

Security and Integrity of documents by using novel algorithm of Stegnography

Dr. Mukesh Kumar, Head of Computer Science Department, Hindu College, Sonapat

Abstract— In Stegnography, one message is hidden inside another, without disclosing the existence of the hidden message, it apparent to an observer that this message contains a hidden message. Moreover, the information hidden by a watermarking system is always associated with the object to be protected or its owner while steganographic systems just hide information. On the other hand, cryptography can be defined as the study of secret writing (i.e. concealing (hide) the contents of a secret message by transforming the original message into a form that cannot be easily interpreted by an observer). The method considered here (diffusion and confusion) can be easily used in both applications. The hidden message can be transformed into a diffused form (i.e encrypted) and inserted into the background. The hidden information might have no relation with the text (foreground). At the same time, backgrounds are usually used with documents and so diffused data will not necessarily trigger the attention of an observer. Moreover, the hidden message is also encrypted which increases the security level of such documents.

In this publication the focus is authenticity and integrity of document by using Stegnography algorithms and methods that can be used to stop counterfeit, reproduction, forgery or tempering of social value documents.

Index Terms: Novel Algorithms for Security of Documents, Technology innovations in Document Security, Revolution in Stegnography,, Strength of Stegnography, Stegnography and corporates

INTRODUCTION

Before discuss any more on the topic it is necessary to know know about security in actual; Securities as a form of protections that provide reliability & improve security under impose of conditions. In our opinion the conditions which allow us to prevent from vulnerability & hazards and also safeguard the documents from unauthorized manipulation or unauthenticated access. We present the hypothetical design and implementation of facts that incorporates methods and algorithms to provide a high degree of safety and security while remaining will be fully transparent and easily usable.

Documents that contain sensitive or important information are often subject to fraud, fake and tampering. Single attempt toward security of documents throughout history people have tried to develop methods to secure the information during exchange. History people have used the Stegnography technique during delivery of secrete messages by using milk, juice, urine to write the message and will be darken after heated the message paper.

Due to the innovative and revolutionary changes in technology, document clone, tempering with data and counterfeiting of document also increasing day by day hence security is the main issue in the revolutionary electronic world. The revolution in technology are accountable for a wide range of industrial and commercial advances. Security algorithms have influenced certain industry standards that are constantly being analyzed, scrutinized and improved upon. So It is an innovative way and there will be always a scope to improve upon always.

1 PROBLEM DEScriptioIN

With the advent of technique and technologies, it is easier for a person to create a fake copy of the legal document very

easily. High resolution scanner and printer made people capable of preparing almost perfect copy of the document. So, we need a technique by which we can verify the authenticity or truthiness of the document or image. To refine Stenography and cryptography technique, we have embedded key and verify the genuineness / truthiness of document concern with high social value like financial, property or any legal documents. We have used some techniques / methods to provide security and to verify genuineness of the document in all respect. All world trade is in counterfeit goods, amounting to an annual value that exceeds 5000 crores. It is necessary to stop these vulnerabilities to detect and prevent counterfeiting.

Hand written signature is the most widely form of personal identification as well as for document verification, especially for cashing cheques, legal deeds, wills etc However, for several reasons the task of verifying human signature can't be considered for pattern recognition because signature samples from the same person may be similar but not identical. A person's signature often changes radically during their life. We can see much variability in signature according to their age, time and psychological or mental state.

So to remove all such vulnerabilities form the system to maintain the integrity of documents, we can use such methods that verify the legitimacy of the social value document.

2 MOTIVATION

It is well known that the tampering with information like scam, indication, alter etc. etc. or document fraud, such as currency note fraud, clone, counterfeit, cracking in the fruitful information. The main aim of the algorithm and methods are:

- High Level Integrity of document
- Cryptographic security do not depend on algorithm secrecy
- Adaptable to the diversified applications
- Economical implementation on hardware / software
- Efficient on high data transfer rate

3 LEGAL CONSIDERATION

Legal aspects are taken into account when considering technologies are used to detect reproduction, counterfeit and tempering of documents.

