

Annual ADFSL Conference on Digital Forensics, Security and Law

2015 Proceedings

May 19th, 4:15 PM

# A New Cyber Forensic Philosophy for Digital Watermarks in the Context of Copyright Laws

Vinod P. Bhattathiripad Cyber Forensic Consultant, GJ Software Forensics, vinodpolpaya@gmail.com

Sneha Sudhakaran Cyber Forensic Consultant, GJ Software Forensics, sneha14888@gmail.com

Roshna K. Thalayaniyil College of Engineering, Kallooppara,Kerala India, roshnakhalidt@gmail.com

Follow this and additional works at: https://commons.erau.edu/adfsl

Part of the Aviation Safety and Security Commons, Computer Law Commons, Defense and Security Studies Commons, Forensic Science and Technology Commons, Information Security Commons, National Security Law Commons, OS and Networks Commons, Other Computer Sciences Commons, and the Social Control, Law, Crime, and Deviance Commons

#### **Scholarly Commons Citation**

Bhattathiripad, Vinod P.; Sudhakaran, Sneha; and Thalayaniyil, Roshna K., "A New Cyber Forensic Philosophy for Digital Watermarks in the Context of Copyright Laws" (2015). *Annual ADFSL Conference on Digital Forensics, Security and Law.* 1.

https://commons.erau.edu/adfsl/2015/tuesday/1

This Peer Reviewed Paper is brought to you for free and open access by the Conferences at Scholarly Commons. It has been accepted for inclusion in Annual ADFSL Conference on Digital Forensics, Security and Law by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.





### A NEW CYBER FORENSIC PHILOSOPHY FOR DIGITAL WATERMARKS IN THE CONTEXT OF COPYRIGHT LAWS

Vinod Polpaya Bhattathiripad Ph D Cyber Forensic Consultant GJ Software Forensics Kozhikode - 673004, Kerala, India vinodpolpaya@gmail.com

Sneha Sudhakaran Cyber Forensic Consultant GJ Software Forensics Kozhikode - 673004, Kerala, India

sneha14888@gmail.com

Roshna Khalid Thalayaniyil College of Engineering Kallooppara, Kerala India

roshnakhalidt@gmail.com

#### **ABSTRACT**

The objective of this paper is to propose a new cyber forensic philosophy for watermark in the context of copyright laws for the benefit of the forensic community and the judiciary worldwide. The paper first briefly introduces various types of watermarks, and then situates watermarks in the context of the ideaexpression dichotomy and the copyright laws. It then explains the forensic importance of watermarks and proposes a forensic philosophy for them in the context of copyright laws. Finally, the paper stresses the vital need to incorporate watermarks in the forensic tests to establish software copyright infringement and also urges the judiciary systems worldwide to study and legalize the evidential aspects of digital watermarks in the context of copyright laws.

**Keywords:** Digital Watermarks, Software Copyright, Idea-Expression Dichotomy, Programming Blunders, Copyright Infringement, AFC, POSAR

#### 1. INTRODUCTION

Software can be copyright protected. When an infringement of the copyright is suspected, the copyright owner has every moral and legal right to ensure the exclusivity of their property rights to the software. It is only natural that when such rights have been flagrantly violated, particularly for commercial profits (and uses), the injured parties will invariably resort to legal measures both for the protection of their property and for the restitution of damages involved therein. Such an issue can trigger a legal battle.

In the process of legally establishing copyright infringement, the watermark (contained in the software) can play an important role. In order to use watermark as an evidence to establish the criminal activity behind the infringement allegation, both the forensic procedure (used as part of the investigation) and the judge's decisionmaking process need to be sensitive to the forensic role of watermarks.

Although much has been done on the design, programming and implementation aspects of watermarks (Cox et al, 2008), there has not been

any effort from cyber forensic researchers to explain the forensic locus standi and philosophical rationalle of watermarks for the benefit of the entire forensic community and also for the benefit of the judiciary across the world. As a result of this deficiency, a cognitive (or an expertise) gap can exist between the forensic community and the judiciary and the goal of this work is to fill this As several different forms of digital watermarks exist, it is the duty of forensic professionals to explain the forensic roles of various different watermarks separately and then generalize these different roles to form a single forensic philosophy which can be ultimately used by the judiciary for effective decision making in any software copyright infringement litigation.

