Mehran University Research Journal of Engineering and Technology
Vol. 41, No. 1, 100 - 110, January 2022
p-ISSN: 0254-7821, e-ISSN: 2413-7219
DOI: https://doi.org/10.22581/muet1982.2201.10

A Secure Digital Text Watermarking Algorithm for Portable Document Format (PDF)

Umair Khadim[1], Muhammad Munwar Iqbal[2], Muhammad Awais Azam[3]

RECEIVED ON 16.12.2019, ACCEPTED ON 19.05.2020

ABSTRACT

Nowadays, with the rapid development of advanced technologies, an illegal copy of digital documents can be easily generated. The Portable Document Format is the most common and widely used text document on the internet. The copyright protection of these documents is a challenging task. Advanced techniques have been proposed in the past but have not delivered the expected results. These techniques are either robust or imperceptible or have a high capacity, but do not maintain the balance between all parameters. Digital watermarking has been used over the past decade to detect forgery and tampering detection, maintain copyright and authentication. This study proposes a novel approach for Portable Document Format based on document page objects. The Special objects of Portable Document Format are used for watermarking without affecting the content of the original document. The proposed technique preserves the imperceptibility and resists against formatting attacks. The watermark information is extracted with high probability, the proposed technique is robust, secured, imperceptible, and embeds 0.22 KB of the watermark in the host document.

Keywords: Digital Watermarking, Copyright Protection, Imperceptible, Capacity, Information Security, Robustness, Portable Document Format (PDF)

1. INTRODUCTION

Text-based information is disseminated daily on the Internet, for example, in the form of academic documents, electronic books, and e-mails, *etc*. Portable Document Format (PDF) is the prevalent carrier of this information; this is why huge number of documents are shared on the internet in PDF format. Due to the advancements in digital technology content can be easily redistributed, copied, and stored. On the other hand, unauthorized copying and illegal distribution of digital contents created problems for data owners for ensuring the copyright, millions of dollars spent by booksellers, and publishers [1]. A massive contribution of mobile devices and social media is in the sharing of digital content. Various illegal activities violate the copyright protection of digital content, such as forgery, tampering, and illegal copying [2-4].

The protection of digital text documents has been seriously ignored in the past. However, it is the most dominant part of the Internet, newspapers, articles, e-books, legal documents, and magazines [5]. For digital content specifically, the text documents copyright protection is the need of time and cannot be neglected. In past steganography, cryptography and information hiding techniques were used to solve the copyright issues [6]. Nowadays, digital watermarking gives a better solution for copyright protection with secret

[1] Department of Software Engineering, University of Kotli Azad Jammu and Kashmir 11100, Pakistan
Email: umair_khadim@uokajk.edu.pk (Corresponding Author).
[2] Department of Computer Science, University of Engineering and Technology, Taxila 47050, Pakistan
Email: munwar.iq@uettaxila.edu.pk
[3] Department of Computer Engineering, University of Engineering and Technology, Taxila, Pakistan
Email: awais.azam@uettaxila.edu.pk

information called watermark embedded in digital content [7, 8]. The watermark is used for ownership verification when illegal use of digital content happens. The major application of digital watermarking is presented in Fig, 1. The most common use of digital watermarking is copyright protection. It is used to provide authentication of digital contents or preserve the ownership verification.

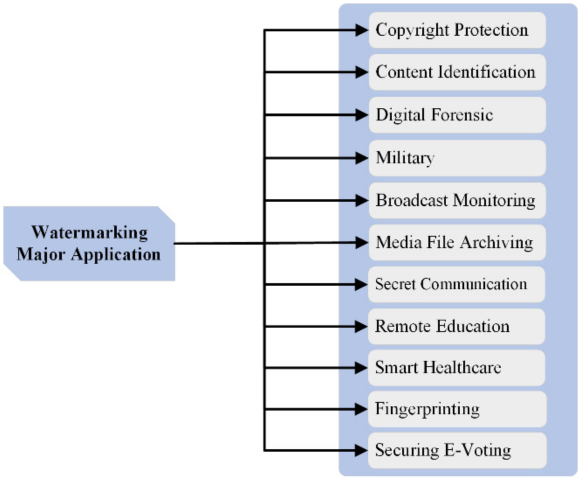


Fig. 1: Watermarking major applications

PDF is an accessible file format and developed by Adobe System Inc. [9-11]. Unauthorized persons can easily modify digital documents with the help of advanced technologies. The Portable Document File format includes some security mechanisms which can be broken with advanced technologies. These modified documents can be illegally distributed through the internet. This fact suggests the need to develop an effective authentication system [12, 13]. In this paper, we have proposed a secure digital watermark algorithm for PDF documents. Our technique is based on PDF file page objects. The proposed system is secure, robust, incorporates large embedding capacity, and ensures the visual imperceptibility of the document.

Our main contributions in this research are as under:

- We develop a secure digital watermark system for PDF documents copyright protection, which prevents illegal distribution, reproduction, and manipulation.
- The proposed watermarking technique does not disturb the digital content and calls zero watermark technology because watermark information is embedded in PDF file objects.
- The proposed technique can incorporate large embedding capacity.

The rest of the paper is organized as follows. In Section 2, the work related detail is presented, the PDF file structure is deliberated in Section 3. Section 4 is about the proposed model. The experimental results and analysis are elaborated in Section 5, and the conclusion and future work are presented in Section 6.

2. RELATED WORK

Digital watermarking is a hot area of research, which can be categorized into text, image, audio, and video. PDF text documents are considered in this research for watermarking. At present, several methods are proposed for watermarking, which are based on syntax, semantic, format based, and so on. Zhong *et al.* [14] proposed a system which changes each word preset distance in PDF files for embedding data in the right margin. Two concepts neighbor difference, and environment equal is proposed, which reveal the statistical properties of spaces are shown in Fig. 2.

