International Journal of
Science and Engineering Investigations vol. 10, issue 111, April 2021
Received on April 22, 2021 ISSN: 2251-8843

Securing PDF Documents Using Cryptography and Digital Watermarking Techniques

Adetokunbo A. Adenowo[1], Mary A. Adedoyin[2], Saheed Adedamola Adeyemi[3]
[1,2,3]Department of Electronic and Computer Engineering, Lagos State University, Epe Campus, Lagos, Nigeria
([1]adetokyom@yahoo.com, [2]mary.adedoyin@lasu.edu.ng, [3]adeyemisaheed@gmail.com)

***Abstract*-** The internet enables the interconnectivity of devices, thereby serves as information transmission medium between devices. The medium is known to be highly porous due to many hackers and spoofing, thus exposing itself and its digital contents to tampering or manipulation. This work therefore concerns a software-based process that authenticates a portable document format (PDF) document, as well as ensure its originality or integrity using a hybrid of cryptography and digital watermarking. The software authentication aspect utilized the advanced encryption standard (AES), while the integrity or originality aspect was addressed using visible watermarking. The software was written using C sharp (C#), an object-oriented programming language. Thus, it authenticates and integrate a PDF document by applying a digital watermarking together with a dynamic alphanumeric authentication password to prevent unauthorized guest from accessing, manipulating or tampering with the information on the document. The software was evaluated using a questionnaire survey filled by persons chosen randomly. Outcome of the survey proved that the software is robust and was able to secure documents from attacks by malicious guests. The software also guarantees integrity of the PDF documents.

***Keywords*-** *Advanced Encryption Standard (AES), Cryptography, Watermarking, Integrity, Authentication*

I. INTRODUCTION

Data security has been a huge challenge facing transmission of data from one destination to another using the World Wide Web. Communication security experts have been in search for the best way to secure data during transmission without compromising its integrity or accessed by unauthorized users. Some of the ways that have been in use for transmitting or sending data include:

- Mailing: this is the process of sending document through local post office.
- Fax technologies: This involve sending digital document by dialing the recipient fax number and input the intending file to send [1], [2].
- Email: This is one of the latest technologies of sending digital documents to intending recipient through the latter unique email identity [3].

Despite the above means of sending data, there have being numerous cases of data or information been compromised before getting to the recipient, either through hacking or spoofing. These data or information insecurity challenges, informed the development of some security frameworks or schemes such as cryptography and digital watermarking technologies.

Cryptography is famous and significantly utilized because of its age and versatile functions. It changes lucid message into a garbled structure, utilizing a key to scramble the message. There are different types of cryptography frameworks [4-6]. In symmetric framework, same key is utilized to encode and unscramble the message. Information control is quicker because of the brevity in the length of the key, dissimilar to asymmetric framework which utilizes public key to scramble and private key to unscramble. In asymmetric framework, the primary key is different from the public key used. A typical example of symmetric framework is the Advance Encryption Standard (AES) [7], it is one of the strongest encryption standards with 256 bits key and has 14 rounds of encryption which makes it survive brute force or any form of penetration attack. Due to the security strength of the latter, this work adopts the AES for the authentication phase of securing a PDF document.

On the other hand, digital watermarking is deployed to ascertain genuinely the particular origin of an information. Two types of digital watermarking have been mentioned in the literatures: the visible and invisible watermarking [8-10]. Both are robust and secure, although invisible watermarking tends to attract more attention compared to visible watermarking. Digital watermarking can be designed using discrete cosine transform (DCT), discrete wavelength transform (DST) [11] or using object-oriented language which is considered in this current work (see [12] and [13]).

World Wide Web, also referred to as the Internet, serves as a medium of transmitting information and has been in existence since the early 90 [14]. Despite its huge benefits to humanity, its week point is exploited by ill-intentioned persons. It's known to be highly porous due to many hackers and spoofing; a major reason data/information security should take a front role in preventing unwanted users viewing digital contents. It is in view of the foregoing that this work is conceived. Hence, security of data or information is hereby considered from the aspects of integrity and authentication.

In this work, integrity is achieved by using visible watermark string, "*Lagos State University Confidential*", at an angle 45 degree across the document. Integrity shows the true ownership and also determines the originality of a sensitive material [15]. On the other hand, authentication is the process of applying key to the information [15]. The key could be public or private key; in some cases, it could involve asymmetric approach. Notwithstanding the choice of type of key or approach, the goal is to prevent unwanted access to information that requires security.

