IOSR Journal of Applied Physics (IOSR-JAP)
e-ISSN: 2278-4861.Volume 11, Issue 5 Ser. II (Sep. – Oct. 2019), PP 51-61
www.iosrjournals.org

Detailed Examination of Information Hiding Techniques for Copyright Protection of Text Documents

Aru, Okereke Eze and Ananaba, Ephraim Chianalamoke
Department of Computer Engineering
Michael Okpara University of Agriculture, Umudike, Umuahia-Abia State-Nigeria
Corresponding Author: Aru, Okereke Eze

Abstract: *In view of the current trend in ICT evolution with incessant usage of web and other online services, copying, sharing, storing and transmitting information is by far done through digital media. It is no surprise that countless methods of protecting such data have evolved, but the greatest disadvantage in digital text transmission is the fact that inestimable digital copies of the data/file can be made and is liable of tainted. This leads to the global problem of copyright protection, copy protection and content authentication. This work critically examines information hiding techniques of text document and their consequences. Various watermarking system properties and text steganography techniques characteristics were also highlighted coupled with different types of attacks and their possible defenses. The relative analytical method of data conversion from qualitative data to quantitative data was applied and finally the work was concluded by examining text as an essential media of information transfer which needs complete protection.*

Keywords*: Watermarking, Steganography, Attack, Text document image, ICT, etc.*

Date of Submission: 28-10-2019 Date of Acceptance: 12-11-2019

I. Introduction

The idea of secret communication is as old as communication itself. While information hiding techniques have received a tremendous attention recently, its application goes back to Greek times. According to Greek historian Herodotus, the famous Greek tyrant Histiaeus, while in prison, used unusual method to send message to his son-in-law. He shaved the head of a slave to tattoo a message on his scalp. Histiaeus then waited until the hair grew back on slave's head prior to sending him off to his son in-law. The second story also came from Herodotus, which claims that a soldier named Demeratus needed to send a message to Sparta that Xerxes intended to invade Greece (Sabu M Thampi, 2002). Back then, the writing medium was written on wax-covered tablet. Demeratus removed the wax from the tablet, wrote the secret message on the underlying wood, recovered the tablet with wax to make it appear as a blank tablet and finally sent the document without being detected (Petitcolas, F.A.P. et al, 1999).

The spread of the internet and subsequent evolution of communication technologies, electronic publishing has become an essential theme and in future organizations, offices seem to be paperless. Currently, several studies are in process to organize and perform some ideas such as e-government, e-commerce, and digital libraries. Electronic publishing has many advantages, but it has some threats such as illegal using of electronic information, copyrighted materials, and changing the data (Dittmann J, et al; 2001). Some of the protective measures include authenticity, integrality, confidentiality, and copyright protections are vital to prevent plagiarism issue. Recognizing the rightful owner of a digital data such as an image or a video, digital watermarking methods are adequate (Saba T et'al 2014).

Manipulating downloaded protected text document and then reusing it without control is very simple today, therefore copyright management/policy is vital. Many organizations (the publishers and book sellers etc), have spent millions of dollars to ensure copyrights (Cheddad A., et al, 2010). To protect the digital information against copyright problems, the most vital part is protecting the text as opposed to some other parts such as images, audio, and video. Protection of text document is a serious international problem. Hence, protection solution is necessary to protect the text in editing, reproducing, and modifying processes (Muhsin ZF et al, 2014).

Digital watermarking is a method of concealing data and combined the copyright information and text document in an unnoticeable way (Zeng F et al, 2012 & Cheng Y et al, 2014). It is very hard separating the watermark from information because digital watermarked information is hidden for other viewers except the intended receiver (genuine owner). Digital watermark is said to be a verification code that will remain beside the data, which may be (visible or invisible), and it is inside the data before and after any process like decryption; however, conventional cryptographic techniques were not as such (Brassil JT, et al, 1999). Digital watermark is one of the solutions and its invisibility is a vital aspect in the protection of intellectual property in digital format. The copyright information that copyright protection puts into the digital data becomes intact, and when there is any question about the digital data, the information refers to the original owner.

Moreover, the original buyer could be identified while the person traces an unauthorized copy of object. Also, another advantage of watermarking is to avoid copyright violations such as illegal redistribution or remaking through applying these methods (Zhang SR, et al, 2014). Various different techniques to protect images, audio and videos objects have been successfully implemented; but however, the methods of protecting texts are not sufficient. Texts documents are the necessary channels that are widely used to transfer information. Plain text is the main part of many kinds of data such as newspapers, classified documents of various kinds, legal documents, and its copyright protection is vital. Digital watermarking, copyright protection methods are very important and should not be missed or ignored. Digital document text watermarking is a process to enter a digital watermark deeply into a digital text and keep some information about the text owner. The main objectives of document text watermarking algorithm is of binary nature, the difference between fore ground and background, word patterning, text meaning, fluency, author writing style, and language rules, and it should be addressed (Haron H, et al 2011). Once copyright protection is embedded in a text, its communication is secured. All of the texts characteristics such as fluency and value of the text should be kept intact during any change in the text, whereas short documents are very hard to protect because of its low capacity to merge with watermark. These factors; Language, grammar, conventions, and communication media affects text watermarking algorithms (Rehman A, et al, 2014). Recently, several text watermarking patterns such as using pre-supposition based, text image, syntactic tree based, synonym based, noun–verb based, word and sentence based, typographical error based and acronym based, have been suggested. Past works on text watermarking could be categorized in three main classes including an image-based approach, a syntactic approach, and a semantic approach.

