Rational Enforcement of Digital Oblivion

Josep Domingo-Ferrer
Universitat Rovira i Virgili
Department of Computer Engineering and Maths
UNESCO Chair in Data Privacy
Av. Països Catalans 26
E-43007 Tarragona, Catalonia
josep.domingo@urv.cat

ABSTRACT

Digital storage in the information society allows perfect and unlimited remembering. Yet, the right of an individual to enforce oblivion for pieces of information about her is part of her fundamental right to privacy. We propose a solution to digital forgetting based on anonymously fingerprinting expiration dates. In our solution, people who learn information about an individual are rationally interested in helping the individual enforce her oblivion policy. Thanks to this rational involvement, even services for content spreading like Facebook or YouTube would be interested in fingerprinting downloads, thereby effectively enforcing the right of content owners to canceling content.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*privacy*

General Terms

Privacy, Digital oblivion, Digital forgetting, Coprivacy, Game theory, Fingerprinting

1. INTRODUCTION

The information society rests on the availability of huge, cheap and perfect digital storage. Piles of information about everything and everyone are being gathered continuously and stored digitally. Such a wealth of data can be thought of as an external memory available to all of us: by checking it, we can remember a lot and other people can remember a lot about us. While such perfect and comprehensive remembering may be felt as an advantage when *we* are the ones who remember, it clearly is a shortcoming when it allows *others* to remember things about *us* which we would prefer to be forgotten (embarrassing party pictures, etc.). Perfect remembering may not even be good when we are the ones who remember: *e.g.* remembering wrongs too well for our entire lifetime may be a hindrance to our happiness. For the above reasons, the right of an individual to enforce oblivion for pieces of information about her is increasingly being regarded as part of her fundamental right to privacy (*e.g.* see [18] and references therein).

PAIS 2011 March 25, 2011, Uppsala, Sweden

Some potential responses to privacy-encroaching digital remembering are reviewed and criticized in [18]. We briefly summarize this analysis below:

- *Individual responses.* Digital abstinence is one possible individual response, whereby the individual is assumed to be able to adjust her information sharing behavior. However, getting people to constrain what they wish to share is difficult. Another individual attitude is cognitive adjustment: it consists of living with comprehensive digital memory, while at the same time limiting the influence digital memory has on our decision-making. The idea is to accept that people can change and restrict ourselves to looking at their most recent behavior and opinions. Cognitive adjustment may be too ambitious a goal, because humans cannot help taking into account the information at their disposal.

- *Legal responses.* Information privacy rights enforced by law could limit what information and for how long is stored on other individuals. It is not clear whether the legal system is able to effectively enforce privacy rights and whether the individuals are willing to bring actions against perpetrators.

- *Technology responses.* It has been suggested to enforce digital rights management (DRM) to enable an individual to control her private information [4, 17]. The idea is that individuals would add meta-data to their personal information detailing who can use it, for what purpose and price. Besides technical objections to the very notion of DRM, it seems quite unrealistic to assume that an average individual will spend a few hours detailing usage policies and will keep the above meta-information up-to-date.

The author of [18] proposes expiration dates as a simple solution to enforce digital forgetting. Information stored in digital memory is associated with expiration dates set by the information creator (the user first recording the information). The author goes further to assume that digital storage devices could be made to automatically delete information that has reached or exceeded its expiry date. Alternative approaches based on employing smart cards on the user side to process encrypted content [6] could also be envisioned, whereby the smart card would not decrypt the content after its expiration date.

1.1 Contribution and plan of this paper

We adopt here the solution based on expiration dates, but we focus on its secure technical enforcement. First, we need to *embed* in the information the expiration date associated with it, in order to prevent such an expiration date from being altered or suppressed. Second, storage devices which automatically delete information

past its expiration date do not currently exist; worse yet, placing trust in the hardware (storage devices, smart cards, etc.) to implement information protection policies has proven to be a flawed approach: *e.g.*, hardware copy prevention mechanisms for CDs and DVDs were easily bypassed.

We specify protocols whereby a content creator can embed an expiration date in the content, spread the content and trace whoever is using and/or transferring the content after the expiration date has passed. Dishonest receiver tracing is made possible as follows:

- The content carries a different fingerprint for each receiver, so that unlawful content usage and spreading can be traced; we state protocols based on anonymous fingerprinting, whereby the content forwarder does not learn the real identity of honest receivers nor does she see the fingerprinted content these receive (which prevents the forwarder from framing honest receivers by spreading herself content fingerprinted with a receiver's identity);

- The sender does not need to fingerprint and send the content individually to each receiver; doing so would be not only extremely cumbersome, but also quite unrealistic, as we cannot prevent one receiver from forwarding the content to other receivers;

- Receivers are rationally interested to collaborate in fingerprinting the content they forward to other interested receivers. With our simple solution, even services for content spreading like Facebook or YouTube would be interested in fingerprinting downloads, thereby effectively enforcing the right of content owners to canceling content. Note that the current Terms of Service of those services [11, 25] only guarantee that they will not further spread the content after the user withdraws it from the service, but they take no responsibility for further spreading conducted by third parties who obtained the content while it was posted.

Section 2 gives some background on game theory, coprivacy, watermarking, fingerprinting and anonymous fingerprinting. Section 3 describes the protocols and justifies their security. Section 4 argues the rational involvement by peers in game-theoretic terms and shows that the spreading protocol is a generally coprivate one. Section 5 summarizes conclusions and future research issues.