Thus, this paper proposes to secure digital contents, specifically PDF documents, using cryptography to authenticate and visible digital watermarking to ensure integrity. Accordingly, a software is proposed, using C sharp (C#) object oriented programming language, to implement both the cryptographic authentication, as well as the digital watermarking. Verification of the proposed software is carried out using a questionnaire (see the appendix for the software codes).

II. RELATED WORKS

Sharma et al. [15] proposed a secure image hiding algorithm by using cryptography and steganography. The proposed solution is based on the ground that the internet medium and its digital contents are experiencing increasing attacks due to the dynamism of technology, despite a lot of data securities have been developed in past decades. According to this paper, the secret image is first encrypted using Blowfish algorithm technique and the encrypted image is embedded with a video steganography by using Least Significant Bit (LSB). Blowfish algorithm was compared with data encryption standard (DES), triple data encryption standard (TDES or 3DES) and Rivest cipher 6 (RC6) algorithms using two computers with different specifications (i.e. P-II 266MHZ and P-4 2.4GHZ machines). The result of the comparison shows that blowfish has a better performance, hence its deployment. On the other hand, LSB was used to embed video on the image. It could, therefore, be deduced that the application of cryptography and steganography on data can create a strong security for information transferred through the internet. The blowfish encryption algorithm could provide a better security; also, additional application of LSB stenographic embedding on an encrypted image could further secure the information and make it difficult or hard for third party interference.

In Varshney [16], an investigation into attacks on digital watermarking systems and their categorization was carried out. The attack on digital watermarking was seen as a major concern. Effectively securing digital multimedia from manipulation or attacks was considered very important because digital watermarking can be used in copyright protection for multimedia information. It should be noted that the recipient should be able to decode the watermarked information. Watermark attacks causes impair of the message that is been transferred. It has been classified into various types such as: active attacks, passive attacks, collusion attacks, forgery attacks and simple/waveform/noise attacks. Also, there is detection-disabling attack or synchronization attacks, ambiguity attacks, removal attacks, protocol attacks, copy attacks and geometric attacks. The geometric attack is similar to removal attack, just that in geometric, there is distortion where the watermark does not synchronize with the host. The paper concluded on the need to consider all the various attacks when designing an algorithm for digital watermarking, especially in transmission and storage of data. Also, the paper argue that the main requirement of watermarking is between robustness and the attacker.

Similarly, Aru and Ananaba [17] discussed about secret communication and how to secure information from unwanted recipient. Due to evolution, the ways of securing important information started changing; the introduction of computer and internet brought about essential changes, thus eradicating paper to paperless or electronic publishing. Plagiarism became another threat, leading to protective measures such as integrity, confidentiality, authenticity and copyright [18]. Downloaded materials are manipulated by attackers; their use was another problem which led to a great need for copyright management policy and punishment. In other to prevent plagiarism, enhance copyright management protection and the originality of a transmitted document, the authors proposed watermarking as means of securing digital information such as text, image, audio and video. Watermark properties that are considered include: Capacity, Fragility, Robustness, Security and imperceptibility. Digital watermarking can be used to conceal data by combining the copyright information and text in an unobservable manner [19-209]. Also, it can be used to identify the original owner and to avoid copyright violation, such as illegal redistribution [21].

In the same vein, Sinchita et al. [22] talked about the rapid growth of the internet and networks techniques, and multimedia data transformation. According to the authors, multimedia data is, without pains, copied and altered, thus leading to copyright protection increasing geometrically. It is therefore the imperceptible marking of multimedia data that shows ownership. According to the research, digital watermarking has been introduced as technique for copyright protection of multimedia data. Digital watermarking embeds the copyright information into multimedia data unnoticeably. The paper informed that digital watermarking is being used for copyright protection, finger protection, fingerprinting, copy protection, and broadcast monitoring. Common types of signals that can be watermarked are images, music clips and digital video. The paper shed more light on the application of digital watermarking to still images. According to the paper, the major technical challenge is the design of a highly robust digital watermarking technique, which discourages copyright infringement by making the process of watermarking removal tedious and expensive.