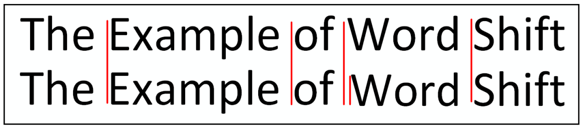


Fig. 2: A simple example of word shift

In [15, 16], words and paragraph spaces are utilized for watermarking. The main drawback of these techniques is if the spaces are removed between text, then embedding bits will be ruined. Lingjun *et al.* [17] introduced a technique for PDF document based on document structure, where PDF document page objects are used for hiding watermarking signals. Kuribayashi *et al.* [16] introduced a new method based on the justified text in PDF document. In this method, first the secret information is compressed by Huffman coding and then embedded into some particular lines of the original document. Bitar *et al.* [18] proposed a new technique for Portable Document Format. It is based on Quantization Index Modulation (QIM). Secret information bits are embedded into characters' x-coordinate values. Liu *et al.* [19] introduced three algorithms that used the incremental update feature of PDF document. The first algorithm embeds data by altering some text state variables and other embed data through writing incremental updates for objects, and the third one uses a cross-reference section for embedding. Zhang *et al*. [20] suggest an algorithm, which is based on PDF document structure. Photographer software is used to generating a font file of the same characters. Simin *et al*. [21] proposed a novel algorithm based on PDF document page objects. The watermark information is embedded in PDF page structure objects.

Hakak *et al.* [22] presented a complete framework regarding the automatic authentication and distribution of the digital Quran and Hadith verses. The verification process is divided into two phases, security and verification. Wen *et al.* [23] proposed an algorithm for XML documents to hide information. In that method, a functional dependency is used for the XML file as a function for Zero-Watermark. The proposed method performs well in alternation attacks, compression attacks, reorganization attacks, and selection attacks. Xiao *et al.* [24] suggest a novel method for embedding information in text, which is based on font code that embeds a watermark into text by disrupting text characters glyphs while retaining text content. The glyph recognition method is also presented to restore the information that is embedded in the encrypted document. Feng *et al.* [25] establish an automatic framework for PDF documents that can check whether the document contains any clues about author information.

Hatoum *et al*. [26] proposed a watermarking technique that is based on the quantization step. The proposed technique is robust against formatting attacks in terms of robustness. Kuribayashi *et al*. [27] utilize the spaces between words and paragraphs for embedding the watermark information. The author counts the length of spaces as one-dimensional feature vector for watermark embedding. The proposed technique is not robust against the formatting attacks because if the spaces are removed, then watermarked information is also removed.

Nursiah *et al*. [28] introduced a technique that is based on glyph positioning coordinate values. The reverse zero-run length coding technique is adopted to suppress bit stream size increment. Tyagi *et al.* [29] proposed an extraction process that provides authentication of received stego-cover file such that only the desired file is acknowledged for the extraction process. Otherwise fake file is discarded by a recipient.

3. PDF FILE STRUCTURE

PDF is the most critical file format and created in 1993 by Adobe Systems. The primary PDF file structure consists of Header, Body, Cross-Reference Table, and Trailer, as shown in Fig. 3. The header is the first line of PDF file that includes the version number. The body can hold all data, which can be shown in PDF viewer and supports eight types of objects. These objects are Null, String, Integer, Boolean, Array, Name, Stream, and Dictionary. The cross-reference table is a core element of PDF document and provides a binary offset from the beginning of the file [30-32]. The responsibility of the cross-reference table is that it contains the reference of all objects in a PDF file. It begins with keyword “xref” and the next lines are exactly 20 bytes long as shown in Fig. 4.

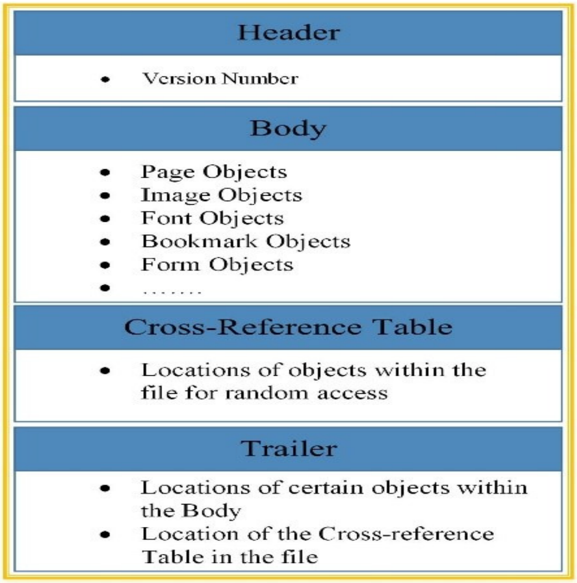
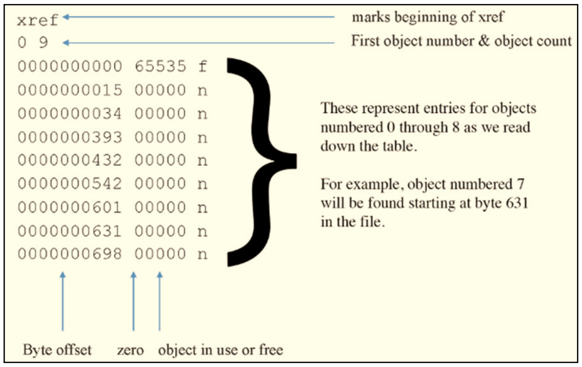

Header
• Version Number
Body
• Page Objects
• Image Objects
• Font Objects
• Bookmark Objects
• Form Objects
• ....
Cross-Reference Table
• Locations of objects within the file for random access
Trailer
• Locations of certain objects within the Body
• Location of the Cross-reference Table in the file

Fig. 3: The basic structure of PDF File [33]

The keyword “xref” indicates the beginning marks, the list starts at object 0, and the next number is a count of cross-reference table objects. The Trailer is used to find the cross-reference table. Each object is linked with that table which acts as a dictionary. The example of a PDF file syntax object is given in Fig. 5, where a unique number is starting with “obj” to “endobj” assigned to each object. The script and the information for displaying text, figure and images appear between “stream” and “endstream”. “BT” signifies Began Text and “ET” denotes End Text. “If”, “Td” and “Ti” are some operators to represent the text document, where “Tf” is used for font size and text style. “Td” signifies the offset of the current line, and “Tj” is used to show characters and spaces between them.



