Legal Aspects of Watermarking Search Services

Martin Steinebach

Fraunhofer SIT,Germany
steinebach@sit.fraunhofer.de

Biography

Dr.-Ing. Martin Steinebach is a researcher at Fraunhofer SIT and director of the CASED application lab. His main research interest is digital audio watermarking. He has developed algorithms for mp2, MIDI and PCM data watermarking, content fragile watermarking and invertible audio watermarking. Dr. Steinebach studied computer science at the Technical University of Darmstadt, where he completed his diploma thesis on copyright protection for digital audio in 1999. In 2003 he received his PhD from the Technical University of Darmstadt for his work on digital audio watermarking.

Legal Aspects of Watermarking Search Services

Martin Steinebach[1], Patrick Wolf[2], Jee-Un Kim[3], and Jens Engelhardt[4]

[1] Fraunhofer SIT, Rheinstrasse 75, 64295 Darmstadt, Germany
steinebach@sit.fraunhofer.de
[2] CoSee GmbH, Rheinstrasse 75, 64295 Darmstadt, Germany
patrick.wolf@cosee.biz
[3] Bartels Kim Wollenhaupt Rechtsanwälte, Invalidenstr. 115, 10115 Berlin, Germany
kim@allmedialaw.de
[4] Notos Rechtsanwaelte Steuerberater, Heidelberger Straße 6, 64283 Darmstadt, Germany jens.engelhardt@notos.de

Abstract. As digital watermarking has become an accepted means for copyright protection in the domain of virtual goods, additional services based on the basic mechanisms of embedding and retrieving watermarks are introduced. Automatic searching for potentially marked content and subsequent downloading and trying to retrieve the watermarks is one prominent example. The advantage of an automated service in comparison to manual handling is commonly accepted. A content owner will provide a search service provider the name of the work he wants to be monitored in the Internet, and at a later time he will receive a list with search results and retrieved watermarks. But one important drawback is the challenging legal situation. At the end, such a system is a tool automatically downloading large amounts of potentially copyrighted material, at the same time in some Internet domains distributing such content. Only after downloading and analyzing the content the provider of such a system can be sure the content is the one he is searching for. Wrong file names, fakes and similar issues can lead to downloading content no search order was provided for. So a search service provider would always be the potential target of law suits from third party content owners. In this work we discuss the legal aspects of this problem with respect to the different parties involved in the process. And this only one example of legal challenges in digital watermarking. Others are privacy concerns when combining media files and customer data or the actual value of using customer IDs as evidence when finding marked copies in the Internet. In this work we will discuss all legal challenges mentioned above and provide potential solution strategies developed in the PlugMark project.

1 Background

In this section we describe the environment for which the legal aspects need to be discussed.

1.1 Transaction watermarking

Transaction watermarking [1] means that copies of media files given to reviewers or sold to online shop customers are marked with in individual ID at the stage of downloading. This allows to retrieve the watermark at a later point and trace back the copy to the original reviewer or customer.

The challenge for transaction watermarking has always been to get hold on a copy of the media used in an illegal manner. Therefore new business models appeared, providing searching for marked content as a service to rights holders.

1.2 Searching as a Service

A search service provider [2] needs a suitable strategy to offer the service for searching for potentially marked media files. For cost efficient services, blind crawling and analyzing huge amounts of content is not acceptable. But when the Internet is used as a distribution channel for media, any access technology will be created in such way that it allows an effective search for content. Therefore a search for suspect media will be most effective if it is conducted in the same way as any user would perform this search.

The information available to users about the media they search for is very rudimentary. It may be the title of the work or the name of an artist. Most of the time, it is also clear what type of medium (audio, video, etc.) they are searching for. All this is basically textual metadata. The same set of information is used by search service providers to search for potentially marked content. Title, media type and similar keywords and other metadata is defined by the rights holder using the search service.

After defining the search criteria, the Internet can be searched for matching media. The Internet is not one homogeneous network. It can be broken up into smaller spaces. The World Wide Web (WWW) is an example of such a subspace. All points in this subspace are interconnected through http. There are also subspaces within subspaces, like all points in the WWW that can be directly reached via Google. Other examples of subspaces are file sharing networks or Gnutella. Each of these subspaces can contain media and have their own special access methods to their media. An Internet-wide search needs to be able to take these specialties into account.

Everyone regularly using Internet search engines will have experienced the varying quality of the search results provided by queries. Depending on the quality of the search terms and the availability of the desired information, wrong results or a vast number of results may occur. In some cases results seem to be the information requested, but after a closer look turn out to be something different or fake.

1.3 Finding as a Risk

While accessing the wrong data with a standard search engine will only end up in visiting the wrong address in the WWW, for a search service provider the risk is much higher.

