TDDC03 Projects, Spring 2004

Anonymous fingerprinting

Anders Bovin and Feihong Lin

Supervisor: Jacob Löfvenberg

Abstract. In this report we describe anonymous fingerprinting and how it can be used. We begin with a basic description of fingerprinting and then a bit more in detail with anonymous fingerprinting. In our opinion fingerprinting in general can be hard to use with many copies of the data. Anonymous fingerprinting on the other hand can profitably be used in E-commerce.

Table of contents

1 Introduction

In this report we will describe how fingerprinting works, what advantages and disadvantages it has and how it can be used.

1.1 Background

Every since the introduction of the Internet to peoples home the rate of illegal copying has increased. Faster processors, bigger hard drives and faster Internet connections have made it possible to copy digital media (music, movies and books) and computer software (programs, games and operating systems) without any significant loss of quality. The Internet has also made it possible to distribute illegal copies much faster to much more people. The introduction of file sharing programs (such as Napster, Kaazaa or DirectConnect) made the distribution even easier.
The producers, artists and authors loose a lot of money on this and have therefore tried different ways of protecting their products. For example one common way for protecting computer software is serial numbers. But serial numbers can't be used on media like audio, movies or texts. Here you would need a kind of protection that is part of the media and doesn't ruin it for the user. The protection should also be hard to remove or even better ruin the media if it's removed. Another way of reducing the distribution of illegal copies is to increase the risk of getting caught. This can be done by adding traces in the product that will lead back to the purchaser and creator of the illegal copy. Fingerprinting is one solution that could be used to protect media such as mentioned above.

1.2 Approach

In the report we will start with a short description of different fingerprinting methods. We will try to find out which major advantages and disadvantages there are with fingerprinting in general and also which benefits are granted or lost with the different methods. After this we will make a deeper study of anonymous fingerprinting and compare it with the results found above. The description and analyses of anonymous fingerprinting will be more extensive then the once made on the other methods. Finally we will describe some different anonymous fingerprinting protocols.

1.3 Method

This report is based on a literature study. The information has been collected mainly from books and articles.

1.4 Chapter description

Chapter 1 is a short introduction to the report.

Chapter 2 discusses fingerprinting in general describing different methods advantages and disadvantages.

Chapter 3 goes deeper into anonymous fingerprinting.

Chapter 4 consists of descriptions of some different anonymous fingerprinting protocol.

Chapter 5 is a summary of conclusions made in the report.

1.5 Terminology

The following terminology will be used:

Mark – refers to a piece of data that is coded into a part of the fingerprint. For example a word in a text that is changed into a synonym to make an unique copy of the text is a mark.

We have also chosen to write he and his instead of he/she and his/hers as we will mostly refer to people making illegal copies and this people are commonly considered to be younger male characters.

2 Fingerprinting

Fingerprinting is one of many ways to protect your digital objects or software. The basic idea with fingerprinting is to make every copy unique. Adding a special ID to every object does this. The ID should be hidden within the significant part of the data so it can't be removed without ruining the object.

2.1 What is fingerprinting

Fingerprinting as a technique is several hundred years old. It was for example used to protect logarithm tables from illegal copying. The idea then was to introduce small errors in insignificant digits (i.e. last decimal) of log x for a few random x. A different set of x's was chosen for every copy of the table. Illegal copies could then be traced back to its owner. [1]
Another way of doing fingerprinting could be changing words to synonyms in a text. For example, take the following string: *"We produce fantastic cars"* could be changed to *"We construct amazing automobiles"*. The two strings say the same thing but using different words. By just changing between the synonyms we can create

eight different strings and by adding more synonyms or more text we would be able to create even more unique copies. This cannot of course be done to all kind of objects, for example changing words in poetry will most likely ruin it.
According to Pfitzmann and Schunter in [2] a good fingerprinting system should at least fulfil the following requirements:

- **The data must be tolerant to errors:** On the one hand, the marks must not decrease the usefulness of the copy to the buyer. On the other hand, the buyer should not be able to derive from the redundancy of the data where the marks are.
- **Collusion-tolerance:** Even if dishonest buyers have up to a certain number of copies, they should not be able to find all marks by comparing the copies. In particular, the fingerprints must have a common intersection. In other words a given number of copies should have marks that cannot be found by comparing the copies.
- **Tolerance to additional errors:** If a dishonest user adds some noise to the copy, the fingerprint should still be recognisable, unless there is so much noise that the copy as such is useless. In other words, the fingerprint should tolerate a greater level of noise than the data. In particular, it should hold for lossy data compression.