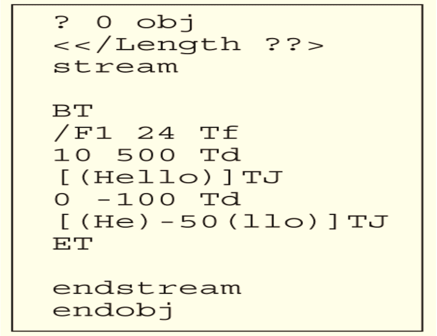
```
xref                          marks beginning of xref
0 9                           First object number & object count
0000000000 65535 f
0000000015 00000 n
0000000034 00000 n
0000000393 00000 n
0000000432 00000 n
0000000542 00000 n
0000000601 00000 n
0000000631 00000 n
0000000698 00000 n
```

These represent entries for objects numbered 0 through 8 as we read down the table.

For example, object numbered 7 will be found starting at byte 631 in the file.

Byte offset zero object in use or free

Fig. 4: The Cross-Reference Table [33]



```
? 0 obj
<</Length ??>
stream

BT
/F1 24 Tf
10 500 Td
[(Hello)]TJ
0 -100 Td
[(He)-50(llo)]TJ
ET

endstream
endobj
```

Fig. 5: A PDF file object example

The PDF document structure is shown in Fig. 6, and the tree structure provides PDF applications for consumers using limited memory [34]. The document catalog consists of the article threads page tree, named destinations, interactive form, and online hierarchy.

4. SPREAD TRANSFORM DITHER MODULATION (STDM)

STDM is applied to embed the secret information in PDF document. The bits of secret message sm ϵ (0,1) are embedded in PDF document “T”. Therefore, according to embedding “sm” bits, two different dither quantizers are applied. Embed the “sm” bit 0, Q_0 is used, as shown in equation (1).

$$Q_0(T,\Delta) = \left[\frac{(T-d_0)}{\Delta}\right]\Delta + d_0 \qquad (1)$$

where Δ denotes the size of the quantization step. Each bit of “sm” is embedded into “T” without any distortion, which is the most significant advantage of STDM. P is the projection vector of the host signal. The quantized single is specified as shown in (2).

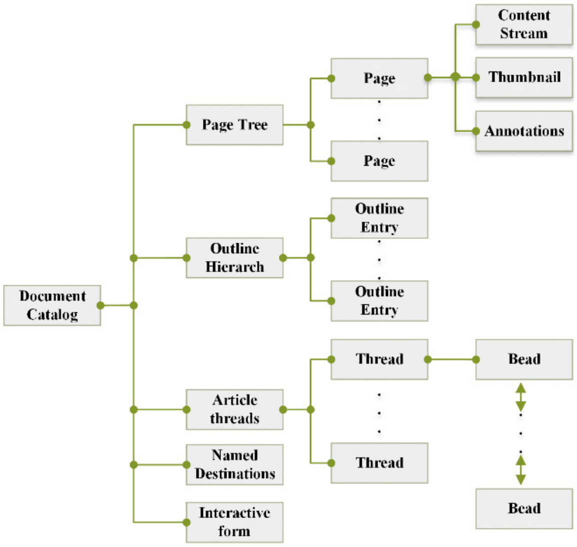


Fig. 6: The Structure of PDF Document

$$T' = T + (Q_{sm}(T^x p, \Delta) - T^x p)p \quad sm \rhd \{0,1\} \qquad (2)$$

where we can re-write equation (2) as.

$$T' = T + \left(\left(\left[\frac{(T^x p)-d_{sm}}{\Delta}\right]\Delta + d_m\right) - T^x p\right)p \qquad (3)$$

5. ADVANCED ENCRYPTION STANDARD (AES) ALGORITHM FOR ENCRYPTION AND DECRYPTION

Encryption and decryption are not our primary task, so any encryption algorithm can be used. We use AES symmetric 256-bit key for encryption and decryption. Its main purpose is to protect important information. In other words, hiding information from unauthorized persons. Encryption is applied to secure the secret message. If any high-level attack is applied on the watermarked document and if anyone acquires the message so cannot read the actual message or ownership detail.

6. ZERO WATERMARKING

A technique is called Zero Watermarking, if it does not change the original content during watermarking. In our proposed technique PDF page objects mentioned in Table 1 are used, and the attribute values in the content stream can be modified for watermarking. The text status operators can be displayed outside of text objects, and the values they set are preserved for text objects in a single content stream. Each page of a document is represented by one or more content streams, which include the page objects. PDF document has a lot of objects which have different types. The attribute values of the content stream object, which can incorporate the values of the numbers are shown in Table 1. Every object attribute has an operator keyword like Tc, Tw, Tz, TL, Tf, Tr, Ts, *etc*. The attribute values in the content stream are modified for watermarking, and it will not effect the entire document contents.