2. BACKGROUND

2.1 Basics of game theory

A game is a protocol between a set of N *players*, $\{P^1, \cdots, P^N\}$. Each player P^i has her own *set of possible strategies*, say S_i. To play the game, each player P^i selects a strategy $s_i \in S_i$. We use $s = (s_1, \cdots, s_N)$ to denote the vector of strategies selected by the players and $S = \Pi_i S_i$ to denote the set of all possible ways in which players can pick strategies.

The vector of strategies $s \in S$ selected by the players determines the outcome for each player, which can be a payoff or a cost. In general, the outcome will be different for different players. To specify the game, we need to give, for each player, a preference ordering on these outcomes by giving a complete, transitive, reflexive binary relation on the set of all strategy vectors S. The simplest way to assign preferences is by assigning, for each player, a value for each outcome representing the payoff of the outcome (a negative payoff can be used to represent a cost). A function whereby player P^i assigns a payoff to each outcome is called a utility function and is denoted by $u_i : S \longrightarrow \mathbb{R}$.

For a strategy vector $s \in S$, we use s_i to denote the strategy chosen by player P^i and s_{-i} to denote the $(N-1)$-dimensional vector of the strategies played by all other players. With this notation, the utility $u_i(s)$ can also be expressed as $u_i(s_i, s_{-i})$.

A strategy vector $s \in S$ is a *dominant strategy solution* if, for each player P^i and each alternate strategy vector $s' \in S$, it holds that

$$u_i(s_i, s'_{-i}) \geq u_i(s'_i, s'_{-i}) \quad (1)$$

In plain words, a dominant strategy s is the best strategy for each player P^i, independently of the strategies played by all other players.

A strategy vector $s \in S$ is said to be a *Nash equilibrium* if, for all players P^i and each alternate strategy $s'_i \in S_i$, it holds that

$$u_i(s_i, s_{-i}) \geq u_i(s'_i, s_{-i})$$

In plain words, no player P^i can change her chosen strategy from s_i to s'_i and thereby improve her payoff, assuming that all other players stick to the strategies they have chosen in s. A Nash equilibrium is self-enforcing in the sense that, once the players are playing such a solution, it is in every player's best interest to stick to her strategy. Clearly, a dominant strategy solution is a Nash equilibrium. Moreover, if the solution is strictly dominant (*i.e.* when the inequality in Expression (1) is strict), it is also the unique Nash equilibrium. See [20] for further background on game theory.

2.2 Coprivacy

We recall this privacy paradigm, introduced in [9]. For each strategy of a player P^i, say $s^i_j \in S^i$, let $u_i(s^i_j)$ be the privacy utility of s^i_j for P^i, where a higher utility means higher overall privacy preservation for P^i versus the other players. Further, let

$$s^{i*}_{u_i} = \arg\max_{s^i_j \in S^i} u_i(s^i_j)$$

be the best strategy for P^i.

DEFINITION 1 (COPRIVACY). *Let Π be a game with peer players $P^1, \cdots, P^N$, and an optional system player P^0. Each player may have leaked a different amount of private information to the rest of players before the game starts. The game is as follows: i) P^1 selects one player P^k with $k \in \{0\} \cup \{2, \cdots, N\}$ and submits a request to P^k; ii) If $k = 0$, P^0 always processes P^1's request; if $k > 1$, P^k decides whether to process P^1's request (which may involve accessing the system player on P^1's behalf) or reject it. The players' strategies are $S^0 = \{s^0_1\}$ (process P^1's request); $S^1 = \{s^1_0, s^1_2, \cdots, s^1_N\}$, where s^1_j means that P^1 selects P^j; for $i > 1$, $S^i = \{s^i_1, s^i_2\}$, where s^i_1 means processing P^1's request and s^i_2 rejecting it. Game Π is said to be* coprivate *with respect to the set $U = (u_1, \cdots, u_N)$ of privacy utility functions if $s^{1*}_{u_1} = s^1_k$ for some $k > 1$ such that $s^{k*}_{u_k} = s^k_1$, that is, if a peer P^k exists such that (s^1_k, s^k_1) is a pure strategy Nash equilibrium between P^1 and P^k.*

Similarly, *general coprivacy* can be defined by replacing in Definition 1 the set U of privacy utility functions by a set $\mathbf{U} = (\mathbf{u}_1, \cdots, \mathbf{u}_N)$ of general utility functions combining privacy with other desirable properties like functionality or security.

2.3 Watermarking, fingerprinting and anonymous fingerprinting

Let $D(0) \in \{0,1\}^*$ denote some digital content (bit-string) some of whose bits can be changed in such a way that the result

remains "close" to $D(0)$ (where "close" means "with a similar utility"), but, without knowing which particular bits were changed, altering a "good portion" of these bits is impossible without rendering the content useless. The changed bits are usually called a mark or watermark; if bits are changed differently for each user receiving the content, the mark can also be called fingerprint. The algorithm used to embed a mark while satisfying the previous two conditions is called a watermarking algorithm; to embed a fingerprint can also be termed "to fingerprint". The second requirement above is actually the marking assumption stated in [2]. Finally, let $\mathcal{D}$ denote the set of all "close copies" of $D(0)$. See [5] for more information on watermarking and fingerprinting.

Whether the original content $D(0)$ is needed for mark/fingerprint recovery depends on the method used for embedding: a watermarking method is said to allow blind detection if only the marked content $\tilde{D}$ and the embedded mark are needed for recovery, in addition to the key used for embedding; methods which need also $D(0)$ are called informed watermarking. In return for their smaller flexibility, informed methods tend to be more robust to content manipulation.

