# An Effective Technique to Hide Encrypted Text inside HTML Document

Ali Mo'allin Abdullahi

Geyrte1@mail.com

Faculty of computing

SIMAD University

***Abstract***

***Steganography can be defined as the art or science of hiding information in cover media for the purpose of making it secrete information. There are many techniques used in steganography like image steganography, audio steganography and text steganography. In this paper we are going to explore an effective technique to hide encrypted text in HTML document, HTML is language regarded as the standard publishing language of the World Wide Web and it can process both plain and format text within it. We were used two that contains two processes and implemented by C#. Firstly is to encrypt the text file using one of the most efficient of encryption algorithms, the AES algorithm, and then to hide encrypted data in HTML document. And the other one is to extract the encrypted secrete file from the HTML webpage and then to decrypt the extracted file using the same key and method of the first algorithm. In order to make secrete message undetectable and more reliable encryption technique where used, encryption form of the text gives the secrete data extra layer of security, encryption gains its roots from cryptography. Cryptography means or it is the study of converting format of the text (the secrete data/information) in to another non-readable form to hide the real message from third part.***

***Keyword: Steganography, cryptography, HTML, Text Steganography, Advanced Encryption Standard (AES).***

## I. INTRODUCTION

Information security sometimes called info sec is to keep data from unauthorized access use disclosure, disruption, modification, personal inspection, recording or destruction it is not matter whether this data is physical or electronic.

Information security plays significant role in keeping personal and organizational data, we often hear news about security threats and incidents like server hacking defacement of websites and unauthorized copying of digital materials .[1]

The field of security, that concern hiding data or changing its form called security through obscurity, this kind of security it is not reliable and it is not considerable as security system since they need to be tested and validated and it is seen by everyone which who have knowledge about crypto analyst and get it but if they cannot find a way to break it considered secure, this weakness of obscurity system can be reduced using cryptography with steganography which hides data in other cover and so it draws no suspicious and no father attention [2].

### *A. Cryptography*

Cryptography is the art or science of making ciphers and it is where computer science connect with mathematics, cryptography give us tools and technology that enabling us to protect distributed systems [3].

Cryptography hides the data by changing it into unreadable form therefore the input text that goes for encryption process is called plaintext, and the output text of encryption process is called cipher text [4].

There are number of cryptography primitives or basic building blocks such as stream cipher, hash function, and block cipher; when the block have same keys for both encryption and decryption process is called shared key (also known as secrete key or symmetric) but when there are different keys they called public key or asymmetric, and through this two ways we can make stronger cipher [3].

## B. *Information hiding*

Today it is easy for people to communicate each other due to the broadband installed all over the world, this communication increase with rapidly increasing of technology and services in every day, which makes easy a lot of data or information to be transmitted over the internet and within this data there can be a lot amount of secrete information in the cover of this transmitted data e.g. text, audio, video and etc this is called information hiding [5].

This technology used to make the secret information invisible and it consist of three processes, embedding, transmitting and extracting [6]. The figure below shows the general model of information hiding process.

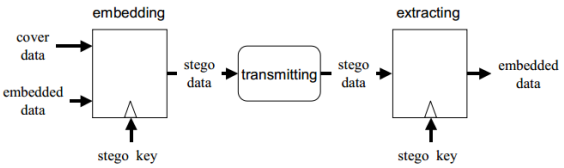


Fig. 1. Information hiding

Information hiding can be categorized into two broad categories which each part have its subcategories..