III. METHODOLOGY

In view of the numerous security challenges reported in existing information security literatures, involving the abuse of integrity and copyright of digital contents, the need to authenticate and ensure the integrity of the information contained in these digital contents becomes necessary. Thus,

the conception of a software-based process that can both authenticate a PDF Document, as well as ensure its originality or integrity using a hybrid of cryptography and digital watermarking. The software is written using C# language within Microsoft visual studio environment. The program has two initiation stages which are: (i) the point when the authentication password is initiated; and (ii) the point where the password is applied to the PDF document. The implementation process is illustrated in the flowchart diagram (see figure 1) and captured in the algorithm below.

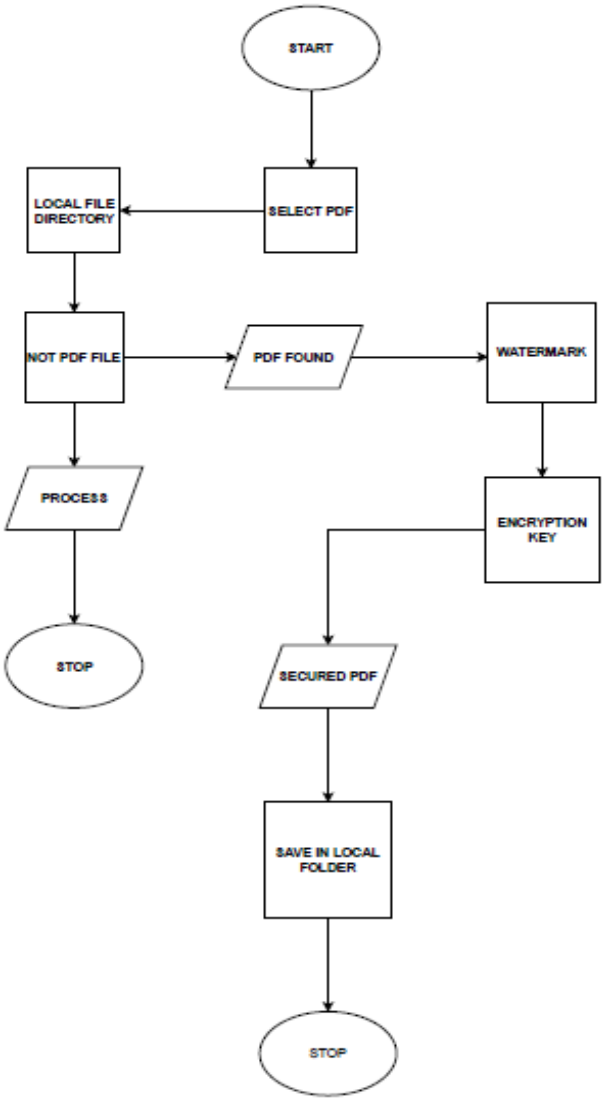

Figure 1. Authentication and Integrity Process flowchart diagram

Algorithm: PDF Authentication and Integrity Process

1. The C sharp software loads on the visual studio which it was designed on.
2. The programs prompt to select the target PDF file which the intended information format is saved with.
3. Within the local folder where the PDF file is, select the file. If the file highlighted or selected is not a PDF file, it will deselect or disallow it and hence the process stops.
4. If it's a PDF file, it selects the file and prompts for the application of the watermark which is ticked.
5. This leads to the inputting of an encryption key to authenticate the action. This encryption key is dynamic and can be on any length. Its 128bits key which is very secured and standardized.
6. The secured PDF is saved and stored in another local folder different from where it was selected and ready to be sent to the recipient.

Also, figure 2 below represents the screenshot of the C# development environment for the authentication and integrity software (see appendix for the software codes), showing the two initiation points from is caption below:

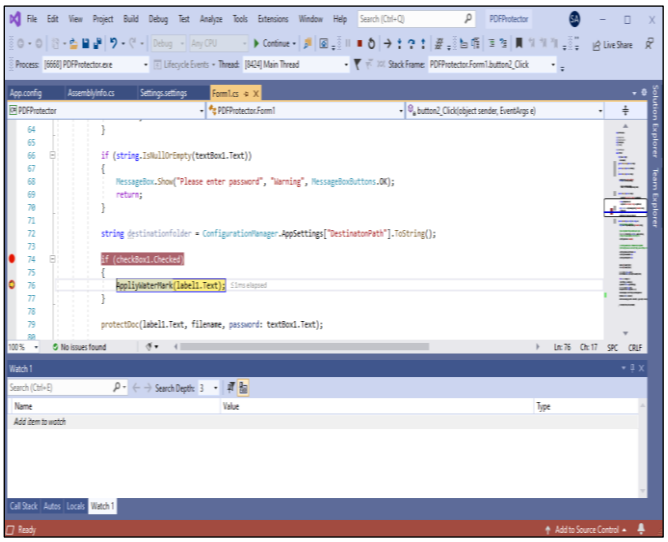

Figure 2. C sharp (C#) application environment

A questionnaire was developed and administered randomly to solicit target users views. This was undertaken to ascertain the software efficiency in satisfying both the authenticity and integrity requirements of the PDF document being verified. Below in figure 3 is a sample of the questionnaire administered.

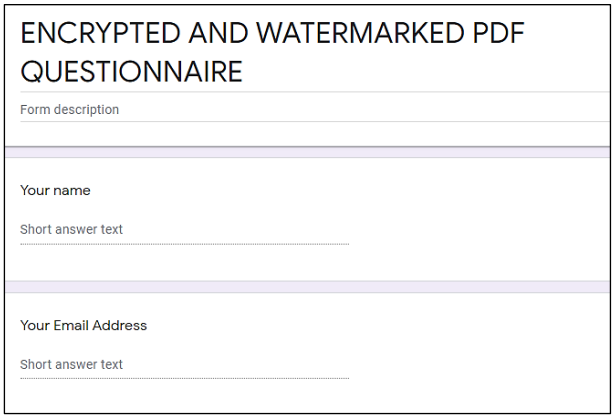

Figure 3. Questionnaire used for the research

IV. RESULTS AND DISCUSSION

Questionnaires were sent out for users' acceptance test and the developed application scripts authentication and integrity test. Below are results from feedbacks received from users and questionnaire sample.

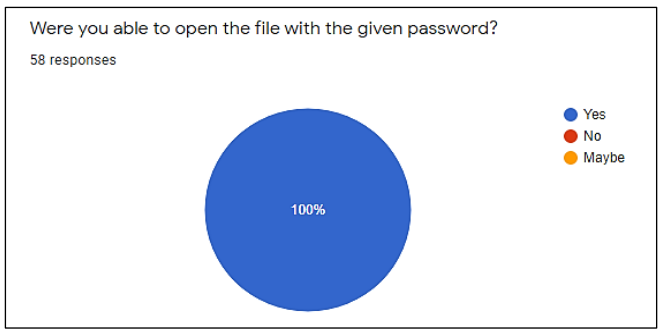

Figure 4. Pie chart of feedbacks received based on authentication test

Figure 4 shows the percentage of responses received from those that were able to use the given password to open the digital document sent to them for review and test. Generally, these results show that the secret keys were genuine and functional. It shows that the authentication test was passed.

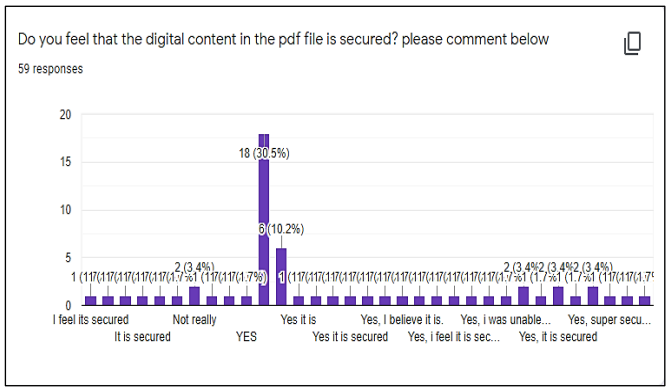

Figure 5. Histogram graph of feedback received on the Integrity test

Considering the histogram graph of figure 5, it can be deduced that the responses showed that it was secured which means that digital content can be secured using this program. Some of the users tried to edit the content but got an error which proves that it can be deployed in the proposed field for securing information. Thus, this proves that the Integrity test was passed.