3.1 DEFINITION OF DIGITAL EVIDENCE

According to the Scientific Working Group on Digital Evidence (SWGDE), Digital Evidence is "information of probative value that is stored or transmitted in binary form". Therefore, according to this definition, evidence is not only limited to that found on computers but may also extend to include evidence on digital devices such as telecommunication or electronic multimedia devices. Furthermore, digital evidence is not only limited to traditional computer crimes such as hacking and intrusion, but also extends to include every crime like counterfeit, forgery, clone, tempering, misuse of intellectual property or social value documents etc. etc. in which digital evidence can be found.

3.2 LAW ON DIGITAL EVIDENCE

The expansion / propagation of computers, the social influence of information technology and the ability to store information in electronic form have all required Indian law to be amended to include provisions on the appreciation of digital evidence. In India the Information Technology Act 2008 has been revised to stop cyber crime activities of electronic commerce together with amendments to the Indian Evidence Act 1872, the Indian Penal Code 1860 and the Banker's Book Evidence Act 1891 provides the legislative framework for transactions in electronic world.

3.3 DIGITAL DOCUMENTS AS LEGAL EVIDENCE

It is well known that visual evidence (still image, documents) is often crucial for determining the result of testing. The most recent techniques in digital photography and video allow obtaining high-quality images that can be provided as valuable proofs, but there is reasonable doubt about the integrity and authenticity of such document / images. Of course, it is possible to tamper with images in analog format, but it is much easier to do it with digital images (nowadays, any average user has access to powerful editing tools). In fact it has been proved that digital renovate / remodel or repair techniques were used to cause to look guilty the innocents. Currently, it is enough to suggest that a digital image could have been maliciously modified in order to reject it as a valid proof, if there is no reliable mechanism for proving its authenticity than sys-

tem will be degraded and crime will be increased automatically

3.4 FORENSIC ANALYSIS OF DIGITAL DOCUMENT

Due to the simplicity for altering or tampering with digital images and social value documents, the work of the forensic scientist as a specialist in digital imagery becomes particularly relevant when using digital images as legal evidence. The forensic scientist establishes whether a certain image is authentic or not. The most important techniques for proving authenticity are hashing, steganography, CGH, data glyph and digital watermarking. Being awareness of these legal needs, many companies have incorporated these technological measures as an added value in their digital document authentication solutions. In spite of this, the introduction of watermarking technology in the legal machinery is not so easy. Watermarks may be seen by certain tools for authenticating digital images and documents.

At the embedding side: from the very instant that the watermark is embedded in the original content, a reasonable doubt may arise: is the watermarked image indeed the same as the original image? Obviously, the watermark must have been embedded in an invisible manner, but the algorithms will manipulate the images and any tempering will be exposed of being analyzed and certified.

At the decoding side: a serious concern is the need for the application of certain transformations to the image under analysis in order to properly decode the embedded watermark. Moreover, the answer of the watermark detector / decoder cannot be regarded as 100% reliable. There exist always no null, false negative and false positive probabilities, and potential decoding error, but the theoretical performance measures are not always representative of all real scenarios. In this regard it is interesting to note that, although watermarking technologies have experimented great advances during the last decade, in general they have not been tested at such extent. This being said, it seems that we still have a long way to go before watermarking technologies gain full legal acknowledgement.

4 STEGANOGRAPHY SECURITY METHOD

4.1 INTRODUCTION

Steganography has been used throughout history for secret communications. Criminals have always sought ways to conceal their activity in real or physical space. The same is true in virtual, or cyber space. The SARC (Steganography Analysis and Research Centre) has developed state-of-the-art steganography detection and extraction capabilities that address the needs of digital investigation specialists and information technology security personnel in law enforcement, government, military, intelligence, and the private sector. By providing a national repository of steganography application hash values, or fingerprints, and developing the most advanced

steganalysis tools, techniques, and procedures to find and extract hidden information, the SARC is rapidly evolving into a high-value asset to computer forensic examiners who wish to conduct steganalysis on seized media.