7. PROPOSED MODEL

A zero-text watermarking algorithm is proposed for PDF documents, which is based on PDF page objects. It is shown in Fig. 7. In our proposed technique, the objects and properties of the document are used for storing the secret information. This technique is considered robust when formatting attacks are applied to the digital contents. The watermark information is not deleted or changed because it is stored in the document's objects or properties.

7.1 Watermark Embedding

A secret message is embedded into PDF documents page objects as watermark information. The copyright and authentication of PDF documents are proved through the watermark. Algorithm 1 describes the complete embedding process of watermarking.

Table 1: PDF document text state objects

Operator	Operands	Description
Tc	charSpace	Tc is used for character spacing, a number expressed in unscaled text space units.
Tw	wordSpace	Tw is used for word spacing and expressed a number in unscaled text units.
Tz	scale	Tz set horizontal scaling, and it's a number that specifies the percentage of normal width.
TL	leading	Tl is used to set text leading and specifies a number.
Tf	font size	Tf indicates the font size, which is in numbers.
Tr	render	Tr specifies text rendering mode, which shall be an integer.
Ts	rise	Ts directs text rise, which is in number

The secret message is first encrypted through the AES encryption algorithm with the private key, which can enhance the hidden data or watermark security. After encryption, the encrypted data is converted to a binary string and then binary to numbers. The mostly PDF document page objects belong to integer types, and numbers can easily be embedded in those objects. The secret information, which is in number form, divided into three equal groups. If different attacks applied to PDF document, the watermark information is recovered from other groups in the worst case. A document translator is used to getting the content stream of PDF files. In last, these classified groups are embedded in suitable page objects of the original PDF file, also known as the watermark PDF document. The

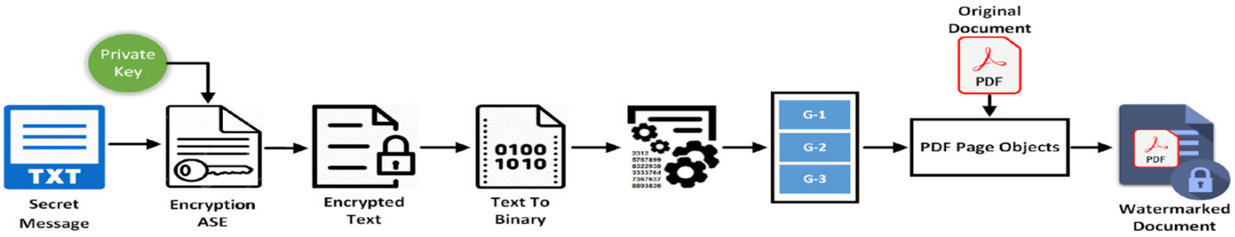


Fig. 7: Proposed model for digital text watermarking in PDF documents

complete watermark embedding process is given in Algorithm 1, where a secret message “sm” is embedded in PDF document objects without any distortion.

Algorithm 1: Embedding watermark in PDF document

```
Input: PDF file “T”, Secret Message “M” and Key “K”
Output: Watermarked PDF file “T´”
Data: $\hat{E}$, $B_n$, $\tilde{N}$, $W\hat{G}$, D[obj]
Variable declaration:
        $\hat{E}$ = Encrypt message, $B_n$ = Binary numbers,
        $\tilde{N}$ = Numbers, $W\hat{G}$ = Watermark groups ($\hat{G}_1$, $\hat{G}_2$, $\hat{G}_3$)
Variable initialization:
        i = 0, D[obj] = 0
Start:
        $\hat{E}$ = AES [M, K]
        $B_n$ = [ $\hat{E}$ ]
        $\tilde{N}$ = [ $B_n$ ]
        $W\hat{G}$ = [ $\tilde{N}$/3 ]
            for i = 0 to D[obj] do
                D[obj]←Find document objects
                    if Len | D[obj] | > 0
                        $W\hat{G}_n$ → D[obj]
                        i ++
                    end if
            end for
        Save Document
        Watermarked PDF file “ T´ ”
end
```

7.2 Watermark Verification

The extracted watermark is checked and compared to the original watermark for document authentication. The details about watermark extraction or verification are as follows: the document translator is used to read PDF document contents in binary form. The page objects of the document are identified, which contain the watermark. The extracted watermark information, which is in the form of numbers, is converted into binaries then characters. The AES decryption algorithm is applied to decrypt the extracted message which is known as a watermark or secret message. After decryption, the retrieved message is compared to the original message that proves the document authentication. The complete watermark extraction process is given in Algorithm 2. The verification of extracted message EM can be performed by using (4).

$$EM = \text{avg}\ \min_{sm>\{0,1\}} |T'P - Q_m(T'p, \Delta)| \tag{4}$$

Algorithm 2: Watermark extraction from PDF document

```
Input: Watermarked PDF file “T´” and Key “K”
Output: PDF file “T”, Secret Message “M”
Data: $D_{\hat{E}}$, $B_n$, $C_{har}$, $\tilde{N}$, $W\hat{G}$, D[obj]
Variable declaration:
        $D_{\hat{E}}$ = Decrypt message, $B_n$ = Binary numbers,
        $C_{har}$ = Characters, $\tilde{N}$ = Numbers, $W\hat{G}$ =
        Watermark groups ($\hat{G}_1$, $\hat{G}_2$, $\hat{G}_3$)
Variable initialization:
        i = 0, D[obj] = 0
Start:
        for i = 0 to D[obj] do
            D[obj]←Find document objects
                $W\hat{G}_n$ ← D[obj]
                i ++
        end for
        $\tilde{N}$ = [$W\hat{G}$]
        $B_n$ = [ $\tilde{N}$ ]
        $C_{har}$ = [ $B_n$ ]
        $D_{\hat{E}}$ = AES [ $C_{har}$, K]
        M ← [$D_{\hat{E}}$]
end
```

8. EXPERIMENTAL RESULTS

In experiments, the PDF file is created using Microsoft Word 2016 with font-family Times New Roman and font size 13pt. To compress and decompress the PDF document “pdftk” toolkit is used before embedding watermark. The experiments are carried on Core i3-3110, Windows operating system. The watermark embedded in the document is “The document copyrights belong to Umair Khadim (umair_khadim@live.com)”.