The main disadvantage of fingerprinting is an indirect result of what makes it good. To be able to trace who made an illegal copy you have to keep track of who got which fingerprint. For a larger scale of copies you would need to maintain a large database. But not only the rows in the database grows as you make more copies, also the size of the fingerprint grows as you would need more marks to make every copy unique.
Another problem is that many people are not that keen on register themselves in databases. It is considered to be both time-consuming and a risk of violation on personal integrity, especially if companies share their databases among each other.

2.2 Fingerprinting methods

There are many techniques for implementing fingerprinting, all with different advantages and disadvantages. Some of them are Basic (symmetric) fingerprinting, Asymmetric fingerprinting and Traitor-tracing fingerprinting that will be described in 2.2.1-2.2.3 and Anonymous fingerprinting that will be described in Chapter 3.

2.2.1 Basic fingerprinting

This is the basic idea of fingerprinting. The idea of fingerprinting is that a unique set of values is assigned to every copy of the object. If an illegal copy is found it can be traced back to its origin by using the unique values. If the creator of the illegal copy has made changes to the copy the values still can be used to calculate the probability that a given user is the creator of the illegal copy.
The problems come when more than one user is involved in creating the illegal copy. This makes it harder to trace the copy and it also possible that the trace back instead leads to an innocent user.
For example user 1 gets the following string *"We produce fantastic cars"*, user 2 gets the string *"We construct amazing cars"* and user 3 gets the string *"We produce amazing cars"*. All the strings are different from each other and if one of the users crates an illegal copy of his string it can be traced back to him. But if user 1 and 2 compares their strings before making an illegal copy based on data from both strings their copy could end up looking exactly like user 3's string and then framing him for the illegal copy.
To get around this problem collusion-secure fingerprinting is introduced. It addresses the problem that occurs when more then one user is involved in creating the illegal copy. By adding marks to the code that can't be recreated by a collusion of c users you could protect innocent users.
The problem with this technique is first that you have to choose for up to how many users in the collusion you want to protect against up to a maximum of the number of copies minus one (the last copy being your private copy). A collusion of all users being involved is most unlikely to occur because you have to be involved yourself and if so you don't need fingerprinting to know which other copies are involved.
The second problem then is that for every increase in c the number of marks that is needed increases. For a large number of copies it's very hard to make a collusion-secure fingerprinting system for a large c, as you have to have a lot of unique marks. This is needed to make it frame-proof so that an innocent user cannot be accused for being part of a collusion. Frame-proof means that a group of users cannot construct a copy, by comparing their own copies, that is identical to an innocent users copy.
For example user 1 gets the string *"We produce fantastic cars"*, user 2 gets the string *"We construct amazing cars"* and user 3 gets the string *"We produce amazing automobiles"*. This example is frame-proof for collusions up to two users as no combination of pairs can construct a string identical to the third user using their own strings.

2.2.2 Asymmetric fingerprinting

With the above-mentioned techniques there still can be problems. Consider a seller that sells a fingerprinted object to a buyer. This object is unique and only the buyer and the seller know the code. This opens up the possibility for the seller to create illegal copies and then try to frame the buyer.
Asymmetric fingerprinting on the other hand creates a copy that only the user knows about. The fingerprint is created by first the seller encrypts the object using the buyers public key and than the buyer decrypts it using his secret key. After the deal is done the buyer have a uniquely, and tied to him, fingerprinted copy of the data. The seller does not get this copy. From the public key gained in the key exchange the seller cannot create a copy identical to the one that the buyer has. But on the other hand, if the buyer distributes illegal copies of the data, the seller can identify them and trace them back to the buyer.

