What a Digital Forensics Investigator should know about Steganalysis of Digital Content

Ugo Reyne
B00122680

School of Informatics
Department of Informatics and Engineering
Institute of Technology, Blanchardstown
Dublin 15.

Word Minimum for assignment: 3,500 **Actual Word Count: 3,537**

Bachelor of Science
Computing in Information Technology
December 12th, 2018

Table of Contents

Table of Figures

Abstract

Steganography, a Greek word for “covered writing” is an art of hiding data inside data. It aims to avoid getting suspicion on the transmission of hidden data. And Steganalysis aims to discover and render hidden data by steganography, it is widely used by Digital Forensics Investigators.

Digital Forensics Investigator, also called Forensics Analyst, are professionals that work with law enforcement agencies to retrieve information from any types of data storage. Steganography and Steganalysis have become a huge part of forensic cases. It is then very important for Forensics Analysts to have some knowledge in steganography and steganalysis.

This paper is an overview of what a Digital Forensics Investigator should know about Steganalysis of Digital Content?

Firstly, the steganography is introduced with some steganography methods on different file types. Then steganalysis is explained, such as the work of forensics and at the end, some security issues are raised up.

Keywords: Steganography, Steganalysis, Forensic, Analyst, Null Ciphers, Image Steganography, Audio Steganography, HTML Steganography, File Header Steganography, Operating System Steganography

Introduction

Steganography, a Greek word for “covered writing”, is an art of hiding data inside another object. This art has been used for a long time, even before the invention of the computer. The term steganography has been created at the end of the 15th century. At the start, messages were hidden on the back of wax writing tables, on rabbits’ stomachs or even tattooed on slaves’ heads. But now that the computer has been popular, the capability of hiding information has also increased, and it is now a lot more digital. This paper focuses more on its digital aspect.

The steganography is the fact to hide messages, but it does not cover that a communication is made. The container of the hidden message is called the carrier. Carriers can take the aspect of text, image, audio, video or any other aspect that can hold a secret message. And the hidden message can be directly a plain text or anything that can represent a bit stream.

In contrary to cryptography, steganography simply hides the message under another one. Cryptography converts ordinary plain text into unreadable text. Steganography is very closely related to Watermarking, only their goal changes. Watermarking hides messages that are related to the content of the carrier and Steganography hides messages that have no relations with the carrier. For example, Watermark is used to identify ownership of copyright.

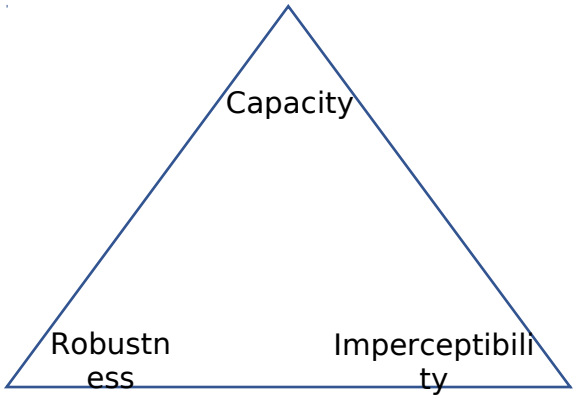

Figure 1. Triangle of steganography

As the “Figure 1” shows, the steganography is divided into three important parts, the capacity, the robustness, and the imperceptibility that represent a triangle. The capacity represents the secret data stored in the carrier. The robustness is the ability of the secret data to stay undamaged if the steganography object is manipulated (cropping, scaling, blurring and more). And the imperceptibility makes sure that the algorithm used is not damaging the carrier to avoid attraction on the hidden message. If one of those three triangle parts is too important, then the others will have less attention and may compromise the hidden message. In every steganography work, there will always be one of those three parameters that have less attention than the two others and could compromise it.