Steganalysis in Steganography Document

One of the principal weaknesses of all encryption systems is that the form of the output data (the cipher-text), if intercepted, alerts the intruder to the fact that the information being transmitted may have some importance and that it is therefore worth attacking and attempting to decrypt it. For example, if a post office worker observed a locked box passing through the post office; it would be natural for them to wonder what might be inside. It would also be natural to assume that the contents of the box would have a value in proportion with the strength of the box/lock. These aspects of cipher text transmission can be used to propagate disinformation, achieved by encrypting information that is specifically designed to be intercepted and decrypted. In this case, we assume that the intercept will be attacked, decrypted and the information retrieved. The key to this approach is to make sure that the cipher text is relatively strong (but not too strong!) and that the information extracted is good quality in terms of providing the attacker with 'intelligence' that is perceived to be valuable and compatible with their expectations, i.e. Information that reflects the concerns / interests of the individual(s) and / or organizations that encrypted the data. This approach provides the interceptor with a 'honey pot' designed to maximize their confidence especially when they have had to put a significant amount of work in to 'extracting it'. The trick is to make sure that this process is not too hard or too easy. 'Too hard' will defeat the object of the exercise as the attacker might give up; 'too easy', and the attacker will suspect a set-up!

In addition to providing an attacker with a honey-pot for the dissemination of disinformation it is of significant value if a method can be found that allows the real information to be transmitted by embedding it in non-sensitive information after (or otherwise) it has been encrypted, e.g. Hide the cipher text. This is known as Steganography which is con-

cerned with developing methods of writing hidden messages in such a way that no one apart from the intended recipient knows of the existence of the message in contrast to cryptography in which the existence of the message itself is not disguised but the content is obscured. This provides a significant advantage over cryptography alone is that messages do not attract attention to themselves, to messengers, or to recipients. No matter how well plaintext is encrypted (i.e. How unbreakable it is), by default, a cipher-text will arouse suspicion and may in itself be incriminating, as in some countries encryption is illegal. Steganography is equivalent to transforming the 'strong box' into some other object that will pass through without being noticed - an 'egg-box', for example.

The word 'Steganography' is of Greek origin and means 'covered', or 'hidden writing'. In general, a steganographic message appears as something else known as a cover text. By way of a simple illustrative example, suppose we want transmit the phrase

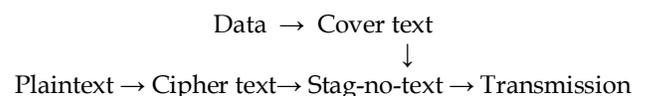
The quick brown fox
Which is encrypted to produce the cipher stream
syoahfsuyTebhsiaulemNG

This is clearly a scrambled version of a message with no apparent meaning to the order of the letters from which it is composed. Thus, it is typical of an intercept that might be attacked because of the very nature of its incomprehensibility. However, suppose that the cipher stream above could be re-cast to produce the phrase

Be caution of dogs

If this phrase is intercepted it may not be immediately obvious that there is alternative information associated with such an apparently innocuous message, i.e. If intercepted, it is not clear whether or not it is worth initiating an attack.

The conversion of a cipher text to another plaintext form is called Stegotext conversion and is based on the use of Cover text. Some cover text must first be invented and the cipher text mapped on to it in some way to produce the stegotext. This can involve the use of any attribute that is readily available such as letter size, spacing, typeface, or other characteristics of a cover text, manipulated in such a way as to carry a hidden message. The basic principle is given below:



Note that this approach does not necessarily require the use of plaintext to ciphertext conversion as illustrated above and that plaintext can be converted into stegotext directly. A simple approach to is to use a mask to delete all characters in a message except those that are to be read by the recipient of the message. For example, consider the following

host image without having to resort to quantization.

Steganography is often used for digital watermarking. This is where the plaintext, which acts as a simple identifier containing information such as ownership, copyright and so on, is hidden in an image so that its source can be tracked or verified. This is equivalent to hiding a 2-bit image in a host image as illustrated in Figure which uses the same method as discussed above. In this example, a columnar transposition cipher has been used to encrypt this sentence using the key-word: Steganography. This grid is given by

11	12	03	04	01	07	08	05	10	02	09	06
t	H	e		q	u	i	c	K		b	r
w	N		f	o	x		J	U	m	p	s
o	V	e	r	t	h	e		L	a	z	y
I	I	t	t	l	e		d	o	g		i
	T	h	i	s		t	e	X	t		t
e		k	e	y		w	o	R	d		i
u		e	d		i	s		S	t	e	g
n	s										a
	o	g	r	a	p	h	Y				

and the ciphertext is : qotlsy a _magtdt_ e_ _ethkeg_frtiedrcj_deo_yrsyitig_uxhe_ _ipi_e_twshbpz_ _ _e_kuloxrs_twol_eunhvit_soo_ _nhsa.

