DOI: 10.24996/ijs.2023.64.3.31

ISSN: 0067-2904

Constructing a Digital Certificate Authentication System for Classified Documents

Saif Ahmed Shaker[1], Ayad Ghazi Nasir[2], Faez Hassan Ali[3]

[1]Iraqi Commission for Computer and Informatics/Informatics Institute for Postgraduate Studies
[2] Ministry of Education, Iraq
[3] Mustansiriyah University, College of Science, Math. Dept., Iraq.

Received: 28/6/2022 Accepted: 6/8/2022 Published: 30/3/2023

Abstract

In this paper, we introduce a new digital authentication certification system to keep the classified documents' information safe. The proposed system is a steganography system divided into two subsystems, the first subsystem is responsible for embedding the information about the person, and it works in the main foundation that issues the documents, while the second subsystem is found in the beneficiary directorates to extract the true information of the person.

Keywords: Steganography, Steganalysis, Digital images.

بناء نظام مصادقة الشهادات الرقمية للمستندات المصنفة

سيف احمد شاكر[1]*, اياد غازي ناصر[2], فائز حسن علي[3]

[1]الحاسبات, معهد المعلوماتية للدراسات العليا, الهيئة العراقية للحاسبات والمعلوماتية, بغداد, العراق
[2] الرياضيات, العلوم, وزارة التربية, بغداد, العراق
[3]الرياضيات, العلوم, الجامعة المستنصرية, بغداد, العراق

الخلاصة

في هذا البحث ، سنقدم نظامًا جديدًا لإصدار شهادات المصادقة الرقمية للحفاظ على معلومات المستندات المصنفة في أمان. النظام المقترح هو نظام إخفاء المعلومات مقسم إلى نظامين فرعيين ، النظام الفرعي الأول مسؤول عن تضمين المعلومات حول الشخص ويعمل في المؤسسة الرئيسية التي تصدر المستندات ، بينما يوجد النظام الفرعي الثاني في المديريات المستفيدة لاستخراج المعلومات الصحيحة الشخص .

1. Introduction

Because electronic transactions involving electronic documents are conducted without physical touch and without face-to-face interaction. It is challenging to verify the counterpart's identity and the sincerity of the transaction's intention, and there is a risk of e-document misuse by fraudsters. It is not just simple to falsify or fake electronic documents throughout the procedure for distribution, and it's tough to show who created the document. However, it is also challenging to keep the contents of the sent data are kept a secret [1].

*Email: ms202010628@iips.icci.edu.iq

Information hiding is a class of procedures for embedding data into a variety of media formats, including text, image, and audio based on the RGB scale levels of pixel values. A human observer should be unable to see the embedded data. The information concealing mechanism, according to Provo, eliminates unnecessary bits from the medium coverage. Redundant bits are ones that can be altered without compromising the cover integrity medium. After that, the embedding mechanism chooses which bits will be replaced with data from the hidden message. Johnson et al. have shown that there are a variety of techniques for concealing information in photographs. One of these techniques is to keep extra information hidden in spare space in file headers. The manipulation of compression algorithms, as well as the alteration of carrier attributes such as brightness, contrast, or colours, are all examples of embedding strategies (this is depended on the type of image) [2].

The hieroglyphic language, which consists of a series of symbols that indicate a message, was used by the ancient Egyptians to communicate secretly. Although the message appears to be a sketch of an image, it may include a concealed message. Only a legitimate individual who knows what to look for could detect concealed information in Hieroglyphic. Following the Egyptians and the Greeks steganography was used, or "hidden writing," from which the term "steganography" was derived (Kipper, 2004; Johnson and Katzenbeisser, 1999). In general, steganography techniques conceal a message under a cover, such as a text, image, or audio file in a style that appears benign and hence does not arouse suspicion. The basic goal of steganography is to prevent an opponent from discovering secret communications, instead of preventing an adversary from decoding a secret communication (Johnson and Katzenbeisser, 1999). As a result, whether a plaintext is leaked or not. The goal of steganography is defeated if suspicion is generated when using any steganographic technique (Kessler, 2004; Martin et al., 2005). Text, image, audio, and other steganographic cover types are the most common categories used in today's schemes [3].