Before getting into the forensic philosophy of watermarks, a quick overview of digital watermarks will help readers to situate this work properly.

#### 2. OVERVIEW OF WATERMARKS

File watermarking is not uncommon in the digital world. It is a widely used mechanism worldwide in order to protect the ownership of a digital file,

Page 87 © 2015 ADFSL

including software. A digital watermark (or, simply a watermark) in a digital file (whether it is a text or image or an audio or a video file) is some kind of electronic thumb impression introduced by the owner into the file for easy establishment of his / her creativity (Nagra et al, 2002).

Since any digital file has a source code (or a hex dump) as part of it (see fig 1), file watermarking virtually becomes a process of embedding some kind of information into the source code (of the file) for the purpose of introducing some degree of personalization (or identity) into the source code (Cox et al, 2008). When a watermark is embedded into any digital file, the source code of the watermark also gets embedded into the source code of the digital file (see fig 2 & 3).

Watermarks can exist in different forms like text, image, audio and video (and also combinations of these forms). The best way to further explain a watermark is to quickly demonstrate the technicalities of an image file, first using its non-watermarked form and then, its watermarked form.

There is a general feeling that a watermark is always a single, identifiable and easily separable entity in a watermarked file and that a watermarked file always differs from its nonwatermarked form by only a few hexdumps. This is not true. Most watermarks do not remain as single, identifiable and easily separable entities in the watermarked file. Also, the hexdump of any non-watermarked image (for example, see fig 2) differs in a big way from that of the watermarked form of the same image (see fig 3) and this difference can be easily verified by comparing the corresponding hex values in fig 2 and fig 3. This big difference is because the watermarking algorithm not only inserts the hex values of watermarks into the original (non-watermarked) image but also modifies most hex values of the original image. In the same manner, the hexdumps of any particular non-watermarked audio, video or a text file also differ largely in the same fashion from those of the watermarked form of the file.

Just as there are different forms of digital files (image, audio, video and text form and also their different combinations), watermarks can also exist in many forms. Further, watermarks can be classified in many different ways based on several factors. An overview of two sample classifications

will help readers to situate the forensic aspects of watermarks properly. Based on their techniques of generation, watermarks are classified into two types and they are static watermarks (which are embedded as code segments within the source code of a digital file) and dynamic watermarks (which are watermarks generated during the runtime with the help of code segments embedded within the source code of a digital file) (collberg and Thomborson, 1999). Again, based on the roles played by different persons involved in the development of the software, watermarks can be classified as authorship mark, fingerprinting mark, validation mark and licensing mark (which are unique identities of the author, distributor, publisher and consumer, respectively, of the software that contains the watermark) (Nagra et al, 2002).

Every watermark has certain desirable features like effectiveness (or the correctness and aptness of the intended purpose of the watermark), integrity (or the watermark's ability to not to interfere with the performance of the source code), fidelity (or how closely the watermark accurately or truthfully helps to identify the 'owner' of the software), robustness (or the watermark's ability to withstand any kind of alteration of the content of the file in which the watermark is embedded) etc. (Nagra et al, 2002; Cox et al, 2008; Marcella and Menendez, 2008).

#### 3.THE IDEA-EXPRESSION DICHOTOMY AND WATERMARKS

The idea-expression dichotomy (Walker, 1996) provides an excellent theoretical perspective to look at and further delineate watermarks embedded as part of the source code of any software. Any software is (or consists of) a collection of code segments and each code segment is an expression of one or more ideas. This being so, software, as a whole, can be considered a collection of expressions of one or more ideas.



Figure 1. A JPEG file and its source code in C, generated using the HxD tool. (Only the beginning and the end of the C code are shown here and the hidden portion is indicated by a thick white space) (Picture courtesy: Kadalundi Mangrove Reserve preserved by Kerala Forests, Kozhikode district, India)

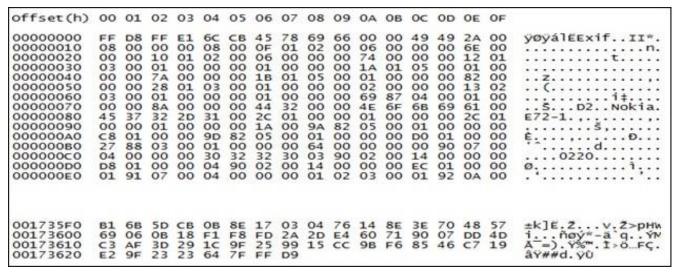


Figure 2. The hexdump (generated using the HxD tool) of the non-watermarked JPEG image shown in Figure.1