Watermark embeds in text image through customizing interline or word gaps in the middle of lines and words (Brassil JT, et al, 1999). The syntactic structure of text is made and used in syntactic approach to embed watermark bits through some transformation like passivization, clefting, and topicalization (Atallah M. J, et al, 2001). Finally, in semantic approach, synonyms, acronyms, words spelling, presupposition, and text meanings as text's semantics are considered as watermark in text (J. M. Topkara et al, 2006). The algorithms used in text watermarking by using binary text image are not strong against retyping. In syntactic approaches, researchers mixed algorithms with the Natural Language Processing (NLP). These algorithms are more effective, but the main issue is slow research progress in NLP and inefficient syntactic analyzers.

Various categories of information security systems are shown in the diagram below. The cryptography and information hiding are security systems that are used to protect data from manipulators, crackers, hackers, and spies. Normally, most of the vicious users want to leave traces from cuts, manipulations, and infections.

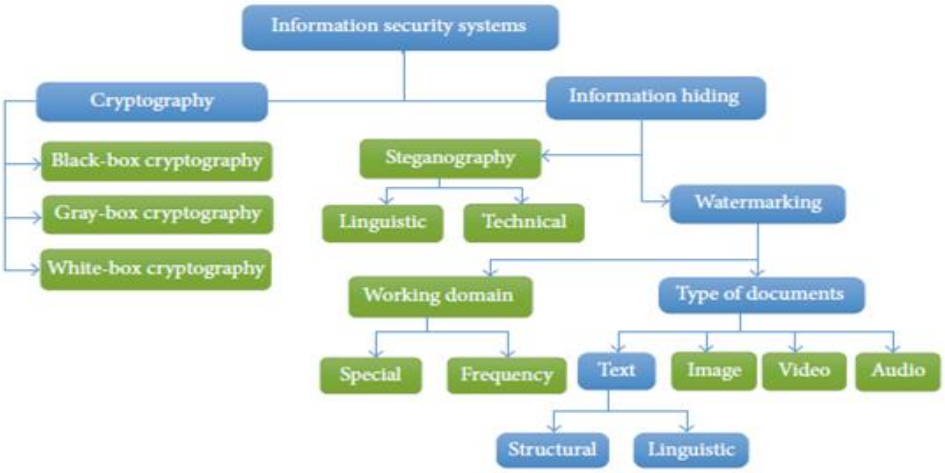


Figure 1.0: *Different categories of information security systems.* (N. A. A. S. Al-Maweri, et al 2016).

II. Review Of Related Work

This work reviews some of the related works in information hiding techniques. There have been many new and powerful steganography techniques reported in many literatures. These include the spatial and the transform domain techniques. The advantages of the spatial domain techniques are their simplicity and computational speed. The disadvantage is their low ability to bear the attacks i.e., they are less robust. The transform domain techniques have the advantage of high capability to face the attacks i.e., they are more robust. However, the methods of this type are computationally complex and hence may be slower.

Also, text watermarking includes the architecture, watermarking categories, applications, evaluation criteria, and attacks. Text Watermarking Architecture, as shown in the diagram below, digital text watermarking includes watermark embedding and watermark extraction.

Watermark Embedding: The embedding phase of text watermarking algorithm involves three stages. Firstly; generating a watermark string which includes the owner's name or other pieces of information (e.g., author and publisher).Secondly, the watermark string is converted to a binary string, which is modified by a hash function according to an optional key, and then an invisible watermark string is generated for embedding it into special locations in the cover text. Finally, it is inserted into special locations where the watermark string will not be affected by attacks (M. Pal, 2016), & (M. H. Alkawaz, et al, 2016).

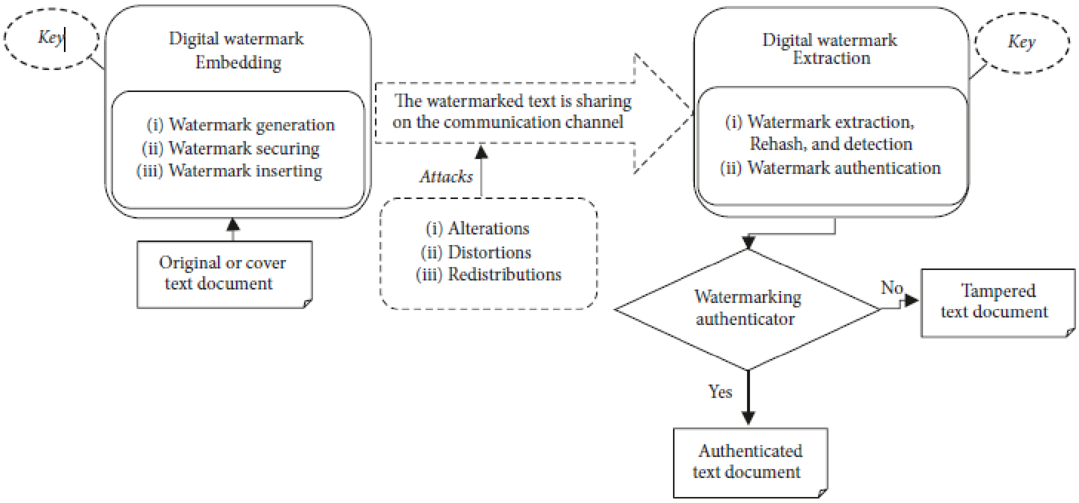

Key
Digital watermark Embedding
(i) Watermark generation
(ii) Watermark securing
(iii) Watermark inserting
Original or cover text document
The watermarked text is sharing on the communication channel
Attacks
(i) Alterations
(ii) Distortions
(iii) Redistributions
Digital watermark Extraction
Key
(i) Watermark extraction, Rehash, and detection
(ii) Watermark authentication
Watermarking authenticator
No
Tampered text document
Yes
Authenticated text document

Figure 2.0: *Digital text watermarking (embedding and extraction) architecture.* (M. H. Alkawaz, et al, 2016).

