Intellectual Property Rights Protection in Peer to Peer Networks

Georgios Stylios[1] and Dimitrios Tsolis[2]

[1] Department of Applications of Informatics in Management and Economy, TEI of Ionion, Greece
stylios@teiion.gr
http://epdo.teiion.gr
[2] Computer Engineering and Informatics Department, University of Patras, Greece
dtsolis@upatras.gr
http://www.ceid.upatras.gr

Abstract. Peer to Peer Networks are oftenly used by internet users to share and distribute digital content (images, audio and video) which is in most of cases protected by the Intellectual Property Rights (IPR) legislation. This fact threatens e-inclusion and Internet democracy as a whole as it forces organizations to block access to valuable content. This paper claims that IPR protection and P2P can be complementary. Specifically, a P2P infrastructure is presented which allows broad digital content exchange while on the same time supports data and copyright protection through watermarking technologies.

Keywords: Computer networks, copyright protection, peer to peer networks, digital image processing, watermarking.

1 Introduction

Peer to peer networking is supported by suitable software which enables a computer to locate a content file (text, image, video, sound, software etc.) on another networked device and copy the encoded data to its own hard drive. P2P technology often is used to reproduce and distribute copyrighted content without authorization of rights owners. Except for digital music and video the P2P infrastructure is also used to make and distribute illegal copies of digital content which lies under the protection of the Intellectual Property Rights (IPR) legislation. For this reason the short history of P2P technology and software has been one of constant controversy by many in the content industry. The content owners are feeling even more threatened by the broad and unregulated exchange of digital content in P2P environments [2].

Proposed solutions for the organizations to the problem of IPR protection in the internet include simple business models and marketing strategies for organizations and corporations which are the digital content owner and application of complex technologies and systems [3]. These solutions tend to lock valuable educational and cultural content which is accessed only from private and restricted numbers of users threatening e-inclusion and the Internet democracy as a whole.

A.B. Sideridis and Ch. Z. Patrikakis (Eds.): e-Democracy 2009, LNICST 26, pp. 171–178, 2010.

As a general protection measure for copyright violations through digital technologies including P2P, copyright owners often uses digital watermarking techniques to encrypt and watermark content or otherwise Digital Rights Management technologies to restrict access, totally blocking digital content to be accessed through the Internet and the P2P software infrastructure.

This paper claims that watermarking, Digital Rights Management (DRM) and P2P can be quite complementary. Specifically, a P2P network infrastructure is presented which allows broad digital content exchange while on the same time supports data protection and copyright protection through watermarking technologies. In brief, the platform is functioning mainly for digital images and is tracking all the watermarked image files which are distributed and copied through the P2P network. The challenge is the algorithmic complexity of detecting multiple watermarking keys in the P2P network effectively and quickly, especially when thousands of image files are concerned. This is managed by an optimization detection algorithm which allows effective watermarking key detection in optimal P2P hops.

Equivalent systems, which combine watermarking, DRM and P2P technologies do not yet exist in practice but only in theory. Certain methodologies and strategies have been proposed for exploiting P2P technologies in DRM and vice versa [9]. The proposed system is setting a new basis for the close cooperation of the two different scientific areas of DRM and P2P aiming at exploiting the distributed computing nature of P2P networks for efficient digital rights protection and management.

2 IPR Protection – Watermarking and Keys

In this section the copyright protection part of the P2P infrastructure is presented which is mainly based on a watermarking algorithm for digital images which produces the correspondent watermarking keys distributed within the P2P environment.

2.1 Copyright Protection through Watermarking

The copyright protection systems main objectives are to provide an appropriate information infrastructure which supports rights management for the digital content and for the transactions taking place and on the same time protects the digital images and their copyright though robust watermarking techniques.

The watermarking techniques are playing a very important role in such systems mainly because they provide the protection means for proving the identification of the copyright owner and detecting unauthorized use of digital content [4]. Towards this functionality, watermarking algorithms are casting keys to the digital content (in most of cases invisible keys) which when detected prove the copyright ownership of the digital content [5].

In case of digital content transactions a very large number of digital images are being exchanged through networks and the Internet for which the legality

of their future use is highly improbable. The situation is even more difficult in P2P network infrastructures through which digital content is being exchanged based on specialized stand alone applications which exchange digital files of all kinds (and not only images).

A proposed solution is to apply a watermarking algorithm which produces sufficient information which is distributed to the P2P nodes. This information consists mainly of the watermarking key and other data relating to the digital image itself.

2.2 Generating Keys with the Watermarking Algorithm

Generally, a watermark is a narrow band signal, which is embedded to the wide band signal of a digital image [1]. In our case spread Spectrum techniques are being used and are methods by which energy generated at one or more discrete frequencies is deliberately spread or distributed in time or frequency domains.