This paper is now going to focus on what a Digital Forensics Investigator should know about Steganalysis of Digital Content. Firstly, the steganography will be introduced with some steganography methods on different file types. Then steganalysis will be explained, such as the work of forensics and at the end, some security issues will be raised up.

1. Steganography methods

Steganography can be performed on any digital objects that contain bit streams. It can be injected into common files like text, images or audio files. Video files are simply a succession of images, so same as the image's files, data can be hidden in videos. Videos can even store larger data on the images that it is composed of. A less common carrier is the protocol one. This one is using the layers of the OSI network model by using the network transmission, like TCP/IP protocols.

1.1. A simple steganography processes

Here is a simple steganography process on the "Figure 2" to make it easier to understand:

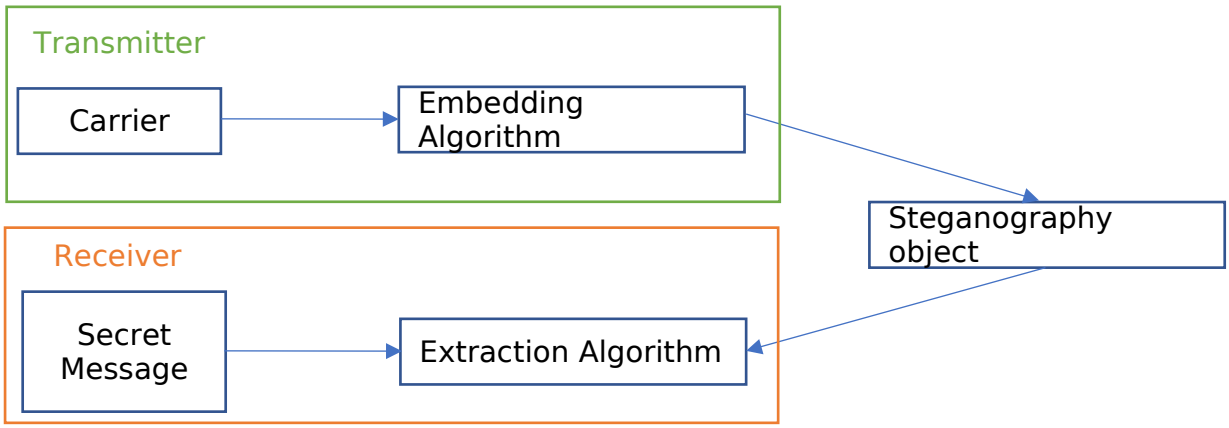

Figure 2. Simple steganography process

The transmitter takes an object, called a carrier. He uses an algorithm to embed a secret message into the carrier, at this moment a steganography object is created with the secret message. Then the transmitter sends the steganography object using any transporter, like online social networks, to the receiver. When the receiver gets the steganography object, he uses another algorithm to extract the secret message from the object received. The transmitter needs to be careful that no one gets the original carrier and the steganography object, otherwise, the algorithm used could be compromised and the secret information could be found more easily by unwanted persons. In some algorithms, keys can be

entered while embedding the secret message and this key will be asked to extract it. The key protects a bit more the secret message, but in many cases, the key can be bypassed.

1.2. Null Ciphers

Also known as concealment cipher, an old technic to encrypt plaintexts. It consists to take only some letters from a full text to produce another one. The Null Cipher can be used in different ways: by taking the first or the last letter of each word, by taking letters or words from certain positions, by using patterns and more. There are many other ways to use this technic.

