Improved Privacy-Preserving P2P Multimedia Distribution Based on Recombined Fingerprints

David Megías, *Member, IEEE*

***Abstract*—Anonymous fingerprint has been suggested as a convenient solution for the legal distribution of multimedia contents with copyright protection whilst preserving the privacy of buyers, whose identities are only revealed in case of illegal re-distribution. However, most of the existing anonymous fingerprinting protocols are impractical for two main reasons: 1) the use of complex time-consuming protocols and/or homomorphic encryption of the content, and 2) a unicast approach for distribution that does not scale for a large number of buyers. This paper stems from a previous proposal of recombined fingerprints which overcomes some of these drawbacks. However, the recombined fingerprint approach requires a complex graph search for traitor tracing, which needs the participation of other buyers, and honest proxies in its P2P distribution scenario. This paper focuses on removing these disadvantages resulting in an efficient, scalable, privacy-preserving and P2P-based fingerprinting system.**

***Index Terms*—P2P content distribution, anonymous fingerprinting, re-distribution tracing, recombined fingerprints**

I. Introduction

LEGAL distribution of multimedia contents is a recurrent topic of research. Broadband home Internet access has enabled the sustained growth of e-commerce, including direct downloads of multimedia contents. However, copyright infringement is one of the most relevant threats to the content industry.

Fingerprinting emerged [1] as a technological solution to avoid illegal content re-distribution. Basically, fingerprinting consists of embedding an imperceptible mark –fingerprint– in the distributed content (which may be audio, still images or video) to identify the content buyer. The embedded mark is different for each buyer, but the content must stay perceptually identical for all buyers. In case of illegal re-distribution, the embedded mark allows the identification of the re-distributor by means of a *traitor tracing* system, making it possible to take subsequent legal actions. Although fingerprinting techniques have been available for nearly two decades, the first few proposals in this field are far from nowadays' requirements such as scalability for thousands or millions of potential buyers and the preservation of buyers' privacy.

Most fingerprinting systems can be classified in three categories [3], namely symmetric, asymmetric and anonymous schemes. In symmetric schemes, the merchant is the one who embeds the fingerprint into the content and forwards the result to the buyer; hence, the buyer cannot be formally accused of illegal re-distribution, since the merchant also had access to the fingerprinted content and could be responsible for the re-distribution. In asymmetric fingerprinting, the merchant does not have access to the fingerprinted copy, but he can recover the fingerprint in case of illegal re-distribution and thereby identify the offending buyer. In anonymous fingerprinting, in addition to asymmetry, the buyer preserves her anonymity (privacy) and hence she cannot be linked to the purchase of a specific content, unless she participates in an illegal re-distribution. Anonymous fingerprinting [16] is, thus, the most convenient strategy to protect both the buyers' privacy and the owner's rights, since it guarantees the following properties: 1) only the buyer obtains the fingerprinted copy of the content, making it impossible for the merchant to accuse her of unlawful re-distribution, and 2) it preserves the anonymity of the buyers' identities with respect to the merchant.

As scalability is concerned, the unicast approach in which the merchant establishes a connection with each single buyer is not a convenient strategy. However, broadcast distribution is not suitable for fingerprinting applications since different fingerprints are required for different buyers in order to guarantee traceability. Peer-to-peer (P2P) distribution can be the answer to this difficulty, as proposed in this paper, since this technique blends some of the advantages of the unicast and multicast solutions. In fact, some content distributors are already operating under the P2P paradigm, such as [15].

Many anonymous fingerprinting schemes exploit the homomorphic property of public-key cryptography [14], [11], [19], [10]. These schemes allow embedding the fingerprint in the encrypted domain (with the public key of the buyer) in such a way that only the buyer obtains the decrypted fingerprinted content after using her private key. However, developing a practical system using this idea appears difficult, because public-key encryption expands data and substantially increases the communication bandwidth required for transfers [9]. Furthermore, homomorphic encryption constrains the type of mathematical operations which can be performed on the content for embedding, making it difficult to use the more advanced and robust techniques in the data hiding literature. In addition, the application of this idea in a distributed scenario (such as P2P networks) is not simple, since embedding would have to be performed by peer buyers, requiring a complex and supervised protocol.

Other approaches for anonymous fingerprinting [16], [17], [3], [2], [9], [4] do not exploit homomorphic encryption in this way, but either 1) require highly demanding technologies such as public-key encryption of the contents, secure multiparty protocols, commitment protocols or zero-knowledge proofs, among others, incurring prohibitive computational and

D. Megías is with the Estudis d'Informàtica Multimèdia i Telecomunicació, Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona, Catalonia, Spain.
E-mail: dmegias@uoc.edu.

communicational costs; or 2) are based on theoretical secure embedding algorithms for which no proof of existence is available.

Very few anonymous fingerprinting schemes with P2P distribution have been suggested so far [7], [12], [13]. In [7], game theory is applied to develop a fingerprinting scheme where embedding occurs between peer buyers, but this approach requires multi-party secure protocols between buyers which may be difficult to apply in a real scenario. The proposal in [12], [13] is more attractive, since embedding occurs only for a few seed buyers and the fingerprint of the other buyers are automatically generated as a recombination of the fingerprints of their "parents" in a graph distribution scenario. However, the traitor tracing protocol presented in those references requires an expensive graph search and disturbs a few honest buyers who must co-operate with the authority to identify the source of an illegal re-distribution. This is a relevant inconvenient, not only for the associated computational cost and the nuisance caused to honest buyers, but also because it represents a weakness of the proposal in [12], [13]. As shown in Section II-B, the participation of other buyers in the tracing protocol can lead to situations in which some illegal re-distributors may be untraceable even if no malicious behavior occurs. In addition, the distribution protocols proposed in [12], [13] rely on the honest behavior of proxies.

This paper reviews the main features of the proposal suggested in [12], [13], highlights its main drawbacks, and suggests several significant improvements to achieve a more efficient and practical system, especially as traitor tracing is concerned, since it avoids the situations in which illegal re-distributors cannot be traced with the proposal of [12], [13]. Furthermore, better security properties against potentially malicious proxies are obtained.

Although the system proposed in this paper uses public-key encryption in the distribution and traitor tracing protocols, it must be taken into account that this encryption is only applied to short bit strings, such as the binary fingerprints and hashes, not to the content. The fragments of the content are encrypted using symmetric cryptography, which is much more efficient.

