Efficient Smart-Card Based Anonymous Fingerprinting [⋆]

Josep Domingo-Ferrer and Jordi Herrera-Joancomartí

Universitat Rovira i Virgili, Department of Computer Science, Autovia de Salou s/n,
E-43006 Tarragona, Catalonia, e-mail {jdomingo,jherrera}@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 is too high for practical implementation on standard computers, let alone smart cards. A scheme for anonymous fingerprinting is presented in this paper where all buyer computations can be performed by the buyer's smart card.

Keywords: Cryptographic protocols for IC cards, Electronic commerce, Anonymous fingerprinting, Intellectual property protection.

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, a tuple 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 n-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, 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

[⋆] This work is partly supported by the Spanish CICYT under grant no. TEL98-0699-C02-02.

is honest. Recently ([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 [Domi98], 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][Domi98] is that, being based on secure multiparty computation ([Chau88a]), their complexity is much too high to be implementable on standard computers, let alone smart cards.

1.1 Our result

Whereas algorithms for secure anonymous payment exist than can be implemented on a smart card on the buyer's side, no efficient anonymous fingerprinting algorithms exist in the literature where the buyer's computation can be carried out by the buyer's smart card. We describe in this paper a new construction for anonymous fingerprinting which keeps the buyer computation simple enough to be implemented by a smart card. In this way, all security functions needed to buy copyrighted information can be performed by a card.

Section 2 describes the new construction. Section 3 analyzes its security. Section 4 is a conclusion and a sketch of future work.

2 Anonymous fingerprinting without secure multiparty computation

In this section, a fingerprinting scheme is presented which provides anonymity and has the advantage of avoiding the secure two-party computation needed in previous asymmetric and anonymous fingerprinting schemes.

2.1 Merchandise initialization

Let $H()$ be a cryptographically strong block hash function. Let n, l and u be nonnegative integer security parameters agreed upon by all parties, where $l < u < n$.

The merchant M splits the information $item$ to be fingerprinted into n disjoint subitems $item_1$, $item_2$, $\cdots$, $item_n$ of similar length all of which must be concatenated to reconstruct the original $item$. In addition, subitems $item_1, \cdots, item_u$ contain one mark (in the sense of Section 1), *i.e.* there exist two slightly different versions $item'_i, item''_i$ of each subitem $item_i$, for $i = 1$ to u.

Note 1 (Marking redundancy). The existence of two versions of $item_i$ allows to embed one bit in the subitem. To make subitem marking resilient to intentional modification, a redundancy scheme may be used. A simple redundancy scheme can be to replicate the bit value embedded in a subitem an odd number $m > 1$ of times so that the m-bit vector $(0, 0, \cdots, 0)$ is embedded in $item'_i$ instead of the bit 0, and the m-bit vector $(1, 1, \cdots, 1)$ is embedded in $item''_i$ instead of the bit 1. To extract the value embedded in a redistributed subitem $item^{red}_i$, all m marks are examined by M and a majority decision is made to determine whether the value is 0 or 1. Note that, by the Marking Assumption, dishonest buyers can only locate and delete marks by comparing their copies, so a single buyer is unlikely to modify a majority of marks while preserving the usefulness of the information in the subitem. Two colluding buyers can delete or alter all marks in the subitem if and only if one of them was given $item'_i$ and the other $item''_i$.

2.2 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 B (in fact B is the buyer's smart card) and the registration authority R have ElGamal-like public-key pairs ([ElGa85]). The buyer's secret key is x_B and his public key is $y_B = g^{x_B}$. The registration authority R uses its secret key to issue certificates which can be verified using R's public key. The public keys of R and all buyers are assumed to be known and certified.

Protocol 1

1. *R chooses a random nonce $x_r \in \mathbf{Z}_p$ and sends $y_r = g^{x_r}$ to B.*
2. *B chooses secret random s_1 and s_2 in $\mathbf{Z}_p$ 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 R. B convinces R 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 B computes an ElGamal public key $y_1 = g^{s_1} \pmod p$ and sends it to R.*
3. *R checks that $S_1 S_2 = y_B^{x_r}$ and $y_1^{x_r} = S_1$. R returns to B a certificate $Cert(y_1)$. The certificate states the correctness of y_1.*

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

2.3 Fingerprinting

Fingerprinting is in some respects similar to secure contract signing. The following protocol exploits such a similarity.