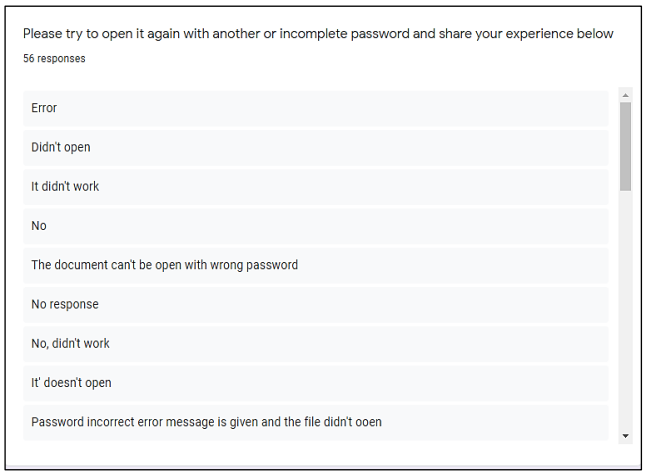

Figure 6. These feedbacks proves the authentication of the program

From feedback received as shown in figure 6, it can be deduced that while trying to put another password which is different from the one used in authenticating, it produced an error. This proves that without the dynamic password that has been configured, no user can change it. Also if after some time of using the correct password, if you log out, it will request the password again.

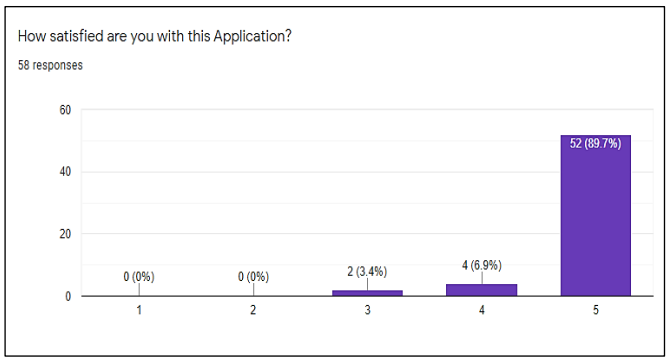

Figure 7. Feedback received based on the application

From figure 7, 89.7% of the people who tested the software were satisfied with the result they got; they chose excellent (i.e. rank 5). The graph showed that 6.9% of the users believed it was very good, while 3.4% believed it's just good. The highest and most satisfactory result which is "excellent," indicated that almost all the people that tested the software can implement it on any desired document.

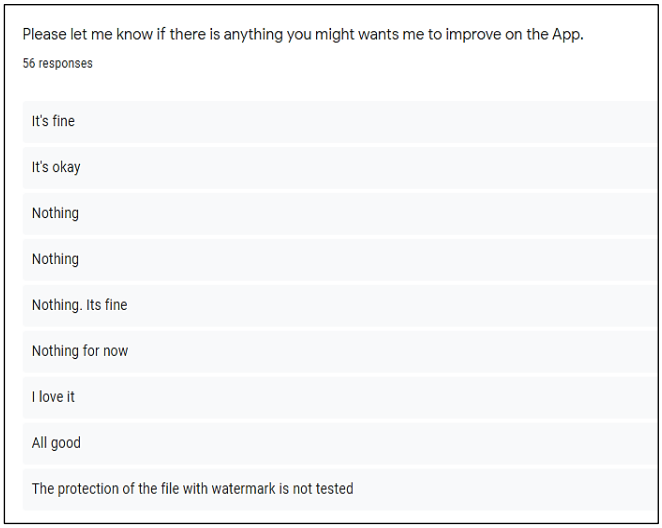

Figure 8. Feedback received on Future adaptive and correctional upgrade

Figure 8 showed that the software has successfully fulfilled its purpose, although some few suggested that since technology is not constant, one should expect a normal update with time. Also, figure 9 presents a dummy page that portrays how a blank PDF page will look after watermarked by the software application.