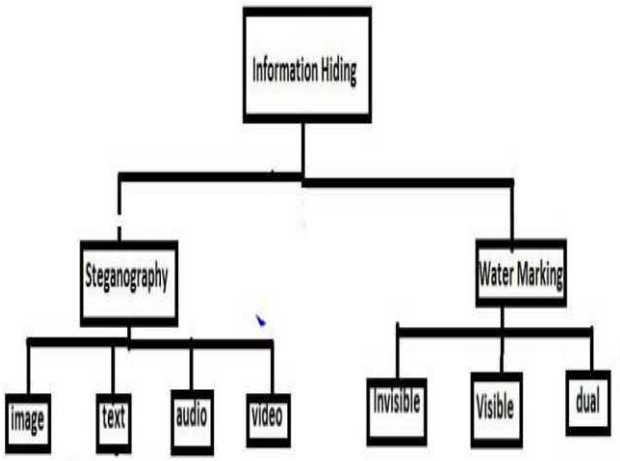


Fig. 2. Broad categories of information hiding

### 1. *Watermarking*

In the last decades many data has been transmitted through world wide web (WWW) in generally these data were insecure and simply it can be copied and damaged, so the requirement of Intellectual properties rights (IPR) protection was raised, digital watermarking is one of the techniques that can solve these problems and it has become a famous technique that can protect the IPR.

Until now the protection techniques are un-countable as more techniques were developed and many ways were suggested, the digital data such as image, video and audio can be processed by using digital watermarking procedure to protect unauthorized copying of this materials, the main advantage of watermarking is to protect digital files to the owners particularly it is so important for keeping against removal of the hiding mark [7].

In generally all the various types of digital watermarking techniques "visible and invisible" have a simple ideas that is to hide a set of materials inside the owner's data and then publishing it therefore the owner of that data have an ability to indicate the proprietary of that data. All types of watermarking are similar in terms of the requirement needed for data quality and safety.

### 2. *Steganography*

Steganography is an art or science of concealing secrete message, inside one of text, audio, video files and etc, the word steganography is a Greek word that means "covered writing [8]. Steganography allows user to hide your sensitive data through one of text, audio, and video files, it also added one layer of security than the cryptography since it hides the existence of the data so there is no opportunity to suspect that there is something hidden [5].

#### *a) Types of steganography application:*

There are many types of steganography in this section we will go a short look of different types of steganography including image, audio and text steganography.

- *Images steganography*

Hiding data inside image is popular technique nowadays an image which carries a secret message inside of it can be spread through the World Wide Web and image steganography represents the hidden text inside image without knowing to any one not evolved in the communication process therefore the image remains intact after hidden the secret data [9].

Image steganography is a good candidate of carrying hidden information as image contains many redundant areas to be replaced by the secret message [7]. The most common method used in image steganography is least-significant bit (LSB), using this method we can able to conceal the bits of the data directly in the cover image as deterministic sequence so the result of modulating the least significant bits is imperceptible for human because the amplitude of difference are so small and not visible [9].

- *Audio steganography*

Ear of the human can hear frequency spectrum between 80 to 20,000 HZ which make very difficult and challenging task hiding information in audio of the human because human auditory system able to hear even small variation in pieces of music and speech so it is very low to hide something in the noisy or echo of HAS [5]. Audio steganography is a steganography method that uses audio files as a cover media to hide the secrete data [8].

Audio steganography allows you to hide secrete information within it while sampling rate has directly relation with the amount of data to be hidden [10].

echo hiding is one of method that uses echo to hide secrete message, three parameters of echo method are amplitude, decay rate and offset of the original signal, you can hide zero or one bits of secrete message between the delay of the original sound and echo but this method it quite difficult and complex also its good for audio files where there is no additional degradation.

- *Text steganography*

Text steganography is one of most and large Steganographic types which we can define it as hiding text inside another text and it is difficulty according to other steganography types because of less redundant of the text [11].

Text steganography is used to hide secret information in text for both plaintext and formatted text. We can make text steganography by changing the structure of text or changing the special characteristics of the text, the purpose of this methods is to develop the alteration that are reliable, decodable and invisible change[11].

*I. Webpage text steganography*

HTML and XML files can also be used to hide bits

*A) XML web pages:*

Extensible Mark-up Language (XML) is plat-form independent and universal language which allows users how the data can be stored, transferred and exchanged electronically, it derived from generalized mark-up language (GML) over the World Wide Web, you can transmit XML page between different platforms [12].

