TDDC03 Projects, Spring 2006

# Practicality of Asymmetric Fingerprinting for Digital Data

Feng Yuan

Supervisor: Tina Lindgren

# Practicality of Asymmetric Fingerprinting for Digital Data

Feng Yuan
Department of Computer and Information Science
Linköping University, Sweden
fenyu170@student.liu.se

## Abstract

*The idea of fingerprinting of digital data has been around for some time. The principal goal behind this concept is to deter people from distributing illegal copies and therefore allow the original copyright holders to receive the royalties from the data.*

*This paper will look at fingerprinting and existing fingerprinting concepts. In particular we will focus on the asymmetric fingerprinting model and analyse its applicability and practicality in the real world scenario.*

## 1 Introduction

The continued rapid expansion of the internet and development of faster and cheaper ways to communicate data has led the increased distribution of illegal digital data. The architecture of P2P networks, most notably BitTorrent [10] has made distribution of such illegal digital data readily available. It is estimated that BitTorrent accounts for 35% of all internet traffic [1]. With such a staggering amount of flow of illegal traffic, loss of revenue for original copyright owners is certainly not a trivial matter.

One of the ways to address this issue is digital fingerprinting. If every piece of data had a unique fingerprint and if we know who that data originally belongs to we can trace the illegally distributed data back to party responsible.

There are many proposed methods of implementing this digital fingerprinting. This paper will address some of the methods and have a look at their applicability and feasibility in the real world environment.

### 1.1 Background

One instance of fingerprinting of data, although not digital, dates several hundred years and was used in logarithmic table. Slight variations in the insignificant figures allow for identification of individual copies. If copies were illegally distributed the variations in the insignificant figures acts as a fingerprinting and can be traced back to the illegal distributor [4].

### 1.2 Problem

Most of the proposed fingerprinting methods are mainly theoretical, even so called practical applications are limited to a much idealised environment. We shall address the issues that may arise when applying the proposed fingerprinting schemes on different types of digital data and compare the differences.

### 1.3 Terminology

Throughout this paper we will use the following entities and terms when describing scenarios.

- Merchant – This entity sells digital data
- Customer – This entity buys digital data from merchants
- Traitor – This entity illegally redistributes digital data.
- Pirates – This entity obtains illegally redistributed data from traitors.
- Fingerprinting Authority – This entity is a trusted third party that embeds the fingerprint
- Registration Authority – This entity is a trusted third party that holds the aliases of customers used in anonymous purchasing of digital data.
- Fingerprint – This is the identifier in digital data which make it unique from all others.
- Codewode – A particular string or representation which is an element of fingerprinting alphabet.

## 2. Fingerprinting

The concept of fingerprinting is to make every copy of a piece of digital data unique so that we identify the copy if it is ever illegally distributed. The fingerprint should be embedded into the data using a robust method such as those described in [7,3]. The fingerprint should be part of the digital data such that it can't be removed without rendering the data useless or at least significantly degraded.

### 2.1 Basic Fingerprinting

Also known as symmetric fingerprinting, as both the customer and merchant both has access the digital data with the fingerprint.

Suppose we want to distribute the following data:

"The fantastic new technology that we in New Tech. Inc. have discovered is of profound importance to new generations of cars."

The underlined text can be interchanged with other text and thus form the codeword.

Substitutions:

| Word | Binary | Word | Binary |
|---|---|---|---|
| fantastic | 0 | amazing | 1 |
| discovered | 0 | invented | 1 |
| profound | 0 | great | 1 |
| New | 0 | future | 1 |
| Cars | 0 | automobiles | 1 |

A fingerprint of 10101 would yield:

"The amazing new technology that we in New Tech. Inc. have discovered is of great importance to new generations of automobiles."

A fingerprint of 01010 would yield:

"The fantastic new technology that we in New Tech. Inc. have invented is of profound importance to future generations of cars."[5]

Two problems exist with this model. Firstly we must choose a suitable marking for the particular type of digital data. As in the example above interchanging word with synonyms can work as long as it does not change the idea of the message being conveyed. However when we dealing with multimedia data such as pictures, video and audio, completely different methods must be used.

Another problem is the collusion of traitors.