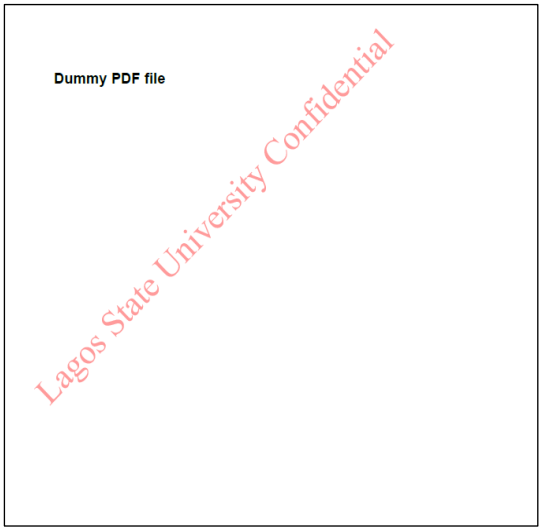

Figure 9. Dummy page of a watermarked page using this program

Figure 10 shows a digital content (i.e. a text-filled page) with watermark across it; the watermark did not obscure the message that is written in the digital file. This reflects the actual look and feel of any PDF document after processing through the proposed software application that implements a hybrid approach. In addition, this document cannot be copied, pasted or printed unless in possession of the master key used for the configuration. This key is different from the private key and it is required to open the document.

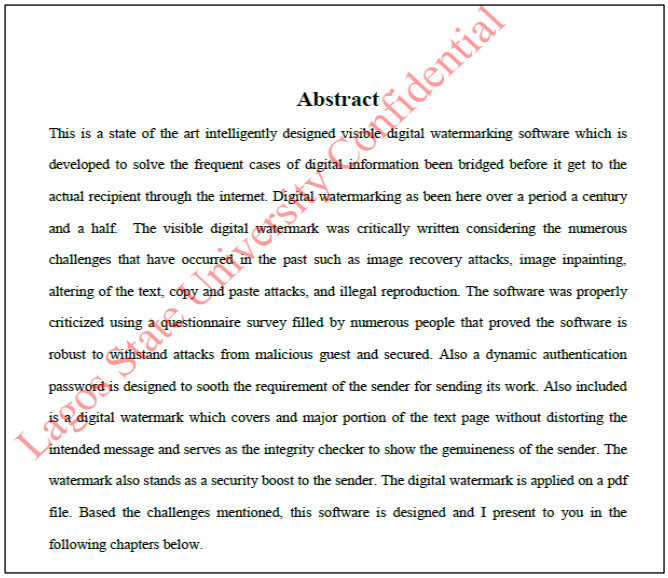

Figure 10. The watermarking of a page to show adequate integrity

V. CONCLUSION

Current research work demonstrates that an encoded document could be sent without its trustworthiness being undermined. The combined use of cryptography and watermarking can ensure the authenticity and integrity of a PDF document if processed with the software application proposed in this research. The implementation undertaken in this work ensures that the advanced watermark is planned at a point of 45 degrees; it takes a sensible piece of the picture text which further forestalls tampering by invaders or hackers. The watermark doesn't darken the content of the PDF document; consequently, it tends to be noticeable, an advantage of utilizing visible digital watermarking. Also, the advance encryption standard (AES) deployed in this work utilizes 128bits key; thus, it prevents brute force and makes document hacking to be impossible or at least difficult. It gives the system a 10 rounds key strength.

However, the feedbacks from the trial runs suggest that there are rooms for improvement of the application to further establish the security of any PDF document. Also, the availability of rich security-related literatures, provides the window to explore different advanced watermarking techniques in combination with cryptography schemes to further secure documents. Further work could be undertaken to utilize the proposed application to secure multimedia contents (that incorporate picture, video, and sound document). Tracking mechanism can be introduced to show whether a document have been opened or not. Also, 192 or 256 bits keys with 12 and 14 rounds respectively for more secured encryption could be explored.

ACKNOWLEDGEMENT

This is to formally recognize colleagues that provided support during the course of this research work. Also, acknowledged are family members for their sacrifices during the tough period.