Here is an example of the implementation of a Null Cipher, a message written by a prisoner:

> SALUDOS LOVED ONE SO TODAY I HEARD FROM UNCLE MOE OVER THE PHONE. HE TOLD ME THAT YOU AND ME GO THE SAME BIRTHDAY. HE SAYS YOUR TIME THERE TESTED YOUR STRENGTH SO STAY POSITIVE AT SUCH TIMES. I'M FOR ALL THAT CLEAN LIVING! METHAMPHETAMINES WAS MY DOWN FALL. THE PROGRAM I'M STARTING THE NINTH IS ONE I HEARD OF A COUPLE WEEKS BEFORE SEPTEMBER THROUGH MY COUNSELOR BARRIOS. BUT MY MEDICAL INSURANCE COVERAGE DENIES THEY COVER IT. I'M USING MY TIME TO CHECK AND IF THE INSURANCE AGENT DENIES STILL MY COVERAGE I'M GETTING TOGETHER PAPERWORK SAYING I TESTED FOR THIS TREATMENT REQUIRED ON THE CHILD CUSTODY. THE NINTH WILL MEAN I HAVE TESTED MY DETERMINATION TO CHANGE. ON THE NEXT FREE WEEKEND THE KIDS ARE COMING, BUT FIRST I GOTTA SHOW CAROLINA I'M STAYING OUT OF TROUBLE WAITING TO GET MYSELF ADMITTED ON THE PROGRAM. THE SUPPORTING PAPERWORK THAT THE FAMILY COURTS GOT WILL ALSO PROVE THERE'S NO REASON NEITHER FOR A WITNESS ON MY CHILDREN'S VISITS. OF COURSE MY BRO HAS HIS MIND MADE UP OF RECENT THAT ALL THIS DRUG USAGE DON'T CONCERN OUR VISITS. I THINK THAT MY KIDS FEEL I NEED THEIR LOVE IF I'M GONNA BE COOL. GUILTY FEELINGS RISE ON ACCOUNT OF

THE MISTAKES I COULD WRITEUP. FOR DAYS I'M HERE. HE GOT A GOOD HEART. SHOULD YOU BE HAVING PROBLEMS BE ASSURED THAT WHEN YOU HIT THE STREETS WE'LL BE CONSIDERING YOU...

Figure 3. Example of a Null Cipher

Hopefully, the FBI found the message hidden behind this text before it was transmitted to the receiver. On this example, by taking every fifth word of the message, we can find this secret message:

TODAY MOE TOLD ME HE TESTED POSITIVE FOR METHAMPHETAMINES
THE NINTH OF SEPTEMBER BUT DENIES USING AND DENIES GETTING
TESTED ON NINTH
TESTED ON FIRST
I'M WAITING ON PAPERWORK
GOT NO WITNESS OF HIS RECENT USAGE
I FEEL IF GUILTY OF WRITEUP HE SHOULD BE HIT

Figure 4. Secret Message hid in the example

1.3. Images

Today images are often used as a carrier. It is now easier to spread an image than a text over the web to bring less suspicion over it. Multiple technics exists to embed a message inside an image without changing the visible properties to draw less attention.

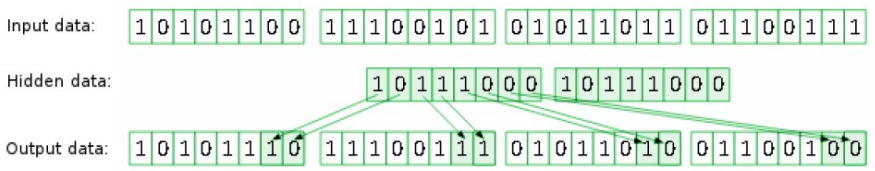

Figure 5. Least significant bits message hiding

The simplest algorithm to hide a message in an image is to use the Least Significant Bits (LSB). The lowest bits values of a pixel in an image. Each bit of the message is stored in the lowest bits of each pixel. For each pixel, grayscale images are only coded on 8 bits and colored images are coded on 24 bits. Eight for the red, eight for the green and eight for the blue. Each of those eight bits represents a level that goes from 0 to 255, where 255 represent the native color. The “Figure 5” shows an example of the use of an LSB algorithm. Each bit of the data to hide is replacing two of the lowest bits on each pixel.