Digital watermarking having three key constraints is described either. The affiliation between steganography parameters is displayed in Fig. 8. These parameters include robustness, imperceptibility, and capacity (payload) [35].

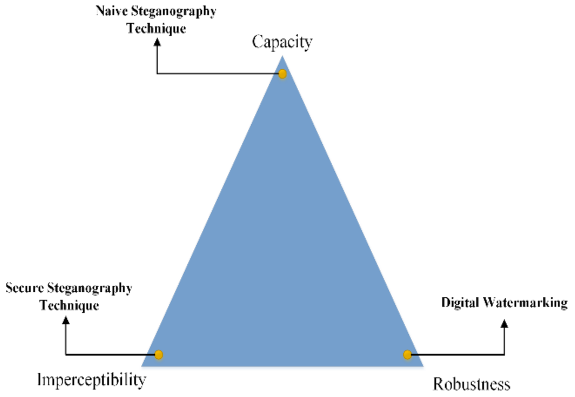


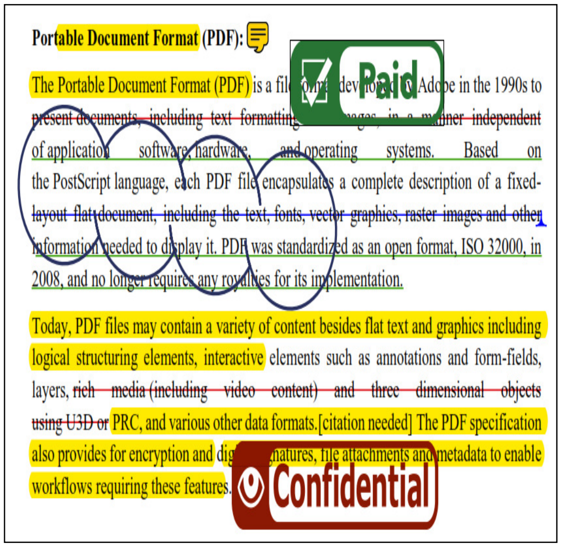
Fig. 8: Information security model [36]

8.1 Robustness

A number of PDF editors are available, which can edit the PDF files online and offline. Various attacks are applied to PDF document in order to examine whether the proposed technique is robust or not. We added comments and marks in the watermark PDF document, as shown in Fig. 9. After inserting comments and marks, we tried to extract the watermark information from it. The experiments show that after applying attacks on watermarked document, the accuracy of the extracted watermark is 99.9%.

The formatting attacks did not affect the watermark information, because PDF document page objects are used for embedding watermark. Interactive forms are the particular type used in PDF documents and appropriate to collect user information in the PDF document. They authorize users to edit, write, modify or delete the information in PDF files on a specific location. We tested interactive forms, and the editing option did not damage or affect the watermark information from PDF documents. Therefore, through experimental results, we can see that the proposed algorithm is robust against formatting attacks.

The hiding capacity of the proposed system is measured using equation (5), where "C" indicates capacity, N_{Bits}(SM) means the secret message size in bits, and W_D(KB) defines the watermarked document size in KBs. As compared with existing techniques, the proposed system improves the embedding capacity size in KBs. As compared with existing techniques, the proposed system improves the embedding capacity [36, 37].


Portable Document Format (PDF):

The Portable Document Format (PDF) is a file format developed by Adobe in the 1990s to ~~present documents, including text formatting and images, in a manner independent~~ of application software, hardware, and operating systems. Based on the PostScript language, each PDF file encapsulates a complete description of a fixed-~~layout flat document, including the text, fonts, vector graphics, raster images and other~~ information needed to display it. PDF was standardized as an open format, ISO 32000, in 2008, and no longer requires any royalties for its implementation.

Today, PDF files may contain a variety of content besides flat text and graphics including logical structuring elements, interactive elements such as annotations and form-fields, layers, ~~rich media (including video content) and three dimensional objects using U3D or~~ PRC, and various other data formats.[citation needed] The PDF specification also provides for encryption and digital signatures, file attachments and metadata to enable workflows requiring these features.

Fig. 9: Applying marks and annotated in a watermarked PDF file using Nitro Pro 8. Capacity

$$C = \frac{N_{Bits}(SM)}{W_D(Kb)} \times 100 \qquad (5)$$

The length of the watermark information is 0.22 KB. The capacity results of the proposed technique are compared with [21], where the author claims that they can embed 0.10 KB of watermark information in PDF document.

Table 2 presents the embedding capacity of secret messages and measures the change in the original and watermarked document, where SM denotes secret message, T represents the original document size, T´ signifies watermarked document size, and W represents the change in both documents. Twenty different document samples with different sizes are used for embedding watermark (secret) information. The comparison of both the original and watermarked document is presented in Fig. 10.

After embedding watermark information in the original document, a slight change is measured in the watermarked document. The 3D representation of capacity analysis and change in the document is represented in Fig. 11.

Secret message (watermarked) embedding capacity is measured in KB, where 0.22 KB of data is embedded in the host document. Original and watermarked document change is measured, which is 0.13%, as revealed in Fig. 12.