An asymmetric or public-key watermarking method is one in which different keys are used for watermark embedding and recovery, in such a way that the recovery key can be made public without compromising the secrecy of the embedding key. This allows public verifiability of the embedded mark. See [5] for more background on watermarking algorithms.

Anonymous fingerprinting is a special type of fingerprinting in which the fingerprinter does neither see the fingerprinted content nor the receiver's identity except in case of unlawful redistribution. Anonymous fingerprinting procedures were proposed in [22, 7, 8, 3, 1], among others. We take the following model from [3].

DEFINITION 2. *An anonymous fingerprinting scheme involves a merchant (*i.e. *a fingerprinter), a buyer (*i.e. *a receiver) and a registration center. Let c denote the maximal size of a collusion of buyers against which the scheme is secure. An anonymous fingerprinting scheme consists of the following five procedures.*

FKG-RC: *A probabilistic key setup algorithm for the registration center. Its outputs are the center's secret key x_C and its public key y_C, which is published in an authenticated manner.*

FReg: *A probabilistic two-party protocol (*FReg-RC, FReg-B*) between the registration center and the buyer. Their common input is the buyer's identity ID_B and the center's public key y_C. The center's secret input is its secret key x_C. The buyer's output consists of some secret x_B and related information y_B. The center obtains and stores y_B and ID_B.*

FPrint: *A two-party protocol* (FPrint-M, FPrint-B) *between the merchant and the buyer. Their common input consists of y_C. The merchant's secret input is $D(0)$ and a transaction number j and her output is a transaction record r_j. The buyer's secret input is x_B and y_B and her output consists of a copy $D(B) \in \mathcal{D}$.*

FRec: *This may be a protocol or an algorithm whose purpose is to recover the identity of a/the fraudulent buyer responsible for the redistribution of a version $\tilde{D} \in \mathcal{D}$:*

- *It is a two-party protocol between the merchant and the registration center if the merchant needs the help of the registration center. The merchant's input is a copy $\tilde{D} \in \mathcal{D}$, all transaction records r_i and perhaps the original content $D(0)$. The center's input consists of its secret key x_C and its list of y_B's and ID_B's. The merchant's output is a/the fraudulent buyer's identity together with a proof p that this buyer indeed bought a copy of $D(0)$, or $\perp$ in case of failure* (e.g., *if more than c buyers colluded to produce $\tilde{D}$).*
- *It is an algorithm run by the merchant alone if the merchant can determine a/the fraudulent's buyer identity with just $\tilde{D}$, all transaction records r_i and perhaps the original content $D(0)$ (if the underlying watermarking is not blind).*

FVer: *A verification algorithm, that takes as input the identity ID_B of an accused buyer, the public key y_C of the registration center, and a proof p, and outputs 1 iff the proof is valid.*

The solution in [3] guarantees the following properties:

Correctness: All protocols terminate successfully whenever players are honest (no matter how other players behaved in other protocols).

Anonymity and unlinkability: Without obtaining a particular $D(B)$, the merchant –even when colluding with the registration center– cannot identify a buyer (anonymity). Furthermore, the merchant is not able to tell whether two purchases were made by the same buyer (unlinkability).

Protection of innocent buyers: No coalition of buyers, the merchant, and the registration center is able to generate a proof $\tilde{p}$ such that $\mathsf{FVer}(ID_B, y_C, \tilde{p}) = 1$, if buyer ID_B was not present in the coalition.

Revocability and collusion resistance: Any collusion of up to c buyers aiming at producing a version $\hat{D} \in \mathcal{D}$ from which none of them can be re-identified will fail: from $\hat{D}$ the merchant will obtain enough information to identify at least one collusion member.

3. A PROTOCOL SUITE FOR DIGITAL OBLIVION

Assume that a player P^0 has a content $D(0)$ for which she wishes to enforce an expiration date $T(0)$. Let $\mathcal{P}^1$ be the set of peers to whom P^0 directly releases a fingerprinted version of $D(0)$. Let $\mathcal{P}^i$ be the set of peers which receive a fingerprinted version of $D(0)$ after i hops, that is, peers in $\mathcal{P}^i$ receive a fingerprinted version of $D(0)$ from some player in $\mathcal{P}^{i-1}$. Note that the sets $\mathcal{P}^i$ are not fixed in advance. The membership of a certain peer P to one of those sets is established when P obtains a fingerprinted version of $D(0)$ from another peer P': if P' belongs to $\mathcal{P}^{i-1}$, then P belongs to $\mathcal{P}^i$. Assuming that the number of hops is limited to N, the following protocol can be used for P^0 to spread $D(0)$ while enforcing expiration date $T(0)$. In the protocol, players are only known by pseudonym, not by their real identities.

PROTOCOL 1 (SPREADING WITH EXPIRATION DATE).