2.2.3 Traitor-tracing fingerprinting

Traitor-tracing fingerprinting were first introduced in [3]. Then it was suggested to be used in tracing abusers in encrypted broadcasting such as pay-per-view TV. It should also provide legal evidence of the abuse.
Traitor-tracing fingerprinting was later improved in [4] by adding asymmetric schemes to the previous symmetric. This to avoid the same problems that asymmetric fingerprinting deals with (i.e. dishonest providers).
The traitor-tracing fingerprint schemes differ from the other fingerprinting schemes mentioned above as it does not prevent or deter from redistribution of the data, but rather focus on prevention of decryption possibilities. The fingerprint is in the decryption keys, not in the actual data.
For example with pay-per-view TV, traitor-tracing fingerprinting does not prevent the user from decrypting the original broadcast and then rebroadcasting it. But it does prevent the pirate from creating a decryption box from his private keys and distributing the box. If he does so, the private keys can than be traced back to the pirate.

3 Anonymous fingerprinting

One of the major disadvantages with the above mentioned fingerprinting methods are that in order to work the buyer of a product must identify himself and be registered for the purchase. Many buyers are against this as the data registered could be used in a non-intended way. For example from seeing what kind of books or CDs a person buys you could find out much information about the person. This information could then be used to send directed advertisements to the buyer.
To get around this problem anonymous fingerprinting was introduced by Pfitzmann and Waidner in [5]. The basic idea here is that the buyers do not have to identify themselves for the purchase, but if necessary the merchant can trace any traitors. The schemes are asymmetric so that honest buyers stay anonymous.
To achieve this the buyer needs a set of keys, one secret and one public. For more information about creation and usage of keys see a book about Cryptography, for example [6]. The public key serves as an identifier so that the buyer can sign something under identity in a protocol. In other words the buyer can prove that he is the legitimate buyer. The merchant cannot see who the buyer really is, just that he has been approved by the registration center.
The user then has to register at a registration center. A good choice here would be the buyer's bank as the bank has to be involved anyway. This because any payment with digital cash has to be paid through a bank and the user will only be anonymous among this banks client.
The keys are then used together with the merchant's to create an asymmetric anonymous fingerprinted copy of the data to the user. The public key is also given to arbiters (The arbiters are the one that should be convinced in a trial). This so it can be used to identify the person behind any illegal copy.

4 Anonymous fingerprinting methods

In this chapter we will describe some different anonymous fingerprinting methods.

4.1 Basic anonymous fingerprinting

This method's main focus is on registration, fingerprinting, identification and trial. The buyer only needs to identify to the registration center and never to the merchant.

4.1.1 Method description

The involved parties in this method are merchants, buyers, registration center and arbiters. The method has seven steps:

1) **Registration Center Key Distribution:** First the registration center generates a pair of keys, one private and one secret key. The public key is then reliably distributed to all merchants, buyers and arbiters registered at the registration center. The secret key is of course kept secret.

2) **Registration:** When a new buyer wants to register at the registration center the buyer and the registration center exchange their public keys. After the key exchange registration records are created for both the registration center and for the buyer

3) **Data initialisation:** The data initialisation is done by an algorithm on each data item the merchant wants to sell. The merchant can set an upper bound on the number of times it can be sold. A data record is created for every data item.

4) **Fingerprinting:** The fingerprinting of the data is done through a two-party protocol. By using the registration record created in step 2 and the data record created in step 3, a fingerprinted copy is created and given to the user. The fingerprinting is done in the same way as asymmetric fingerprinting (Chapter 2.2.2), but instead of the public key the registration record is used: This is the last step as long as no illegal copy of the data is discovered. The merchant records the sale in a purchase record.

5) **Identification:** If an illegal copy of the fingerprinted data created in step 4 is discovered the merchant can try to identify the user who made the purchase of the data. This is done together with the registration center. They compare the fingerprint from the copied data with the registration and purchase records. In this way the buyer can be identified and taken to trial.