Given the situation a rights owner orders a search for a term describing his work, he also allows the search service provider to access the work to retrieve the watermark from it. This includes downloading from direct links as FTP or sharehosters, but also from peer to peer networks. In such networks, like the Torrent network or eDonkey, downloading often means also distributing, as active downloads require the user to share the content already downloaded to other users. As long as the search only provides the work of the rights owner as a result, the search service provider is allowed to download and distribute the content in this process.

But chances are high, that due to wrong naming, similar titles or fakes some search results will be works of third parties. The search service provider needs to download the works to recognize them as wrong results. But doing so requires distribution of works he has no search order for by the rights owner. This could end up in the owner of the false search result to see the search service provider as a party illegally downloading and distributing his content.

2 State of the Art

In this section we briefly discuss the various basic technologies for watermarking and search services mentioned in the previous section. This includes the act of individualization by watermarking and the search process.

2.1 Transaction Watermarking

Digital watermarking is a flexible technology used in a broad range of applications. The first watermarking approaches were directed at owner authentication or copyright protection, where an owner or producer identification is embedded to prove ownership or source of the cover. Today digital watermarking is used to identify single copies of an original media file by e.g. embedding a transaction code, a customer ID or a simple continuous number into each copy. Whenever a copy is found, the watermark can be retrieved and the source can be identified by the embedded individual information. With digital watermarking restrictive protection solutions are not necessary to secure digital media. The customer is free to use and consume the media data he bought in any way and on any device he likes. But if he passes the content into illegal environments and copies are found, he can be identified. As the individual watermark is embedded during the sales process of the work, this is called transaction watermarking.

2.2 Searching for Content

When the Internet is used as a distribution channel for media, any access technology will be created in such way that it allows an effective search for content. Otherwise such an access technology would fail its purpose. Therefore a search for suspect media will be most effective if it is conducted in the same way as any user would perform this search. The information available to users about

the media they search for is very rudimentary. It may be the title of the work or the name of an artist. Most of the time, it is also clear what type of medium (audio, video, etc.) they are searching for. All this is basically textual metadata. In addition, P2P networks work with cryptographic hashes uniquely identifying the medium as a whole or chunks thereof. Logical operators (and, or, not) are used to form more complex criteria. So, when searching for Harry Potter audio books, a search criterion could be ("Harry Potter audio book") or a combination of text and media type ("Harry Potter" AND MEDIA_TYPE="audio"). After defining the search criteria, the Internet can be searched for matching media.

2.3 Content Identification

One important aspect of downloading content previously searched for is the question if one actually downloads the desired content or something else, either by accident or by incorrect naming. The latter can be result of a so-called fake, which is quite common in Internet file sharing, especially in new and popular content. Fakes can be injected into downloading platforms by service agencies fighting illegal downloads to frustrate file sharing users. In that case usually unusable content will be downloaded, which carries no risk for the Search Service, only overhead. By in some cases also users remame other works, especially porn, into popular download titles. In that cases the download turns out not to be the desired content, but nonetheless copyrighted material.

Identification before downloading is almost impossible: Hash values for content don't help, because a Search Service will search for new content where the hash is not yet known. User comments also are to slow to appear, chances are high that content is removed at the time sufficient user comments are available to decide if the content is a fake or not.

At the end only after downloading the decision about the nature of the content can be made. The most common strategies would be either manual inspection or robust hashes, the latter choice is preferable as it fits in the desired automation of the Search Service.

2.4 Download Spaces

The Internet is not one homogeneous network. It can be broken up into smaller spaces. The "World Wide Web" (WWW) is an example of such a subspace. All points in this subspace are interconnected through http. There are also subspaces within subspaces, like all points in the "WWW" that can be directly reached via "Google". Other examples of subspaces are "file sharing networks" or "Gnutella". Each of these subspaces can contain media and have their own special access methods to their media. An Internet-wide search needs to be able to take these specialties into account. An interesting fact is that some subspaces require other subspaces for accessing media. The most prominent examples is the BitTorrent network where no search functionality for content is implemented. The network requires access to the WWW to browse web pages providing a "catalogue" for content available within BitTorrent. It is also important to know that

in some subspaces like BitTorrent and many file sharing networks downloading also means distributing, as during the download process of a work the client of the user also acts as a distributor of the already downloaded parts of the content.

3 Legal Discussion

After defining the scenario and introducing the technology used there, in this section we discuss various legal issues raised in that context. This includes the act of watermarking as well as downloading content while searching.