This process alters slightly on the image quality. But as the qualities of images are usually higher than the human vision, the degradation is not noticeable by a human being. Up to half of the bits can be taken to store some secret data on an image without starting to really alter the image. Of course, the fewer bits modified and less the image is altered.

This technic brings the possibility of hiding an image inside another image by using half of the pixels bits.

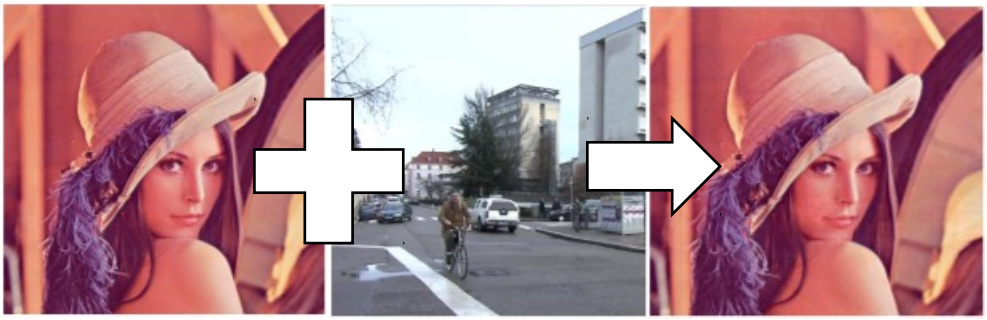

Figure 6. Hiding an image in an image

The “Figure 6” is an example of an image hidden in another image. The image on the left is the carrier, the one in the middle is the secret message and the one on the right is the result of the algorithm. The steganography image generated has a slightly lower quality than the original carrier. This is due to the replacement of the Least Significant Bits by the pixels of the hidden image which represents here half of the pixel’s bits.

To get less attention, a masking and filtering technic can be used. This technic consists to find significant areas on the carrier where the message will be less noticeable and more integrated into the image.

1.4. Audio

Audio files can also be the carrier of a secret message. Digitized sample audio files contain 8 bits, so as the same way as on the image, the Least Significant Bits of the carrier can be replaced by the message.

Another algorithm for audio carriers, phase coding. Phase coding exploits the fact that the phase changes in audio signals are not recognized by the Human Auditory System. The message bits are converted into phase shifts that are adjusted to preserve the phases between segments.

1.5. HTML files

With the internet expansion in various life and work aspects, HTML can be widely used for steganography, it is very easy to hide a secret communication under a basic one. Tons of web pages are available on the internet and could hide data in many ways.

Hiding messages in comments or at the end of the HTML tags, simple ways but it is not shown by web browsers. Creating white spaces, Ordering the case of tags letters or adding some special characters don’t affect the web page, messages could also be hidden like this and it would even be more difficult to find out the message as it is not written in formal letters. Some algorithms analyze the source code to hide data inside the tags by implementing one of the technics. To reverse this type of steganography, analyzing the source code is the key.

1.6. File Headers

Many files structures have headers. Some information is stored in there, such as the file type. But some values of the header are insignificant and can be replaced. A secret message can, therefore, be placed in the header of the carrier. The file header is widely used in steganography with TCP/IP packets.

1.7. Operating Systems

In the same principle as the file header, in operating systems with file partitions in FAT16 for example. A cluster with a size of 32 kilobytes, if a file of 4 kilobytes is stored in it, then this means that the others 28 kilobytes are wasted and could contain a secret information that the operating system will not show. Again, in an operating system, a special partition can be created with an information that when the system starts normally, the partition is not shown.

These are few technics used on different types of file, but lots of more complex algorithm exists. With the technics explained here, it is easy to find the messages hidden and destroy it if wanted. The Least Significant Bits technics used on images, with a simple compression or modification of the carrier the message could be altered. For example, an employee wants to take some information out of the company he works for, if the company is compressing everything that gets out of it, this steganography method will not work, or the secret data will be damaged.