Suppose traitor A has a copy of a piece of particular software with codeword: *101001*, and traitor B has a copy with codeword: *100011*. They can detect the difference in the fingerprint and generate other codewords, i.e. *101011*. If then they distribute a copy of software with this codeword, the copy may be traced back to an unsuspecting and innocent third party [5].

Ideally we would like to have a fingerprinting scheme where it would be impossible for colluding traitors to generate new codewords and frame innocent parties.

### 2.2 Asymmetric Fingerprinting

Up until now we assumed that traitor is always an end user, i.e. a rogue customer. However it is also quite conceivable that the traitor can also be a rouge merchant. Even if the digital data employed the use of a fingerprinting scheme that is totally collusion resistant and frameproof, the merchant can simple distribute the data themselves and blame the customer.

The concept of asymmetric fingerprinting comes into play when we realise that we do not live in an ideal would and it is conceivable that the merchant can also be a traitor.

Asymmetric fingerprinting works in very much the same way as asymmetric, or public key cryptography [6].

In the practical use of asymmetric fingerprinting we must introduce at least one other Trusted Third Party, the reason behind this will be shown.

In a simplified explanation of this:

1. The customer composes a message consisting of the description of the product and their public key.
2. The message is encrypted using the Fingerprinting Authority's public key.
3. The encrypted message is sent to the merchant along with the description of the product.
4. After the ecommerce part of the transaction is satisfied the encrypted message and product is forwarded to the Fingerprinting Authority.
5. The Fingerprinting Authority decrypts the message and checks the description of the product against the product.

6. The Fingerprinting Authority chooses a codeword from a collusion free code list. If the product has never been marked a code list is generated with appropriate parameters as specified in [4], otherwise the next code word in the list is used.
7. The Fingerprinting Authority embeds the code word system using proposed schemes [3, 7].
8. The customer and code word is stored in the Fingerprinting Authority's database.
9. The copy embedded with the fingerprint is encrypted with a random key and the random key is encrypted and sent to the customer using their public key [8].

The trusted third party ensures that the digital data embedded with the codeword corresponding to the customer is only delivered to that customer,

The marked copy is not encrypted using the customer's public key as asymmetric encryption is inefficient for large quantities of data [6].

If a copy of the digital data is found to be distributing on networks such as BitTorrent or Direct Connect [9] tracing the identity of the traitor is as shown.

1. The merchant sends the illegal copy to the Fingerprinting Authority.
2. The Fingerprinting Authority recovers the embedded mark and decodes the codeword.
3. The codeword is checked against the Fingerprinting Authority's database and the traitor is found.

## 2.3 Anonymous Fingerprinting

Many customers are concerned with maintaining privacy when making purchases. Schemes such as asymmetric fingerprinting make is possible to rule out, or at least reduce the possibility of a traitor being a merchant however privacy is another matter. One scheme proposed by [7], incorporates both asymmetric and anonymous fingerprinting. It may seem quite trivial as we basically introduce another trusted third party into the scheme. We introduce the concept of a Registration Authority. The Registration Authority would have a database of customers and their aliases. So instead of the customer sending their real identities to the merchant and Fingerprinting Authority, they send their aliases. During the step of ecommerce schemes such as those proposed in [11, 12] can be used to maintain the customer's anonymity.

# 3. Practicality Overview

Although the use of asymmetric and anonymous fingerprinting appears very appealing we shall have a look at why they are not as commonly deployed in the real world environment. For the purpose of the convenience from this point on in the paper we shall regard asymmetric fingerprinting as having incorporated anonymous fingerprinting as well, thus we have two trusted third parties, the Fingerprinting Authority and the Registration Authority.

The main advantage for using asymmetric fingerprinting is that both parties, customer and merchant with no knowledge of each other can trust each other. Using trusted third parties is more focused towards small merchants, customers may be unwilling to pass on their financial details. We are already using trusted third parties for payments, such as PayPal and PayMate. Using a Fingerprinting Authority will allow the user to have confidence that the merchant will not redistribute the digital data and blame the customer.