In the work published in 2011, Kekre et al (2011) suggested a steganalysis method based on advantages generated from an image's co-occurrence matrix. For categorization, two different distance metrics are used: absolute distance and Euclidean distance. For LSB concealing, this approach outperforms earlier steganalysis efforts. It can be used with both grayscale and colour images. In colour photos, Euclidean distances outperform Absolute distance by 265 percent, and in grayscale images, they outperform Absolute distance by roughly 329 percent. Colour images have a detection accuracy that is almost 18 percent greater than grayscale images. In the absolute distance, they are virtually same, and in Euclidean distance, they are nearly identical. When embedding rates are low, there is a sense of supremacy. When compared to feature vectors without the diagonal d0, feature vectors with the diagonal d0 yield poor results.

(Islam, Siddiqa, Uddin, Mandal, and Hossain, 2014) suggested a filtering-based algorithm that employs LSB in a unique method. Instead of secret data bits, the technique embeds hints regarding whether or not a pixel includes concealed data in the LSBs. First, they determine if the lighter or darker pixels are more numerous. Because a pixel is made up of three colours, each of which is represented by a single byte, pixels with two- or three-bit values of 1 are regarded as lighter. As a result, darker pixels have two or three MSBs with a value of zero. The method hides the secret info in the darker and lighter pixels with a higher count. Furthermore, only those pixels that match a requirement that is part of the hiding procedure are used to contain hidden data. The algorithm calculates the MSBs of each pixel's three colours’ decimal value Pn. Pn would be a number between 0 and 7. The LSB of the third byte is set to 1 as an indicator that this byte includes a secret bit if the value of the bit with an index equal to Pn equals the secret bit inside the third byte of the pixel; otherwise, the LSB is set to 0.

Salih and Al-Jarrah (2015) published a study that provided many steganography approaches to assure secure internet use; nevertheless, these methods cannot yet check the presence of attacks in hidden messages. As a result, this paper describes the invention of Bi-LSB, an enhanced (LSB) technique that addresses the classic LSB techniques' low security and capacity issues. The suggested technique is assessed by first inserting an AWGN into the stego-file before extracting the embedded messages to see how it affects the PSNR values, and then comparing the extracted message with the original one to check for integrity. The results reveal that when the AWGN assault is included, the PSNR values drop significantly. The goal of image steganography is to conceal data within image files (Gowda and Sulakhe, 2016). LSB-based approaches for hiding data within photos have been proposed by a number of researchers [4].

Using fuzzy identity-based encryption, Xu and Nie (2018) [5] devised a time-domain identity-based technique that allowed multiple sensitive data to be merged in a cover-image for different individuals. Several attributes were allocated to the various messages, and the extraction process can only extract the appropriate information if the appropriate attribute is used. Although the method allows for hierarchical retrieval of information in steganography, embedding both traits and information lowers the system's capacity to keep data (IWT).

Valandar et al. (2019) [6] introduced a steganographic system based on a 3D sine chaotic map and “integer wavelet transform” (IWT). The IWT broke down the colour picture into four sub-bands. Each subband was separated into non-overlapping 16 by 16 blocks, which served as embeddable zones for sensitive data. The sensitive information zones are determined via chaotic maps.

With the goal of achieving a high payload with less distortion on the stegoimage, Kadhim et al., 2020 [7] in the “dual-tree complex wavelet transform”, we adopted an indirect concealment method. For categorizing favourable photos with expected high payload as cover-image, a K-nearest neighbour machine learning model was used. Using a 100-image dataset for training, the images were classed as textured or smooth. The dual-tree complex wavelet coefficient held the hidden information. With a proposed security solution to mitigate steganalyst assaults, the system attained a PSNR of 50.55dB and an MSE of 0.1810, according to the experimented results. The effectiveness of the transform domain used to translate the stego-image into its coefficient is equally important in achieving a high payload. This is because, among other things, a real-time program interacts with a variety of images compressed using different algorithms, variations in light intensity, environmental conditions, texture, and noise from the acquisition source. As a result, the machine learning dataset must be expanded in order to learn all of the possible images for the steganography system, which may or may not be possible.

The important points of this study are listed below: The second portion includes the image file format and the most significant models. With steganography and steganalysis, information concealment is described in section three. The problem statement, which will be covered in this paper, is addressed in paragraph four. In paragraph five, the suggested authentication scheme is discussed. Finally, we examine the most relevant conclusions and recommendations in section six.