The rest of this paper is organized as follows. Section II summarizes the references that are the basis of the proposed scheme and details the drawbacks related to the system described in those references. Section III describes the improvements proposed for the method, whose improved features and efficiency aspects are discussed in Section IV. A security and privacy analysis of the new system is presented in Section V. Finally, Section VI summarizes the conclusions of this work.

II. Previous work

The proposal presented in this paper stems from the fingerprinting system described in [12], [13], which introduced the concept of automatically recombined (also called DNA-inspired) fingerprints in P2P networks. The next sections present the main features and drawbacks of the previous work.

A. Main features of the previous work

The main features of the referred method are the following:

- The content is divided into several ordered fragments and each of them is embedded separately with a random binary sequence. The binary sequence for each fragment is called segment and the concatenation of all segments forms the whole fingerprint.
- The merchant distributes different copies to a reduced set of M seed buyers. The fingerprints of these buyers are such that their segments have low pair-wise correlations.
- The buyers other than the seed ones engage on P2P transfers of the content in such a way that each new buyer obtains fragments from at least two other buyers. The total number of buyers is $N \gg M$.
- The communication between peer buyers is anonymous through an onion routing-like protocol using a proxy.
- The fingerprint of each new buyer is built as a recombination of the segments of its parents.
- Proxies know the pseudonyms of source and destination buyers and they have access to the symmetric keys used for encrypting the multimedia content.
- A transaction record is created by a transaction monitor to keep track of each transfer between peer buyers. These records do not contain the embedded fingerprints, but only an encrypted hash of them. The fingerprints' hashes are encrypted in such a way that the private key of at least one parent is required for obtaining their cleartext.
- The real identities of buyers are known only by the merchant. The transaction monitor records buyers' pseudonyms.
- In case of illegal re-distribution, a search is required through the distribution graph. The search starts from the seed buyers and is directed by a correlation function between the traced fingerprint and the fingerprints of the tested buyers. These tested buyers must co-operate with a tracing authority to compute the correlation between their fingerprint and the one extracted from the illegally re-distributed file. The fingerprints' hashes recorded in the transaction monitor are enough to prevent buyers from cheating in this step.
- At each step of the traitor tracing protocol, the buyer with maximum correlation is chosen as the most likely ancestor of the illegal re-distributor. This criterion is mostly right, but some incorrect choices may occur during the search process, requiring the exhaustion of a subgraph and backtracking.
- The search ends when perfect correlation is found between the fingerprint of the tested buyer and that of the illegally re-distributed file. If a buyer refuses to take a correlation test, the hash recorded in the transaction monitor can be used as evidence against her.
- A method to defeat the collusion of several buyers is also described.

The automatic construction of fingerprints by re-combining *segments* of the parent buyers' fingerprints is depicted in Fig.1. It is worth pointing out the difference between the terms "fragments" and "segments" as used in this system. A "segment" refers to each of the fixed-sized pieces that form the whole fingerprint embedded in a content, whereas the term

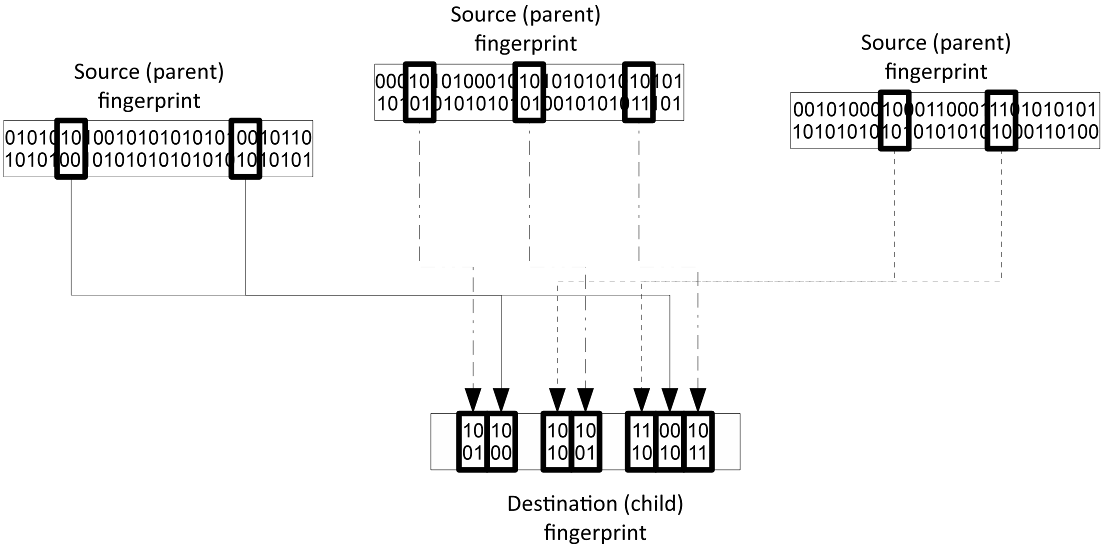

Fig. 1: Automatic recombined fingerprint construction

"fragment" is used for the different pieces of the content itself, each of which contains a segment embedded into it.

As detailed in [13], each child is interested in obtaining fragments from more than one parent, and each parent is interested in *not* providing all the fragments to the same child: the reason is that if two peers A and B get exactly the same copy of the content, then A could be held responsible for any unlawful content re-distributed by B and viceversa. This mutual interest of children and parents to protect each others' privacy is known as the co-privacy property, as defined in [7].

This recombination (or DNA-inspired) fingerprinting scheme has remarkable advantages for both buyers and the merchant: 1) no one has access to the full fingerprints of buyers; 2) transactions between buyers are fully anonymous; and 3) a tracing protocol makes it possible to identify the source of an illegal re-distribution.

B. Drawbacks of the previous work

Despite its advantages, the scheme also has some drawbacks, the most remarkable being: 1) the tracing process may be cumbersome and requires the participation of a few innocent buyers; 2) the fraction of tested buyers in a search is not known (although it is proven to asymptotically decrease to zero as the number of buyers increases if no backtracking occurs); and 3) although it is recommended that more than one proxy is used for each download, the proxies could still collude to reconstruct the whole fingerprinted copy of a buyer and illegally re-distribute that copy, making it possible to frame an innocent buyer. All three drawbacks are quite inconvenient, especially the third one that may lead to charges on an innocent user.