The main negative aspect with asymmetric fingerprinting is the resources required for distribution of digital data. There may be robust and efficient ways of embedding fingerprints into digital data such as movies, but there is still the issue of encrypting the data to send to the customer. If it is a popular title the computing power needed would be very significant as numerous copies would have to be encrypted and distributed. Right now we essentially need two escrow networks in which we have complete trust, which raises the point of accountability. Who is accountable for maintaining the list of the customers and their public keys? Would we have to introduce another trusted third party? Who would be designated as these trusted third parties and what would be the level and limitation of their trustworthiness.

Another interest aspect is the feasibility of the schemes used to mark the digital data, its effectiveness for different types of digital data, its ability to withstand attacks.

## 3.1 Images

There has been a lot of research done involving the watermarking of images. Some

of the proposed schemes are non-oblivious; to simply put it an image with a text caption at the bottom is of such category. Another category is oblivious watermarking, where the mark is invisible and embedded as part of the image itself. However such schemes are vulnerable to a number of attacks. The various schemes described in [13, 14, 15, 2] proposes numerous ways of embedding fingerprints into the images which does not allow for collusion to occur. However some of those schemes do not work so well under scaling and geometric distortion attacks, whilst others simply do not even address it issue. One scheme proposed in [7] does take these considerations into account and address these issues to some degree of success. They are addressed to a point where fingerprints can be recovered from the images of scaled and partially rotated images. However they failed to mention or address possible attacks by cropping images and changing the colour palate of the image, i.e. converting a 32bit image to an 8bit image.

## 4. Summary

With the widespread use and distribution of illegal digital data is it necessary for original copyright owner to employ schemes which minimises their losses due to piracy. Fingerprint is one possible viable solution that may deter the distribution of illegal data. However if fingerprinting is to be accepted we must find more robust way of watermarking digital data and efficient ways of distribution data in a trusted environment.

## References


[1] A. Pasick, "LIVEWIRE - File-sharing network thrives beneath the radar", http://in.tech.yahoo.com/041103/137/2ho4i.html, 2004

[2] C.-S. Lu and H.-Y. Mark Liao, "An oblivious and robust watermarking scheme using communications-with-side-information mechanism", in International Conference on Information Technology: Coding and Computing - ITCC'2001, IEEE Computer Society, 2001, pp. 103-107.

[3] J. Domingo-Ferrer and F. Sebe, "Enhancing watermark robustness through mixture of watermarked digital objects", in *IEEE Intl. Conf. on Information Technology: Coding and Computing-ITCC'2002*, IEEE Computer Society, pp. 85-89, 2002.

[4] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data", in *Advances in Cryptology-CRYPTO'95*, LNCS 963, Springer-Verlag, pp. 452-465, 1995.

[5] J. Löfvenberg, "Copyright protection", LiU TDDC03 Lecture Slides, 2006

[6] W. Trappe and L. Washington, "Introduction to Cryptography with Coding Theory", Second Edition, pp. 164-200, 2006

[7] F. Sebe and J. Domingo-Ferrer, "Oblivious image watermarking robust against scaling and geometric distortions", LNCS 2200, Vol. *Information Security*, Springer-Verlag, pp. 420-432, 2001.

[8] A. Martínez-Balleste, F. Sebé, J. Domingo-Ferrer, M. Soriano, "Practical Asymmetric Fingerprinting with a TTP", IEEE Computer Society, 2003

[9] http://www.neo-modus.com/

[10] http://bittorrent.com/

[11] S. Brands, "Untraceable off-line cash in wallets with observers", in *Advances in Cryptology-CRYPTO'93*, LNCS-773, Springer-Verlag, pp. 302-318, 1993.

[12] D. Chaum, "Blind signatures for untraceable payments" in *Advances in Cryptology-CRYPTO'82*, Plenum Press, pp. 199-203, 1983.

[13] M. Ramkumar and A. N. Akansu, "A robust oblivious watermarking scheme", in International Conference on Image Processing - ICIP'2000, IEEE Signal Processing Society, 2000.

[14] M. Caramma, R. Lancini, F. Mapelli and S. Tubaro, "A blind & readable watermarking scheme for color images", in International Conference on Image Processing - ICIP'2000, IEEE Signal Processing Society, 2000.

[15] C.-S. Lu and H.-Y. Mark Liao, "Oblivious cocktail watermarking by sparse shrinkage: a regional- and global-based scheme", in International Conference Image Processing - ICIP'2000, IEEE Signal Processing Society, 2000.