APPENDIX

```
using iDiTect.Pdf.IO;
using PdfSharp.Pdf;
using PdfSharp.Pdf.IO;
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Configuration;
using System.Data;
using System.Diagnostics;
using System.Drawing;
using System.IO;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
namespace PDFProtector
{
   public partial class Form1 : Form
   {
      string filename = string.Empty;
      public Form1()
      {
         InitializeComponent();
      }
         private void textBox2_TextChanged(object sender, EventArgs e)
      {
       }
         private void panel1_Paint(object sender, PaintEventArgs e)
      {
        }
         private void button1_Click(object sender, EventArgs e)
      {
         // var openFileDialog = new OpenFileDialog();
         openFileDialog1.Title = "Select A file";
         openFileDialog1.Filter = "Pdf Files (*.pdf)|*.pdf";
         var dialogResult = openFileDialog1.ShowDialog();
      //var dialogResult = openFileDialog1.ShowDialog();
         if (dialogResult == DialogResult.OK)
         {
            label1.Text = openFileDialog1.FileName;
            filename = System.IO.Path.GetFileName(openFileDialog1.FileName);
         }
      }
         private void button2_Click(object sender, EventArgs e)
      {
         if (string.IsNullOrEmpty(label1.Text))
         {
            MessageBox.Show("Please select a file", "Warning", MessageBoxButtons.OK);
            return;
         }
         if (string.IsNullOrEmpty(textBox1.Text))
        {
            MessageBox.Show("Please enter password", "Warning", MessageBoxButtons.OK);
            return;
         }
         string destinationfolder = ConfigurationManager.AppSettings["DestinatonPath"].ToString();
         if (checkBox1.Checked)
         {
            AppliyWaterMark(label1.Text);
         }
            protectDoc(label1.Text, filename, password: textBox1.Text);
         }
          private void protectDoc(string source, string fileDest, string password)
        {
            // Get a fresh copy of the sample PDF file
            //const string filenameSource = "HelloWorld.pdf";
            string filenameDest = ConfigurationManager.AppSettings["DestinatonPath"].ToString();
            File.Copy(Path.Combine(source),
              Path.Combine(filenameDest, fileDest), true);
       // Open an existing document. Providing an unrequired password is ignored.
            PdfDocument document = PdfReader.Open($"{filenameDest}{fileDest}", "some text");
            PdfSharp.Pdf.Security.PdfSecuritySettings securitySettings = document.SecuritySettings;
            // Setting one of the passwords automatically sets the security level to
            // PdfDocumentSecurityLevel.Encrypted128Bit.
            securitySettings.UserPassword = "tesco";
            securitySettings.OwnerPassword = password;
            // Don't use 40 bit encryption unless needed for compatibility
            //securitySettings.DocumentSecurityLevel = PdfDocumentSecurityLevel.Encrypted40Bit;
            // Restrict some rights.
            securitySettings.PermitAccessibilityExtractContent = false;
            securitySettings.PermitAnnotations = false;
            securitySettings.PermitAssembleDocument = false;
            securitySettings.PermitExtractContent = false;
            securitySettings.PermitFormsFill = true;
            securitySettings.PermitFullQualityPrint = false;
            securitySettings.PermitModifyDocument = true;
            securitySettings.PermitPrint = false;
           // Save the document...
            document.Save($"{filenameDest}{fileDest}");
            // ...and start a viewer.
            Process.Start($"{filenameDest}{fileDest}");
        }
          private void AppliyWaterMark(string filePath)
        {
          iDiTect.Pdf.IO.PdfFile  pdfFile = new PdfFile();
           iDiTect.Pdf.PdfDocument document = pdfFile.Import(File.ReadAllBytes(filePath));
            //Get first page of pdf
            iDiTect.Pdf.PdfPage page = document.Pages[0];
            PageContentBuilder builder = new PageContentBuilder(page);
            //Create a block with watermark text
```

```
            iDiTect.Pdf.Editing.Block block = new Block();
            //Set text color and font size
            block.GraphicState.FillColor = new RgbColor(100, 255, 0,
0);
            block.TextState.FontSize = 50;
            //Set a text format.
            block.HorizontalAlignment                              =
iDiTect.Pdf.Editing.Flow.HorizontalAlignment.Center;
            block.VerticalAlignment                                =
iDiTect.Pdf.Editing.Flow.VerticalAlignment.Center;
            block.InsertText("Lagos State University");
            System.Windows.Size needSize = block.Measure();
            //Set watermark text position and rotation
            builder.Position.Translate((page.Size.Width                -
needSize.Width) / 2, (page.Size.Height - needSize.Height) / 2);
            builder.Position.Rotate(-45);
            builder.DrawBlock(block);
            File.WriteAllBytes(filePath, pdfFile.Export(document));
        }
    } }
```

REFERENCE

[1] Olson, D. A. (2019). Faxed: The Rise and Fall of the Fax Machine. By Jonathan Coopersmith.

[2] Peterson, A. M. (1991). Facsimile (fax) technology. *Hospital pharmacy*, *26*(2), 110-2.

[3] Orman, H. (2015). Encrypted Email: The History and Technology of Message Privacy. Springer International Publishing.

[4] Sharma, S., & Gupta, Y. (2017). Study on cryptography and techniques. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2(1).