Protocol 2

1. *B sends $y_1, Cert(y_1)$ and text to M, where text is a string identifying the purchase. B computes an ElGamal signature sig on text with the secret key s_1; sig is sent to M.*

2. *M verifies the certificate on y_1 and the signature sig on text.*
3. *For $i = 1$ to l, M sends one message out of the two messages $item'_i$ and $item''_i$ using the 1-2 provably secure oblivious transfer sketched in [Berg85]. If $item^*_i$ is the output of the oblivious transfer, it should be similar to the original meaningful $item_i$, so it should be easy for B to tell it from junk.*
4. *B computes an ElGamal signature $sig^*_{(l)}$ on $H(item^*_{(l)})$ using the key s_1, where*
$$item^*_{(l)} = item^*_1||item^*_2||\cdots||item^*_l \tag{1}$$
B returns $H(item^*_{(l)})$ and $sig^*_{(l)}$ to M. B proves to M in zero-knowledge that $H(item^*_{(l)})$ was correctly computed, i.e. *that it was computed based on the outputs $item^*_i$ of the oblivious transfers, for $i = 1$ to l. If B fails to return $H(item^*_{(l)})$ and the zero-knowledge proof, then M quits the fingerprinting protocol.*
5. *For $i = l + 1$ to u, M sends one message out of the two messages $item'_i$ and $item''_i$ using the 1-2 provably secure oblivious transfer sketched in [Berg85].*
6. *B computes an ElGamal signature $sig^*_{(u)}$ on $H(item^*_{(u)})$ using the key s_1, where*
$$item^*_{(u)} = item^*_{l+1}||item^*_{l+2}||\cdots||item^*_u \tag{2}$$
B returns $H(item^*_{(u)})$ and $sig^*_{(u)}$ to M. B proves to M in zero-knowledge that $H(item^*_{(u)})$ was correctly computed, i.e. *that it was computed based on the outputs $item^*_i$ of the oblivious transfers, for $i = l+1$ to u. If B fails to return $H(item^*_{(u)})$ and the zero-knowledge proof, then M quits the fingerprinting protocol.*
7. *M sends $item_{u+1}||item_{u+2}||\cdots||item_n$ to B.*

The following remarks on the security parameters n, l and u are in order:

- As will be justified in Section 3, it should be infeasible for M to figure out the values of $item^*_{(l)}$ and $item^*_{(u)}$ corresponding to a buyer B. Thus the sizes 2^l and 2^{u-l} of the spaces where $item^*_{(l)}$, respectively $item^*_{(u)}$, take values should be large enough (a good choice could be $l \geq 64$, $u - l \geq 64$ which implies $n \geq 128$).
- The buyer B can obtain up to l subitems from M and then quit Protocol 2 before Step 4. This means that up to l subitems can be obtained without copyright protection. *Thus l should be small as compared to n.*
- Once Step 6 is run, the first u subitems are copyright protected, but no protection is provided for the last $n - u$ subitems. *Thus u should not be much smaller than n.* However, the last $n-u$ subitems should contain enough information to deter B from quitting the Protocol 2 before Step 6, which in the worst case would leave subitems $l + 1$ up to u unprotected.

From the above remarks, a good choice would be to take $l = 64$, u such that $u-l \geq 64$, and n such that the last $n-u$ subitems contain the minimum amount of information needed to deter a standard buyer B from quitting Protocol 2 before Step 6.

If the computational resources of B are a bottleneck (as it may happen when B is a smart card), then a possibility is to suppress Steps 5 and 6 from Protocol 2 and send $item_{l+1}, \cdots, item_n$ in Step 7. In this case, the wisest choice is probably to take $l \approx n/2$.

Note 2 (Collusion-resistance). The information embedded in the fingerprinted copy is determined as the successive 1-2 oblivious transfers progress. If B takes part in the fingerprinting protocol through a tamper-proof device such as a smart card, then the 1-2 oblivious transfer from [Berg85] could be replaced by oblivious transfers where B's smart card can *choose* between $item'_i, item''_i$ (the choice remains unknown to M, see [Crép88]). Otherwise put, the card chooses the bit b_i to be embedded during the i-th oblivious transfer. B. Then B's card could be programmed to choose the embedded bits as a random codeword of a c-secure code ([Bone95]), which would provide protection against buyer collusions. B should not learn the codeword chosen by his card.

2.4 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 M 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 subitems which are (perhaps modified) copies of $item^*_1, \cdots, item^*_u$, for some fingerprinted version $item^*$ of $item$.

It must be noted that no redistributor identification can be requested by M if Protocol 2 is quit before Step 4 is run (M gets no receipt). In the following two-party identification protocol between the merchant M and the registration authority R, we will assume that Step 4 of Protocol 2 was run and that $item^{red}$ contains enough (perhaps modified) copies of subitems among $item^*_1, \cdots, item^*_l$ to allow reconstruction of $item^*_{(l)}$ by M:

Protocol 3

1. *Upon detecting a redistributed $item^{red}$, M determines whether*

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

holds for some information item on sale. If yes, M uses the redundancy scheme to recover the bit value that was embedded in each subitem of $item^{red}$ (notice that $item^{red}$ may not exactly correspond to any fingerprinted $item^*$). If the redundancy scheme is sufficient and $item^{red}$ contains enough (perhaps modified) subitems among $item^*_1, \cdots, item^*_l$, M can reconstruct in this way *the correct fingerprinted* $item^*_{(l)}$ *from which* $item^{red}$ *was derived (if a few subitems in* $item^*_{(l)}$ *had been suppressed or a majority of their marks had been modified in* $item^{red}$*, such subitems can be reconstructed by* M *using exhaustive search).*