2. Image Files [8]

A picture is described as a collection of integers, which typically represent varying light intensities in various areas of the image. The numerical representation is in the form of a lattice, with individual points referred to as "pixels." Row by row, pixels are shown horizontally. The number of bits assigned to each pixel in a colour scheme is referred to as bit depth. Furthermore, the colour scheme's smallest bit depth is 8, suggesting that each pixel's colour is represented by eight bits. 8 bits per pixel are used in both monochrome and grayscale images, Up to 256 different colours or shades of grey can be displayed using these bits. Another thing to keep in mind is that most digital colour images are saved in 24-bit formats and employ the RGB colour model. “Almost every colour variation in a 24-bit image is derived from three basic colour terms: red, green, and blue, each of which is represented by eight bits”. Thus, the number of various shades of red, green, and blue in each given pixel can approach 256. There are over 16 million colour combinations, resulting in over 16 million hues. “The graphics interchange format (GIF), joint photographic experts' group (JPEG), and, to a lesser extent, the portable network graphics (PNG) format are the most popular picture formats used only on the internet”. The crucial point to mention here is that the most steganographic algorithms try to take advantage of the structure of these formats. However, due to its simplicity and uncomplicated data structure, certain literary contributions use the bitmap format (BMP). We will look at PNG and BMP images in this paper.

2.1 Portable Networks Graphics (PNG) Images [9]

The newly designed “portable networks graphics (PNG)” is versatile and has a number of advantages over previous Internet standard photo file formats that makes it an appealing option for digital teaching resources. The PNG format allows for lossless compression, as well as gamma and chronicity correction, so files can be opened, edited and saved frequently. The two-dimensional interlacing characteristics of the PNG format allow a picture to fill in from top to bottom and right to left, allowing for faster retrieval than other formats. Additionally, images can be seen in their original settings, and metadata (data-related information) can be added to files. The PNG format includes a network-friendly, patent-free, lossless compression approach that is really cross-platform, as well as a number of other advantages for multimedia and Web-based radiologic teaching. The overwhelming support of PNG by the World Wide Web (www) Consortium, the leading Web browsers and visual manipulation software companies suggest that it will play a larger role in multimedia educational file generation in the future.

2.2 Bitmap (BMP) Images [10]

Only on the Windows platform is the “Bitmap (BMP) file format”. Bitmap graphics are created with this format. Unlike other file formats, which store image data from top to bottom and pixels in Red, Green, Blue order, the BMP format stores picture data from bottom to top and pixels in Blue, Green, Red order. Because of this, BMP images may appear to be drawn from the bottom up if memory is restricted. The BMP files are often big since they do not support compression. When saving a file in the BMP format, we add the "BMP" file extension to the end of the file name.

3. Information Hiding [11]

A new type of covert communication technology is information concealment technique. Multimedia items such as audio are used in the bulk of today's information-hiding systems. It is usually more difficult to encode secret messages in digital sound.

3.1 Steganography [12]

Steganography is the study of encoding hidden data in a suitable multimedia carrier, such as an image, audio, or video file. It is based on the concept that if the feature is visible, the point of attack is obvious, hence the goal is always to hide the fact that the embedded data exists. Steganography can be used for a variety of purposes. However, it, like any other science, can be misused for nefarious purposes. It has risen to the forefront of current security tactics as a result of significant increases in processing capacity, increased security awareness among individuals, groups, agencies, and government, and intellectual endeavor. The major qualities that distinguish steganography from comparable techniques like watermarking and encryption are undetectability, robustness (resistance to various image processing methods and compression), and capacity of the hidden data.

3.2 Steganalysis [13]

Steganography is the polar opposite of steganalysis. We begin by attempting to detect steganographic content in a digital device, and then we try to figure out what the hidden message is. From this perspective. There are two types of steganalysis: passive and aggressive steganalysis. The goal of passive steganalysis is to determine the steganographic embedding algorithm and categorize a cover medium as stego, whereas active steganalysis aims to calculate the length of the embedded message and, preferably, to extract it from the cover medium.

Depending on the information accessible to the steganalyst, attacks can take a variety of forms:

1. In a stego-only attack, just the stego-object is available for analysis. Only the stego-carrier and the stego-secret carrier's information for example, are available.
2. Known attack’s cover: There are pattern differences between the cover‘s original object and the object‘s stego. The image’s original and the image with hidden data, for example, are both available and may be compared.