1. *P^0 embeds the expiration date $T(0)$ into $D(0)$ using a blind and robust asymmetric watermarking algorithm such that $T(0)$ can be recovered by anyone seeing the watermarked content but it cannot be altered without substantially damaging the quality of the content. Public recoverability of $T(0)$ is necessary to preclude disputes about the expiration date.*

2. *For any P^{j_1} to whom P^0 wishes to forward the content, P^0 and P^{j_1} run an anonymous fingerprinting scheme such as the one described in Section 2.3 where the underlying watermarking is* blind*. The secret input by P^0 is the watermarked content obtained in Step 1. After running* FReg *and* FPrint*, P^{j_1} obtains a fingerprinted copy $D(0, j_1)$ of the content and P^0 obtains a transaction record $r(0, j_1)$ (partially or totally consisting of information input by P^0 herself like the transaction number).*

3. **for** $i = 1$ **to** $N - 1$ **do**

   for all *pairs* $(P^{j_i}, P^{j_{i+1}})$ *such that* $P^{j_i} \in \mathcal{P}^i$ *and* P^{j_i} *wishes to forward the content to* $P^{j_{i+1}}$ **do**

   (a) Engage in the same anonymous fingerprinting scheme described in Step 2 whereby $P^{j_{i+1}}$ obtains a fingerprinted version

   $$D(0, j_1, j_2, \cdots, j_{i+1})$$

   and P^{j_i} obtains a transaction record $r(j_i, j_{i+1})$ (partially or totally consisting of information input by P^{j_i} herself like the transaction number);

   (b) P^{j_i} sends $r(j_i, j_{i+1})$ to P^0.

   end for

   end for

In Protocol 1 we are implicitly assuming that the asymmetric watermarking scheme used in Step 1 and the watermarking scheme underlying the anonymous fingerprinting scheme allow embedding $T(0)$ and at least the N successive fingerprints corresponding to the N hops in such a way that:

1. It still holds that the resulting $D(0, j_1, \cdots, j_N)$ is close to $D(0)$, that is $D(0, j_1, \cdots, j_N) \in \mathcal{D}$.

2. Embedding a new fingerprint does not destroy the previously embedded fingerprints. In fact, this results from the aforementioned marking assumption and the previous assumption on the "closeness" of $D(0)$ and $D(0, j_1, \cdots, j_N)$.

If the maximum N for which the above holds is chosen, then players will be rationally interested in not performing more than N hops in Protocol 1: indeed, in the same way as more hops would damage the embedded fingerprints, they would damage the perceptual quality of the content, which would no longer belong to $\mathcal{D}$.

The need for blind watermarking and for P^{j_i} to return the transaction record to P^0 will be justified in Section 3.1 below.

3.1 Security analysis

Systematically detecting *all* copies used beyond the expiration date is beyond the scope of this paper and it is probably an infeasible task in general; *e.g.* if expired copies are not publicly used, it will be very difficult to spot them. Hence, in this section, we assume that *one* copy of the content $\tilde{D} \in \mathcal{D}$ is detected by P^0 after the expiration date $T(0)$ fingerprinted in it (for example, P^0 finds $\tilde{D}$ posted on the web or some social media). What we target is *oblivion in the public domain*, that is, identifying who makes public use of expired content; private oblivion is certainly impossible to enforce, because each individual's private memories are her own. We focus on how P^0 can identify the redistributor's identity, that is, whose content the detected one is.

P^0 runs the following protocol to establish the identity of the redistributor. Whereas the names P^{j_i} are pseudonyms of the players, the protocol allows to find real identities of players. For $i = 1$ to N, let R^i be the set of i-hop transaction records received by P^0 in Protocol 1 for $i = 1$ up to N, that is,

$$R^i = \{r(j_{i-1}, j_i) : \forall P^{j_{i-1}} \in \mathcal{P}^{i-1}, P^{j_i} \in \mathcal{P}^i\}$$

where $P^{j_0} := P^0$ and $r(j_0, j_1) := r(0, j_1)$. Further, let $R^i(P^{j_{i-1}})$ be the subset of i-hop transaction records originated by player $P^{j_{i-1}}$, that is,

$$R^i(P^{j_{i-1}}) = \{r(j_{i-1}, j_i) : \forall P^{j_i} \in \mathcal{P}^i\}$$

PROTOCOL 2 (REDISTRIBUTOR TRACKING).

1. *Set $i := 1$ and $finished :=$* **false***;*

2. **while** $finished =$ **false and** $i \leq N$ **do**

   if $R^i(P^{j_{i-1}}) \neq \emptyset$ **then**

   (a) Sample without replacement one transaction record r' from $R^i(P^{j_{i-1}})$;

   (b) Run with the registration center the FRec *protocol with inputs the redistributed content $\tilde{D}$ and r';*

   (c) **if** *the real identity of a player P^{j_i} can be successfully recovered* **then**

   i. Request the set $R^{i+1}(P^{j_i})$ from P^{j_i};

   ii. $i := i + 1$;

   else *finished:=* **true***;*

   end while

3. *Output the real identity of $P^{j_{i-1}}$ as the redistributor's identity.*

Some remarks on the above identity recovery process follow:

- The anonymous fingerprinting scheme used must be based on an underlying blind watermarking method. Indeed, by construction of Protocol 1, unless the redistributor belongs to $\mathcal{P}^1$, P^0 does not know the original unmarked content corresponding to $\tilde{D}$. Imagine that the redistributor is some $P^{j_i} \in \mathcal{P}^i(P^{j_{i-1}})$ where $\mathcal{P}^i(P^{j_{i-1}})$ is the set of i-hop players who received the content from $P^{j_{i-1}}$, with $1 < i \leq N$. In that case, $\tilde{D} = D(0, j_1, \cdots, j_i)$ and the original unmarked content is $D(0, j_1, \cdots, j_{i-1})$, only known to $P^{j_{i-1}}$, not to P^0.