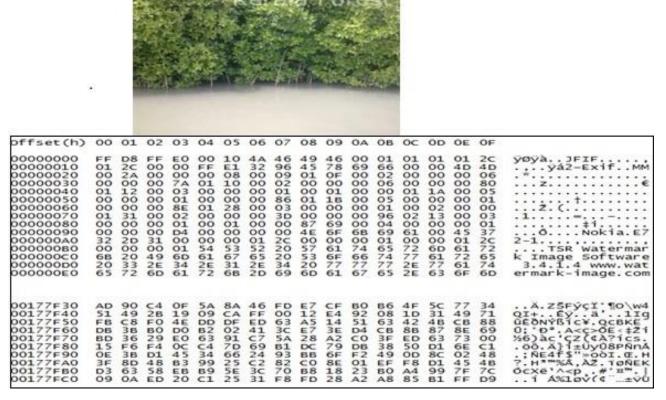


Figure 3. The watermarked form of the image shown in Figure.1 and its hexdump (First, the picture was watermarked using the tool TSR Image Watermark and then the corresponding hexdump was generated using the HxD tool)

From the idea-expression perspective, any watermark (embedded as part of any software) is a genuine idea which is properly expressed in a manner that does not adversely affect the syntax (or sometimes even the semantics) of the software. It is a part of the source code (of the software) which is not a functional requirement of the software. In other words, watermark in any software is part of the requirements marking and identifying the original ownership of the software and not part of the requirements of the potential users of the software.

The above explanation of watermarks in terms of the idea-expression dichotomy clearly opens the door to linking watermarks directly to copyright infringements of any software because the ideaexpression perspective is the basis of formulation of software copyright laws of several countries (Newman, 1999; Hollar, 2002). The ideaexpression basis of copyright laws of several countries (especially the US copyright laws) says that if there is only one way or a limited number of (exclusive) ways of effectively expressing an idea, this idea and its expression tend to "merge" (Walker, 1996) and in such instances an idea and its expression are not protectable through copyright (Hollaar, 2002). In other words, if the same idea can be realized through more than a limited number of expressions, all such different realizations are protected by copyright laws. Thus, if the idea behind the expressions in a watermark (which is embedded in any particular copyrighted software) can be expressed in more than a limited number of ways, then the copyright obtained for the software can extend to the watermark contained in it. Thus, watermarks are directly linked to copyright.

This link requires further explanation. Even if the copyright of the main software can be extendable to the watermark contained in it as well, the copyright may not be extendable to all the elements of the watermark. This non-extendability is because a watermark can contain several legally unprotectable elements such as globally common mnemonics, names and expressions, globally shared notations, codes or expressions due to shared nature of technology, and globally common functional area elements.

If all the elements in the watermark are unprotectable, then the copyright obtained for the software will not be extendible to the watermark contained in the software. Finally, if there is at least one protectable element in the watermark contained in software, the copyright of any software will extend to the watermark contained therein as well.

To summarize, watermarks can be perceived in terms of idea-expression dichotomy, and thus, can be directly linked to copyright and can also be an indicator of software copyright infringement.

## 4. FORENSIC IMPORTANCE OF COPYRIGHTED WATERMARKS

Despite their apparent functionally irrelevant and thus innocuous status in any software, watermarks, when copyrighted (that means, when there is at least one protectable element in a watermark), can be of great value / assistance to the cyber forensic expert and a discussion of this evidence is the prime objective of this article.

The approach to the forensic importance of a watermark can be best done in the context of the concept of programming blunders (Bhattathiripad, 2012). A programming blunder has been defined as a "variable in a program or a code segment in a program ... which is .... unnecessary for the user's functionality". Looking from this definitional point of view, a watermark is technically (or can be explained in terms of) a programming blunder because a watermark in any software is not part of the functional requirements of the software or (in is unnecessary other words) for functionality. The locus standi and functionality of watermarks can thus be best situated through their inclusion in the category of blunders.