Today, lots of new complex algorithm are created and are almost undetectable. For example, FontCode, a new algorithm that embeds information in text using Glyph perturbation. This new algorithm creates nearly undetectable steganography with most of the tools present on the internet. But steganalysis tools also evolves and tries to detect most complicated steganography algorithms. Since the steganography algorithm evolves very fast, it is very important to stay up to date with all the algorithms.

1.8. Tools to perform steganography

Here are some great tools to embed data and extract it. Each tool supports different formats of files. Steganos a tool for audio and video that supports mp3, m4a, avi, mpeg, mov and exe files. S-Tools and Camouflage are more for images like bmp, gif or jpeg pictures. StegHide and Invisible Secrets are tools for images and video files like jpeg, bmp and wav files. On most of the tools now, an encryption key can be entered and will be needed to extract the information. This key adds a protection to the data hidden, but as it has been said earlier in this paper, the key can sometimes be easily bypassed.

2. Digital Forensics Investigator

Digital Forensics Investigator, also called Forensics Analyst, are professionals that work with law enforcement agencies to retrieve information from any types of data storage. Often the work is done on corrupted or damaged devices which doesn't make the job easier. Forensics Analysts examines, recover, analyze and preserve data that could be used as evidence in criminal prosecutions. Reports are produced to explain exactly how they got the evidence and every step that has been made to get them. A Forensic Investigator is always impartial and needs to keep his neutrality and will never judge or convict. For this job, lots of knowledge (not only on the network and hard drives) and patience is needed. The chain of custody should never be broken, at any time the Analyst should know where the evidence is and keep them safe. Today steganography becomes more and more present on the internet and on local machines in digital forensics cases.

3. Steganalysis in forensics

Steganalysis, a research discipline that consists to detect steganography, to discover messages hidden. A steganalysis does not include the extraction of the message.

Different phases need to be realized for a forensic steganalysis. Firstly, the detection phase, where the files are scanned and checked if they contain any information. Nowadays with the most advanced algorithm, it can be hard to find if a file is a carrier that holds a message. That's what makes this phase complicated, keep scanning a file deeply that maybe contains nothing or stopping the scan on a file that contains a message. When a file has been notified suspicious the steganalysis is not finished for forensics, an extraction needs to be performed. If on the detection phase the tool used to hide the message could be identified, the extraction can probably be made with the same tool. A secret key can be added to the embedding algorithm, it is then more difficult but can sometimes be bypassed easily. If no keys have been entered, then the extraction is simple.

3.1. Detection

There is multiple methods to detect the use of steganography, the visual detection for all the images and videos (JPEG, BMP, GIF and more), The audible detection for all the audio files (WAV, MPEG, MP3 and more), the statistical detection to detect changing patterns in pixels or LSB, and the structural detection to examine the size, date and time, contents and checksum of the files.

Because the human eye can usually not detect steganography, deeper searches need to be realized. To detect and extract steganography files, many tools exist like FTK Imager. Those tools try to find all the suspicious files by manipulating them. Lots of tools for the steganography forensics use are listed by the National Institute of Standards and Technology that maintains a list of digital signatures in the National Software Reference Library.

To find if a file hides an information, checking for any anomaly by comparing properties and checksum of the original and steganography file is the first thing to do when both files are present. The file signature is another thing to look and can give information about the embedding tool used. When using a hex editor, by simply searching the term “steg*” or related terms, steganography could be identified in the results.

3.2. Tools for Steganalysis

If the original file and the steganography file are both found, then the tool WinHex is very powerful. It compares files and gives reports, it can search for differences, convert between ASCII and Hexadecimal and adds more functionalities for forensic analysis.

But if only the steganography file is found, then the tool StegSpy or an equivalent one is recommended, it is a tool that checks the signature of files. It checks for the program used on this file, it also searches for steganography signatures and can even give the location of the secret data inside the steganography file.