[5] Goyal, S. (2012). A Survey on the Applications of Cryptography. *International Journal of Science and Technology*, *1*(3).

[6] Damghani, H., Hosseinian, H., & Damghani, L. (2019, May). Cryptography review in IoT. In *2019 4th Conference on Technology In Electrical and Computer Engineering (ETECH2019)*.

[7] Rijmen, V., & Daemen, J. (2001). Advanced encryption standard. Proceedings of Federal Information Processing Standards Publications, National Institute of Standards and Technology, 19-22.

[8] Chawla, G., Saini, R., & Yadav, R. (2012). Classification of watermarking based upon various parameters. *International Journal of Computer Applications & Information Technology*, *1*(II).

[9] Saini, L. K., & Shrivastava, V. (2014). A survey of digital watermarking techniques and its applications. *arXiv preprint arXiv:1407.4735*.

[10] Ghantasala, G. P. A Study on Features, Types, Applications and Techniques of Digital Image Watermarking.

[11] Thanki, R., Kothari, A., & Trivedi, D. (2019). Hybrid and blind watermarking scheme in DCuT–RDWT domain. *Journal of Information Security and Applications*, *46*, 231-249.

[12] Tao, B., & Dickinson, B. (1997, April). Adaptive watermarking in the DCT domain. In *1997 IEEE International conference on acoustics, speech, and signal processing* (Vol. 4, pp. 2985-2988). IEEE.

[13] Fang, H., Zhou, H., Ma, Z., Zhang, W., & Yu, N. (2019). A robust image watermarking scheme in DCT domain based on adaptive texture direction quantization. *Multimedia Tools and Applications*, *78*(7), 8075-8089.

[14] Taleby Ahvanooey, M., Li, Q., Shim, H. J., & Huang, Y. (2018). A comparative analysis of information hiding techniques for copyright protection of text documents. *Security and Communication Networks*, *2018*.

[15] Sharma, M. H., MithleshArya, M., & Goyal, M. D. (2013). Secure image hiding algorithm using cryptography and steganography. *IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN*, 2278-0661.

[16] Varshney, Y. (2017). Attacks on digital watermarks: classification, implications, benchmarks. *Int J Emerg Technol (Special Issue NCETST-2017)*, *8*(1), 229-235.

[17] Aru, O. E. & Ananaba, E.C. (2019) "Detailed Examination of Information Hiding Techniques for Copyright Protection of Text Documents" IOSR Journal of Applied Physics (IOSR-JAP) e-ISSN: 2278-4861.Volume 11, Issue 5 Ser. II, PP 51-61 www.iosrjournals.org.

[18] Saba, T., Rehman, A., & Elarbi-Boudihir, M. (2014). Methods and strategies on off-line cursive touched characters segmentation: a directional review. *Artificial Intelligence Review*, *42*(4), 1047-1066.

[19] Zeng, F., & Deng, X. (2012). Reversible visible image watermarking: Model evaluation and application. *International Journal of Advancements in Computing Technology*, *4*(10), 118-124.

[20] Cheng, Y., Zhang, J., Gong, X., Wan, H., & Liu, X. (2014, November). Research on polymorphism and inertial reading application in text watermarking algorithm. In *2014 Ninth International Conference on Broadband and Wireless Computing, Communication and Applications* (pp. 89-95). IEEE.

[21] Zhang, S. R., Yao, Z., Meng, X. C., & Liu, C. C. (2014, June). New digital text watermarking algorithm based on new-defined characters. In *2014 International symposium on computer, consumer and control* (pp. 713-716). IEEE.).

[22] Sinchita, B., Roy T., Sriya Aishwarya, T., & Sinha (2018). *Digital Watermarking In Image Processing Using Python*, Department Of Electronics & Communication Engineering Rcc Institute Of Information Technology Affiliated To Maulana Abul Kalam Azad University Of Technology, West Bengal Canal South Road, Beliaghata, Kolkata – 700015, 2018.

How to Cite this Article:

Adenowo, A. A., Adedoyin, M. A. & Adeyemi, S. A. (2021). Securing PDF Documents Using Cryptography and Digital Watermarking Techniques. International Journal of Science and Engineering Investigations (IJSEI), 10(111), 58-64. http://www.ijsei.com/papers/ijsei-1011121-08.pdf