- Unless a player redistributes her received content, she is the only one to know that content, because anonymous fingerprinting is also asymmetric: the fingerprinter does not see the fingerprinted content. Therefore, no one can frame an honest receiver P^{j_i} by accusing her of redistributing $D(0, j_1, \cdots, j_i)$, because the latter content is known to no one but P^{j_i}.

- P^0 only wants to obtain the identity of the *last* player who fingerprinted the redistributed content. There is a trade-off between anonymity preservation and search efficiency:

  – Since spreading is potentially exponential, Protocol 2 tries to reduce the search space and the computational burden for P^0, even if this implies anonymity loss for all players along the path between P^0 and the redistributor.

– Since P^0 receives all transaction records in Protocol 1, if preserving the anonymity of all honest players is more important than reducing the search space, Protocol 2 could be modified for P^0 to start searching backwards: first search $\mathcal{P}^N$ by trying FRec with the redistributed content and all N-hop transaction records; if the redistributor was not found then search $\mathcal{P}^{N-1}$, etc.; the search would stop as soon as a real identity (the redistributor's) could be successfully recovered. The first and only recovered identity would be the redistributor's; the registration center would refuse its necessary collaboration in FRec if P^0 tried to identify other players after the redistributor had been found. This approach prioritizes anonymity preservation but it can implode P^0, who must try a virtually exponential number of transaction records.

- Protocol 2 works even if some peers fail to send the transaction record to P^0 in Protocol 1. Assume that P^{j_i} and $P^{j_{i+1}}$ engage on a transfer of content and that this transfer takes place without any fingerprinting or without P^{j_i} sending the resulting transaction record to P^0. Since P^0 does not have the transaction record for the transfer to $P^{j_{i+1}}$, P^0 will not be able to identify $P^{j_{i+1}}$ in case $P^{j_{i+1}}$ performs unauthorized redistribution. By following the chain of transaction records, Protocol 2 will accuse P^{j_i} of redistribution. Note that P^{j_i} is indeed guilty for not having correctly followed Protocol 1 and thus can be held liable for the redistribution performed by $P^{j_{i+1}}$ or by anyone in $\mathcal{P}^{i+2}(P^{j_{i+1}})$. A particular case of the above is when P^{j_i} publishes the content on the web or otherwise releases it without keeping track of who accesses it. In that case, P^{j_i} will be accused of redistribution.

3.2 Practical instantiation

Blind and robust asymmetric watermarking schemes for Step 1 of Protocol 1 can be found, for example, in [12, 13, 14, 24].

The audio watermarking scheme described in [19] can be used as a scheme underlying the anonymous fingerprinting in Protocol 1, because it satisfies the requirements listed above. The scheme is blind, so that it is possible to extract the embedded mark from a marked audio object without knowing the original unmarked audio object.

Also, the scheme tolerates embedding several successive fingerprintings without significant damage to the content utility or the previous fingerprints. We have used the Objective Difference Grade (ODG) based on the ITU-R Recommendation standard BS 1387 [15, 23] to evaluate the damage inflicted by successive fingerprintings. This standard makes it possible to evaluate the transparency of the fingerprinting scheme by comparing the perceptual quality of the marked files with respect to the original content $D(0)$. The ODG values are in the range $[-4, 0]$, where 0 means imperceptible, -1 means perceptible but not annoying, -2 means slightly annoying, -3 means annoying and -4 means very annoying. In order to evaluate the ODG, we have used the Opera software by Opticom [21]. The imperceptibility results are shown in Table 1. As it can be noticed, the transparency slowly decreases for each successive receiver in Protocol 1. However, even with 5 embedded fingerprints, the ODG result is much closer to 0 (imperceptible) than -1 (perceptible but not annoying); hence, even in this worst case, the perceptual quality achieved by Protocol 1 with the scheme [19] can be regarded as very satisfactory.

Furthermore, the scheme can be adapted for anonymous fingerprinting. The adaptation is described in detail in [10] and summarized next. The scheme uses a double embedding strategy:

Table 1: Imperceptibility results with five successive fingerprintings

Content	# fingerprints	ODG
$D(0, j_1)$	1	0.000
$D(0, j_1, j_2)$	2	-0.004
$D(0, j_1, j_2, j_3)$	3	-0.034
$D(0, j_1, j_2, j_3, j_4)$	4	-0.115
$D(0, j_1, j_2, j_3, j_4, j_5)$	5	-0.193

- A time-domain synchronization watermark ("SYN") is embedded for fast search of the information watermark position;
- A frequency-domain information watermark is embedded next to the SYN marks.

This double embedding strategy makes it possible to embed the transaction records $r(j_i, j_{i+1})$ and the receiver related information y_B in different domains. The transaction records can be embedded as synchronization marks in the time domain with different bit strings, and the related information y_B can be embedded more robustly in the frequency domain. This scheme has the additional advantage of a very fast search of transaction records; extracting an embedded transaction record from a portion of audio takes less time than playing that portion.

In order to preserve anonymity and make the registration center necessary for redistributor identification (as mandated by Protocol 2), we let y_B be the receiver identity encrypted under the registration center's public key. To obtain unlinkability, a random nonce is appended to the receiver's identity before encrypting it under the registration center's public key. Embedding next to y_B a hash of x_B (or x_B encrypted with the public key of the receiver) has the additional advantage of thwarting a collusion of the sender P^0 and the registration center, who would not be able to produce a correctly fingerprinted copy of the content corresponding to any receiver.