On NTFS file systems, multiple Alternate Data Streams can be present on the steganography files. Alternate Data Streams allow files to have multiple streams of data. Files have at least one data stream and by default, it is called $DATA. On windows, data streams are hidden and difficult to find.

Here is an example, file.txt:hiddenFile.txt. The secret message is written in hiddenFile.txt and file.txt stays empty. Windows operating system is showing the only file.txt and will never tell that a file hiddenFile.txt exists. But some tools like LNS, LADS, NTFS ADS Check and others can check if Alternate Data Streams are present on files.

4. Security issues

Steganography should be used only for legitimate business, but since steganography has become easy to perform by simply using tools, nobody can really stop its illegal use. But things can be made to avoid or reduce its illegal use, especially in an organisation.

Organizations are always trying to limit their employees from installing software, a good thing to do to avoid the installation of steganography tools. But many employees can bypass this restriction. Organizations can also encrypt the data that is getting out and in from their network. By doing this, they can damage or even destroy steganography messages during their transmission through the network. This is actually a very good thing to do, but it can’t be applied to the global world network, otherwise, it would also block legitimate and legal use of steganography.

Laws make it very complicated to manage those transmissions, but organizations have the rights to do that on their network, and they should take the steganography more seriously to avoid unwanted transmissions. Today steganography needs to be taken very seriously. Recently security researches have found that some advertisement images on the Internet contained a malware hidden in it. People need to be even more aware than before on the internet contents, everyone now should be aware of the thing that nobody can notice without any examination.

Conclusion

The steganography has a huge history of illegal and legal uses. The abuse use of it will continue to increase. As legal restrictions are difficult to enforce, organizational security should act to limit the use of steganography. It should be part of the organizational policies and procedures.

This paper provided an overview of what a Digital Forensics Investigator should know about Steganalysis of Digital Content. The steganography and steganalysis in forensics have been introduced, with some simple algorithm methods. The security aspect has also been overviewed.

More complex embedding algorithm and more complex steganalysis tools are created every day as a continuous process. It is very important as a forensic steganalysis to stay up to date to know where and how information can be hidden and how to extract it. Nowadays Digital Forensics Analysts consider Steganalysis as a routine part of the examination. An investigator might need clues from other aspects of the case to point on the right steganography files. Steganography could sometimes be hard to find without any clues.

Cryptography is also very important in forensics. The steganography and cryptography combined could make it even more complicated to retrieve hidden data. A Digital Forensics Investigator should then also have lots of knowledge in cryptography. For a Forensic Analyst, knowledge is the key, an analyst should never limit his knowledge to the minimum required.

Reference List

- Artz, D., “Digital Steganography: Hiding Data within Data”, 2001
- CHANG XIAO, CHENG ZHANG, CHANGXI ZHENG, “FontCode: Embedding Information in Text Documents using Glyph Perturbation”, 2017
- Jamil, T., “Steganography: The art of hiding information is plain sight”, 1999
- M. Pavanil, S. Naganjaneyulu, C. Nagaraju, “A Survey on LSB Based Steganography Methods”, 2013
- Provos, N. & Honeyman, P., “Hide and Seek: An introduction to steganography”, 2003
- Sara Khosravi, Mashallah Abbasi Dezfoli, Mohammad Hossein Yektaie, “A new steganography method based HIOP (Higher Intensity Of Pixel) algorithm and Strassen's matrix multiplication”, 2011
- Stuti Goel, Arun Rana, Manpreet Kaur, “A Review of Comparison Techniques of Image Steganography”, 2013
- Wang, H & Wang, S, “Cyber warfare: Steganography vs. Steganalysis”, 2004
- Wingate, J. “Digital Steganography: Threat or Hype?”, 2007
- Yin-cheng qi, liang ye, chong liu “Wavelet domain audio steganalysis for multiplicative embedding model”, 2009