In particular, this technique employs pseudorandom number sequences (noise signals) to determine and control the spreading pattern of the signal across the allotted bandwidth. The noise signal can be used to exactly reconstruct the original data at the receiving end, by multiplying it by the same pseudorandom sequence: this process, known as "de-spreading", mathematically constitutes a correlation of the transmitted pseudorandom number sequence with the receivers assumed sequence [10]. Thus, if the signal is distorted by some process that damages only a fraction of the frequencies, such as a band-pass filter or addition of band limited noise, the encrypted information will still be identifiable [11]. Furthermore, high frequencies are appropriate for rendering the watermarked message invisible but are inefficient in terms of robustness, whereas low frequencies are appropriate with regards to robustness but are useless because of the unacceptable visual impact.

In our case, the embedding of a robust multibit watermark is accomplished through casting several zero-bit watermarks onto specified coefficients. The image watermark, a random sequence of Gaussian distribution in our case, is casted multiple times onto the selected coefficients preserving the same sequence length but shifting the start point of casting by one place.

Actually the final watermark that is embedded into the image is not a single sequence but many different sequences generated with different seeds. These sequences are casted, one after the other, on the mid coefficients of the image, using the additive rule mentioned above and begging from successive starting points. If all sequences where to be casted, beginning from the same starting point, then, besides the severe robustness reduction resulting from the weak correlation, the possibility of false positive detector response would dramatically increase, since every number that has participated as a seed during the sequence generation procedure, will be estimated by the detector as a valid watermark key. Shifting the starting point by one degree for every sequence casting ensures that the false positive rate will remain in very small level due to the artificial desynchronisation introduced. Every single random sequence of Gaussian distribution is generated using a different number as the seed for the Gaussian sequence generator. It is

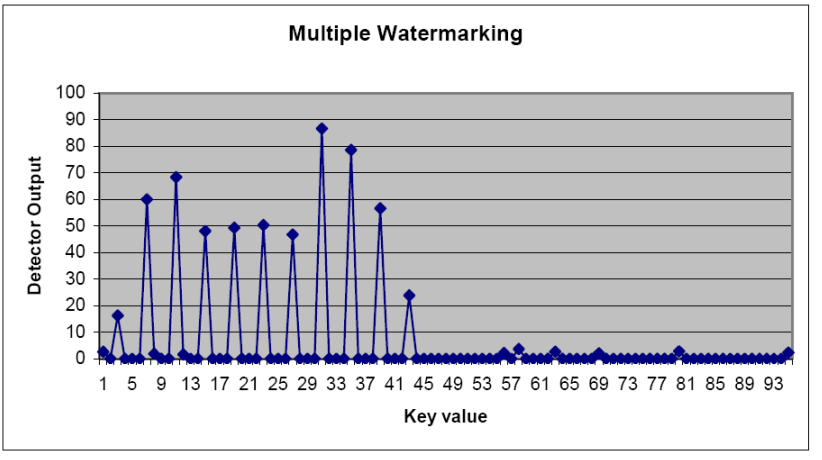

Fig. 1. Multiple Watermarking Keys per Image

important to differentiate the sequences in order not to mislead the detection mechanism, since it is based on the correlation between the extracted sequence and the sequence produced with the watermark key.

The watermark key is responsible both for the generation of the first sequence and the construction of a vector, containing the rest of the numbers that will serve as the corresponding seeds. The placement of several Gaussian sequences into the image content can model, under specific conventions, a multi-bit watermark. The detection of a zero-bit watermark is interpreted as if the bit value of the specified bit is set to one. On the contrary, failure of the detector to detect the zero-bit watermark leads to the conclusion of a zero bit value. Thus, in order for a message to be casted into the image content, it is initially encoded using the binary system and applied afterwards in the sense of zero-bit watermarks using the embedding mechanism and according to the derived bit sequence.

2.3 Watermarking Keys and the P2P Network

In this section a watermarking algorithm has been presented which is robust enough to facilitate data and copyright protection for the digital images while at the same time produces sufficient information which is distributed and stored to the P2P nodes. This information consists mainly of the watermarking key.

Taking into consideration that for each digital image a set of watermarking keys are being used for copyright protection, the next step towards an efficient P2P environment which supports digital rights management is to use these keys as an information for retrieving the copyright status of each image transacted through the P2P network. For this reason, the watermarking keys are being stored in the independent network Peers. The copyright owner can use the watermarking key as query information to track down its digital images and their use. The issue is how quickly and efficiently the Peer that contains the under

inspection key is being located taking into account that thousands of digital images could exist in the P2P network and multiple watermarking keys could exist in a digital image. The solution proposed is a scalable and robust data indexing structure based on a Nested Balanced Distributed Tree (NBDT). The next section presents the NBDT P2P Network.