If the sender P^0 finds a version of the audio file illegally redistributed on the Internet, she can search for the transaction records in the time domain (fast search) and then extract the information y_B related to the malicious receiver. This information (y_B) will then be sent to the registration center in order to identify the illegal redistributor.

4. RATIONAL INVOLVEMENT OF PLAYERS: COPRIVACY

In this section, we show how to motivate the players to rationally play their corresponding roles specified in the protocols. Showing that players have no interest in deviating is especially necessary in peer-to-peer (P2P) protocols whose correct operation depends on the commitment of the players.

P^0 has an obvious interest in correctly following Protocol 1. If she deviates by not correctly participating in the anonymous fingerprinting process, the entire expiration date enforcement does not even start. Let s^0 be the strategy whereby P^0 follows Protocol 1. Let us now analyze the strategies of the other players.

Each player $P^{j_i} \neq P^0$ has at least the following possible strategies with respect to P^0 and $P^{j_{i+1}} \in \mathcal{P}^{i+1}(P^{j_i})$:

$s_0^{j_i}$**:** Correctly follow Protocol 1 by engaging in anonymous fingerprinting with $P^{j_{i+1}}$ and returning transaction record $r(j_i, j_{i+1})$

to P^0.

$s_1^{j_i}$: Deviate from Protocol 1 by engaging in anonymous fingerprinting with $P^{j_{i+1}}$ but *not* returning $r(j_i, j_{i+1})$ to P^0.

$s_2^{j_i}$: Deviate from Protocol 1 by anonymously forwarding to $P^{j_{i+1}}$ the content without fingerprinting and returning a fake transaction record $r(j_i, j_{i+1})$ to P^0.

$s_3^{j_i}$: Deviate from Protocol 1 by anonymously forwarding the content without fingerprinting and not returning any transaction record.

$s_4^{j_i}$: Deviate from Protocol 1 by not forwarding the content and not returning any transaction record.

We next go through an exercise of mechanism design (see Chap. 23 of [16]), to find how Protocols 1 and 2 need to be modified to ensure that, for any player P^{j_i}, her rational choice is strategy $s_0^{j_i}$.

4.1 Utility without reward or punishment

Consider the following payoffs:

d_{j_i}: Payoff that P^{j_i} derives from obtaining $D(0, j_1, \cdots j_i)$, *without losing her anonymity*. That is, d_{j_i} combines the functionality payoff of P^{j_i} obtaining the content and the privacy payoff of P^{j_i} preserving her anonymity thanks to anonymous fingerprinting with $P^{j_{i-1}}$. If P^{j_i} pays a fee or reward for obtaining the content, this fee or reward must be deduced from d_{j_i}.

$-v_{j_i}$: Negative payoff (that is, cost) that P^{j_i} incurs from engaging in anonymous fingerprinting with $P^{j_{i+1}}$; this cost may be quantified in terms of computation and communication.

$-w_{j_i}$: Negative payoff (that is, communication cost) that P^{j_i} incurs for returning the transaction record $r(j_i, j_{i+1})$ to P^0.

$-f_{j_i}$: Negative payoff (communication cost) that P^{j_i} incurs from anonymously forwarding the unfingerprinted content to $P^{j_{i+1}}$. Obviously $f_{j_i} \leq v_{j_i}$ because forwarding unfingerprinted content is simpler than anonymously fingerprinting it.

If there are no other payoffs (like reward earned for following the protocol or punishment incurred for not following it), the general utility functions of the above strategies are the following:

$$\mathbf{u}_{j_i}(s_0^{j_i}) = d_{j_i} - v_{j_i} - w_{j_i}$$

$$\mathbf{u}_{j_i}(s_1^{j_i}) = d_{j_i} - v_{j_i}$$

$$\mathbf{u}_{j_i}(s_2^{j_i}) = d_{j_i} - f_{j_i} - w_{j_i}$$

$$\mathbf{u}_{j_i}(s_3^{j_i}) = d_{j_i} - f_{j_i}$$

$$\mathbf{u}_{j_i}(s_4^{j_i}) = d_{j_i}$$

Clearly, strategy $s_4^{j_i}$ has the maximum utility. Considering a sequence of players $P^0, P^{j_1}, \cdots, P^{j_N}$ with $P^{j_i} \in \mathcal{P}^i(P^{j_{i-1}})$ for $i = 1$ to N, the dominant strategy solution of the game is

$$(s^0, s_4^{j_1}, -, \cdots, -)$$

In plain words, the rational equilibrium is for P^0 to conduct anonymous fingerprinting of the content with P^{j_1} and for P^{j_1} to acquire the anonymously fingerprinted content and do nothing else. The strategies of the rest of players are irrelevant, because they receive no content.

4.2 Utility with reward and no punishment

In an attempt to induce rational players to correctly follow Protocol 1, we can think of introducing a reward for a player who forwards the content to other players. $P^{j_{i+1}}$ pays a reward g_{j_{i+1},j_i} to P^{j_i} upon receiving the content. $P^{j_{i+1}}$ discounts g_{j_{i+1},j_i} from her payoff $d_{j_{i+1}}$.