Text Documents Attacks: Currently, many users can easily utilize various digital text files such as articles, books, and online news. This is as a result of the availability of open access to these text documents, which has warranted unauthorized attacks such as alterations, copy, distortions, and redistributions are simultaneously on the increase, hence, text watermarking can be used as a security tool to prove the accuracy and originality the of text documents (M. H. Alkawaz, et al, 2016).

Watermark Extraction: Globally, watermarked text documents are shared through various communication medium such as email, web, or social media over the internet. Obviously, it is essential to authenticate the originality of the text documents. The two (2) different terms are used for this phase, that is, extraction and detection. Although authors often regarded both as similar functions in some literature, we can distinguish them in this way: whereas the extraction discovers the watermark string from the watermarked document and authenticates its integrity, the watermark detection verifies the existence of the watermark string from the watermarked text.

Text Watermarking Categories: During past two decades, many types of research have been carried out based on structural (format based), linguistic, scanned-image watermarking and frequency of words in the cover text. Herein, we consider those methods which are focused on modifying the structure and content of the cover text. Bashardoost M. et al 2015, cited that the linguistic technique concerns with the special features of the text content that can be changed in a specific language, and moreover, the structural technique concerns with the layout or format of the cover text that can be modified.

Steganography:- Steganography or Stego as it often referred to in the IT community, literally means, "Covered writing" which is derived from the Greek language. Steganography is defined as follows, "Steganography is the art and science of communicating in a way which hides the existence of the communication. The goal of Steganography is to hide messages inside other harmless messages in a way that does not allow any enemy to even detect that there is a second message present". In a digital world, Steganography and cryptography are both intended to protect information from unwanted parties. Both Steganography and Cryptography are excellent means by which to accomplish this task but neither technology alone is perfect and both can be broken. In-consideration of this reason, that most experts would suggest using both to add multiple layers of security.

Image Steganography: It has been proved that hiding in the frequency domain rather than in the time domain would provide better security for steganography. In fixed capacity mode, different LSBs of different high frequency sub-bands are used for embedding, based on the embedding rule to get different quality of the stego image.

All complex cover blocks are replaced by the complex secret message blocks. Provos, N, et al 2001, proposed a spatial domain technique that adopts the interpolation and the edge detection algorithm for embedding. Image interpolation algorithm enlarges the cover image to accommodate large amount of secret information while edge detection improves the quality of the stego image. Also, a robust color image adaptive steganography using integer wavelets, which embeds the payload in every 4x4 blocks of the low frequency sub-band of cover image. Two pixels of the secret image are hidden at a time in the coefficients, one on either sides of the principal diagonal.

Audio Steganography: In audio steganography too, the transform domain methods produce great results in terms of audio quality, capacity, and resistance against attacks. (Johnson, N.F, 1998, discuss about different audio steganography techniques in wavelet domain. The literature survey articulates various methods in the transform domain. The Audio steganography techniques must be robust against the common attacks such as amplification, filtering, resampling, re-quantization, noise addition, and compression.

Video Steganography this is an extension of the image steganography. There are different image steganography techniques which is applicable to videos. In-view of the dynamic contents of the video, the privilege of detecting the hidden data are less compared to the images. The common attacks for videos are compression, rotation, frame rate variation, frame exchange, insertion, and deletion of frames etc. The videos provide incredible hiding capacity and offer new dimensions for data hiding such as hiding messages in motion vectors and audio components. The information can be hidden in either the compressed video format or uncompressed video format. P. Singh et al, 2013, used Flash Videos for steganography because of its simple file structure, its relatively small size compared to other video file formats.

Lew CH et al. 2013, proposed a video watermarking scheme, which embeds the watermarks in the specific sub-bands of the wavelet transform. Also, Johnson, N.F. et al. 1998, proposed a video steganography technique to hide the secret messages in the motion vectors of the cover media during H.264 compression process. The above compressed video steganography techniques did not accomplish good payload capacity. Compared to the size of the video the capacity was too less. The capacity can be improved upon if the information is hidden in many frames. However, the challenge is the selection of frames, which can resist the compression.

III. Information Hiding

The hiding and retrieval of a digital media in information communication is referred to as information hiding. The digital media is ranged from a video, an audio to an image, or a simple text. The information hiding is a general term that covers several sub-categories including three major disciplines. These three groups are cryptography, watermarking, and steganography (Al-Haj AM, 2010). As it could be seen in Figure below that it depends on application domain, watermarking could be strong or breakable.

The evolution of internet and technology innovations with a large variety of applications, information hiding techniques in many applications are very important. Other types of digital data like audio, video, text, and images are reinforced with hidden specific marks. Trademark identification or specific mark could be altered such as a code, writer's ID, or anything that needs to reach to the destination safely, as to prevent illegal copying; information hiding is a great asset (Lung JWJ et al, 2014).