Table 2: Watermark embedding capacity and change in document size

Sample	SM (KB)	T (KB)	T´ (KB)	W %	Sample	SM (KB)	T (KB)	T´ (KB)	W %
1	0.01	13.44	13.45	0.08	11	0.12	13.46	13.47	0.12
2	0.02	13.46	13.47	0.09	12	0.13	13.49	13.48	0.09
3	0.03	13.45	13.46	0.08	13	0.14	13.51	13.52	0.08
4	0.05	13.43	13.44	0.10	14	0.15	13.52	13.53	0.10
5	0.06	13.40	13.43	0.11	15	0.16	13.53	13.54	0.11
6	0.07	13.45	13.46	0.10	16	0.18	13.54	13.55	0.10
7	0.08	13.48	13.49	0.10	17	0.19	13.55	13.56	0.10
8	0.09	13.49	13.51	0.11	18	0.20	13.49	13.50	0.11
9	0.10	13.47	13.49	0.10	19	0.21	13.48	13.49	0.10
10	0.11	13.41	13.43	0.12	20	0.23	13.50	13.51	0.10

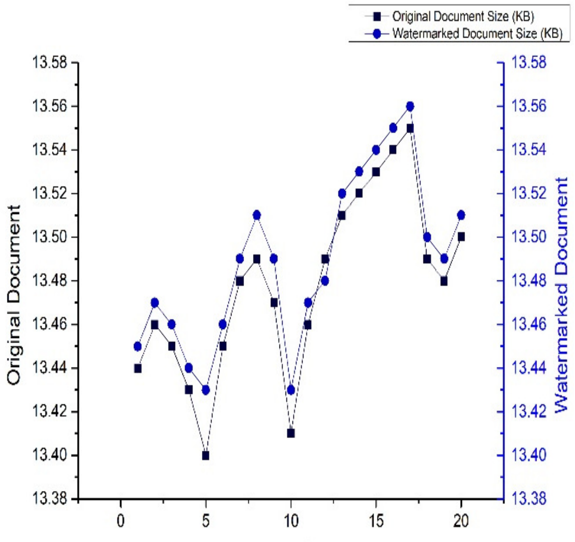


Fig. 10: Comparison of both documents

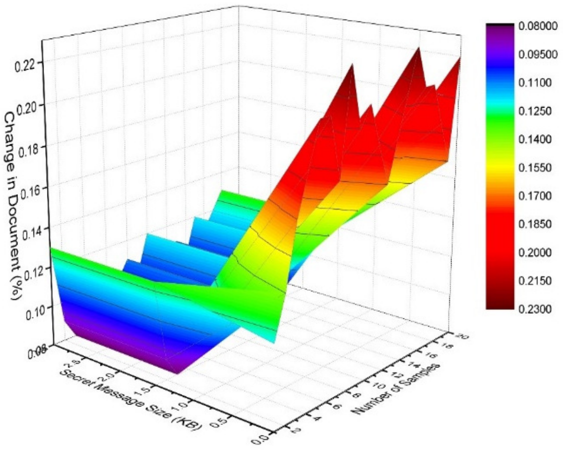


Fig. 11: 3D representation of capacity analysis

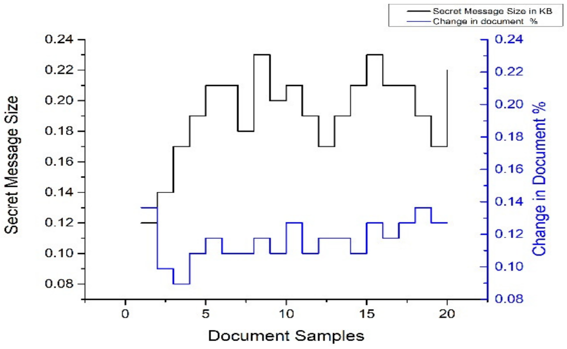
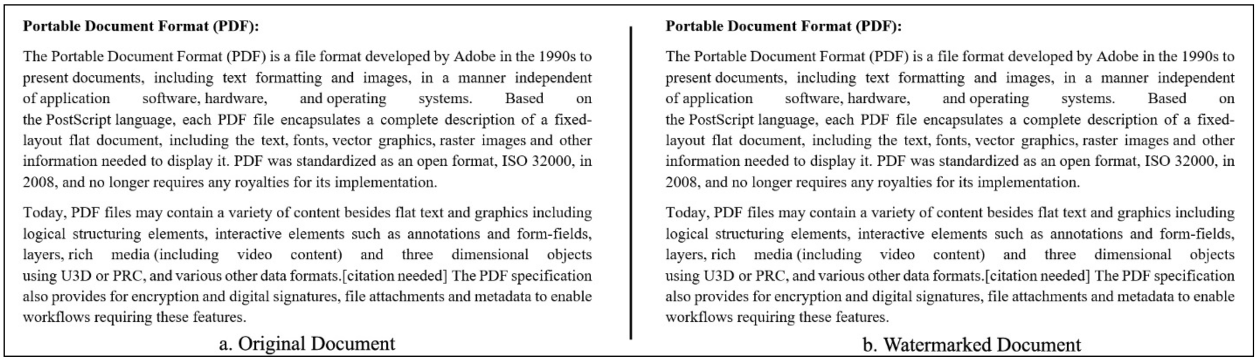


Fig. 12: Size of secret message and change in the document

8.2 Imperceptibility

The imperceptibility means that the watermark information could not feel the audience, or the watermark should not affect the original text. The authorized agency can only detect the watermark only through special processing. The watermark embeds the imperceptibly into the PDF file object without affecting the original documents contents. The original and watermarked document is presented in Fig. 13, where watermark information is embedded in PDF file objects and did not modify the contents of watermarked documents. In this work, four different components of PDF file structure are used for watermarking, which include the header, body, cross-reference table, and trailer. The proposed scheme used STDM for watermark embedding in PDF file objects. Experimental results illustrated that the proposed scheme is robust, interceptible and improves payload as compared to previous techniques.