3 NBDT P2P Network

NBDT provides a tree-like structure for the P2P network upon which watermarking key-based searching can be performed. In terms of bandwidth usage, searching scales very well since no broadcasting or other bandwidth consuming activities take place during searches. Since all searches are key based there are two possibilities: either (a) each host implements the same algorithm, that translates a keyword to a binary key or (b) another service provides the binary key. This service accepts keyword based queries and can respond with the corresponding key. The second approach is more precise. It is also possible to use a more centralized implementation for such a service. From now on we assume that the key is available. This section describes an algorithm for the first case.

The structure was built by repeating the same tree-structure in each group of nodes having the same ancestor, and doing this recursively [6]. This structure may be imposed through another set of pointers. The innermost level of nesting will be characterized by having a tree-structure, in which no more than two nodes share the same direct ancestor. The figure 2 illustrates a simple example (for the sake of clarity we have omitted from the picture the links between nodes with the same ancestor). Thus, multiple independent tree structures are imposed on the collection of nodes inserted. Each element inserted contains pointers to its representatives in each of the trees it belongs to.

Let an initial given sequence of w-bit keys belonging in universe K=[0,2w-1], where an unknown density. At initialization step we choose as peer representatives the 1st key, the lnKst key, the 2lnKst key and so on, meaning that each node with label i (1<i<N) stores ordered keys that belong in range [(i-1)lnK,..ilnK-1], where N=K/lnK the number of peers. Note that during update operations; it is not at all obvious how to bound the load of the N peers, since new w-bit keys with w¿w may be appeared in the system and K must exceed. For this purpose we will model the insertions/deletions as the combinatorial game of bins and balls presented in [10]: Modeling the insertions/deletions of keys in this way, the load of each peer becomes Q(polygonN) in expected case with high probability. Obviously, peers representatives early described have also been chosen according to this game. We also assume that each key is distinct and as a result the probability of collisions is zero. Each key is stored atmost in O(loglogN) levels. We also equip each peer with the table LSI (Left Spine Index). This table stores pointers to the peers of the left-most spine (for example in figure 2 the peers 1, 2, 4 and 8 are pointed by the LSI table of peer 5) and as a consequence its maximum length is O(loglogN). Furthermore, each peer of the left-most spine is equipped with the table CI (Collection Index). CI stores pointers to the collections of peers presented at the same

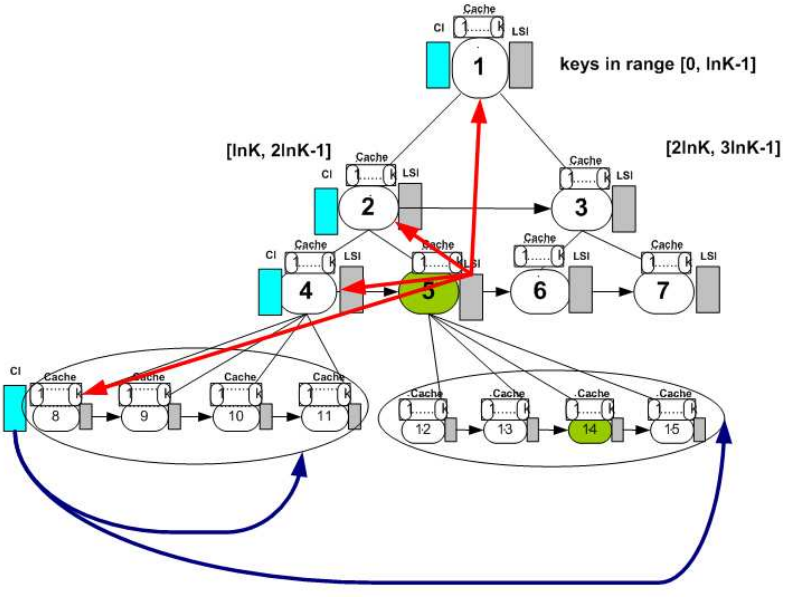

Fig. 2. The NBDT P2P System

level (see in figure 2 the CI table of peer 8). Peers having same father belong to the same collection. For example in the figure 2, peers 8, 9, 10, and 11 constitute a collection of peers. Its obvious that the maximum length of CI table is OpN. For example in figure 2 we are located at (green) node 5 and we are looking for a key k in [13lnn, 14lnn-1]. In other words we are looking for (green) node 14. As shown in [12], the whole searching process requires an optimal number of O(loglogN) hops or lookup messages for detecting the watermarking key and that is also validated using the proposed simulator.