Bibliography

- Amandeep Kaur, Rupinder Kaur, Navdeep Kumar, “A Review on Image Steganography Techniques”, 2015,
  - (https://pdfs.semanticscholar.org/a0a0/570c1c88ef0461b9fab9c07de6cb60c9bbf5.pdf)
- Rogerie Grégory, “La stéganographie”
  - (http://www.univ-orleans.fr/mapmo/membres/louchet/teaching/timo/Rogerie.pdf)
- Champakamala .B.S, Padmini.K, Radhika .D. K Asst Professors, “Least Significant Bit algorithm for image steganography”
  - (http://ijact.org/volume3issue4/IJ0340004.pdf)
- Tanmay Sinha Roy, “IMAGE STEGANOGRAPHY USING LSB BIT-PLANE SUBSTITUTION”, 2016
  - (https://www.irjet.net/archives/V3/i12/IRJET-V3I1247.pdf)
- SadhanaRathore, “STEGANOGRAPHY: BASICS AND DIGITAL FORENSICS”, 2015
  - (http://ijsetr.org/wp-content/uploads/2015/07/IJSETR-VOL-4-ISSUE-7-2589-2593.pdf)
- Neil F. Johnson, Sushil Jajodia, “Steganalysis: The Investigation of Hidden Information”, 1998
  - (https://cse.buffalo.edu/courses/cse725/peter/Johnson_1997.pdf)
- Gary C. Kessler, “An Overview of Steganography for the Computer Forensics Examiner”, 2015
  - (https://www.garykessler.net/library/fsc_stego.html)
- Masoud Nosrati, Ronak Karimi, Mehdi Hariri, “An introduction to steganography methods”, 2011
  - (https://pdfs.semanticscholar.org/2331/1184a7b078945f519e8bf89c719fed7b1f81.pdf)
- John R. Vacca, K Rudolph, “System Forensics, Investigation, and Response”
  - (https://books.google.ie/books?id=astqv8hRnT0C&pg=PA155&lpg=PA155&dq=steganography+investigation&source=bl&ots=Jn4wegzREF&sig=zJVjjDqVLT4UQIqb15Dcw6huVyU&hl=fr&sa=X&ved=2ahUKEwj-

hbqWyvfeAhVqDcAKHSQjBHI4ChDoATAJegQIChAB#v=onepage&q&f=false)

- Jayaram P, Ranganatha H R, Anupama H S, “INFORMATION HIDING USING AUDIO STEGANOGRAPHY – A SURVEY”, 2011
  - (http://aircconline.com/ijma/V3N3/3311ijma08.pdf)
- Vikas Yadav, Vaishali Ingale, Ashwini Sapkal, Geeta Patil, “CRYPTOGRAPHIC STEGANOGRAPHY”, 2014
  - (https://airccj.org/CSCP/vol4/csit42603.pdf)
- Gunjan, Er. Madan Lal, “Investigation of Various Image Steganography Techniques in Spatial Domain”, 2016
  - (http://ijcert.org/ems/ijcert_papers/V3I614.pdf)
- Nanhay Singh, Bhoopesh Singh Bhati, R. S. Raw, “DIGITAL IMAGE STEGANALYSIS FOR COMPUTER FORENSIC INVESTIGATION”, 2012
  - (https://airccj.org/CSCP/vol2/csit2217.pdf)
- Merrill Warkentin, Ernst Bekkering, Mark B. Schmidt, “Steganography: Forensic, Security, and Legal Issues”, 2008
  - (https://commons.erau.edu/jdfsl/vol3/iss2/2/)
- Chintan Dhanani, Krunal Panchal, “HTML Steganography using Relative links & Multi web-page Embedment”, 2014
  - (https://www.ijedr.org/papers/IJEDR1402108.pdf)
- Mohammad Shirali Shahreza, “A New Method for Steganography in HTML Files”, 2006
  - (https://link.springer.com/chapter/10.1007/1-4020-5261-8_39)