In this case, the utility functions of the four strategies of P^i are:

$$\mathbf{u}_{j_i}(s_0^{j_i}) = d_{j_i} + g_{j_{i+1},j_i} - v_{j_i} - w_{j_i}$$

$$\mathbf{u}_{j_i}(s_1^{j_i}) = d_{j_i} + g_{j_{i+1},j_i} - v_{j_i}$$

$$\mathbf{u}_{j_i}(s_2^{j_i}) = d_{j_i} + g_{j_{i+1},j_i} - f_{j_i} - w_{j_i}$$

$$\mathbf{u}_{j_i}(s_3^{j_i}) = d_{j_i} + g_{j_{i+1},j_i} - f_{j_i}$$

$$\mathbf{u}_{j_i}(s_4^{j_i}) = d_{j_i}$$

If the reward is sufficient to cover the costs of P^{j_i} anonymously forwarding the content without fingerprinting, that is, if $g_{j_{i+1},j_i} \geq f_{j_i}$, then $s_3^{j_i}$ has the maximum utility. In these conditions, the dominant strategy solution of the game is

$$(s^0, s_3^{j_1}, \cdots, s_3^{j_{N-1}}, s_4^{j_N})$$

In plain words, the rational equilibrium is for P^0 to correctly follow Protocol 1 with P^{j_1} and for P^{j_i} $(i = 1, \cdots, N-1)$ to anonymously forward the content to $P^{j_{i+1}}$ without fingerprinting and without returning any transaction records to P^0. The strategy of P^N can only be $s_4^{j_N}$, because P^{j_N} is not supposed to forward the content any further.

4.3 Utility with reward and punishment

A punishment mechanism is needed to motivate players P^{j_1} through $P^{j_{N-1}}$ to return valid transaction records to P^0 (and therefore correctly engage in anonymous fingerprinting with the next player).

Let $-p_{j_i}$ be the expected negative payoff (punishment) that P^{j_i} incurs when accused of redistribution as a result of not having returned a valid transaction record $r(j_i, j_{i+1})$. This is actually an *expected* negative payoff, computed as the probability of being accused times the cost of being accused: such cost may be a fine, the utility loss of being disgraced vs P^0, etc. We can recompute the utilities of the four strategies available to P^{j_i}:

$$\mathbf{u}_{j_i}(s_0^{j_i}) = d_{j_i} + g_{j_{i+1},j_i} - v_{j_i} - w_{j_i}$$

$$\mathbf{u}_{j_i}(s_1^{j_i}) = d_{j_i} + g_{j_{i+1},j_i} - v_{j_i} - p_{j_i}$$

$$\mathbf{u}_{j_i}(s_2^{j_i}) = d_{j_i} + g_{j_i,j_{i+1}} - f_{j_i} - w_{j_i} - p_{j_i} \tag{2}$$

$$\mathbf{u}_i(s_3^i) = d_{j_i} + g_{j_{i+1},j_i} - f_{j_i} - p_{j_i}$$

$$\mathbf{u}_i(s_4^i) = d_{j_i}$$

Assume like above that $g_{j_{i+1},j_i} \geq f_{j_i}$. Assume also that

$$v_{j_i} + w_{j_i} < p_{j_i} \tag{3}$$

that is, anonymously fingerprinting the content and returning the correct transaction record is less costly than the punishment. Then,

$$\mathbf{u}_{j_i}(s_1^{j_i}) \leq \mathbf{u}_{j_i}(s_0^{j_i})$$

$$\mathbf{u}_{j_i}(s_2^{j_i}) \leq \mathbf{u}_{j_i}(s_0^{j_i})$$

$$\mathbf{u}_{j_i}(s_3^{j_i}) \leq \mathbf{u}_{j_i}(s_0^{j_i})$$

With the above constraints, P^{j_i} will rationally choose $\mathbf{u}_{j_i}(s_0^{j_i})$ if the reward is sufficient to cover the costs of anonymous fingerprinting and returning the transaction record, *i.e.* if

$$g_{j_i,j_{i+1}} \geq v_i + w_i \tag{4}$$

Otherwise, P^{j_i} will choose $\mathbf{u}_{j_i}(s_4^{j_i})$ (no forwarding). At any rate, the strategies consisting of forwarding unfingerprinted content will not be chosen by P^{j_i}. Therefore, even if $P^{j_{i+1}}$ has no rational interest in getting a fingerprinted content, she will have to increase the reward to meet Inequality (4) if she wants to be forwarded the content at all. Under conditions (3) and (4), the dominant strategy solution of the game is

$$(s^0, s_0^{j_1}, \cdots, s_0^{j_{N-1}}, s_3^{j_N})$$

In plain words, with the proposed modifications, we have succeeded in inducing a rational behavior for P^0 and players P^{j_i} ($i = 1, \cdots, N-1$) which consists of correctly following Protocol 1. The strategy of P^{j_N} can only be $s_4^{j_N}$, because P^{j_N} is not supposed to forward the content any further.

LEMMA 1. *With the utility functions defined in Equations (2) and the constraints (3) and (4), in Protocol 1 there is general coprivacy between P^{j_i} and $P^{j_{i+1}}$ for any $i \in \{0, \cdots N-1\}$. Additionally, the equilibrium strategy between P^{j_i} and $P^{j_{i+1}}$ results in increased redistribution tracing capabilities for P^0.*

PROOF. With the utilities in this section, the dominant strategy solution has been shown to be the one in which every player P^{j_i} plays $s_0^{j_i}$. Note that $s_0^{j_i}$ yields the maximal possible payoff $d_{j_{i+1}}$ for $P^{j_{i+1}}$: indeed, $P^{j_{i+1}}$ obtains the content while preserving her anonymity (thanks to anonymous fingerprinting). Now, whatever the strategy chosen by $P^{j_{i+1}}$, the general utility function $\mathbf{u}_{j_{i+1}}$ monotonically increases with $d_{j_{i+1}}$.