6) **Enforced identification:** This step is only taken if the registration center refuses to take part in the identification of the user, making step 5 impossible. Instead the merchant can turn to the arbiters for help. The identification is done in the same way as in step 5, but with the arbiters taking the registration centers place. The merchant get the identity of the user and the arbiters get a output saying if the buyer or the registration center is guilty. The merchant can then take the guilty part to trial.

7) **Trail:** Here the merchants present their proof of a buyer guilty of producing illegal copies of the purchased data. The arbiters can check if the buyer is guilty or not. Also other parties can be taken to trial (i.e. the registration center).

For a more detailed description of the method and its security properties see [5].

4.1.2 Security advantages

An honest buyer can never be identified, this because a copy of the purchased data is needed in the identification. As long as the buyer doesn't distribute his copy he cannot be identified. If the buyer makes several purchases this purchases should be unlinkable even by a collusion of all merchants, the registration center and the other buyers.
The merchants gain the possibility to identify the buyer who has distributed an illegal copy, as long as it's not created by a collusion of buyers over a certain size. The size is set when the fingerprint is created. They also get proof to win the corresponding trial.
The method also guaranties that honest arbiters cannot decide an honest registration center guilty.

4.2 Coin-Based anonymous fingerprinting

In this section we introduce an anonymous asymmetric fingerprinting scheme based on the principles of "coin". The main idea of this model is the redistributing of a data item should correspond to double-spending 'coins', which is used as a cryptographic primitive and have no monetary value. It is fairly efficient if all the operations are efficient computations with modular multiplications and exponentiations. But it does not offer direct non-repudiation. Direct non-repudiation means the merchant alone has enough information to convince the arbiter that the buyer has distributed the copy illegally. In other words, the accused buyer has to participate in any trial protocol to deny charge. Also the buyer has to prepare the coin, which makes the model not so efficient to use in practice, before fingerprinting so as to reveal the dishonest buyer's identity.

4.2.1 Method description

The involved parties in the coin-based anonymous fingerprinting are merchants, buyers, registration center and arbiters. In this method there are four main protocols: Key distribution, registration, fingerprinting, identification and trial. The method works as following:

1) Key Distribution and Registration

1.1) Registration center key distribution: The registration center chooses a hash function that can convert a variable-size input string into a fixed-size output string. It also generates a pair of keys (one secrete key and one public key) and a group generator that can be hashed using a hash function (for example, Hash: G_q* G_q*G_q*G_q---Z_q^*). It then

publishes the group generator, its public key and hash. The secret key is kept secret.

1.2) Opening a one-time account: When a new buyer wants to open an account at the registration center, he generates a key pair, one public and one secret key. Then he sends his public key and the claimed common identity to the registration center. The registration center creates a registration record by combining the buyer's public key and the identity.

1.3) Withdrawal: The registration center plays the role of the bank and issues coin to registering buyer. The coin is represented by a unique piece of digital information (for example: the identify information of the buyer) with a corresponding signature of the registration center.

2) Fingerprinting: The buyer firstly presents the coin to the merchant. If the coin is valid, the merchant will be convinced that it contains information that will allow the registration center to retrieve the buyer's identity. Then the buyer embeds the identifying information contained in the coin into the sold data item. Later the merchant gives the fingerprinted data item to the buyer and stores the purchase record of the sale.

3) Identification: If the merchant finds a redistributed copy of the fingerprinted data item, he tries to trace the identity of the buyer who purchases the data item. This is a protocol performed between a merchant and the registration center. They compare the fingerprint from the copied data item with the registration and purchase records. Then the identity of the buyer can be trace.

4) Trail: The merchant tries to convince an arbiter that a buyer redistributed the fingerprinted data item by showing the proof. The arbiter can decide whether the buyer is guilt or not by recovering the identity of the buyer.

For the detailed definitions of the protocols and security properties see [7].

4.3 Anonymous fingerprinting using group signatures

In this section we introduce an anonymous asymmetric fingerprinting using group signature.

4.3.1 Method description

The involved parties in the anonymous fingerprinting using group signatures are merchants, buyers, registration center, arbiters and revocation manager. In this method there are five main protocols: Registration center key distribution, registration, fingerprinting, identification and trial. The protocols work as following:

1) **Registration center key distribution:** The registration center generates a pair of keys (one secrete key and one public key) and publishes the public key.