Even so, unlike a typical programming blunder, watermark is neither unintentional nor accidental. Rather, it is an intentional 'programming blunder', introduced into the software by its developer for a specific purpose. In general, every watermark is an intentionally introduced software element and is technically an intentional programming blunder. Because watermarks are intentionally introduced code segments in any software, the three etiological factors of programming blunders (see Bhattathiripad, 2012) are not sufficient enough to explain the etiology of watermarks. All the

existing etiological factors of programming blunders assume that programming blunders can happen only due to inability or inattention of the programmer (or the quality engineer) to completely remove those statements that are not required for user's functionality. This also means that the existing etiological aspects programming blunders do not consider the possibility of programming blunders happening due to software developer's intentional effort to introduce (into a software), a code segment (like a watermark) which is not required for user's functionality.

While doing a juxtaposed comparison of two sets of software to establish possible copyright infringement, the existence of a particular watermark in identical contexts in both the complainant's and the defendant's versions can be a more positive indication of illegal copying (than other kinds of blunders), as the watermark was deliberately inserted into but not carelessly leftover in the complainant's version. It is highly unlikely that two programmers will design and insert exactly same watermarks exactly in the same position and exactly in the same way, and this elevates the similarity into potential evidence of copyright infringement.

Thus, most watermarks can provide direct evidence (or at least probable, corroborative or supporting evidence) to establish copyright infringement more decisively than other programming blunders. In the absence of other direct evidence of copyright infringement, watermarks can form the only basis of the expert opinion to the judiciary about the possibility of copyright infringement.

#### 5. WATERMARKS AS EVIDENCE IN COPYRIGHT INFRINGEMENT FORENSIC TEST

The importance of watermarks has not been given any role or status in the forensic procedure of the Abstraction-Filtration-Comparison (AFC) test (which is the only judiciary-accepted procedure for establishing software copyright infringement in the US) (Bhattathiripad, 2014). Watermarks are not even considered during this test because during the abstraction of the software, only the functionally active or relevant parts (of the two

sets of software) will be considered for abstraction and used for further investigation (Hollaar, 2002). As a result, the functionally irrelevant parts (or those items that are irrelevant for user's functionality, like watermarks) may not be considered for abstraction. In such case of unfortunate non-consideration, the watermarks will not be available for final comparison and this unavailability certainly adversely affects the rigour of the AFC test and thus, can affect its reliability.

Hence, this paper proposes that, along with the AFC test results, the evidence concerning watermarks, if any, should also be identified and gathered separately by the forensic expert, before the final findings and inferences are presented to the court.

The software forensic research community is encouraged to take on this proposal and find ways to incorporate watermarks in the AFC test.

The judiciary systems worldwide also need to be encouraged to study and legalize the evidential aspects of digital watermarks in the context of copyright laws. Some preliminary suggestions are presented below.

During the forensic analysis as part of any software copyright infringement litigation, any watermark (embedded into a software package by the developer and identified and detected by the forensic expert) needs to be considered as a separate program segment. In other words, during the forensic test in any copyright infringement litigation, the embedded watermark needs to be first separated<sup>1</sup> from the main software and then

<sup>&</sup>lt;sup>1</sup> The task of separation of the source code of a watermark from the source code of the main software (or any digitally watermarked file) can be easy if and only if the source code of the watermark can be perfectly identified in the original source code as a single unit of code segments. To put it clearer, the task of separation of watermarks from an image / audio / video file can be complicated and strenuous, for many reasons. Two such potential reasons are (a) the hex values of the watermark get fragmented (as against an identifiable single unit) in the ocean of hex values of any watermarked image / audio / video file and (b) the watermarking algorithm not only inserts the hex values of watermarks into the original (non-watermarked) image / audio / video file but also modifies a few, if not all, hex values of the original. Even so, this task of

subjected to the forensic test separately. This is in order to ensure that the watermark has (or does not have) protectable elements. The ultimate goal here is to establish whether the copyright of the main software is (or is not) extendable to the watermark as well. For instance, if the test used is AFC, then the watermarks in both the software packages need to be separated first, and then separately abstracted. Subsequently, the unprotectable elements in both watermarks need to be filtered out and removed. Finally, the comparable elements in the remaining "golden nuggets" (Walker, 1996) need to be compared and the resulting evidence (or evidence of infringement of protectable elements) needs to be reported to the court. If the test used is POSAR (Bhattathiripad, 2014), watermarks need to be separately subjected to this 5-stage forensic test process and the resulting evidence<sup>2</sup> need to be reported to the court.