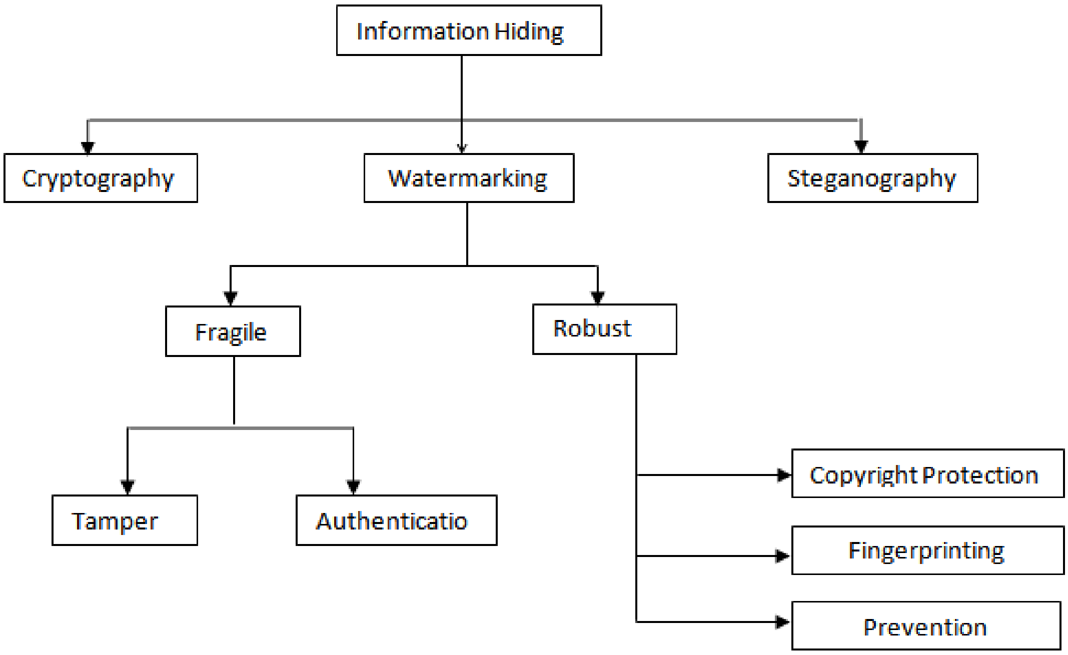


Figure 3.0: *Information Hiding (*Saba T, et al; 2014)

IV. Watermarking Properties

The properties watermark depends majorly on applications; a proper watermark is expected to exhibit the below characteristics as. Also, there are many references that deliberated about the identifications of watermarks (Wolfgang RB, et al, 1999). Robustness, being imperceptible, security, being undetectable and capacity is the main characteristics of watermarking. Achieving all of these properties with one watermarking system is not practicable. Considering the application domain, all watermarking methods choose some of them and give up the others (Nodehi A, et al, 2014).

Capacity: The maximum number of bits that a watermarking method could encode in a unit of time is capacity that shows the watermark size, and it should be high for any watermark shape. It varies and depends on the type of object; for example, broadcast monitoring needs a very high capacity.

Fragility: Sometimes it is better to be on the opposite side of robustness to provide fragility in order to enable the watermark to overcome some transformations. For instance, in content authentication, watermarks should be saved. Designing a fragile watermarking technique is hard because of this feature communicated online and are referred as hazard that could.

Robustness: When designing a watermarking strategy, considering the attacks resistance is a pivotal matter. Being powerful in any intentional or unintentional process from any watermarked data such as audios or videos is an important requirement for any watermarking system. A robust watermark is the one that could overcome attacks of the object even if it is changed without permission. Alternatively, a robust watermark should not be failed unless a large amount of marked data is ignored. In designing process of a watermarking system, it is considered to attack and intended applications that could answer or correspond (Ahmad AM, et al, 2014).

Security: Security is another important parameter for a watermarking technique. Watermarks should conceal against any looking or touching efforts without permission. The definition of security is that watermarks should be hidden and its payload in a watermarking system. The existence and the amount of a watermark should not be visible for any foreign person.

Imperceptibility: This implies that one should enter the watermark into object in a way that is imperceptible. This is a very important need of watermarking. All the characteristics of the basic and watermarked data should resemble each other, and they should be identical in everything. The watermark should not change the original data enough to maintain the high quality.

V. Text Watermarking Applications.

Text watermarking techniques is applicable in various applications. Below are some of the most important watermarking applications viz;

(a) **DIGITAL COPYRIGHT PROTECTION (PROOF OF OWNERSHIP):** Text watermarking provides passive protection tools for digital documents as to hinder illegal replicating or copying text document content. For instance, if copies of a watermarked document/file (e.g., PDF, Docx, Latex, and RTF) is made by someone, therefore, the reversibility of watermarking techniques can be used to prove the ownership of the copied text documents (M. H. Alkawaz, et al 2016).

(b) **ACCESS CONTROL (COPY CONTROL):** Publishers and the content providers are seeking more reliable ways to control / limit copy or access to their cherished documents, and simultaneously, desired to make the documents accessible on the Internet in order to obtain more income. Text watermarking is a desirable technique on the online systems that provide access control to avert illegal copy or restrain the possible number of times of copying / duplicating the original text (R. Petrovi´c et al, 2007 and P. Singh et al 2013).

(c) **TAMPER PROOFING:** It is noticeable that a huge number of text documents are available online for selling or reading for users. Hence, these documents are totally prone to be exposed to a number of attacks (e.g., copy, unauthorized access, and redistribution). Therefore, text watermarking can be used as a delicate tool for tamper proofing of the watermarked texts against attacks. (M. Bashardoost et al 2017).

(d) **TEXT CONTENT AUTHENTICATION:** Articles and newspapers published online in form of plain text documents has brought several issues related to authenticating the integrity of these documents. To verify the integrity of plain text documents, Text watermarking can be applied as an authentication tool (R. Petrovi´c et al, 2007).

(e) **FORGERY DETECTION (PREVENTION):** P. Singh et al 2013, quipped that Plagiarism and reproduction of text documents are serious forgery activities and are swiftly on the increase. Before online publication, Text watermarking can be used as a forgery detection tool by embedding a watermark in the original text. Thus, it can prove the plagiarism and reproduction of the watermarked texts (M. H. Alkawaz et al, 2016).