Portable Document Format (PDF):

The Portable Document Format (PDF) is a file format developed by Adobe in the 1990s to present documents, including text formatting and images, in a manner independent of application software, hardware, and operating systems. Based on the PostScript language, each PDF file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, vector graphics, raster images and other information needed to display it. PDF was standardized as an open format, ISO 32000, in 2008, and no longer requires any royalties for its implementation.

Today, PDF files may contain a variety of content besides flat text and graphics including logical structuring elements, interactive elements such as annotations and form-fields, layers, rich media (including video content) and three dimensional objects using U3D or PRC, and various other data formats.[citation needed] The PDF specification also provides for encryption and digital signatures, file attachments and metadata to enable workflows requiring these features.

a. Original Document

Portable Document Format (PDF):

The Portable Document Format (PDF) is a file format developed by Adobe in the 1990s to present documents, including text formatting and images, in a manner independent of application software, hardware, and operating systems. Based on the PostScript language, each PDF file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, vector graphics, raster images and other information needed to display it. PDF was standardized as an open format, ISO 32000, in 2008, and no longer requires any royalties for its implementation.

Today, PDF files may contain a variety of content besides flat text and graphics including logical structuring elements, interactive elements such as annotations and form-fields, layers, rich media (including video content) and three dimensional objects using U3D or PRC, and various other data formats.[citation needed] The PDF specification also provides for encryption and digital signatures, file attachments and metadata to enable workflows requiring these features.

b. Watermarked Document

Fig. 13: Comparison of the original and watermarked document

9. CONCLUSION

In this study, we have proposed a digital text watermarking algorithm for Portable Document Format. That is based on page objects Portable Document Format. The special properties of Portable Document Format, which include page objects are exploited for embedding watermark information. The experimental results prove that after applying the formatting attacks watermark information is successfully extracted from the document. The proposed algorithm does not affect the original contents of the document. The proposed technique reports excellent results against robustness and impermeability. The proposed technique is superior to other similar methods in terms of imperceptibility, robustness, and capacity. In future work, we will improve the embedding capacity of watermark information and design a secure watermarking system for printed documents.

ACKNOWLEDGEMENT

We would like to thank Computer Science Department, University of Engineering and Technology, Taxila for providing testing facilities for the research work published in this article.

REFERENCES

1. Alkawaz M.H., Sulong G., Saba T., Almazyad A.S., Rehman A., "Concise analysis of current text automation and watermarking approaches", *Security and Communication Networks*, Vol. 9, pp. 6366-8378, 2016.
2. Rizzo S.G., Bertini F., Montesi D., "Text Authorship Verification through Watermarking", Proceedings of the European Intelligence and Security Informatics Conference (EISIC), pp. 168-171, Uppsala, Sweden, 17-19 August 2016.
3. Iqbal M.M., Khadam U., Han K.J., Han J., Jabbar S., "A Robust Digital Watermarking Algorithm for Text Document Copyright Protection based on Feature Coding", *Proceedings of the 15th International Wireless Communications and Mobile Computing Conference*, pp. 1940-1945, Tangier, Morroco, 24-28 June 2019.
4. Khadam U., Iqbal M.M., Azam M.A., Khalid S., Rho S., Chilamkurti N., "Digital Watermarking Technique for Text Document Protection Using Data Mining Analysis", *IEEE Access*, Vol. 7: pp. 64955-64965, 2019.
5. Khadam U., Iqbal M.M., Habib M.A., Han K., "A Watermarking Technique Based on File Page Objects for PDF", *Proceedings of the Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, pp. 1-5, Victoria, B.C. Canada, 21-23 August 2019.
6. Kamaruddin, N.S., Kamsin A., Por L.Y., Rahman H., "A Review of Text Watermarking: Theory, Methods, and Applications", *IEEE Access*, Vol. 6, pp. 8011-8028, 2018.
7. Khadim U., Khan A., Ahmad B., Khan A., "Information hiding in text to improve performance for word document", *International Journal of Technology and Research*, Vol. 3, No.3, pp. 50-55, 2015.
8. Hatoum M.W., Darazi R., Couchot J.-F.. "Blind PDF Document Watermarking Robust Against PCA and ICA Attacks", *Proceedings of the 15th*

International Joint Conference on e-Bussiness and Telecommunications, pp. 420-427, Porto, Portugal, 2018.