XML allows other specialist field such as chemistry, finance or environmental data collection to define XML schema to develop the mark-up language for the exchange of specialized data unique for their field, and an attribute of XML allows additional information about elements of the data to be defined within the element definition [4].

*B) HTML Document*

HTML “Hyper Text Markup Language” is a fundamental building of WebPages containing tags which are surrounded two angle parenthesis and the information between two tags which you can see in the internet browsers, father more it is easy to use with other languages such as java and xml and it does not need specialized software for programming.

Hiding text in HTML document is one of the text steganography approaches that uses the organization and characteristics of the tags and attributes of the webpage to hide some bits of secrete information which make easy to exchange the secrete message between the recipient however, users of internet is concerned for the information of the web page, based on this organization and characteristics, most steganography algorithms over web pages deal with the coding of the web pages not the page`s information or content [12].

## II. PURPOSED METHODS

### A) *Advanced encryption standard AES*

This technique is one of the cryptography algorithms which is used to encrypt the secrete message (plain text) into another form called cipher text, the cipher text can decrypt or decipher only who possess a secrete message e.g. authorized users.

This algorithm developed by two Belgian mathematician joandaemen and Vincent rijmen is the last standard adapted by NIST in 2001.AES is symmetric encryption algorithm which uses same keys for both encryption and decryption process and it uses a block size of 128 bits, the .NET frame work which we will use in this system have a common abstract base class Symmetric Algorithm which shared by AES algorithm and algorithms like DES and Triple DES, this method also uses cipher block chaining (CBC) as default and PKCS7 mode as default padding mode.

some authors recommended using this method when you are developing a new system and require to encrypt data with key of 255 bits (32 bytes), the key and iv (initialization vector) are set because it passed the encryption and decryption methods and it must be the same for both encryption and decryption process, the key must be secrete but iv have not to be.

Although this system is mainly concern is hiding text in html webpage but we will encrypt the text before hiding it to add extra security layer by using hybrid method "cryptography with steganography".

### B) *HTML attribute order algorithm*

HTML can contain a lot of tags and each tag contains numerous amount of attributes and the order of attributes in the tag does not affect the output which appears in the web browsers, for example the attribute order of <body background = "image.jbeg" bgcolor = "#FFFFFF"> can hide 0 bit and <body bgcolor = "#FFFFFF" background = "image.jbeg"> can hide 1.

In the above example if the sequence of attributes is (background, bgcolor) then it hides 0 and if the sequence of attributes (bgcolor, background) then it hides 1. This method basically contains three components hiding process, generation key and extraction process. The key is formed as rows and columns this combination of rows and columns are generated through scanning of HTML page it contains two types of attributes corresponding to two columns which are primary attributes and secondary attributes.

#### a) *Basic Procedure*

The hiding process scans each attribute of each html tag, and checks to see whether that attribute exists in the primary attribute field of the key file, if yes, its corresponding secondary attribute is searched in the corresponding html tag, if found, then this combination of attribute is used to hide a bit, if not, skip this attribute. The hiding of a bit is determined by the order of the attributes in the attribute combination. If primary attribute is followed by a secondary attribute, it can hide a bit 1; else it can hide a bit 0. The extractor program extracts the message from stego text by first identifying the attribute combinations that hides a bit and then finding the bit corresponding to the order of those attributes. If primary attribute is followed by secondary attribute, a bit 1 is detected, else a 0 is detected.

## III. IMPLEMENTATION

Based on the proposed algorithms we have developed system using C# to implement the algorithms. This system contains two windows one for hiding processes each contains two steps.

### A. *Hiding process*

- *Step one*: first you have to upload both the HTML File that you want to hide within and the secrete message (txt file) and then to encrypt the secrete message write or key which contains digits not less than 9 characters that will be needed in the extraction process.
- *Step two:* second step is to hide the encrypted file (the secrete message) which was already

encrypted in the previous step and last shows message tells whether the hiding process completed or not.