The fact that some honest buyers are required to participate in the traitor tracing protocol has several implications. To begin with, in some cases, a few buyers may have perfectly honest reasons for not collaborating in traitor tracing. For example, they may have lost the copy of the content being traced either because of file removal or hardware damage. In such a situation, there are some distribution scenarios that may lead to one or more buyers being impossible to trace. This is illustrated in Fig. 2. Imagine that buyers B_{13} and B_{17} are unable to collaborate with the traitor tracing protocol, either because of honest or malicious reasons, and that they cannot provide their private keys (used in the distribution system). If the illegal re-distributor was B_{22}, since her only parents are B_{13} and B_{17}, there would be no way of decrypting the

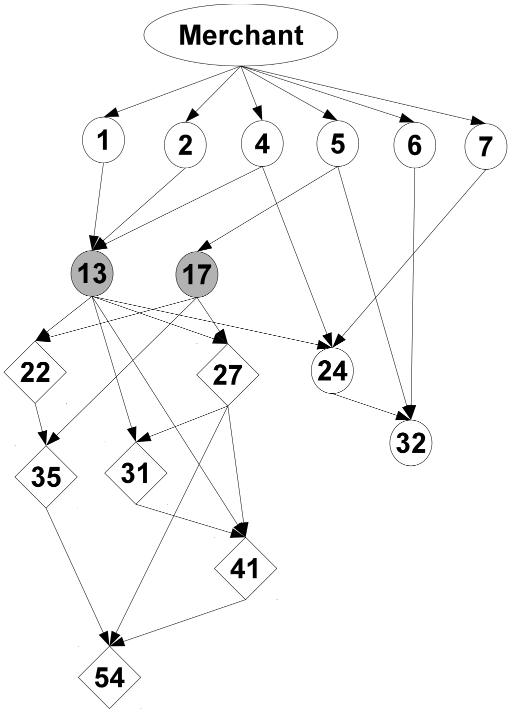

Fig. 2: Possibly untraceable subgraph

fingerprint's hash of B_{22} (stored in the transaction monitor in encrypted form). if B_{22} refused to take part in the protocol (as it is likely in case she was the illegal re-distributor), the merchant would never be able to accuse her. This situation could affect not only a single buyer, but a whole subgraph of buyers if they have B_{13} and B_{17} as the only ancestors. Hence, buyers B_{22}, B_{27}, B_{31}, B_{35}, B_{41} and B_{54} may all be untraceable (if other conditions occur) once B_{13} and B_{17} are unable to participate in the traitor tracing protocol.

Hence, the participation of honest buyers in the traitor tracing protocol is not only a matter of computational and communicational costs. This participation may lead to situations in which an illegal re-distributor cannot be traced by the system. In addition, the involvement of honest buyers entails some other unwanted consequences:

- Firstly, when an honest buyer participates in the traitor tracing protocol, her privacy is, to some extent, surrendered. Although the honest buyer B_i will only be identified with her pseudonym, the fact is that the authority will be aware that B_i has purchased a copy of those particular content. In addition, the authority will certainly have access to other sensitive information about that buyer, such as her IP address.
- Secondly, the participation in the traitor tracing protocol requires that the correlation between the buyer's and the traced fingerprints is computed. Even if this correlation is computed only by the authority (which should be trusted), this means that the particular fingerprint of the buyer is no longer secret (unless a very sophisticated multi-party secure protocol is used for computing the correlation). If the fingerprint of the buyer B_i is available to another party (the authority), if B_i decides to re-distribute the file illegally in the future, it may be difficult for the merchant to take legal action against her. B_i may argue that her fingerprint or even her copy of the content could have been leaked during the traitor tracing protocol, obtaining future immunity for illegal re-distributions.

Avoiding the participation of buyers in traitor tracing is, thus, a relevant issue to improve the system presented in [12], [13] significantly. The next section is focused on providing cryptographic protocols to modify the original scheme and solve the problems described above.

III. Improved system

Several improvements for the system [12], [13] to overcome the aforementioned drawbacks are described in this section.

A. *System model*

The participants in the proposed fingerprinting system are the following:

- Merchant: he distributes copies of the content legally to the seed buyers. Each fragment of the content contains a different segment of the fingerprint embedded into it. The segments have low pair-wise correlations.
- Seed buyers (B_i for $i = 1, \ldots, M$): they receive fingerprinted copies of the contents from the merchant that are used by the P2P distribution system to bootstrap the system. They can be either real or dummy buyers as discussed in [13].
- Other buyers (B_i for $i = M+1, \ldots, N$, with $M \ll N$): they purchase the content and obtain their fingerprinted copies from the P2P distribution system. The content is assembled from fragments obtained from different "parents". Anonymous connections with peer buyers are provided by means of proxies.
- Proxies: they provide anonymous communication between peer buyers by means of a specific protocol analogous to Chaum's mix networks [5] (see below).
- Transaction monitor: it keeps a transaction register for each purchase carried out for each buyer. This transaction register includes an encrypted version of the embedded fingerprints.
- Tracing authority: in case of illegal re-distribution, it participates in the tracing protocol that is used to identify the illegal re-distributor(s).

B. *Security model*

The security assumptions of the proposed system are the following:

- The merchant does not need to be trusted either for distribution or to associate a pseudonym with the identity of a buyer. The protocols for distribution and for traitor tracing described below are proven to work even if the merchant is not trusted.
- Buyers are not trusted and protocols are provided to guarantee that 1) they are transferring authenticated fragments of the content and 2) their anonymity can be revoked in case they re-distribute the content illegally.
- The transaction monitor (or any other single party) will not have access to the cleartext of the fingerprints. This prevents that any single party can frame an innocent buyer.
- As privacy is concerned, the transaction monitor is not trusted and it should only have access to pseudonyms, but not to the buyers' real identities.
- The transaction monitor is trusted as the symmetric keys used for encrypting the fragments are concerned. This means that 1) the transaction monitor stores the key provided by each parent buyer and 2) this key can be retrieved only once from its database (in principle by the child buyer). After this retrieval, the transaction monitor blocks the register and eventually removes it (details are provided in the description of Protocol 1 below).
- The transaction monitor returns the true pseudonym corresponding to an illegal re-distributor in the traitor tracing protocol. However, this trust can be replaced by a collection of signatures provided by the proxies.
- The tracing authority is part of the legal system and shall be trusted. It is not expected that the authority participates in any coalition to frame an innocent buyer or break someone's privacy.
- The communication between the merchant and the seed buyers, and between peer buyers within the P2P distribu-

tion system, must be anonymous using an onion routing-like approach. The fragments of the content are encrypted using symmetric cryptography.