9. Naz F., Khan A., Ahmed M., Khan M.I., Din S., Ahmad A., Jeon G., "Watermarking as a service (WaaS) with anonymity. Multimedia Tools and Applications", *Multimedia Tools and Applications*, Vol. 79, pp. 16051-16075, 2020.
10. Zhang S., Wu Y., Li Q., Li G.. "PDF document watermarking algorithm based on discarded page object", *Proceedings of the 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD),* pp. 3107-3111, Guilin, China, 29-31 July 2017.
11. Sharma K.U., Talan P.P., Nawade P.P., Ali M.S., Sharma A.U., "Digital Watermarking—An Overview and a Possible Solution". In Satapathy S., Joshi A. (Eds.): *Information and Communication Technology for Intelligent Systems, Smart Innovation, Systems and Technologies*, Vol. 107, Springer, Singapore, 2019.
12. Soleymani S.H., Taherinia A.H., "High Capacity Image Data Hiding of Scanned Text Documents Using Improved Quadtree", *arXiv:1803.11286*, 2018.
13. Alotaibi R.A., Elrefaei L.A., "Improved capacity Arabic text watermarking methods based on open word space", *Journal of King Saud University-Computer and Information Sciences*, Vol. 30, No.2, pp. 236-248, 2018.
14. Zhong S., Cheng X., Chen T., "Data Hiding in a Kind of PDF Texts for Secret Communication", *International Journal of Network Security*, Vol. 20, No.1, pp. 17-26, 2007.
15. Por L.Y., Delina B., "Information hiding: A new approach in text steganography", *Proceedings of the 7th WSEAS International Conference on Applied Computer and Applied Computational Science*, Hangzhau, China, 6-8 April 2008.
16. Kuribayashi M., Fukushima T., Funabiki N., "Data hiding for text document in PDF file", *Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Matsue, Japan, 12- 15 August 2017.
17. Lingjun L., Liusheng H., Wei Y., Xinxin Z., Zhenshan Y., Zhili S., "Detection of word shift steganography in PDF document", *Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks*, pp. 1-8, Istanbul, Turkey, 22-25 September 2008.
18. Bitar A.W., Darazi R., Couchot J.-F., Couturier R., "Blind digital watermarking in PDF documents using Spread Transform Dither Modulation", *Multimedia Tools and Applications*, Vol. 76, No.1, pp. 143-161, 2017.
19. Liu H., Li L., Li J., Huang J., "Three novel algorithms for hiding data in pdf files based on incremental updates". In: Shi Y.Q., Kim H.J., Perez-Gonzalez F. (Eds.): *Digital Forensics and Watermarking (IWDW 2011), Lecture Notes in Computer Science*, Vol. 7128, Singapore, Berlin,
20. Zhang S., Li Q., Liu C.-C., Li G., "Hiding new words in a PDF document", *Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pp. 1703-1707, Zhangjiajie, China, 15-17 August 2015.
21. Simin H., Xingming S., Zhangjie F., "A Novel Information Hiding Algorithm Based on Page Object of PDF Document", *Proceedings of the 10th International Symposium on Distributed Computing and Applications to Business, Engineering and Science*, pp. 266-270, Wuxi, China, 14-17 Octoer 2011.
22. Hakak, S., Kamsin A., Veri J., Ritonga R., Herawan T., "A Framework for Authentication of Digital Quran, in Information Systems Design and Intelligent Applications", In Bhateja V., Nguyen B., Nguyen N., Satapathy S., Le D.N., (Eds.): *Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing*, Vol. 672, Springer, Singapore, 2018.
23. Wen Q., Wang Y., Li P., "Two Zero-Watermark methods for XML documents", *Journal of Real-Time Image Processing*", Vol. 14, No. 1, pp. 183-192, 2018.
24. Xiao C., Zhang C., Zheng C., "FontCode: Embedding Information in Text Documents using Glyph Perturbation", *ACM Transactions on Graphics*, Vol. 37, No.2, pp. 1-16, 2018.
25. Feng Y., Liu B., Cui X., Liu C., Kang X., Su J., "A Systematic Method on PDF Privacy Leakage Issues", *Proceedings of the 17th IEEE*

International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE), pp. 1020 -1029, New York, N.Y., U.S.A., 2018.

26. Hatoum M.W., Darazi R., Couchot J.-F., "Normalized blind STDM watermarking schemefor images and PDF documents robust against fixed gain attack", *Multimedia Tools and Applications*, Vol. 79, No.3, pp. 1887-1919, 2020.
27. Kuribayashi M., Fukushima T., Funabiki N., "Robust and Secure Data Hiding for PDF Text Document", *IEICE Transactions on Information and Systems*, Vol. 102, No.1, pp. 41-47, 2019.
28. Nursiah N., Wong K., Kuribayashi M., "Reversible Data Hiding in PDF Document Exploiting Prefix Zeros in Glyph Coordinates", *Proceedings of the Asia-Pacific Signals and Information Processing Association Annual Summit and Conference (APSIPA ASC),* pp. 1298-1302, Lanzhou, China, 18-21 November 2019.
29. Tyagi S., Dwivedi R.K., Saxena A.K., "A novel PDF steganography optimized using segmentation technique", *International Journal of Information Technology*, Vol.12, pp.. 1227-1235, 2020.
30. Sloan, T., Hernandez-Castro J., "Dismantling OpenPuff PDF steganography", *Digital Investigation*, Vol. 25, pp. 90-96. June 2018.
31. Zhong Z., Guo Y., Xu G., "Digital watermarking algorithm based on structure of PDF document", *Journal of Computer Applications*, Vol. 32, No. 10, pp. 2776-2778, 2012.
32. Al Shaikhli I.F., Zeki A.M., Makarim R.H., Pathan A.-S. Khan, "Protection of integrity and ownership of PDF documents using invisible signature", Proceedings of the UKSim 14th International Conference on Computer Modelling and Simulation, pp. 533-537, Cambridge, U.K., 28-30 March 2012.
33. Reference, A.S.I.P., Adobe Systems Incorporated. PDF Reference, 1.6, 2006.
34. Liu, X., Zhang Q., Tang C., Zhao J., Liu J., "A steganographic algorithm for hiding data in PDF files based on equivalent transformation", *Proceedings of the International Symposiums on Information Processing*, pp. 417-421, Moscow, Russia, 23-25 May 2008.
35. Amirtharajan R., Rayappan J.B.B., "Inverted pattern in inverted time domain for icon steganography", *Information Technology Journal*, Vol. 11, pp. 587-595, 2012.
36. Fridrich J., "Applications of data hiding in digital images", . Proceedings of the Fifth International Symposium on Signal Processing and its Applications, (IEEE Cat. No. 99EX359). Brisbane, Australia, 22-25 August, 1999.
37. Naqvi N., Abbasi A.T., Hussain R., Khan M.A., Ahmad B., "Multilayer partially homomorphic encryption text steganography (MLPHE-TS): a zero steganography approach", *Wireless Personal Communications*, Vol. 103, No.2, pp. 1563-1585. 2018.