2. *Once* $item^*_{(l)}$ *has been reconstructed,* M *retrieves the corresponding* $sig^*_{(l)}$*, text, sig,* y_1 *and* $Cert(y_1)$ *from her purchase record. Then* M *sends*

$$proof = [item^*_{(l)}, sig^*_{(l)}, text, sig, y_1, Cert(y_1)] \qquad (3)$$

to R *asking for identification of* B*.*

3. R *computes* $H(item^*_{(l)})$ *and uses the public key* y_1 *to verify that* $sig^*_{(l)}$ *is a signature on* $H(item^*_{(l)})$*. If the verification fails, this means that either* M *is trying to unjustly accuse a buyer or that the redundancy scheme was not sufficient to allow the correct reconstruction of* $item^*_{(l)}$*; in either case, identification fails. If the verification succeeds, the same key* y_1 *is used to verify that sig is a signature on the text identifying the purchase. Finally* R *searches its records to find the buyer* B *who registered the key* y_1 *and returns the name of* B *to* M*.*

If $item^{red}$ does not allow reconstruction of $item^*_{(l)}$ and Step 6 of Protocol 2 was run, then reconstruction of $item^*_{(u)}$ (equation (2)) can be attempted. To do this, take Protocol 3 and replace l by u and $item^*_1, \cdots, item^*_l$ by $item^*_{l+1}, \cdots, item^*_u$. Since usually $u - l > l$, $item^{red}$ is likely to contain more subitems of $item^*_{(u)}$ than of $item^*_{(l)}$.

Note 3. If R refuses to collaborate in Protocol 3, its role can be performed by an arbiter except buyer identification. Replace "identify buyer" by "declare R guilty".

3 Security analysis

We analyze in this section the security of the construction of Section 2.

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

Proof. In registration, R sees S_1, S_2, y_1 and two zero-knowledge proofs. The latter leak no information. Without considering the zero-knowledge proofs, R 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 $S'_1 = y_r^{s'_1}$. Finally, $S'_2 = y_B^{x_r}/S'_1$.

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 and u be the security parameters defined in Protocol 2. Then M must perform an exhaustive search in a space of size* $\min(2^l, 2^{u-l})$ *to unduly identify an honest buyer who correctly followed Protocol 2.*

Proof. In the fingerprinting protocol, M 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 M.

Therefore, M must rely on Protocol 3 to unduly identify an honest buyer. In this protocol, M must figure out either the value $item^*_{(l)}$ or the value $item^*_{(u)}$. Since the oblivious transfer [Berg85] is provably secure and B's proofs on the correctness of $H(item^*_{(l)})$ and $H(item^*_{(u)})$ are zero-knowledge, the only conceivable attack is for M to start trying all possible values for $item^*_{(l)}$ or $item^*_{(u)}$ until one is found such that either $H(item^*_{(l)})$ is the value contained in $sig^*_{(l)}$ or $H(item^*_{(u)})$ is the value contained in $sig^*_{(u)}$. With subitems having two versions each, $item^*_{(l)}$ is uniformly and randomly distributed over a set of 2^l different values, and $item^*_{(u)}$ is uniformly and randomly distributed over a set of 2^{u-l} different values. ◇

Merchant security depends on the marks being preserved. According to the Marking Assumption, marks can only be deleted by buyer collusion.

Definition 2 (Successful collusion). *A collusion of buyers is said to be successful if the colluding buyers manage to delete or modify a majority of marks of enough subitems to render reconstruction of $item^*_{(l)}$ and $item^*_{(u)}$ infeasible in Protocol 3.*

If no c-secure codes are used (see Note 2), then we can only say the following about collusions.

Proposition 3. *The expected percent of subitems whose marks can be deleted by a collusion of c buyers is* $100(1 - 1/2^{c-1})$.

Proof. As pointed out in Subsection 2.1, c colluding buyers can locate (and delete) all marks in the i-th subitem if and only if they can pool both versions $item'_i$ and $item''_i$ of the subitem. 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 $1 - 1/2^{c-1}$. ◇

Merchant security also depends on:

- The choice of the security parameters n, l and u (see Subsection 2.3).
- The resilience of the redundancy scheme used in fingerprinting. Even if one or several buyers cannot locate the marks in a subitem (because they all have the same subitem version), they could attempt a blind modification of the subitem with the hope of destroying its marks. Redundancy schemes are helpful against such blind modifications.
- The kind of similarity relation *sim* used (see Subsection 2.4). If *sim* is very loose, this means that M 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.

4 Conclusion and future research

We have presented a protocol suite for anonymous fingerprinting which is computationally much simpler than previous proposals. The new scheme does not require secure multiparty computation and the buyer's computation can be performed by a smart card.

Future research should be directed to replacing the general zero-knowledge proof used in Protocol 2 with an efficient zero-knowledge proof specially designed for proving the correctness of a hash value.

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. 110-123.

[Domi98] J. Domingo-Ferrer, "Anonymous fingerprinting of electronic information with automatic identification of redistributors", *IEE Electronics Letters*, vol. 34, no. 13, June 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.

[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.