As in the previous example, the host image is required to recover the cipher-text information and is thus the key to the process. The methods discussed above refer to electronic-to-electronic type communications in which there is no loss of information. Steganography and watermarking techniques can be developed for hardcopy data which has a range of applications. These techniques have to be robust to the significant distortions generated by the printing and/or scanning process. A simple approach is to add information to a printed page that is difficult to see. For example, some modern colors laser printers, including those manufactured by HP and Xerox, print tiny yellow dots which are added to each page. The dots are barely visible and contain encoded printer serial numbers, date and time stamps. This facility provides a useful forensics tool for tracking the origins of a printed document which has only relatively recently been disclosed.

5 IMAGE AUTHENTICATION

The image authentication in this thesis is to detect modification and fabrication, and to distinguish the content tampered images from the credible images, in order to deter active attacks on Internet images. As shown by Fig. 4.2, a general image authentication system must consider three factors: the sender, the transmission channel, and the receiver. The sender creates an authenticator, encrypts it, and sends it with the image. Then the image, with the authenticator, is sent via the

unreliable channel. The receiver obtains the possibly corrupt image and authenticator. Using the authenticator generating algorithm, she/he constructs another authenticator from the received image. By comparing the locally generated authenticator with the received one, a decision can be made about whether or not and where the image has been modified.

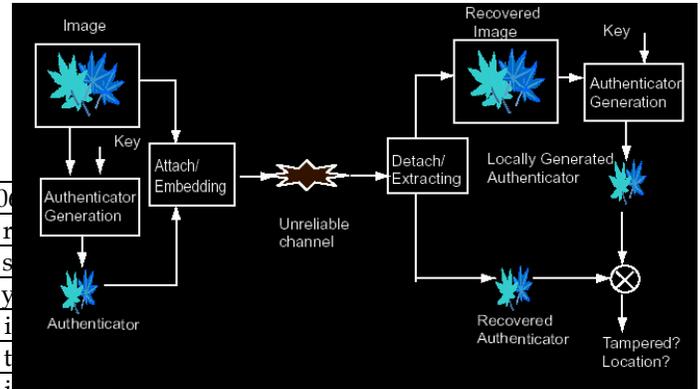


Figure 4.2: Image authentication system

We define I as a original image with dimension $N_x \times N_y$, $N_x \in \mathbb{Z}^+$, $N_y \in \mathbb{Z}$ and I_t as a recovered image, let F be the function to generate authenticator from images. We also let $\{A\}$ be the authenticator of original image, $\{A_{re}\}$ be the received authenticator, and $\{A_r\}$ be the authenticator generated from the recovered image. Therefore, we have:

$$\begin{aligned} \{A\} &= F(I) \\ \{A_r\} &= F(I_r) \end{aligned}$$

To keep $\{A_{re}\}$ equal to $\{A\}$ is very important in conventional "hard" image authentication, because a single message bit difference affects the calculation of the checksum bits and changes each one in roughly half of the cases. If $\{A_{re}\} \neq \{A\}$, the decision is made on the Hamming Distance between $\{A_{re}\}$ and $\{A_r\}$ in the majority of algorithms as follows:

$$DH(\{A_{re}\}, \{A_r\}) < \tau \quad \text{Authentic}$$

$$DH(\{A_{re}\}, \{A_r\}) \geq \tau \quad \text{Tempered}$$

where τ is threshold, and $DH(\dots)$ is the Hamming distance operation. Since the Internet, channel is an open, unreliable lossy communal channel, and traffic and user data are possibly modified from attacks in various forms of eavesdropping and packet sniffing, the received authenticator $\{A_{re}\}$ is hardly the same (bit equality) as the original authenticator $\{A\}$, and the sent image I is not the same as the recovered image I_r . For traditional "hard" image authentication, a modification of a single message bit affects the calculation of the checksum bits and changes each one in roughly half of the cases. Images data can tolerate minor changes, due to the existence of irrelevant signal information. The "loss tolerant" feature of an image is exploited in lossy compression for the reduction in file size. In the likely event of "lossy compression", "occasional" or "low priority" bit losses during transmission, a conventional digital signature and MAC would

fail, since the received image data and the signed data are not identical, but the content of images is still the same. So “soft” image authentication is desired for lossy image communication on the Internet. We attempt to achieve a practical semi-watermarking for “soft” image authentication on the Internet by using an holographic watermarking technique based on Fourier transform domain.