- Proxies are not trusted and the fragments sent through them shall be encrypted in such a way that only the sender and the recipient have access to their cleartext. Malicious proxies may also try to cheat by reporting false fingerprint segments (or not reporting them at all) to the transaction monitor.
- The fingerprints must be constructed with a long enough number of segments to guarantee that recombination will produce different fingerprints for different buyers due to numerical explosion.
- The hashing functions used in the system are secure and cannot be inverted.
- Public-key cryptography is restricted to the encryption of short binary strings, such as fingerprint segments or hashes. The different parties: merchant, transaction monitor, proxies and buyers have a pair of public and private keys to be used in different steps of the protocols.
- A single malicious party shall not be able to construct the fingerprinted copy corresponding to any buyer to frame an honest user of the system. In a similar way, a single malicious party shall not be able to link the identity of some buyer to a particular content unless that user is involved in an illegal re-distribution of the content.

Different attacks that may be mounted against the system proposed in this paper, regarding both security and privacy, are described below. The following assumptions are made:

- The watermarking method used for embedding and detecting the fingerprint is transparent, robust and secure enough for a fingerprinting application. There are hundreds of watermarking algorithms available to be used with the proposed fingerprinting system and watermarking properties are out of the scope of this paper. For example, a robust video watermarking scheme is presented in [18] and a robust audio watermarking scheme is described in [8].
- Collusion attacks are the main topic in most of the research related to fingerprinting. Collusion occurs when several buyers decide to recombine their fingerprinted copies of a given content trying to obtain a new copy in which neither of their fingerprints is detectable. The system suggested in this paper inherits the collusion resistance of the method described in [12], [13] and, hence, no further analysis on collusion attacks is required.

Thus, the main attacks that may be performed on the proposed system are related to either the P2P distribution protocol, the traitor-tracing protocol and the P2P network itself. These attacks may be aimed to break either the security or the privacy properties of the system. The attacks to the cryptographic protocols require that one or more of the involved parties are malicious or that a malicious party tries to mimic the behavior of an honest party in order to gain sensitive information that may be used afterwards. As far as security is concerned, there are two main items to be protected:

- **Buyer frameproofness** is related to the possibility that an innocent buyer is accused of illegal re-distribution of the purchased content.
- **Copyright protection** would be broken if any party obtains a copy of the content whose fingerprint is not included in the fingerprints' database of the transaction monitor (and thus can be re-distributed illegally) or the association of that particular fingerprint with the illegal re-distributor cannot be completed.

On the other hand, privacy would be broken if someone could associate a real identity with the purchase of some specific content.

Attack models to frame honest buyers require that another party is able to obtain either the fingerprint of a buyer, or the fingerprinted copy of the content, in such a way that it can be further re-distributed and, finally, the honest buyer is accused of illegal re-distribution. The following types of attacks would be aimed to frame an innocent buyer:

- Random guess: the merchant has access to all fragments of the seed buyers and he may try to recombine the fragments to produce a new copy of the content. If this copy is re-distributed, there is a possibility that it leads to some innocent buyer that could be framed.
- Buyer authentication attacks: an attacker may impersonate a buyer in the system and try to obtain a fingerprinted copy of the content that would be linked to the impersonated buyer.
- Proxy authentication attacks: an attacker may impersonate a proxy in the system and try to obtain fragments of a content from different buyers in order to build a fake "colluded" copy of the content. This fake "collusion" could produce evidence against one ore more honest buyers that served fragments to the fake proxy.
- Man-in-the-middle attacks: an attacker may try to intercept the traffic between a buyer and one or more of her proxies and keep a copy of all the fragments of the content.
- Database authentication attacks: an attacker may try to obtain the fingerprint of a buyer that is stored in the transaction monitor's database.
- Protocol attacks: one or more of the participants in the protocols (proxies, merchant, buyers, transaction monitor or tracing authority) may be malicious and try to obtain the fingerprint or the fingerprinted content linked to a particular honest buyer.

The security of the system against these attacks is discussed in Section V.

C. Modified P2P distribution protocol

As already remarked, in the original distribution protocol, the fingerprints were not stored in the transaction monitor in order to protect the privacy of the buyers. Only the hash of the fingerprint was stored for each buyer. The fingerprint's hash was encrypted and stored as many times as parents each buyer had, using the public key of the parent for encryption (and also the public key of the transaction monitor). In this way, the participation of at least one parent was required to obtain the cleartext of the fingerprint's hash.

The new proposal is to store *also* the fingerprints of the buyers in an encrypted form. The transaction registers would then be formed as follows:

P_i	Username (pseudonym) of the buyer B_i.
$H(c)$	Perceptual content hash (used for indexing in the content database).
E_{h_i}	Encrypted hash of the buyer's fingerprint.
E_{f_i}	**Encrypted buyer's fingerprint**.
d	Transaction date and time (for billing purposes).

In the original proposal [13], E_{h_i} was stored one time per parent with double encryption, using the public keys of the parent **and** the transaction monitor. In the improved proposal, E_{h_i} is encrypted **only** with the public key of the transaction monitor. Note that having access to the fingerprints' hashes does not allow the transaction monitor to reconstruct any buyer's fingerprint, since a hash function is not invertible, thereby preserving buyer frameproofness.