2) Registration: When a new buyer wants to register at the registration center, he generates a key pair, one public and one secret key. Then he sends his public key and the claimed common identity to the registration center. The registration center creates a registration record by combining the buyer's public key and the identity.

3) Fingerprinting: The revocation manager who can find the identity of a signature origniator generates a key pair, one public key and one secret key. Then he sends the keys pair to the buyer. The public key is then used for issuing the group signature. The secret key together with the registration record (created in the step 2) are embedded into the sold data item to create a fingerprinted data item. Later the merchant gives the fingerprinted data item to the buyer and records the sale in a purchase record.

4) Identification: If the merchant finds a redistributed copy of fingerprinted data item produced by the dishonest buyer, the merchant tries to trace the identity of the buyer. He compares the fingerprint from the redistributed copy with the purchase record. Then he can trace the identity of the dishonest buyer.

5) Trail: The merchants present their proof of a buyer who illegally redistributes the fingerprinted data item. The arbiters can check if the buyer is guilty or not by recovering the identity of buyer.

In this method, buyers need to register only once and can then buy many goods without these transactions being linkable. The merchant need not contact the registration center to identify the buyer who has distributed the copy illegally. And also the revocation manager can trace the signature without any interact with the registration center. So the method is efficient to use in practice.

For the detailed definitions of this method and its security properties see [8].

5 Summary

Fingerprinting is a method for copy protection that has been around for a long time. It was first used to trace people who distributed illegal copy of text. It had a rebirth with the introduction of digital data. The main idea is to make small unique changes, not noticeable for the user, in every copy.

There are many methods for fingerprinting data, as there are many types of data that can be marked

(i.e. texts, audio, pictures etc.). To be able to trace the person behind an illegal copy records of who bought each copy needs to be stored. There is also the problem of collusions of users trying to get around the fingerprint protection by comparing their copies.
Symmetric fingerprinting methods are in our opinion best suited for when the numbers of copies aren't that large. Asymmetric fingerprinting methods on the other hand are more suited to handle a larger number of copies, as the possibility that a collusion of users will frame an innocent user is less likely to occur.
Asymmetric anonymous fingerprinting was then developed to get around the problem of users have to identify themselves at every purchase. By introducing a third party, a registration center, for the user to register their identity at, the user can then make purchases from merchants and still stay anonymous to them. A good idea is to let the users bank be the registration center as the bank would be involved in payment anyway. Anonymous fingerprinting is well suited to be used in e-commerce.

References

[1] D. Boneh and J. Shaw *Collusion-Secure Fingerprinting for Digital Data* IEEE Transactions on Information Theory, vol IT-44, Sep. 1998, pp. 1897—1905

[2] B. Pfitzmann and M. Schunter *Asymmetric fingerprinting* EUROCRYPT 1996 p. 84-95

[3] B. Chor, A. Fiat, M. Naor and B. Pinkas *Tracing Traitors* Proceedings of the 14th Annual International Cryptology Conference on Advances in Cryptology, p.257-270, August 21-25, 1994

[4] B. Pfitzmann *Trial of Traced Traitors IHW'96 - Proc. of the First International Information hiding Workshop, May 30 - June 1. Vol. 1174 1997. pp. 49-64.*

[5] B. Pfitzmann and M. Waidner *Anonymous fingerprinting* Advances in Cryptology – EUROCRYPT 1997, LNCS 1233 p. 88-102, Springer Verlag Berlin

[6] W. Trappe and L. C. Washington *Introduction to Cryptography with coding theory* Prentice-Hall Inc. Upper Saddle River, New Jersey 2002

[7] B. Pfitzmann and M. Schunter *Coin-Based Anonymous Fingerprinting* Eurocrypt'99, LNCS 1434 p. 150-164., Springer-Verlag, Berlin

[8] J. Camenisch, *Efficient Anonymous Fingerprinting with Group Signatures*

IBM Research, Zurich Research Laboratory, CH-8803 Ruschlikon 2000