VI. Possible Attacks On Data Originality

Available watermarking methods are not effective enough to cover data originality against all possible attacks that could damage the data originality. Actually, the watermarking method tolerates the attacks based on the application. In this regard, some important fundamental attacks are described as follows.

i. **Active Attacks:** The first threat is hacker who attempts to eliminate or disrupt the watermark. In some conditions such as owner realizing, ownership proof, fingerprinting, and copy control, this kind of attacks is important because when the mark cannot be identified, the mark is removed (Cox IJ, et al, 1998).

ii. **Passive Attacks:** In passive attacks, the hacker's aim instead of removing the watermark is to find the existence of the watermark in a hidden communication. Actually, owners are not worried about this kind of attack in most of the mentioned application areas because most of the times watermarks are visible and alarm its presence. However, in a hidden communication, it is very important to hide the watermark's existence. In some cases, such kinds of active attacks happened that hackers make several copies from one part of media with different watermarks and then make a copy without watermark (Cox IJ, et al 1998). For fingerprinting applications, being resistant against collusion attacks is vital because it puts different labels for various copies of one specific part. (Saba T, et al, 2014).

iii. **Forgery Attacks:** The aim of the hacker is not removing a watermark but is to incorporate it. In authentication applications, such attacks should be seriously considered as the hacker could make the watermark to modify the media, if the hacker inserts true authentication labels. Also proofing ownership is an important type that is also considered in forgery attacks.

6.1TEXT DOCUMENT WATERMARKING

In ensuring high security, digital watermarks are embedded in all documents available on Internet sources, to use the Internet effectively as a content distribution media and a communication. There are some effective watermarking methods to protect images, audio, and video. Redundancy in images and limitations of human visual system is applied to watermark images. The plain text is generally the most important media over the Internet (Saba T, et al, 2014).

Copyright violators are necessary to protect the plain text in a high level of security. There are not effective and useful algorithms for text watermarking as opposed to the digital watermarking algorithms for image, audio, and video protection. The definition of digital watermarking is to protect digital data against illegal copying and other attacks through inserting a special digital watermark. Digital text watermarking is to realize the genuine owner of information in the process of inserting a digital watermark or extracting it from the data. The rules for text watermarking are similar to the watermarking of images, audio, or video. Abdullah M et al, 2008, proposed that being resistant against various attacks, being hidden for everybody except the owner or writer, and being automatically reproducible with the algorithm of extraction are some factors of a watermark. The plain text includes lower amount of unnecessary data than images, audio, and video, and this is an important matter.

6.2TEXT WATERMARKING CHALLENGES

It is unarguable that plain text is the simplest form of information; it faces numerous challenges with itself when copyright protection is referred. Text does not have much capacity to include watermark such as watermark images and videos. Some important characteristics of text watermarking algorithm are the binary nature with significant changes between front/back colors, statement/sentence, syntax and semantic statistics. There are some characters that a text acquires them from the general shape of watermarking techniques such as imperceptibility, robustness, and security that should be met. All the changes and transformations should keep the original meaning, fluency, grammar, and the text's value that it is the meaning of text and are protected by watermarking to prevent the communication disturbing. In writing especially in literary writing, fluency is an important factor to explain the text meaning clearly and fluently. J. M. Topkara, et al, 2006, quipped that the time, readability is another important factor; to preserve it, watermark must be inserted according to the language statistics. In literature or news, keeping the writer's type of writing is essential in some cases.

Some applications like legal documents, poems, and quotes are very sensitive, so semantic transformations are not performed on them in a random base because it might damage or destroy the semantic text quality and the text's value as well.

6.3POSSIBLE ATTACKS ON TEXT WATERMARKS

Text watermarking methods are not much acceptable in cyber community because of the unclear watermarking methods and poor robustness against attacks. Partial attacks are possible for any attacker and could make its status incomplete; as a result, all possible watermark attacks including unauthorized insert, delete, or detect are analyzed. These attacks with their characteristics are elaborated below:

Reordering: Another way of attackers to remove the watermark is that they change the order of words and sentences to make text's shape different. In texts, the hacker paraphrases the sentences and changes some words with their synonyms. The focus is mainly on changing the writing style, connotation, and the meaning of the text.

Unauthorized detection: Normally, it is required to limit the ability to identify an unauthorized attempt. It is believed that the power of an opponent only realizes that the current mark of a special work is hazardous for a watermarking system's security.

Unauthorized deletion: In this type of attack, the attacker aim is to change the original text by random deletion of some words and sentences and change of the real text. Hence, it is difficult for readers to point out any flaw. In all watermarking techniques, one should consider security issues against these attacks; the attacker should not be able to hide the original watermark. Even if the hacker changes the text, they should not be able to remove watermarks. The extraction algorithm should identify the watermarks.

Volume: The severity of attack is related with the aim of the attacker. If the aim is insertion and deletion of information into and from the text, the severity is not very high. On the other hand, when the attacker wants to change the ownership of one part of the text, then the severity becomes high.

6.4DIGITAL TEXT WATERMARKING APPROACHES

With the advents of the Internet and communication technologies, text watermarking is one new area for study, (Brassil JT, et al, 1999) did first try for copyright protection in 1995; there was an IEEE journal that was planned for publishing about some areas in communication, and it was about the security of electronic publishing trial.

Currently, group of researchers are working in some areas of study and also in text watermarking in many languages include English, Persian Turkish, Korean, Urdu, and Arabic. (Brassil JT, et al, 1999). There are three major approaches to digital text watermarking as listed below:

Image-based approach: This approach is simple and sensitive, and its low capacity makes it difficult to imbed watermark to the text. Some frantic efforts are being made to treat text as image in text watermarking. Watermarks are inserted in the layout and shape of the text image and the introduction some text watermarking in text image (Brassil JT et al 1999). Zhang SR et al, 2014, proposed text watermarking algorithm which is built on the statistics about word categorization and inter-word gaps. This method is composed of few steps. First of all, it categorizes all words in text according to some text characteristics; second, similar words are set in a segment. Finally, further, segments are classified into words of segment.