The improvements to the system stem from the storage of an encrypted version of the buyers fingerprints, E_{f_i}, computed as follows:

- Each fragment of the content shall be transmitted with a fingerprint's segment g_j embedded into it and together with an encrypted version of the segment $E^c_{g_j} \doteq E(g_j, K_c)$, where K_c is the public key of the transaction monitor. A signature can also be included, as detailed in Section V.
- Each proxy selects a set of m contiguous fragments of the content and facilitates the anonymous communication between parents and child for the transmission of those fragments. These m contiguous fragments of the content carry m contiguous segments of the fingerprint embedded into them. The construction of the fingerprint with segments and sets of contiguous segments is shown in Fig. 3.
- The same proxy is required for transferring the m contiguous fragments of the content. In this process, the proxy stores the corresponding encrypted segments $E^c_{g_j}$.
- The proxy concatenates the m contiguous encrypted segments, encrypts the concatenation using the public key of the tracing authority (K_a) and sends the result to the transaction monitor.
- Hence, the transaction monitor stores the following encrypted version of the fingerprint:

$$E_{f_i} = E\left(E^c_{g_1}|E^c_{g_2}|\dots|E^c_{g_m}, K_a\right)|\dots| \\ E\left(E^c_{g_{(L-1)m+1}}|E^c_{g_{(L-1)m+2}}|\dots|E^c_{g_{Lm}}, K_a\right). \quad (1)$$

Note that 1) no proxy has access to the complete sequence of encrypted segments (since at least two proxies must be chosen by each buyer as described in [13]) and 2) the transaction monitor cannot decrypt E_{f_i} without the private key K^s_a of the authority.

In both the original and the modified schemes, the (decrypted) fingerprint's hash h_i would be used in case of collusion of several buyers, as discussed in [13].

The second modification of the transmission protocol refers to the use of symmetric cryptography to encrypt the content in such a way that intermediate routers do not have access to the cleartext of the content. Since communication is anonymous through a proxy, this was achieved, in [13], by means of symmetric encryption with a session key chosen by the child buyer that was forwarded to the proxy and then, from the proxy to the parent buyer. Of course, this implied that the proxy could also decrypt the content and a coalition of all proxies chosen by a child buyer would be enough to reconstruct the full copy of the fingerprinted content and frame an honest buyer. A solution is proposed below.

The new protocol for anonymous communication is also based on using proxies between the parent and the child buyer. The content that is transferred over the proxy is encrypted, again, using symmetric cryptography, but the session key for encrypting the content is shared by parent and child using the transaction monitor as a temporary key database.

Protocol 1 (Anonymous communication):

1) The parent buyer chooses a symmetric (session) key k.
2) The parent chooses a pseudorandom binary sequence r to be used as a handle (primary database key) for k. The space for r should be large enough (e.g. 128 bits) to avoid collisions.
3) The parent buyer sends (r, k) to the transaction monitor, who stores it in a database.
4) The parent buyer sends r to the proxy and the proxy forwards r to the child buyer.
5) The child buyer sends the handle r to the transaction monitor, who replies with the symmetric key k.
6) The transaction monitor blocks the register (r, k) for a given period (timer). When the timer expires, the transaction monitor removes the register from the database.
7) The parent buyer sends the requested fragments, encrypted with k, to the proxy.
8) The proxy forwards all fragments to the child buyer, who can decrypt them using k.

An analysis of the security of this protocol is provided in Section V.

D. Improved traitor tracing protocol

The tracing protocol described in [13] required exploring the distribution graph from the seed buyers following a path directed by a correlation function between the fingerprint of the illegally re-distributed copy and that of the explored buyers. Hence, some innocent buyers were requested to cooperate with the traitor tracing system. With the modifications introduced in the previous section, a simpler traitor tracing protocol is possible.

The new basic traitor tracing protocol (when no collusion occurs) begins with the extraction of the fingerprint of the illegally re-distributed copy by the tracing authority. Then, the authority uses the public key of the transaction monitor and its own public key to produce the encrypted fingerprint which can be efficiently searched in the database of the transaction monitor. Once the pseudonym of the illegal re-distributor is available, it can be associated to a real identity.

g_1	g_2	$\cdots$	g_m	g_{m+1}	g_{m+2}	$\cdots$	g_{2m}	$\cdots$	$g_{(L-1)m+1}$	$g_{(L-1)m+2}$	$\cdots$	g_{Lm}

Fig. 3: Fingerprint's segments (g_j) and sets of m contiguous segments

Protocol 2 (Basic traitor tracing):

1) The fingerprint f of the illegally redistributed content is extracted by the tracing authority using the extraction method and the extraction key (provided by the merchant).
2) The fingerprint's segments g_j are encrypted using the public key of the transaction monitor: $E^c_{g_j} = E(g_j, K_c)$.
3) The encrypted segments are grouped in sets of m consecutive elements which are encrypted using the public key of the authority, thereby obtaining E_f as per (1).
4) E_f is (efficiently) searched in the database of the transaction monitor in order to recover the pseudonym of the illegal re-distributor.
5) The merchant checks his database of clients and retrieves the identity of the traitor corresponding to the pseudonym obtained in the previous step.

In the last step of this protocol, to avoid a malicious merchant from cheating and returning the identity of an innocent buyer, the identity must come together with a signed document that certifies the association between the pseudonym and the buyer's identity. All buyers will have to sign such a document (using their private keys) upon registration in the system. These signatures will be verifiable using the buyers' public keys. In this way, 1) there is no need for the merchant to be trusted and 2) non-repudiation from illegal re-distributors with respect to their pseudonym is obtained.

Note that all fingerprints are kept secret except the one that is being traced (f). The advantages of this traitor tracing system are obvious: the cleartext of the fingerprints of honest buyers is never required, and the traitor tracing protocol is based on a a standard database search (not involving innocent buyers).

A security analysis of this protocol is also provided in Sectin V.

IV. Improved features and efficiency

This section discusses the improvements introduced by the new protocols and the efficiency and scalability properteis of the proposed system.

A. Improved features

As remarked in Section II-B, there are three major drawbacks of the system published in [12], [13]: 1) the tracing system is cumbersome and requires the participation of several honest buyers, 2) the number of tested buyers in the tracing protocol is not known a priori and 3) the system relies on honest proxies for anonymous content distribution. This section discusses how these drawbacks are overcome by the modified system.

The first two drawbacks were a consequence of the involvement of an unknown number of honest buyers in the tracing protocol of [12], [13]. In the modification presented in this paper, the transaction monitor stores an encrypted version of the fingerprint of each buyer, which was not recorded in the original protocol. This encrypted version of the fingerprint makes it possible to trace an illegal re-distributor in the tracing protocol (Protocol 2) without the involvement of any buyer and also without decrypting any single fingerprint. Hence, since no involvement of buyers is required for traitor tracing, the first two drawbacks are directly avoided.

