction of Section 3.

Proposition 1 (Registration security). Protocol 1 provides buyer authentication without compromising the private key x_B of the buyer.

Proof. In registration, the authority Ron sees S_1 , S_2 , y_1 and a zero-knowledge proof. The latter leaks no information. Without considering the zero-knowledge proof, Ron needs no knowledge of x_B to find values S'_1 , S'_2 and y'_1 which are related in the same way as S_1 , S_2 and y_1 . Take a random s'_1 , then compute $y'_1 = g^{s'_1}$ and $S'_1 = y^{s'_1}$. Finally, $S'_2 = y^{x_r}_B/S'_1$.

Now consider the zero-knowledge proofs; imagine that an impersonator not knowing x_B can compute S_1 , S_2 such that he/she can demonstrate possession of $\log_{y_r} S_1$ and $\log_{y_r} S_2$ and $S_1 S_2 = y_r^{x_B}$ holds. Then the impersonator can compute the discrete logarithm x_B . In general, if impersonation is feasible, so is computing discrete logarithms. \diamond

Proposition 2 (Buyer anonymity). Let l_2 be the minimal number of marks to be opened by a suspect buyer in the identification protocol. Then the probability that the merchant identifies an honest buyer who correctly followed Protocol 2 is upper-bounded by 2^{-l_2} .

Proof. In the fingerprinting protocol, Mary sees a pseudonym y_1 , which is related to y_B by the equation $y_1^{x_r}S_2 = y_B^{x_r}$. Even knowledge of $\log_g y_r = x_r$ would not suffice to uniquely determine y_B from y_1 , since S_2 is unknown to Mary.

Thus Mary must rely on Protocol 3 to unduly identify an honest buyer. Suspect but honest buyers are not especially vulnerable since they are given a new fingerprinted copy if they cannot be proven guilty. So the only strategy is for Mary to fabricate an $item^{red}$ with the hope that the $l_3 \geq l_2$ bit positions agreed upon by coin-flipping will contain the same values than the l_3 commitments opened by the buyer. Since the n cots performed by Mary and the buyer during fingerprinting do not allow Mary to learn anything about the buyer's choices b_i (see [Crép95]), the probability of unlawful identification is $2^{-l_3} \leq 2^{-l_2}$. \diamond

Merchant security depends on the marks being preserved. The next proposition shows that, for a non-colluding redistributor to remain undetected, the fingerprinted copy must be modified substantially and randomly.

Proposition 3 (Merchant security). In order to remain undiscovered after the identification protocol, a non-colluding redistributor must modify on average n/l_2 randomly chosen bits of the fingerprinted copy. This number can be made large by choosing $l_2 \ll n$.

Proof. Since the redistributor does not know where the marks are, his only possibility is random search. The probability that modification of one bit of the fingerprinted copy results in modification of one of the l_2 marks opened during the identification protocol is l_2/n . Thus, to ensure modification of one mark, n/l_2 randomly chosen bits of the fingerprinted copy must be modified on average. \diamond

Collusion is another strategy for buyers to delete marks. In Note 3, the use of tamper-proof smart cards was sketched as a way to obtain collusion-secure fingerprinting. If no smart cards are used, then we can only state the following:

Proposition 4. The expected percent of marks that can be deleted by a collusion of c buyers is $100(1-1/2^{c-1})$.

Proof. By the marking assumption, if the *i*-th bit position of *item* contains a mark, c colluding buyers can locate (and delete) this mark if and only if they can pool two bit versions $item_i^0$ and $item_i^1$ such that $item_i^0 \neq item_i^1$. Thanks to the shuffling step in Protocol 2, buyers cannot control which version of the *i*-th bit is delivered to them. Thus, the probability that all c buyers were given the same version is $1/2^{c-1}$. Therefore, the probability that they can pool both versions is $1-1/2^{c-1}$. \diamond

Merchant security also depends on the kind of similarity relation sim used (see Subsection 3.4). If sim is very loose, this means that Mary wishes to identify the original buyer of any redistributed item that vaguely resembles an item on sale; of course, identification may often fail in such cases (the authority Ron is likely to deny identification).