Syntactic approach: Characters, words, and sentences make up a text. Sentences could come in various syntactic structures. One traditional approach for text watermarking is to apply syntactic transformations on text structure to insert watermark. Atallah et al 2001, first used the syntactic structure of text for natural language watermarking scheme, where the syntactic structure is made and transformations are attached to it to insert the watermark deeply and to preserve all inherent properties of the text. To watermark electronic text contents in using the American Standard Code for Information Interchange (ASCII) characters and punctuation in text, Meral HM et al. 2015, suggested the natural language water marking algorithm is necessary in doing the morpho-syntactic alterations to the text. Firstly, the text is changed into a syntactic tree diagram where the classified organizations and the useful dependences are made clear and watermark is inserted. The diagram below exhibits watermarking steps.

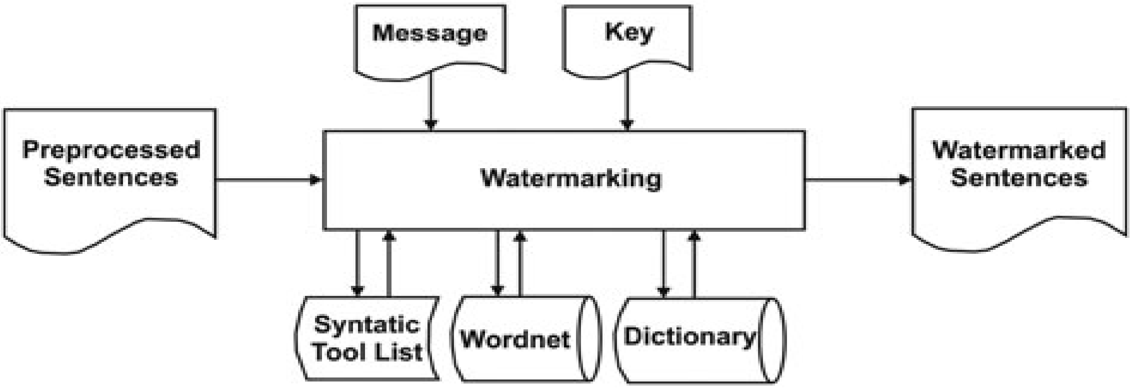


Figure 4.0: Syntactic sentence level watermarking (Jalil Z, et al, 2009; Meral HM et al. 2015).

Semantic approach: Inserting watermark in the semantic watermarking techniques, normally, semantic structure of text is employed. Text contents, verbs, nouns, words and their spellings, acronyms, sentence structure, and grammar rules have been exploited to insert watermark in the text, but these do not prove to be able to be strong and decrease the quality of the text largely, (Atallah et al, 2001).

VII. A Graphical Presentation of Watermakeing And Stenographic Technique In Text Hidding

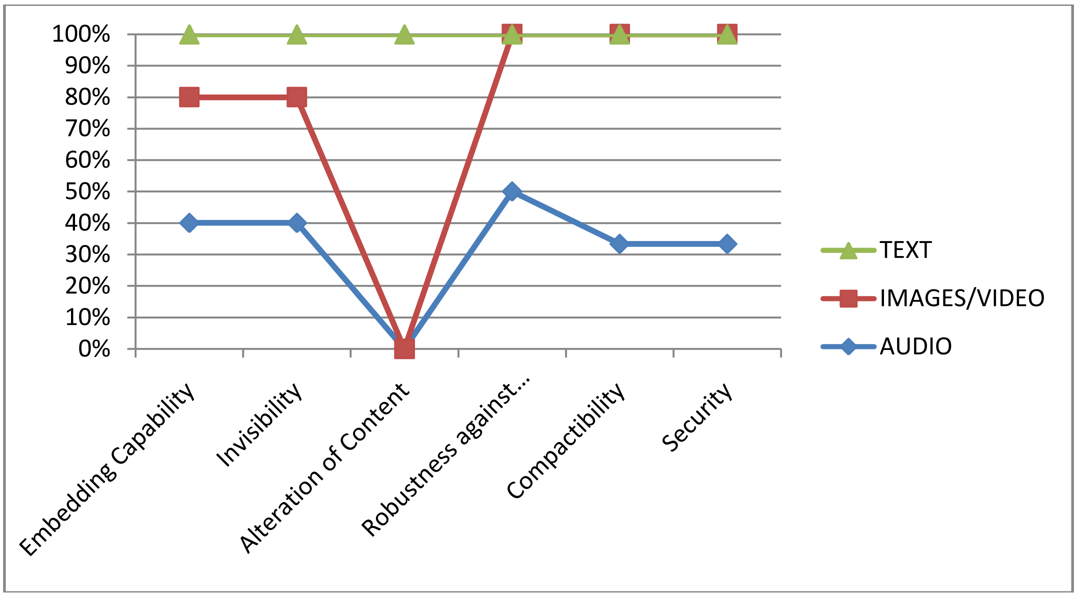
Watermarking ratio compatibility of the audio, images/video and text content

FACTORS	AUDIO	IMAGES/VIDEO	TEXT
Embedding Capability	High	High	Medium
Invisibility	High	High	Medium
Alteration of Content	Low	Low	High
Robustness against Conventional Attack	High	High	Low
Compatibility	Medium	High	Low
Security	Medium	High	Low

Applying the formular of conversion, relative to the analysis of conversion from Qualitative data to Quantitative Data, this implies that:
HIGH = 2, MEDIUM = 1, LOW = 0

Converted watermarking table

FACTORS	AUDIO	IMAGES/VIDEO	TEXT
Embedding Capability	2	2	1
Invisibility	2	2	1
Alteration of Content	0	0	2
Robustness against Conventional Attack	2	2	0
Compatibility	1	2	0
Security	1	2	0