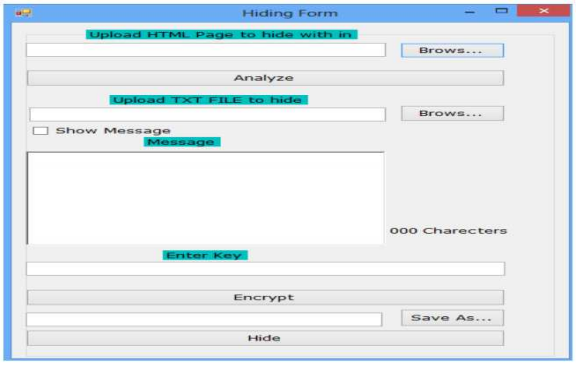


Fig. 3. shows hiding process window form

### *B. Extracting process*

Same to the step one in the hiding process you have to upload the stego-file (html file) that contains the secrete message. Extract the secrete message when this process will complete you will see encrypted secrete file (unreadable text). to decrypt this encrypted secrete message in the html stego-file enter the same as encryption key and the extraction process will complete.

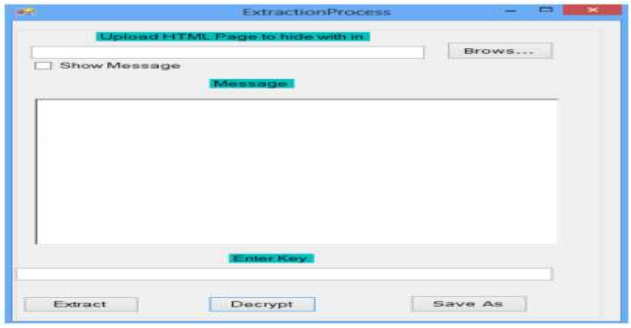


Fig. 4. shows the extraction process

## REFERENCES:


[1] James C. Judge, "Steganography: Past, Present, Future", SANS Institute, 2001

[2] Cole, Hiding in Plain Sight: Steganography and the Art of Covert Communication, Indiana, John Wiley & Sons Inc., 2003.

[3] John F. kennedy, "A Guide to Building Dependable Distributed Systems", National Chiao Tung University Taiwan, 2008.

[4] Rajkumar Yadav, "Study of Information Hiding Techniques and their Counterattacks", International Journal of Computer Science & Communication Networks, 2011.

[5] Peter Bayer Henrik Winderfors, "Steganographic content in streaming media", BlekingeInstitude of Technology, August 2002.

[6] Matsumoto, Inoue, makino,*, "A Proposal on Information Hiding Methods using XML," Symposium on Cryptography and Information Security, Jan. 2000 (in Japanese).

[7] Adel Almohammad, "Steganography-Based Secret and Reliable Communications: Improving Steganographic Capacity and Imperceptibility", Brunel University, August, 2010.

[8] Sumanth Minnakanti Spring, Multimedia Steganography Tool for Hiding Text,Texas A&M University-Corpus ChristiCorpus Christi, TX,2014

[9] MasoudNosrati , RonakKarimi , Mehdi Hariri , "An introduction to steganography methods", Kermanshah University of Medical Science, Kermanshah, Iran , 2011.

[10] W. Bender, "Techniques for data hiding", MIT Media Laboratory, 20 Ames Street, Cambridge, 1996.

[11] Neha Rani, Jyoti Chaudhary, "Text Steganography Techniques: A Review", International Journal of Engineering Trends and Technology (IJETT), July 2013.

[12] L. P o l a k, Z. K o t u l s k I, SENDING HIDDEN DATA THROUGH WWW PAGES:
DETECTION AND PREVENTION" ," ENGINEERING TRANSACTIONS" ,2010.