6 Conclusion and Future Directions

To our best knowledge, we have presented the first construction for anonymous fingerprinting which is completely specified from a computational point of view and is thus readily implementable. Unlike previous proposals, the proposed construction relies only on computationally well-defined primitives. By properly tuning its security parameters, good buyer and merchant protection can be attained. In addition, if combined with smart cards for fingerprinting on the buyer's side, the construction also provides protection against collusions.

Future research should be directed to:

- Implementing all buyer functionality on a smart card. This may require further efficiency improvements.
- Speeding up the whole process. A possible way to speed up the fingerprinting protocol is to modify the protocol for COT proposed in [Crép95] so that what is transferred is not a single bit but an r-bit string. In the protocol described in [Crép95] a privacy amplification function $h: \{0,1\}^m \to \{0,1\}$ is used; to achieve the desired speed-up, one could replace h with another privacy amplification function $h': \{0,1\}^m \to \{0,1\}^r$.

References

- Berg85. R. Berger, R. Peralta and T. Tedrick, "A provably secure oblivious transfer protocol", in Advances in Cryptology-EUROCRYPT'84, LNCS 209. Berlin: Springer-Verlag, 1985, pp. 408-416.
- Blak86. G. R. Blakley, C. Meadows and G. B. Purdy, "Fingerprinting long forgiving messages", in Advances in Cryptology-CRYPTO'85, LNCS 218. Berlin: Springer-Verlag, 1986, pp. 180-189.
- Bone95. D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data", in Advances in Cryptology-CRYPTO'95, LNCS 963. Berlin: Springer-Verlag, 1995, pp. 452-465.
- Chau88a. D. Chaum, I. B. Damgaard and J. van de Graaf, "Multiparty computations ensuring privacy of each party's input and correctness of the result", in *Advances in Cryptology CRYPTO'87*, LNCS 293. Berlin: Springer-Verlag, 1988, pp. 87-119.
- Chau88b. D. Chaum, J.-H. Evertse and J. van de Graaf, "An improved protocol for demonstrating possession of discrete logarithms and some generalizations", in *Advances in Cryptology- EUROCRYPT'87*, LNCS 304. Berlin: Springer-Verlag, 1988, pp. 127-141.
- Crép88. C. Crépeau, "Equivalence between two flavours of oblivious transfer", in Advances in Cryptology-CRYPTO'87, LNCS 293. Berlin: Springer-Verlag, 1988, pp. 350-354.
- Crép
90. C. Crépeau, "Verifiable disclosure of secrets and applications", in Advances in Cryptology-EUROCRYPT'89, LNCS 434. Berlin: Springer-Verlag, 1990, pp. 181-191.
- Crép95. C. Crépeau, J. van de Graaf and A. Tapp, "Committed oblivious transfer and private multi-party computation", in Advances in Cryptology-CRYPTO'95, LNCS 963. Berlin: Springer-Verlag, 1995, pp. 110-123.
- Domi98a. J. Domingo-Ferrer, "Anonymous fingerprinting of electronic information with automatic identification of redistributors", *Electronics Letters*, vol. 34, no. 13, June 1998.
- Domi
98b. J. Domingo-Ferrer and J. Herrera-Joancomartí, "Efficient smart-card based
anonymous fingerprinting", in *Preproceedings of CARDIS'98*. Louvain-la-Neuve: Université Catholique de Louvain, 1998.
- ElGa85. T. ElGamal, "A public-key cryptosystem and a signature scheme based on discrete logarithms", *IEEE Transactions on Information Theory*, vol. IT-31, July 1985, pp. 469-472.
- Gold91. S. Goldwasser and L. Levin, "Fair computation of general functions in presence of moral majority", in Advances in Cryptology-CRYPTO'90, LNCS 537. Berlin: Springer-Verlag, 1991, pp. 77-93.
- Pfit96. B. Pfitzmann and M. Schunter, "Asymmetric fingerprinting", in Advances in Cryptology-EUROCRYPT'96, LNCS 1070. Berlin: Springer-Verlag, 1996, pp. 84-95.
- Pfit97. B. Pfitzmann and M. Waidner, "Anonymous fingerprinting", in *Advances in Cryptology-EUROCRYPT'97*, LNCS 1233. Berlin: Springer-Verlag, 1997, pp. 88-102.
- Rabi
81. M. Rabin, How to Exchange Secrets by Oblivious Transfer, Technical Report
TR-81, Aitken Computation Laboratory, Harvard University, 1981.