Conclusion Approach of Steganography using Cover-text Encryption. In terms of further developments, the principles discussed in this thesis can be used to design an entirely covert encryption system. By inputting any encrypted file as binary data, we can generate a binary image (consisting entirely of pixels with values of 0 or 1). For example, consider the plaintext Cryptology which is encrypted to provide the cipher-text string ydr39bkLP9 and is equivalent to the 7-bit ASCII bit stream

```
11110011100100111001001100110111001
11000101101011100110010100000111001
```

This bit stream is converted into the 9×9 square image with zero padding being used to complete the array as given below:

1	1	1	1	0	0	1	1	1
0	0	1	0	0	1	1	1	0
0	1	0	0	1	1	0	0	1
1	0	1	1	1	0	0	1	1
1	0	0	0	1	0	1	1	0
1	0	1	1	1	0	0	1	1
0	0	1	0	1	0	0	0	0
0	1	1	1	0	0	1	0	0
0	0	0	0	0	0	0	0	0

The binary image is then diffused with a random image field and the output embedded in a cover text through addition using a suitable diffusion-to-confusion ratio (suitable in the sense that the binary image is recovered with no bit errors for the case when the difference

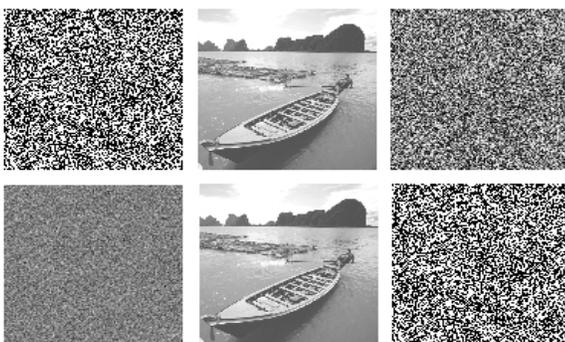


Figure 6.1. From top to bottom and from left to right (all images are 512×512): Binary image of cipher text, cover text (digital image), cipher, diffused image, stegotext after addition of the diffused data (for a confusion-to-diffusion ratio of $r = 0.01$), reconstruction between the cover-text and stegotext is insignificant).

nificant).

The size of the image that is required to implement this method is related to the binary length of the cipher-text. Assuming that the cipher-text and plaintext are of the same size (i.e. no padding is applied to the plaintext before encryption), and, given that the average number of letters per word (in the English language) is 6 (including the space), then a n^2 binary image will provide for approximately $n^2/(7 \times 6)$ words.

An example of using this approach is illustrated in Figure 6.1. The ‘watermark’ (top- left) is a 512×512 binary image obtained from the cipher text of a 6000 word document after encryption with 7-bit ASCII binary conversion. The reconstruction is a bit-for-bit replica of the input and can thus, be decrypted without error. The binary image is first converted back into a bit stream (upto the point beyond which padding is applied) and each consecutive 7-bit block, converted back into the cipher-text which is then decrypted.

In order to enhance the cryptographic strength associated with approach, the cipher shown in Figure 6.1 can be obtained from a genuine random number generator and then encrypted (to secure the data file) using a specified cryptosystem. Clearly, in addition to the receiver of the stegotext requiring the facility to decrypt the reconstruction, in order to obtain this reconstruction, the receiver must have the cipher and the cover-text. The cover-text should be one of a database of images maintained by both parties together with the cipher that is ideally stored in encrypted form. Because the stegotext and cover-text images look identical, the receiver can search through the image data base to select the appropriate cover-text. The whole point of this process is that it provides a way of camouflaging the encrypted data during transmission, the difference between sending the cipher-text and the stegotext being illustrated in Figure 6.1 as digital images. However, in this process, a macro-key is required to be exchanged which is composed of the following:

- the cipher
- the covertext database
- the decryption system

A cover text database is required for two reasons: (i) each time a transmission is undertaken, it is safer to transmit a different stegotext in order not to alert a potential attacker to multiple transmissions of the same data; (ii) a database of images should be stored rather than a single image in order that no apparent significance is given to a single image should the platform (i.e. PC or USB stick, for example) be compromised.