The third drawback was caused by the access of proxies to the symmetric keys used to encrypt the content in the distribution protocol of [12], [13]. This drawback has been circumvented in this paper with the modified protocol for anonymous communication between peer buyers (Protocol 1). The new protocol protects the symmetric keys using a database in the transaction monitor and, thus, the keys are not transmitted through the proxies. In this way, if a proxy tries to access the database and retrieve the symmetric key, the corresponding register will be blocked when the receiver buyer tries to access it, and the malicious access by the proxy would be detected.

B. Efficiency aspects

In this section, the *efficiency* of the proposed fingerprinting system is discussed taking into account different points of view. Although simulations could be provided comparing the proposed system with other works of the literature, it must be taken into account that most of the anonymous fingerprinting protocols proposed so far are centralized and, thus, rely on unicast distribution. Simulated experiments to compare unicast and P2P distribution protocols, both as the CPU and communications costs are taken into account, would produce extremely different results, as discussed below. For this reason, a reasoned discussion of this topic (not illustrated with simulations) is provided in this section.

From the merchant's viewpoint, the efficiency of a multimedia content distribution system often refers to scalability. There are two main reasons why the system proposed in this paper is scalable. The first one is the use of a P2P distribution system, which reduces the distribution cost significantly on the merchant side: the merchant bootstraps the system by feeding only a few seed buyers, whereas the rest of transactions occur between peer buyers. Among the systems mentioned in the introduction, only [7], [12], [13] propose P2P-based distribution systems. In fact, the system of [12], [13] is exactly as efficient as the modification of this paper from the merchant's point of view. On the other hand, the system of [7] requires a multi-party secure protocol between each seed buyer and the merchant to bootstrap the system. Multi-party secure protocols are difficult to implement and require large computational and communication costs.

All the other anonymous fingerprinting systems reported in the introduction [16], [17], [3], [14], [11], [2], [19], [9], [4], [10] are centralized. This means that: 1) a unicast distribution is required for each buyer; and 2) an anonymous protocol for embedding the fingerprint and transmitting the fingerprinted file is run between the merchant, each buyer and (possibly) other parties. Obviously, these solutions are much less scalable for the merchant, and involve much higher computational and communicational costs.

Hence, it can be concluded that the proposed system (as well as that of [12], [13]) is the most efficient solution for a merchant.

As the buyers are concerned, the centralized approaches also involve running expensive protocols, either because of the use of homomorphic encryption of the contents (which require homomorphic decryption in the buyer's side), zero-knowledge proofs, bit commitments or other similar techniques. Only [4] takes the computational costs for the buyers into account and proposes the use of powerful servers for the most expensive operations of the protocols. Neither the system proposed in this paper nor that of [12], [13] involve expensive decryption of the content (since symmetric encryption is used), zero-knowlege proofs or bit commitments. Although the buyers are required to participate in anonymous communication protocols using symmetric encryption and proxies, these protocols are not particularly demanding in communication or computation terms. For a buyer, the cost associated with obtaining a fingerprinted copy of the content will not be higher than that associated to a centralized approach. However, in P2P distribution systems, the buyers themselves become providers of the contents for other buyers and, thus, their bandwidth will be used in further transactions. Hence, P2P-based systems entail an upload cost for each buyer as long as she keeps the content in her multimedia database.

As the proposal of [7] is concerned, it is required that each buyer (except the leaves of the distribution tree) participates in an anonymous fingerprinting protocol with at least another buyer. In such a protocol, a new fingerprint must be embedded for each new buyer, which increases the communication and computational costs for the buyers who become sources of the content. Note that neither the proposed system nor the one described in [12], [13] require that the buyers embed new fingerprints, since the new fingerprints are created through recombination of the segments of existing ones.

In principle, buyers would prefer a centralized distribution system in what regards bandwidthd, since once the file has been obtained, no further communications are required. Nevertheless, the fact that the distribution is much "cheaper" from the merchant's point of view implies that the infrastructure (servers) required by the merchant is much simpler. The reduced costs as compared with centralized distribution and fingerprinting will certainly have some effect in the price of the contents. If buyers enjoy cheaper prices, this would compensate them for the use of their bandwidth required by the P2P distribution protocol.

Another disadvantage of the proposed system, as far as efficiency is concerned, is the need of an anonymous communication protocol, analogous to Chaum's mix networks [5], between the peer buyers. This protocol increases the number of messages in the network as a consequence of the use of proxies. In any case, this is a general disadvantage of any anonymous communication system, since a direct connection between sender and recipient must be avoided.

In short, the proposed system (as well as [12], [13]) can be considered efficient from both the merchant's and the buyers' points of view, but the buyers are requested to become providers of the content for other buyers (which is a standard drawback of any P2P distribution system).

V. Security and privacy analysis

This section analyzes the security and privacy properties of the proposed system according to the security model introduced in Section III-B.

As detailed in Section III-B, attacks to the system may be classified as authentication/impersonation attacks, man-in-the-middle attacks and protocol attacks. Authentication/impersonation attacks should be overcome by using existing secure authentication protocols and are out of the scope of this paper. As man-in-the-middle attacks are concerned, there is no possibility of intercepting and decrypting the messages between a buyer and a proxy, since communications with the transaction monitor and the child buyer should also be attacked in order to obtain the session key used for encrypting the content. If the communication between the child buyer and the transaction monitor (Step 5 of Protocol 1) are strongly authenticated (e.g. using a Public Key Infrastructure), the possibility of a successful man-in-the-middle attack can be neglected.

The following sections deal with the security and privacy of the protocols proposed, firstly taking a formal approach and then with a description of more complex collusion attacks.

A. Formal analysis of the proposed protocols

First of all, the security and privacy properties of Protocols 1 and 2 is analysed by means of two theorems (and their corresponding proofs).

Theorem 1: In Protocol 1, a malicious proxy trying to decrypt the fragments of the content would be detected.

Proof: If a malicious proxy tries to obtain the session key k by sending r to the transaction monitor there are two possibilities:

1) If the child buyer has already retrieved k from the database by sending the handle r to the transaction monitor, the register containing k would be either blocked or removed. Note that the transaction monitor is assumed to be honest for the management of the symmetric keys (see Section III-B).
2) If the child buyer has not retrieved k from the transaction monitor, the proxy will obtain it, but the child buyer will find the corresponding register either blocked or removed. Then, the malicious behavior of the proxy can be reported to the authorities and the transaction monitor and the child buyer have enough information (such as

pseudonyms and IP addresses) to identify the misbehaving proxy. Again, the assumption of honest behavior for the management of symmetric keys (Section III-B) applies.