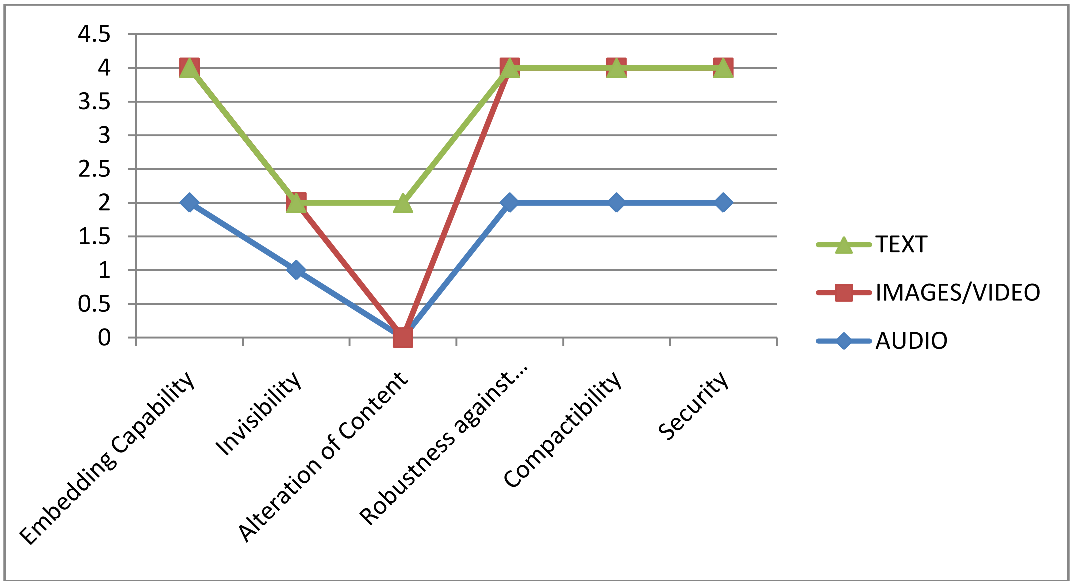
From the graphical analysis, it implies that Text which records 100% is popularly used in watermarking but runs high risk of various attacks of deletion and reordering, than the images / videos with the Audio having the less significant attack on its properties.

Steganograph ratio compatibility of the audio, images/video and text content

FACTORS	AUDIO	IMAGES/VIDEO	TEXT
Embedding Capability	High	High	Medium
Invisibility	Medium	Medium	Low
Alteration of Content	Low	Low	High
Robustness against Conventional Attack	High	High	Low
Compatibility	High	High	Low
Security	High	High	Low

Converted Steganograph table

FACTORS	AUDIO	IMAGES/VIDEO	TEXT
Embedding Capability	2	2	0
Invisibility	1	1	0
Alteration of Content	0	0	2
Robustness against Conventional Attack	2	2	0
Compatibility	2	2	0
Security	2	2	0



From the graphical analysis, implies Steganograph has a high effect in protection of text document. Also, Text and one of the major means of communication and message sharing, runs high risk of various attacks of deletion and reordering, than the images / videos with the Audio having the less significant attack on its properties.

VIII. Conclusion

The topic Detailed Examination of Information Hiding Techniques for Copyright Protection of Text Documents elaborated different types of information hiding techniques. Major characteristics of watermarking such as imperceptibility, robustness, and security are discussed. The works done to reveal the place of watermarking in today's digital world. Especially, the types of attacks that digital document faces in the digital world. Furthermore, text content watermarking; several applications areas, the challenges, and possible attacks was also discussed. The conversion analysis chart shows that text as an essential media of information transfer has high attention for information hiding and protection but also is highly vulnerable and runs the highest risk of attack of deletion, and reordering of its contents.

References

[1]. Abdullah M, Wahab F. Key based text watermarkingof e-text documents in an object based environmentusing z-axis for watermark embedding. WorldAcademy of Science 2008.

[2]. Al-Haj AM. Advanced techniques in multimedia watermarking: image, video and audio applications: image, video and audio applications: IGI Global. 2010

[3]. Atallah MJ, Mcdonough CJ, Raskin V, Nirenburg S. Natural language processing for information assurance and security: an overview and implementations. Proceedings of The 2000 Workshop on New Security Paradigms: ACM. 2001; 51–65.

[4]. Ahmad AM, Sulong G, Rehman A, Alkawaz MH, Saba T. Data hiding based on improved exploiting modification direction method and Huffman coding, Journal of Intelligent Systems 2014; 23(4): 451–459. doi. 10.1515/jisys-2014-0007

[5]. Bashardoost M, Rahim MSM, Hadipour N. A novel zero-watermarking scheme for text document authentication. Journal Teknologi 2015; 75(4): 49–56.

[6]. Brassil JT, Low S, Maxemchuk NF. Copyright protection for the electronic distribution of text documents. Proceedings of the IEEE 1999; 87(7): 1181–1196.

[7]. Cheddad A, Condell J, Curran K, McKevitt P. Digital image steganography: survey and analysis of current methods. Signal Processing 2010; 90(3): 727–752.

[8]. Cheng Y, Zhang J, Gong X, et al. Research on polymorphism and inertial reading application in text watermarking algorithm. IEEE Ninth International Conference Broadband and Wireless Computing, Communication and Applications (BWCCA), 2014; 89–95.

[9]. Cox IJ, Linnartz J-PM. Some general methods for tampering with watermarks. IEEE Journal on Selected Areas in Communications 1998; 16(4): 587–593.

[10]. Dittmann J, Wohlmacher P, Nahrstedt K. Using cryptographic and watermarking algorithms. IEEE Multimedia 2001; 4:54–65.