6 SUGGESTIONS ON STEGANOGRAPHY

6.1 STEGANOGRAPHY OFFERS BRIGHT PROSPECTS FOR COMMERCIAL DATA HIDING:-

As steganography continues on its evolutionary path researchers have exposed platforms where steganographic techniques could be employed to hide information perfectly. Such research efforts have regenerate the research and development efforts oriented towards steganography platforms and steganalysis and a number of researchers are working towards discovering new platforms that troublemaker could potentially use to hide information. Platforms such as images and other multimedia content are expected to be widely used for concealing information. 'Current research in steganography is focused on identifying various platforms through which one can hide information,' notes the analyst of this research service. Although each mode has many benefits, it is very difficult to ascertain the single best platform to send hidden messages. Steganography is capable of mitigating piracy by supporting the copyright marking.

There is a distinct lack of awareness about steganography, particularly among the business community. A major challenge associated with the field is convincing organizations to deploy tools to detect insider use of steganography, which requires complete awareness of the way data could be embedded and various approaches toward detecting this activity. 'With steganography ideally suited for protecting intellectual property, implementing steganalysis and steganography tools could be beneficial for businesses,' says the analyst. 'However, companies are unconvinced / doubtful about adopting steganography as they do not find tangible monetary benefits.'

As the possibility of steganography being used with mainly malicious intent is high, enterprises and national security organizations need to recognize the threats established by steganography and implement the right countermeasures.

Government and security enterprises must take the lead and implement measures to increase the awareness about steganography.

Industry leaders need to work with researchers and channel their R&D efforts toward development of effective technologies/solutions that would provide substantial benefits. Standard bodies could also take assistance from enterprises for developing techniques that could cost effectively improve the technology.

Future of Information Hiding provides an analysis of the trends that are shaping the domain of steganography and offers approaches that could help in its commercialization. Within this analysis, identified the challenges in the industry, as well as the drivers and restraints in the market.

6.2 LEGITIMATE USE

Steganographic techniques have obvious some legitimate uses. The business case for protection of property, real and intellectual is strong. The watermarking of digital media is constantly improving, primarily in an attempt to provide inflexible watermarks or proof of ownership. Individuals or organizations may decide to place personal / private / sensitive information in steganographic carriers. Unfortunately, there are usually better ways to manage this task. One can like these applications to the use of a security device lock on a door. The padlock will keep honest people honest, but those determined to break and enter can simply break a window and gain entry. With advances in steganography, it is possible that this medium could serve as a relatively secure storage / transmission method.

Illegal Use: Other uses for steganography range from the minor to the objectionable. There are claims that child pornography may be creep around inside innocent image or sound files. Although no any occurrence identify in the electronic world but this is entirely possible according to report on high technology crime list of eight common type of computer crime like:

- Criminal communications
- Fraud
- Hacking
- Electronic payments forgery
- Gambling and pornography
- Harassment
- Intellectual property offenses
- Viruses

In examining these criminal activities, one can identify several of these areas where steganography could be used.

7 REFERENCES

1. Mahmoud K W, Blackledge J M, Datta S and Flint J, Print Protection using High Frequency Self similarity Noise, Security, Steganography and Watermarking of Media Contents VI (Eds. E J Delp and P W Wong), Proc. SPIE-IS&T Electronic Imaging, SPIE
2. J. Cox, M. Miller, and J. Bloom, Digital Watermarking, Morgan Kaufmann, 2002.
3. S. Katzenbeisser and F. Petitcolas, Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, Boston, 2000.
4. Petitcolas, A. R, and M. Kuhn, "Information hiding: A survey," IEEE 87(7), pp. 1062-1077, 1999.