Hence, a malicious proxy trying to obtain k from r would be detected, since the register would be blocked either to the proxy or to the child buyer, raising an investigation. This completes the proof. ■

Theorem 2: By applying Protocol 2, an illegal re-distributor can be traced efficiently using a standard database search in the transaction monitor and it is not required to decrypt any of the fingerprints recorded by the transaction monitor. The output of the tracing protocol is the identity of at least one illegal re-distributor.

Proof: If no collusion occurs, the fingerprint f would be first extracted by the tracing authority, which is trusted (Section III-B). Then the tracing authority would compute $E^c_{g_j} = E(g_j, K_c)$ for each segment (using the public key of the transaction monitor), and finally obtain E_f after grouping the segments in sets of m consecutive elements and encrypting these groups with its public key K_a. After that, the transaction monitor, which is also trusted for transaction database search (Section III-B), would output the pseudonym of the illegal re-distributor. The pseudonym can be linked to the real identity by the merchant, who provides also a signed document that associates the real identity and the pseudonym. This completes the proof. ■

In case of collusion of several buyers, the extracted fingerprint would not be a valid codeword of the anti-collusion code used in the scheme. Then, the system described in [13] would be used: the encrypted hash $E_{h_f} = E(h_f, K_c)$ would be searched instead of the encrypted fingerprint, where h_f denotes the hash obtained applying the hash function to the traced fingerprint f. Thus, Protocol 2 would be used with the hash of the fingerprint instead of the fingerprint itself. As described in [13], with a large enough hash space, hash collisions would be almost negligible and a traitor would still be identified in the vast majority of the cases.

The requirement that the transaction monitor is trusted and returns the pseudonym of the buyer associated with the traced fingerprint (and not a different pseudonym) can be relaxed if a signature of the encrypted sets of segments of the fingerprint is provided by the proxies. These signatures can be verified using the public keys of the proxies. In that case, both the signatures and the pseudonyms of the proxies shall also be included in the registers of the transaction database to facilitate the verification of these signatures when required.

B. Collusion attacks on the protocols

This section discusses possible collusion attacks on the proposed protocols.

1) Buyer frameproofness: As already discussed in [13], the merchant is not able to produce any buyer's fingerprint by random guess due to the numerical explosion of the fingerprint space, even with a reduced number of seed buyers. On the other hand, the transaction monitor has access only to the hashes of the fingerprints (not the fingerprints themselves without the private key of the authority). Since the hash function is not invertible, it is not possible for the monitor (even in coalition with the merchant) to reconstruct any buyer's fingerprint. Possible collusions to disclose the specific fingerprint of an innocent buyer are the following:

1) The tracing authority and the transaction monitor.
2) All the proxies (for a transfer) and the transaction monitor.
3) All the proxies (for a transfer) and the merchant.

In the first case, the authority and the transaction monitor may use their private keys to obtain the cleartext of all the fingerprints. However, this possibility can be neglected since at least the authority must be trusted (as described in Section III-B). In the second case, all the segments of the fingerprint could be decrypted using the private key of the transaction monitor, since the malicious proxies would not encrypt them with the public key of the authority. Also, the transaction monitor could collude with the proxies and use the session keys k to decrypt the fragments. Both possibilities would involve at least three malicious parties: all the proxies (two at least per each purchase) and the transaction monitor. In the third case, even if the transaction monitor does not provide her private key, a brute force attack segment by segment would be possible to reconstruct a buyer's fingerprint, because the number of different segments is small for each fragment (equal to M). Again, at least three malicious parties would be required: two (or more) proxies plus the merchant.

Hence, the minimum coalition required to frame an innocent buyer is formed by three malicious parties (or two if one of them is the authority). Note that a coalition of the transaction monitor and the merchant is not enough to obtain the cleartext of any fingerprint. As the proxies encrypt a set of m consecutive segments, and there are M possible values for each segment, the total number of combinations per set of consecutive segments is M^m. This avoids a brute force attack if m is reasonably large. For example, if $M = 10$ and $m = 32$, there would be 10^{32} possible combinations for each set of consecutive segments, what would be enough for security against a brute force attack. If the segments were encrypted one by one (or grouped with a small value of m), the system would be vulnerable against a brute force attack for a collusion of the merchant and the transaction monitor.

2) Copyright protection: In order to ensure copyright protection, it is essential that the fingerprint embedded in each buyers' copy of the content and its encrypted version recorded by the transaction monitor are identical. If there is a way to cheat in the recorded fingerprint, the corresponding buyer would be able to re-distribute her copy illegally without any chance of being detected. As already remarked in [13], the content fragments are signed by the merchant from origin. The same approach can be used here for each encrypted segment of the fingerprint, making it impossible for a proxy to cheat about the fingerprint. The authority and the merchant could verify randomly, with some probability, the signatures of the set of contiguous segments reported by a proxy. If the signature was not verified, the proxy would be accused of forgery. Note that the fingerprints would still be protected since 1) only some sets of contiguous segments would be verified (not the whole fingerprint) and 2) those segments would still be encrypted with the transaction monitor's public key.

However, a proxy may still try to get alternative fragments for the same position of the content by requesting them from different parents. That possibility would allow the proxy to cheat about the true fingerprint of the child buyer, since several correctly signed fragments would be available for him for the same content. This behavior can be avoided in several ways. For example, temporary records can be created in the transaction monitor by the parents to detect if a proxy tries to obtain two alternative fragments for the same content.

3) Buyers' privacy: The identity of a buyer who has purchased a specific content could be revealed by a coalition of two parties: one of the proxies chosen by the buyer and the merchant (who can link her pseudonym to a real identity) or, similarly, the transaction monitor and the merchant. Better privacy could be achieved if, for example, the pseudonyms were encrypted by the proxies using the public key of the tracing authority. In that case, a coalition of the merchant and the transaction monitor would not be enough to break a buyer's privacy, but a coalition of a proxy and the merchant would still be enough. However, the merchant should not be interested, in principle, to break her clients privacy, since privacy would be one of the clear advantages of the proposed distribution system.

Another threat to privacy is the fact that all anonymous communications between each child and each parent occur through a unique proxy. This means that this proxy has access to different pseudonyms (the parents' and the child's). This can be easily circumvented if more proxies are used in Protocol 1 between child and parent. With two proxies, each of them would know only the pseudonym of one of the parties (although they could still collide). With three or more proxies, only two of them would have access to different pseudonyms (either the parents' or the child's). Of course, increasing the number of proxies in each transfer would affect the efficiency of the system, since more communication burden would be required.