3.1 Watermarking and Privacy

Embedding a transaction watermark and storing the transaction id together with user data in a database is an act relevant for user privacy. The rights holder or shop owner creates a file where personal data is stored, potentially for a long time. In Germany, this is not legal without certain additional measures. The file needs to be protected and stored in such a manner that it can be accessed only in the case of an evidence for illegal activities. Or the ability to link the transaction id and the user data must be removed by utilizing a third party for embedding. This third party would receive a request for providing a marked copy of a work to an end user together with a transaction ID. It would then embed a watermark with a random number, and store both transaction ID and random number in a database. Thereby it has no access to the personal data of the customer stored with the transaction code by the rights holder or shop owner. If a marked copy is found, first the third party embedder will retrieve the random number watermark and thereby identify the transaction ID linked with to the copy by a database lookup. It then will pass the transaction ID to the rights holder or shop owner who can identify the customer.

3.2 Copyright Violation by Downloading

As already mentioned in section 2.3., a Search Service always faces the risk of downloading content for which he has been not assigned to do so by the copyright owner of that work. In some cases as explained in section 2.4., this not only means accessing content, but also distributing that content. The act is distribution is often seen to be more important for legal actions against a participant of file sharing networks than mere downloading.

So if a Search Service accidently accesses and distributes copyrighted works of a third party, it could be the target of legal actions as agents of the third party may also monitor file sharing.

But while the act of downloading and distribution of the Search Service is similar to the common content pirate, the intention behind it is different. From a legal perspective this should be sufficient protection for Search Services against being sued by third parties. As the intention of the act is clearly not illegal, a legal instance at least in Germany should decide in favor of the Search Service.

3.3 Customer Identification

Transaction watermarks can be used to trace back to the customer or client to whom the marked content was issued. There are many possible usages for this feature. One of those would be to find the source of illegal distributions of marked content in the internet which is known as traitor-tracing. The question here would therefore be, if the customer who is identifiable after finding marked content by way of the embedded transaction information can be held liable for the copyright violations. Depending on field of usage the watermark can enable to find the "leak" especially when the content in question was to be handled discreetly. In other cases though transaction watermarks will not enable the rights holder to enforce his copyrights against the client. Under German law the right of private copying is a legal limitation on copyright. If therefore a customer buys digital content in an online-shop and gives one copy to his sibling or to a friend, it is not considered a copyright violation. Also, the customer could argue that he has lost his device on which marked content was saved or it was stolen from him. The rights owner would then have to prove that the customer who was identified by the watermark was actually the infringer or at least disturber.

As a rule the rights holder bears the burden of proof regarding the copyright violation. In certain cases the German law provides for a relief of the burden of proof for the rights owner. The secondary burden of proof applies when the claimant is not in the position to know the facts regarding for example the usage of the protected content. In this case the respondent has to conclusively substantiate the circumstances which rule out his liability. In the end, the judge will come to a conclusion by freely assessing the evidence.

3.4 Watermarking at Court

The usage of watermarking as a means to furnish evidence in court is not yet ruled out. A decade ago the development of stable imperceptible watermarks was in the beginning. Due to instabilities and the possibility of manipulations the watermarking method was not considered to be eligible for providing proof of the information it was built to convey. Since then the watermarking technology has advanced at a very fast pace. Watermarks which are not removable by its recipient and are able to persist heavy compressions and format changes, mutilations of the file etc. are slowly introduced into legal proceedings. There are already some decisions by inferior courts which were based among other things on digital watermarks which were used to provide evidence regarding the authorship of a protected work.

3.5 Watermarking as Copyright Violation

Imperceptible watermarks which are embedded in the digital content should provide in general no copyright violation under German law. With advancing technologies watermarks nowadays will in general not be considered a distortion, mutilation or modification of the protected content. The legal situation is

therefore very different from the usage of visible watermarks. What is to be considered though are the authors moral rights, for example the right to be named and also the right not to be named when using watermarks as authentification tools.

4 Summary and Outlook

In our work we discuss the different legal aspects of applying watermarks and searching for watermarked content. None of the issues show a legal thread to the process, while in some cases like potential copyright violations by downloading misnamed content, a certain risk for the Search Service remains. As watermarking at court is still a relatively new technology, the future will clarify the legal situation by actual rulings at court.

Acknowledgments.

This work was supported by the CASED Center for Advanced Security Research Darmstadt, Germany, and by PlugMark, both funded by the German state government of Hesse under the LOEWE grant program.

References

1. I. J. Cox, M. L. Miller, J. A. Bloom, Digital Watermarking. San Mateo, CA: Morgan Kaufmann, ISBN: 1-55860-714-5, 2001.
2. Steinebach, Wolf, On the necessity of finding content before watermark retrieval: Active search strategies for localising watermarked media on the Internet, in: Multimedia Forensics and Security, Chang-Tsun Li (edt.), IGI Global, Hersbey, USA, ISBN: 1599048698 , S. 106-119, 2008