3.3 Steganalytic Methods

It is feasible to construct an image’s stego that within the confines of human perception, it appears to be the same by carefully selecting a suitable image’s cover and tool’s stego. Each of these instruments, however, When utilized electronically, it leaves a “fingerprint or signature” in the image that can be used to alert a viewer to the presence of a hidden message. We will go through visual and statistical analysis in more detail below.

3.3.1 Visual Analysis

The visual assault is a just attack’s stego that takes advantage of the vast majority of steganography application makers' assumption that the “least significant bits” of a file’s cover is random. Depending on a person to determine whether or not a picture presented by a filtering system contains concealed data. The image is filtered to remove elements of the image that obscure the message. The filtering technique produces a photo that just contains the bits that could potentially be utilized to incorporate data. The steganographic embedding function that is assessed determines how the prospective stego picture is filtered. However, because most embedding functions are similar, adapting an existing filtering method to a different steganographic embedding function usually only requires minor changes.

3.3.2 Statistical Analysis

Visual attacks have two significant disadvantages. When a large number of photographs need to be studied, it takes a long time or costs a lot of money because each photo should be filtered, shown and examined by humans. Another significant disadvantage is that the LSBs of some (unaltered) pictures may include random-looking data. The visual attack will fail if such a photo is utilized as the file’s cover. Statistical attacks take advantage of the fact that most steganography applications regard the cover file's They treat LSBs as random data, expecting to be able to overwrite these bits with other random information, comparable to visual assaults (the encrypted secret message). The LSBs of an image are not random, as it is shown by the visual attacks. When a steganography program embeds a bit by overwriting it, the LSBs of a pixel are altered to an adjacent colour value in the palette (or in the RGB cube if the cover file is a true-colour image). These simple tests cannot determine whether or not an image contains a hidden message.

4. Problem Statement

Some students who intend to work in government institutions linked with Iraqi state ministries must bring a sixth-stage graduation certificate to show their graduation, as we are aware. The Iraqi Ministry of Education normally issues this document, which is presented to the student on paper. Of course, some pupils might tamper with or falsify the document. The point is, how can you be certain that this paper is not a forgery?

The competent authorities have taken steps to confirm that the graduation document (which is designated as having a high level of confidentiality) has not been tampered with by writing to the Iraqi Ministry of Education to validate the document's issue. This approach has a number of drawbacks, including:

1. It may not prevent the document from being falsified.
2. As is customary, these correspondences take a long time to complete, which slows down the process.

As a result, an electronic system based on the Internet has been proposed to address these issues by immediately confirming the authenticity of graduation document information and ensuring that it has not been tampered with while maintaining a high rate of completion.

5. Constructing of Digital Certification Authentication System

We will offer a new design of digital certification Authentication system (DCAS) in this work, which consists of two DCASs, the first of which is built for embedding information and the second of which is designed for extracting information. The proposed electronic system (DCAS) is essentially a steganography system that is specialized in embedding important student information. This information is embedded in the student certification document image (CDI), and the certification office (which is specialized in sending real certification documents to other ministries) sends this stego-image of the certification to the beneficiary directorate to ensure that the paper certification document is not a forgery. As a result, the DCAS is separated into two stego-subsystems: the first is for embedding information and is known as DCAS-Embedding (DCAS-Em). This system is mainly only used in the ministry of education's certification office. The DCAS-Extracting (DCAS-Ex), which is employed at the specialized office in the beneficiary directorate, is the second stego-subsystem, which is specialized for extracting hidden information.

Naturally, the DCAS requires an efficient key generator to embed the information in the stego-image at random, thus we use the Efficient Stream Key Generator (ESKG) from [14]. The proposed DCAS is a strategy information security system, which means that the system has achieved its objectives when the information reaches the recipient before the attacker can breach the system and figure out what the information is?.

5.1 DCAS Procedures

When a student completes an official transaction to obtain a graduate certificate in order to present it to a beneficiary directorate. A paper graduating document will be provided by the minister of education. The following are the DCAS office procedures that will begin immediately:

1. The Document Issuance Department (DID) will transfer a copy of the student's paper

document to the Ministry of Education's Document Inspection Office (DIO) when someone requests a graduation document from the Ministry of Education, which is normally addressed to a specific ministry.