VI. Conclusion

The use of automatic recombined fingerprints has been recently suggested in the literature [12], [13], showing remarkable advantages: the fingerprints of buyers are unknown to the merchant (achieving anonymity) and fingerprint embedding is required only for a few seed buyers, whereas the other fingerprints are automatically obtained as a recombination of segments. However, the published system has some shortcomings: 1) it requires an expensive graph search in order to identify an illegal re-distributor, 2) some innocent buyers are requested to co-operate for tracing, and 3) the P2P distribution protocol requires honest proxies. This paper shows that the co-operation of honest buyers in traitor tracing entails several relevant drawbacks that can make the published system fail under some circumstances.

The improvements suggested in this paper overcome the drawbacks of [12], [13] by recording the fingerprints using multiple encryption in such a way that the graph search is replaced by a standard database search, whilst buyers' frameproofness is retained. Also, misbehaving proxies are discouraged by means of random checks by the authority and using a four-party anonymous communication protocol to prevent proxies from accessing the cleartext of the fragments of the content.

The final result is a fingerprinting system that features: 1) efficient and scalable distribution of multimedia contents in P2P networks; 2) efficient traitor tracing of illegal redistributors through a standard database search; 3) privacy preservation and buyer frameproofness; 4) mutual anonymity for merchant and buyers and between peer buyers; 5) collusion resistance; 6) avoidance of fingerprint embedding except for a few seed buyers; and 7) avoidance of (complex) homomorphic (or any type of public-key) encryption of the multimedia content. Further research can be focused on developing a proof of concept of this proposal on a real distribution scenario.

Acknowledgment

This work was partly funded by the Spanish Government through projects TSI200765406-C03-03 "E-AEGIS", TIN2011-27076-C03-02 "CO-PRIVACY" and CONSOLIDER INGENIO 2010 CSD2007-0004 "ARES".

References

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

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

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

[4] C.-C. Chang, H.-C. Tsai, and Y.-P. Hsieh, "An efficient and fair buyer-seller fingerprinting scheme for large scale networks," Computers & Security, vol. 29, pp. 269–277, Mar. 2010.

[5] D. L. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," Communications of the ACM, vol. 24, pp. 84–90. Feb. 1981.

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

[7] J. Domingo-Ferrer and D. Megías, "Distributed multicast of fingerprinted content based on a rational peer-to-peer community," Computer Communications, vol. 36, pp. 542–550, Mar. 2013.

[8] M. Fallahpour and D. Megías, "Secure logarithmic audio watermarking scheme based on the human auditory system," Multimedia Systems, in press.

[9] S. Katzenbeisser, A. Lemma, M. Celik, M. van der Veen, and M. Maas, "A buyer-seller watermarking protocol based on secure embedding," IEEE Trans. on Information Forensics and Security, vol. 3, pp. 783–786, Dec. 2008.

[10] M. Kuribayashi, "On the implementation of spread spectrum fingerprinting in asymmetric cryptographic protocol," EURASIP Journal on Information Security, vol. 2010, pp. 1:1–1:11, Jan. 2010.

[11] C.-L. Lei, P.-L. Yu, P.-L. Tsai, and M.-H. Chan: An efficient and anonymous buyer-seller watermarking protocol. IEEE Transactions on Image Processing, vol. 13, pp. 1618–1626, Dec. 2004.

[12] D. Megías and J. Domingo-Ferrer, "DNA-Inspired Anonymous Fingerprinting for Efficient Peer-To-Peer Content Distribution," Proc. 2013 IEEE Congress on Evolutionary Computation (CEC 2013), pp. 2376–2383, Jun. 2013.

[13] D. Megías and J. Domingo-Ferrer, "Privacy-aware peer-to-peer content distribution using automatically recombined fingerprints," Multimedia Systems, in press.

[14] N. Memon and P. W. Wong, "A buyer-seller watermarking protocol," IEEE Trans. on Image Processing, vol. 10, pp. 643–649, Apr. 2001.

[15] Pando Networks. http://www.pandonetworks.com/p2p.

[16] B. Pfitzmann and M. Waidner, “Anonymous fingerprinting,” Advances in Cryptology-EUROCRYPT’97, LNCS 1233, Springer, pp. 88–102, 1997.

[17] B. Pfitzmann and A.-R. Sadeghi, “Coin-based anonymous fingerprinting,” Advances in Cryptology-EUROCRYPT’99, LNCS 1592, Springer, pp. 150–164, 1999.

[18] R. O. Preda and D. N. Vizireanu, “Robust wavelet-based video watermarking scheme for copyright protection using the human visual system,” Journal of Electronic Imaging, vol. 20, pp. 013022–013022-8, Jan.-Mar. 2011.

[19] J. P. Prins, Z. Erkin, and R. L. Lagendijk, “Anonymous fingerprinting with robust QIM watermarking techniques,” EURASIP Journal on Information Security, vol. 2007, pp. 20:1–20:7, Dec. 2007.

David Megías achieved the PhD in Computer Science in July 2000 at the Universitat Autònoma de Barcelona (UAB). He was an assistant lecturer at the UAB from September 1994 to October 2001. Since October 2001, he is at the Universitat Oberta de Catalunya (UOC) with a permanent position (currently as associate professor). At the UOC, he has held several academic positions, such as director of the Master programme in Free Software (from 2002 to 2009), associate director of the Doctoral programme in Information and Knowledge Society (from 2009 to 2013) and director of the Doctoral programme in Network and Information Technologies (from 2012 to 2013). In January 2014, he was appointed director of the Doctoral School. His current teaching is mostly related to computer networks, information security (watermarking and steganography), free and open source software and to research techniques and methodologies in the field of Network and Information Technologies. His current research interests include security and privacy in multimedia content distribution (mainly in the watermarking and fingerprinting topics), steganography and steganalysis. He has recently started a research line about privacy concerns in different applications of on-line social networks. He has published research papers in numerous international journals and conferences and has participated in several national joint research projects both as a contributor and as a manager (main researcher). He also has experience in international projects, such as the European Network of Excellence of Cryptology funded by the European Commission within the Sixth Framework Programme.