Hence, the best strategy for P^{j_i} results in enhanced general utility for $P^{j_{i+1}}$, whatever $P^{j_{i+1}}$'s strategy.

Additionally, since the best strategy for P^{j_i} is to transfer fingerprinted content, P^0 increases her tracing capability, because she can trace this transfer in case of redistribution (Protocol 2). The lemma follows. □

5. CONCLUSIONS AND FUTURE RESEARCH

We have described protocols whereby expiration dates for digital content can be enforced thanks to rational involvement by all players having received the content. This rational, coprivate approach allows digital oblivion to be effectively enforced.

Future research will be directed to designing automated procedures to obtain an empirical estimation of the rewards and punishment amounts needed in practical cases. We also plan to release as freeware a demonstrator of Protocols 1 and 2.

Acknowledgments and disclaimer

Thanks go to David Megías for suggesting the use of the audio watermarking scheme [19] as a practical instantiation. Insightful comment by Jordi Soria are also acknowledged. This work was partly funded by the Spanish Government through projects TSI2007-65406-C03-01 "E-AEGIS" and CONSOLIDER INGENIO 2010 CSD2007-0004 "ARES", by the Government of Catalonia through grant 2009 SGR 1135 and by the European Commission under FP7 project "DwB". The author is partly supported as an ICREA-Acadèmia researcher by the Government of Catalonia. He holds the UNESCO Chair in Data Privacy, but the views expressed in this paper are his own and do not commit UNESCO.

6. REFERENCES

[1] Y. Bo, L. Piyuan, and Z. Wenzheng. An efficient anonymous fingerprinting protocol. In *Computational Intelligence and Security*, LNCS 4456, Springer, pp. 824-832, 2007.

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

[3] J. Camenisch. Efficient anonymous fingerprinting with group signatures. In *Asiacrypt 2000*, LNCS 1976, Springer, pp. 415-428, 2000.

[4] J. E. Cohen. DRM and privacy. *Berkeley Technology Law Journal*, 18:575-617, 2003.

[5] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker. *Digital Watermarking and Steganography*. Burlington MA: Morgan Kaufmann, 2008.

[6] J. Domingo-Ferrer. Multi-application smart cards and encrypted data processing. *Future Generation Computer Systems*, 13(1):65-74, 1997.

[7] J. Domingo-Ferrer. Anonymous fingerprinting of electronic information with automatic identification of redistributors. *Electronics Letters*, 34(13):1303-1304, 1998.

[8] J. Domingo-Ferrer. Anonymous fingerprinting based on committed oblivious transfer. In *Public Key Cryptography-PKC 1999*, LNCS 1560, Springer, pp. 43-52, 1999.

[9] J. Domingo-Ferrer. Coprivacy: towards a theory of sustainable privacy. In *Privacy in Statistical Databases-PSD 2010*, LNCS 6344, Springer, pp. 258-268, 2010.

[10] J. Domingo-Ferrer, D. Megías and J. Soria. Distributed multicast of fingerprinted content based on a rational peer-to-peer community. Submitted manuscript, 2010.

[11] Facebook Terms of Service, Revision dated October 4, 2010. `http://www.facebook.com/terms.php`

[12] T. Furon and P. Duhamel. An asymmetric public detection watermarking technique. In *Information Hiding IH'99*, LNCS 1768, Springer, pp. 88-100, 2000.

[13] T. Furon and P. Duhamel. An asymmetric watermarking method. *IEEE Transactions on Signal Processing*, 51(4):981-994, 2003.

[14] G.-F. Gui, L.-G. Jiang and C. He. A robust asymmetric watermarking scheme using multiple public watermarks. *IEICE Trans. Fundamentals*, E88-A(7):2026-2029, 2005.

[15] ITU-R. Recommendation BS.1387. Method for objective measurements of perceived audio quality. December 1998.

[16] A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic Theory*, Oxford: Oxford Press, 1995.

[17] V. Mayer-Schönberger. Beyond copyright: managing information rights with DRM. *Denver University Law Review*, 84: 181-198, 2006.

[18] V. Mayer-Schönberger. *The Virtue of Forgetting in the Digital Age*. Princeton and Oxford: Princeton University Press, 2009.

[19] D. Megías, J. Serra-Ruiz, and M. Fallahpour. Efficient self-synchronised blind audio watermarking system based on time domain and FFT amplitude modification. *Signal Processing*, 90(12):3078-3092, 2010.

[20] N. Nisan, T. Roughgarden, É. Tardos, and V. V. Vazirani (eds.). *Algorithmic Game Theory*. Cambridge UK: Cambridge University Press, 2007.

[21] Opera software by Opticom. `http://www.opticom.de/products/opera.html`. Last checked on February 25, 2011.

[22] B. Pfitzmann and M. Waidner. Anonymous fingerprinting. In *Advances in Cryptology-EUROCRYPT'96*, LNCS 1233, Springer, pp. 88-102, 1997.

[23] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandeburg, and B. Feiten. PEAQ – The ITU standard for objective measurement of perceived audio quality. *Journal of the Audio Engineering Society*, 48(1-2):3-29, 2000.

[24] H. Yan-Jun, M. Xiao-Ping and G. Li. A robust public-key image watermarking scheme based on weakness signal detection using chaos system. In *International Conference on Cyberworlds 2008*, pp. 477-480, 2009.

[25] YouTube Terms of Service, Revision dated June 9, 2010. `http://www.youtube.com/static?gl=US&template=terms`