2. The DIO scans the student's certification paper with a scanner device and converts it to a digital image, which is then saved in a secure computer.

3. The DCAS-Em system will be used to embed the following student information in the digital image of the certification document: student's full name, serial number, general directorate, school name, grades, average, organizer staff name, and date.

4. After the competent person has finished embedding information in the CDI, the stego-image will be sent to the recipient directorate.

5. When the CDI is received by the beneficiary directorate, a competent employee will use the DCAS-Ex system to extract the student's concealed information and match it to the information stated on the paper document. If the information recovered from the stego-image matches the information derived from the document paper, the graduation certification certificate is legal and not forged; otherwise, legal action will be taken against the person who fabricated it.

5.2 DCAS-Em System

The DCAS-Em is designed to conceal student information within a document picture. When this system is used by a competent person in the DIO, the system asks for the name of the ministry to which the photograph should be sent. The CDI and student information are then required to hide this information in the CDI. The recipient ministry receives the stego CDI (SCDI). The DCAS-Em system is depicted in Figure 1 as a block diagram.

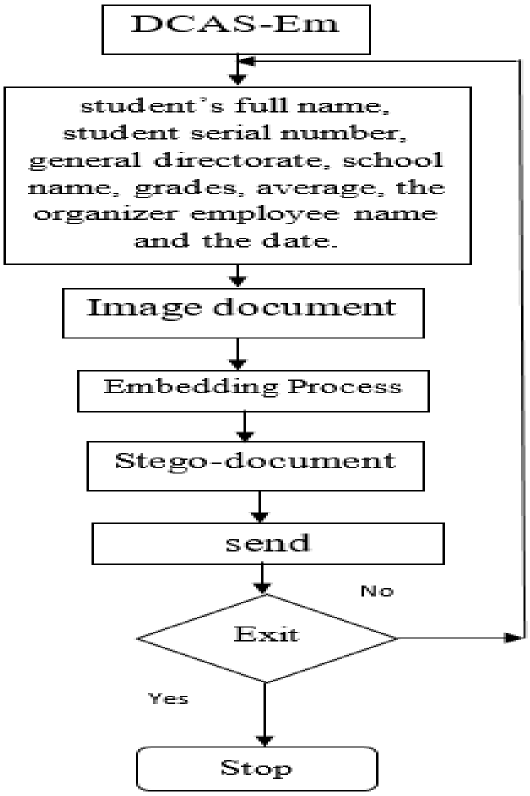
DCAS-Em
student's full name, student serial number, general directorate, school name, grades, average, the organizer employee name and the date.
Image document
Embedding Process
Stego-document
send
No
Exit
Yes
Stop

Figure 1: The diagram's block system of the DCAS-Em.

The algorithm of the DCAS-Embedding is as follows :

```
DCAS-Embedding Algorithm
Step (1): INPUT: Student Information; CDI;
Step (2): CALL ESKG;
Step (3): Embedding Process
Step (4): OUTPUT: Stego-Image (SCDI);
Step (5): END.
```

5.3 DCAS-Ex System

The DCAS-Ex will be completed by a competent employee after obtaining the CDI from the DIO, which contains the person's information. The CDI is required for this system, after which the information is shown on the computer screen. This information will be compared to a paper document, and a competent employee will determine whether or not the certification document paper is phony. The DCAS-Ex system is depicted in Figure (2) as a block diagram.