Although outside the purview of AFC and POSAR, the evidence of copyright infringement of watermark will form part of the evidence of copyright infringement of the main software as well (because watermark is a part of the main software) and sometimes, can turn out to be valuable evidence to establish copyright infringement of the main software.

Before concluding, a note on what a judge expects from a forensic expert would add value to the special attention and consideration given to watermarks. In any software comparison report, what the judge would expect from the forensic expert is a set of details that helps the court in arriving at a decision on the copyrightable aspects of the elements in both software packages (Newman, 1999). So, what is expected in the case of watermarks is not a mere statement on the extendability of the copyright to the watermarks. Rather, the statement should be substantiated and supported by a set of details on the merger aspects

separation is not impossible if the algorithm for separation is sensitive to both the insertions and the modifications done by the watermarking algorithm.

of the ideas and expressions contained in the watermarks.

It needs to be stated here that the future research on forensics of watermarks should not ignore all these complex aspects that determine the status and role of watermarks in copyright cases.

#### 6. CONCLUSION

In the process of legally establishing copyright infringement, the watermark (contained in the software) can play an important role. As any watermark can be considered to be technically a programming blunder, the forensic importance and philosophy of programming blunders (explained in the context of the idea-expression dichotomy) can be extendible to every watermark as well. While doing a juxtaposed comparison of two sets of software to establish possible copyright infringement, the existence of watermarks in identical contexts in both versions can be a more positive indication of illegal copying (than other kinds of blunders), as they were deliberately inserted into and not carelessly leftover in the complainant's version.

In order to use watermark as evidence to establish the criminal activity behind the infringement allegation, both the forensic procedure (used as part of the investigation) and the judge's decision making process need to be sensitive to the forensic role of watermarks. Hence the forensic tests (to establish software copyright infringement) need to be re-designed so as to ensure that possible evidence like watermarks are procedurally collected, forensically analyzed and then properly reported to the court. The forensic report should be substantiated and supported by a set of details on the merger aspects of the ideas and expressions contained in the watermarks. Future research in this area should ensure not to leave out the importance of watermarks as well as their role in establishing software copyright infringement.

#### **REFERENCES**

[1] Bhattathiripad, P. V. (2012), Software Piracy Forensics: A Proposal for Incorporating Dead Codes and other Programming Blunders as Important Evidence in AFC Test, Proceedings of the IEEE 36th International Conference on

<sup>&</sup>lt;sup>2</sup> The evidence set here contains the evidence of infringement of protectable elements along with the evidence of post-piracy modifications and the evidence of infringement of programming blunders as part of the watermark

- Computer Software and Applications Workshops, Turkey
- [2] Bhattathiripad, P. V. (2014). Judiciary-friendly forensics of software copyright infringement, IGI-Global
- [3] Collberg, C., & Thomborson, C. (1999). Software watermarking: Models and dynamic embeddings, Proceedings of the 26th ACM SIGPLAN-SIGACT symposium on Principles of programming languages (pp. 311-324). ACM.
- [4] Cox, I. J., Miller, M. L., & Bloom, J. A. (2008). J. Fridrich, T. Kalker, Digital Watermarking and Steganography, Morgan Kaufmann Publishers
- [5] Hollaar, L. A. (2002) "Legal Protection Of Digital Information", BNA Books, 2002

- [6] Nagra, J., Thomborson, C., & Collberg, C. (2002). A functional taxonomy for software watermarking. In *Australian Computer Science Communications* (Vol. 24, No. 1, pp. 177-186). Australian Computer Society, Inc..
- [7] Marcella, A. J. Jr., and Menendez, D. (2008), Cyber Forensics, Auerbach Publications, page 3
- [8] Newman, J. O. (1999), "New Lyrics For An Old Melody, The Idea/Expression Dichotomy In The Computer Age", 17, Cardozo Arts & Ent. Law Journal, p.691
- [9] Walker, J., (1996), Protectable 'Nuggets': Drawing the Line Between Idea and Expression in computer Program Copyright Protection, 44, Journal of the Copyright Society of USA, Vol 44, Issue