[11]. Haron H, Rehman A, Wulandhari LA, Saba, T. Improved vertex chain code based mapping algorithm for curve length estimation, Journal of Computer Science 2011; 7(5): 736–743. doi.10.3844/jcssp.2011.736.743.

[12]. Jalil Z, Mirza AM. A review of digital watermarking techniques for text documents international conference on information and multimedia technology 2009. doi. 10.1109/ICIMT.2009.11.

[13]. Johnson, N.F., Jajodia, S., "Steg analysis of images created using current steganographic tools", April 1998, URL: http://www.ise.gmu.edu/~njohnson/ihws98/jjgmu.html (11/26/01 17:00)

[14]. J. M. Topkara, U. Topkara, and M. J. Atallah, "Words are not enough: sentence level natural language watermarking," in *Proceedings of the 4th ACM International Workshop on Contents Protection and Security (MCPS '06)*, pp. 37–46, Santa Barbara, Calif, USA, October 2006.

[15]. Lew CH, Woo CS. Using combined pseudo-random number generator with digital text-based watermarking for cryptography application. Proceedings of the International Multi conference of Engineers and Computer Scientists. 2013.

[16]. Lung JWJ, Salam MSH, Rehman A, Rahim MSM, Saba T. Fuzzy phoneme classification using multi-speaker vocal tract length normalization, IETE Technical Review 2014; 31(2): 128–136. doi.10.1080/02564602.2014.892669

[17]. M. Pal, "A survey on digital watermarking and its application," International Journal of Advanced Computer Science and Applications, vol. 7, no. 1, pp. 153–156, 2016.

[18]. Meral HM, Sankur B, Özsoy AS, Güngör T, et al. Natural language watermarking via morpho-syntactic alterations. Computer Speech & Language n.d; 2015, 23(1): 107–125.

[19]. Muhsin ZF, Rehman A, Altameem A, Saba A, Uddin M. Improved quadtree image segmentation approach to region information The imaging science journal 2014; 62(1): 56–62. doi. 10.1179/1743131X13Y.0000000063.

[20]. M. H. Alkawaz, G. Sulong, T. Saba, A. S. Almazyad, and A. Rehman, "Concise analysis of current text automation and watermarking approaches," *Security and Communication Networks*, vol. 9, no. 18, pp. 6365–6378, 2016.

[21]. M. Bashardoost, M. S. Mohd Rahim, T. Saba, and A. Rehman, "Replacement Attack: A New Zero Text Watermarking Attack, *3D Research*, vol. 8, no. 1, article no. 8, 2017

[22]. N. A. A. S. Al-Maweri, R. Ali, W. A. Wan Adnan, A. R. Ramli, and S. M. S. A. Abdul Rahman, "State-of-the-art in techniques of text digital watermarking: Challenges and Limitations," *Journal of Computer Science*, vol. 12, no. 2, pp. 62–80, 2016.

[23]. Nodehi A, Sulong G, Al-Rodhaan M, Al-Dhelaan A,Rehman A, Saba T. Intelligent fuzzy approach forfast fractal image compression, EURASIP Journalon Advances in Signal Processing 2014. doi.10.1186/1687-6180-2014-112.

[24]. P. Singh and R. S. Chadha, "A survey of digital watermarking techniques, applications and attacks," *International Journal of Engineering and Innovative Technology*, vol. 2, no. 9, pp. 165–175, 2013.

[25]. Petitcolas, F.A.P., Anderson, R., Kuhn, M.G., "Information Hiding - A Survey", July1999, URL: http://www.cl.cam.ac.uk/~fapp2/publications/ieee99-infohiding.pdf (11/26/0117:00)

[26]. Provos, N., Honeyman, P., "Detecting Steganographic Content on the Internet", August 2001, http://www.citi.umich.edu/techreports/reports/citi_tr_01-11.pdf (11/26/01 17:00)

[27]. Rehman A, Saba T. Evaluation of artificial intelligent techniques to secure information in enterprises, Artificial Intelligence Review 2014; 42(4): 1029–1044.doi. 10.1007/s10462-012-9372-9

[28]. R. Petrovi´c, B. Tehranchi, and J. M. Winograd, “Security of copy-control watermarks,” in Proceedings of the TELSIKS 2007 - 8th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, pp. 117–126, Serbia, September 2007

[29]. Saba T, Rehman A, Elarbi-Boudihir M. Methods and strategies on off-line cursive touched characters segmentation: a directional review. Artificial Intelligence Review 2014; 42(4): 1047–1066. doi:10.1007/s10462-011-9271-5.

[30]. Sabu M Thampi: Information Hiding Techniques: A Tutorial Review. Department of Computer Science & Engineering LBS College of Engineering, KasaragodKerala- 671542, S. 2016 2002. India

[31]. Zeng F, Deng X. Reversible visible image watermarking: model, evaluation and application. International Journal Of Advancements In Computing Technology 2012; 4(10): 118–124.

[32]. Zhang SR, Yao Z, Meng XC, Liu CC. New digital text watermarking algorithm based on new-defined characters. IEEE International Symposium on Computer, Consumer and Control 2014; 1: 713–716.

[33]. Wolfgang RB, Podilchuk CI, Delp EJ. Perceptual watermarks for digital images and video. Proceedings of the IEEE 1999; 87(7): 1108–1126.

IOSR Journal of Applied Physics (IOSR-JAP) is UGC approved Journal with Sl. No. 5010, Journal no. 49054.

Aru, Okereke Eze " Detailed Examination of Information Hiding Techniques for Copyright Protection of Text Documents.” IOSR Journal of Applied Physics (IOSR-JAP) , vol. 11, no. 5, 2019, pp. 51-61.