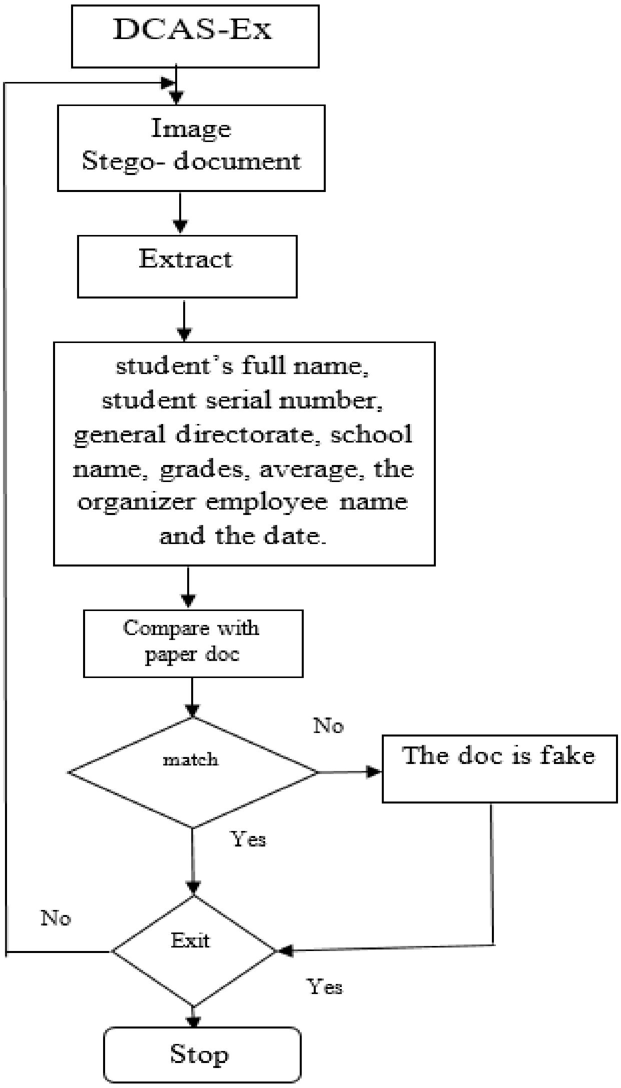
DCAS-Ex
Image
Stego- document
Extract
student's full name, student serial number, general directorate, school name, grades, average, the organizer employee name and the date.
Compare with paper doc
match
No
The doc is fake
Yes
No
Exit
Yes
Stop

Figure 2: The diagram's block system of the DCAS-Ex.

The algorithm of the DCAS-EX is as follows:

DCAS-EX Algorithm

```
Step (1): INPUT: SCDI;
Step (2): CALL ESKG;
Step (3): Extracting Process;
Step (4): Student Information;
Step (5): Compare the Student Information with paper document information;
Step (6): IF Matching the information is legal
          ELSE the information is faked;
Step (7): END.
```

5.4 DCAS Key Management

As we all know, to control the sending and receiving of CDI, any security system requires effective key management. In reality, we can consider the ministry of education to be the center of the DCAS, and the other ministries to be the stations that are tied to the center when it comes to examining the authenticity of certificates. Of course, DCAS is a one-way process (the center will only embed data, while the stations will only extract data), which means the center will send data and the stations will only get data.

Assume that we have n stations (ministries), and we will assign just one key to each document, so we will assign $k_i, i = 1,2, \dots, n$ different keys for each station. When executing the DCAS_Em it requires the name of the ministry, say m_i (station i) then the system will open the m_i Basic Key (BK) file and take the key number j_i, where $j_i = 1,2, \dots, k_i$, then the document number j_i will be processed and sent to the ministry m_i. When the ministry m_i receives the document j_i, the DCAS_Ex will process this document using it j_i key. Of course, there is no connection between the stations with each other. Figure 3 shows the key management of the DCAS.

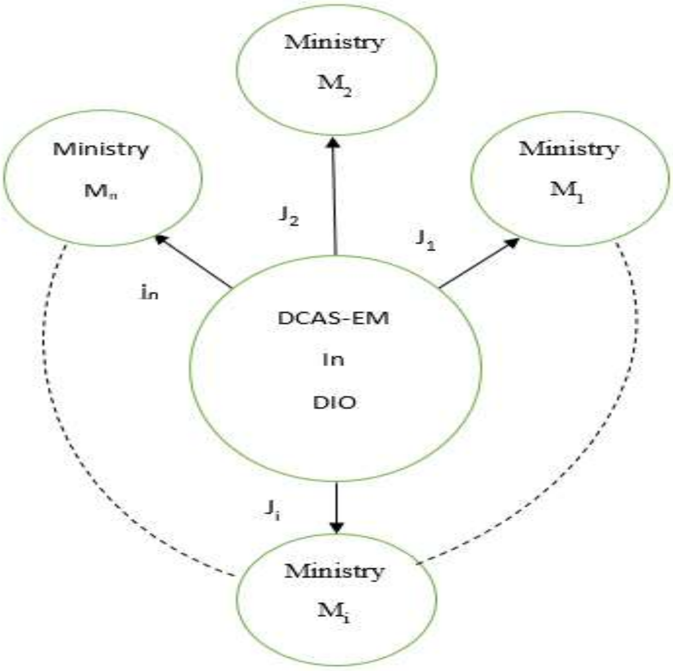

Figure 3 : key management of DCAS

6. Conclusions

We can draw the following conclusions and recommendations from this paper:

1. We use different fundamental keys for each ministry offers several advantages, including ensuring system security by avoiding the use of duplicate keys. We also make certain that

the system's security is maintained in the event that one of those keys is compromised in any ministry. The treatment is to change the ministry's keys.