When we want to insert/delete a key/node from the structure, we initially search for the node that is responsible for it (using a number of O(loglogN) hops in worst-case) and then we simply insert/delete it from the appropriate node.

If new w-bit watermarking keys, with w¿w, request to be inserted into the system, then we have to insert new peers on the network infrastructure and as a result we have to re-organize the whole p2p structure. In practice, such an expensive reorganization is very sparse. The new peers of NBDT are inserted at the end of the whole infrastructure consuming O(1) hops in worst-case. In particular, when a node receives a joining node request it has to forward the join request to the last node. The last node of NBDT infrastructure can be found in O(1) hops in worst-case by using the appropriate LSI and CI indexes.

If the load of some peer becomes zero, we mark as deleted the aforementioned peer. If the number of marked peers is not constant any more then we have to re-organize the whole p2p structure. Based on the basic theorem of [7], if we generate the keys according to smooth distributions, which is a superset of

regular, normal, uniform as well as of real world skew distributions like zipfian, binomial or power law (for details see [8]), we can assure with high probability that the load of each peer never exceeds polylogn size and never becomes zero. The latter means that with high probability split or delete operations will never occur. In other words, the re-organization of the whole P2P structure with high probability will never occur which means that only the O(loglogN) hops are necessary to detect the appropriate watermarking key and no further time is being consumed for structure re-organization.

4 Conclusions

In this paper we focused on a P2P network infrastructure which allows broad digital content exchange while on the same time supports data protection and copyright protection through watermarking technologies.

In brief, a watermarking algorithm casts watermarking keys to the digital images and the same time the watermarking keys are being stored in the independent network Peers. The watermarking algorithm is robust enough to facilitate data and copyright protection for the digital images while at the same time produces sufficient information which is distributed and stored to the P2P nodes (the watermarking key).

The P2P environment which supports digital rights management is achieved through the use of these keys as an information for retrieving the copyright status of each image transacted through the P2P network. The copyright owner can use the watermarking key as query information to track down its digital images and their use. The tracking down solution used is a scalable and robust data indexing structure based on a Nested Balanced Distributed Tree (NBDT). Based in the NBDT system, in the steady state, in a N-node network, each node resolves all lookups via O(loglogN) messages to other nodes. Key updates require only O(loglogN) number of messages in worst-case. Node updates require O(1) number of messages in expected-case with high probability.

The watermarking key detection process withih the P2P framework is very efficient and outperforms the most popular infrastructures used directly for many solutions for P2P information discovery. The key detection process is very important for the copyright owner because when successful the copyright status of each digital image can be retrieved and evaluated.

The future applicability of the proposed infrastructure is strong as it could be used for the creation of P2P environments, supported by GUIs, with which a user could exchange digital files while copyright protection occurs at the same time.

References

1. Cox, I.J., Miller, M.L., Bloom, J.A.: Digital Watermarking. Morgan Kaufmann Publishers, San Francisco (2002)
2. Computer Science and Telecommunications Board, National Research Council. The Digital Dilemma: Intellectual Property in the Information Age, pp. 2–3. National Academy Press, Washington (1999)

3. House of Representatives. Digital Millennium Copyright Act (October 1998)
4. Davis, R.: The Digital Dilemma. Communications of the ACM 44, 80 (2001)
5. Wayner, P.: Disappearing Cryptography - Information Hiding: Steganography and Watermarking, 2nd edn., pp. 291–318. Morgan Kaufmann, San Francisco (2002)
6. Sioutas, S.: NBDT:An efficient P2P indexing scheme for Web Service Discovery. International Journal of Web Engineering and Technologies 4(1), 95–113
7. Kaporis, A., et al.: Improved Bounds for Finger Search on a RAM. In: Di Battista, G., Zwick, U. (eds.) ESA 2003. LNCS, vol. 2832, pp. 325–336. Springer, Heidelberg (2003)
8. Kaporis, A., et al.: Dynamic Interpolation Search Revisited. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds.) ICALP 2006. LNCS, vol. 4051, pp. 382–394. Springer, Heidelberg (2006)
9. Einhorn, M., Rosenblatt, B.: Peer to Peer Networking and Digital Rights Management - How Market Tools Can Solve Copyright Problems. Policy Analysis Journal 534 (2005)
10. Fotopoulos, V., Skodras, A.N.: A Subband DCT Approach to Image Watermarking. In: Proc. X European Signal Processing Conference (EUSIPCO 2000), Tampere, Finland, September 5-8 (2000)
11. Fotopoulos, V., Skodras, A.N.: Image Watermarking for Quality Control Based on Modified Key-Dependent DCT Basis Functions. In: 15th Int. Conf. on Digital Signal Processing (DSP 2007), Cardiff, Wales, UK, July 1-4 (2007)