# Chapter 28
# Deterring Text Document Piracy with Text Watermarking

**Rakesh Kumar Mishra**
*Feroze Gandhi Institute of Engineering and Technology, India*

## ABSTRACT

*Protecting copyright of a digital content is gaining momentum and has been suitably complemented by technological innovations. Literature piracy, though not being given much attention, constitutes a major bulk. Unlike audio and video, text piracy is not complemented by IT solutions except for certain proprietary initiatives. This pursuit embarks on review of technological advancements for Text Copyright Protection along with issues and challenges for their implementation. Appraisal comprises of watermark embedding algorithms and distribution infrastructure. A brief discussion over the document structure, watermark composition and type, classification of algorithms and future direction is also being accomplished. To make approach holistic couple of systems were also studied.*

## INTRODUCTION

Communication network has expanded itself to new dimensions and now is available even to mobile users. The exponential growth of networks and its users has transformed Internet into a principal source of information dissemination. Transformation of the network from wired to wireless has made information distribution process a multi-fold complex activity. Insecure spaces appearing on account of the complexity are usually exploited by plagiarist to make unauthorized accesses and detain data illegally. Plagiarism is a multimillion dollar illegal business and has affected every form of digital contents. Proactive measures against filthy attempts and protecting genuine ownership of authors are the foremost challenges to be ensconced by collaborative researches around the world.

Digital Rights Management (DRM) techniques are now surfacing strongly against plagiarism including Digital Watermarking. In watermarking a proprietary signal is incorporated into host at scaled down strength. Hiding proprietary mark

within creations is differentiated from steganography and cryptography on the basis of relevance of received data to illegal recipient. A dummy host is visible to the interceptor in case of Steganography while cryptography scrambles data using key, hence forcing restriction over those who are key deficient recipient. Watermarking of literary resource is most ignored area (Fetscherin & Schmid, 2003) despite of known fact that text constitutes the largest share of Internet download traffic (Hurtung & Kutter, 1999). Text document image dissemination is a common phenomenon with growing number of "Open Archives and Digital Libraries". These literary assets need to be secured against any misappropriation and right infringement. Watermarking techniques for images are very effective but can not be applied to text document images due to the inherent attributes and the mildest change to host becomes strongly evident. The only left option is to use available for data hiding.

Marking text document image can be classified into three dominant categories. Format marking on one side is of low capacity but on another hand is truly transparent. Space Coding is a variant of Format Marking where data hidden in form of varying spaces is semitransparent. Pixel flipping although ensures high density but is sensitive to noise. With all favours and odds there is a continuous research going on globally for a versatile technique for marking document images which capable of persisting across all media, formats and channel quality.

Next generation watermarking techniques for text document should be made compliant with some minimum standards. Presently, initiatives differ from one another severely and most of these are highly dependent of the language or the script being used. Another concerning factor is that none of the algorithm guarantee a reasonable data hiding capacity and its viability across formats and scripts. Key distribution infrastructure for deciphering the extracted watermark is still an undeclared infrastructure. After all sort of discussions and claims,

the issues related to the file format for marked document has not been decided yet; specifically the metadata content for the marked documents and the compression technique to be deployed for transmission. All-in-all a holistic approach for designing the coder, extractor, watermark and distribution infrastructure is required.

## DIGITAL RIGHTS MANAGEMENT

Digital Rights Management (DRM) is a universal concern. The digital data once decomposed into binaries have same level of vulnerability as that of unsecured content. Legal framework has already been put into action, but once media converts to binary form all the proprietary marks can be eliminated with little effort, obviously using computers. Hence the answer to question "*to whom does this binary belong?*" remains a major encumbrance in citation of rulings against plagiarist.

DRM is a scaffold carved from paraphernalia spanning from standalone solutions to protocols designed as per regulation for upholding the rights of an Author. DRM implemented for digital documents is referred as Enterprise DRM i.e. E-DRM. The Association of American Publishers defines DRM as "*the technologies, tools and processes that protect intellectual property during digital content commerce*". This is the most comprehensive definition encompassing all form of digital content including Image, Video, Text and Audio. Another definition by Einhorn (Fetscherin & Schmid, 2003) appears as "*digital rights management entails the operation of a control system that can monitor, regulate, and price each subsequent use of computer file that contains media contents like photos, video, audio or text.*" this definition describes a holistic characteristic of a DRM which not only helps in protecting the rights of the authors but also thrives for controlling the usage and securing the financial revenue of authors.

Early implementations of DRM trace back to 1996 with the DVD forum applying Content

## Related Content

Designed to Seduce: Epistemically Retrograde Ideation and YouTube's Recommender System
Fabio Tollon (2021). *International Journal of Technoethics (pp. 60-71).*
www.irma-international.org/article/designed-to-seduce/281077

Legal Issues of the French Law on Creation and Internet (Hadopi 1 and 2)
Eleni Metaxa, Miltiadis Sarigiannidisand Dimitris Folinas (2012). *International Journal of Technoethics (pp. 21-36).*
www.irma-international.org/article/legal-issues-french-law-creation/69981

An Overview of Child Abuses in 3D Social Networks and Online Video Games
Miguel A. Garcia-Ruiz, Miguel Vargas Martinand Patrik Olsson (2012). *Investigating Cyber Law and Cyber Ethics: Issues, Impacts and Practices (pp. 101-116).*
www.irma-international.org/chapter/overview-child-abuses-social-networks/59939

Engineering and Environmental Technoethics
Luppicini Rocci (2010). *Technoethics and the Evolving Knowledge Society: Ethical Issues in Technological Design, Research, Development, and Innovation (pp. 146-162).*
www.irma-international.org/chapter/engineering-environmental-technoethics/40606

Navigating the Ethical Complexities: Unveiling the Implications of Social Media Research
Somia Abdul Same'eand Antony Sheela Anmary (2024). *Reviving and Re-Writing Ethics in Social Research For Commoning the Community (pp. 139-158).*
www.irma-international.org/chapter/navigating-the-ethical-complexities/341291