2. Physical and software protection for the DCAS can be provided to safeguard its security against penetration and to limit the circle of responsibility when a system breach is established.
3. The suggested DCAS is cost-effective, as it only requires one or two users with a bachelor's degree in a scientific area such as computer science or mathematics, and they have great confidence and a strong sense of security with a personal computer in a remote office.
4. We recommend that Ministry of Education (or Ministry of Higher Education and Scientific Research) professionals use this system to preserve certified documents and expedite the delivery of documents to the ministries and directorates responsible for this issue.
5. The DCAS can be developed to work in all state departments that accept the principle of correct issuance of any official order or document in order to safeguard information from tampering and forgery in general.

References

[1] N H M. Ali, A M S. Rahma, A S Jamil, "Text Hiding in Color Images Using the Secret Key Transformation Function in GF (2n)", Iraqi Journal of Science, 2015, Vol 56, No.4B, pp: 3240-3245.

[2] A Talib," Corners-based Image Information Hiding Method", Vol. 43 No. 1,(2017): Iraqi Journal for Computers and Informatics .

[3] Z T R AL-Windawi, "Security Enhancement of Image Steganography Using Embedded Integrity Features", Middle East University May, 2017.

[4] X Xu, Q Nie, "Identity-Based Steganography in Spatial Domain", Journal of Computer and Communications, Vol.06 No.03(2018), Article ID:83242,10 pages.

[5] B Y Irani, P Ayubi, F A Jabalkandi, M Y Valandar, M J Barani," Digital image scrambling based on a new one-dimensional coupled Sine map" , Young Researchers and Elite Club, Urmia Branch, Islamic Azad University, Urmia, Iran, ,2019.

[6] I J Kadhim, P Premaratne, P J Vial, 2019, "High Capacity Adaptive Image Steganography with Cover Region Selection using Dual – Tree Complex Wavelet Transform", Journal Pre-proofs, Research (2019), doi: https://doi.org/10.1016/j.cogsys.2019.11.002.

[7] N Hamid, A Yahya, R B Ahmad and O M. Al-Qershi," Image Steganography Techniques: An Overview", International Journal of Computer Science and Security (IJCSS), Volume (6): Issue (3): 2012, pp:168-187.

[8] A. S. Mahdi, A. H. Khidhir, M. A. Hussein, "Image in Image Steganography Based on DCT", Iraqi Journal of Science, 2014, Vol 55, No.4A, pp: 1675-1684.

[9] M Juneja and P S Sandhu, "Designing of Robust Image Steganography Technique Based on LSB Insertion", 2009 International Conference on Advances in Recent Technologies in Communication and Computing,pp:302-305.

[10] R I Yousif, and N H Salman, "Image Compression Based on Arithmetic Coding Algorithm ", Iraqi Journal of Science, 62(1), 329–334.

[11] A Cheddad, J Condell, K Curran and P M Kevitt, "Digital image steganography: Survey and analysis of current methods", Signal Processing 90 (2010) 727–752.

[12] K Karampidis, E Kavallieratoua and G Papadourakis, "A review of image steganalysis techniques for digital forensics", Journal of Information Security and Applications 40 (2018) 217–235.

[13] A G. Naser and F H. Ali, "Detecting the LSB-LFSR Algorithm Stego-Image Using the Statistical Analysis and Visual Attack", AL-Mansour Journal / No.17/ Special Issue 2012, pp:55-75.

[14] S A Shaker, A G Naser and F H Ali, "New Design of Efficient Non-Linear Stream Key Generator", 4th International Scientific Conference of Engineering Sciences and Advances Technologies, (2022), (IICESAT), 3–4 June, 2022, AIP Conference Proceedings (ISSN: 0094-243X, 1551-7616) (Accepted).