Steganography & the Art of Deception: A Comprehensive Survey

Youssef Bassil

LACSC – Lebanese Association for Computational Sciences
Registered under No. 957, 2011, Beirut, Lebanon
youssef.bassil@lacsc.org

***Abstract*—Ever since the beginning of human civilization, mankind had always confidential things to hide or share secretly. Endless methods were devised; an ingenious one is called steganography which refers to secret writing. In essence, steganography is the science of hiding secret data into innocuous-looking mediums in such a way that only the communicating parties are aware of this trick. Steganography maybe started during the Stone Age and greatly evolved during the computer age. Currently, it has many techniques, methods, and applications, making it worth having a closer look at. This paper presents a comprehensive overview on steganography and on its different techniques that have been proposed in the literature during the last decades. It additionally sheds the light on its history before and after the computer age, its various models, requirements, and processes.**

***Keywords*—** *Computer Security, Information Hiding, Steganography*

1. Introduction

Steganography is the art and science of hiding information in ways that only the communicating parties are aware of [1]. Steganography has been exploited throughout history by individuals, military, secret intelligence, and governments to stealthily communicate and transmit secret information without drawing any attraction. Steganography started centuries before the computer age and it lasted till our present days. In modern time, steganography is being implemented using digital computers to embed secret digital data such as text, documents, and binary files into some other form of digital data such as image, video, and audio files. The advantage of steganography over cryptography is that the former conceals secret data into a computer file so that it cannot be seen by eavesdroppers; while, the latter garbles secret data so that it cannot be read by eavesdroppers. As a result, to a third party, garbled data would right away imply a secret communication. However, concealed data would not draw any attention; and therefore, would not raise suspicions that a secret communication is taking place [2]. For this reason, steganography is often regarded as a stealthy method for transmitting sensitive data into total secrecy across public channels in such a way that no one, apart for the communicating parties namely the original sender and the intended receiver, can know about the existence of the communication. The key requirement of a steganography system, that without it, it would be compromised is imperceptibly. Characteristically, imperceptibility refers the fact of hiding secret data inside computer files without changing their size or damaging their visual or audible properties. To do so, steganography often takes advantage of the limited capabilities of the human biological systems. For instance, image steganography is drawn upon the visual limited capabilities of the human visual system (HVS) which cannot detect the slight intensity variation of the pixels of an image. As a result, hiding the secret data into the LSBs of the pixels composing the carrier image would change marginally the color intensities of the image so much so that it would be unnoticeable by a human naked eye. Thus, an unauthorized observer cannot distinguish between the original carrier image and the tampered one, i.e., the one that carries the secret data into the LSBs of its pixels [3]. On the other hand, audio steganography takes advantage of the Human Auditory System (HAS) which cannot hear the slight variation of audio frequencies at the high frequency side of the audible spectrum; and thus, audio steganography can exploit and use this type of frequencies to hide secret data without damaging the quality of the audio file or changing its size.

This paper presents a comprehensive survey on steganography, examining elaborately its history, methods, applications, and state-of-the-art. It first starts by discussing the history of steganography and its applications before and after the advent of digital computers. Then, the various requirements, elements, and methods of hiding are tackled intensely.

International Journal of Latest Trends in Computing
IJLTC, E-ISSN: 2048 - 8688

Afterwards, the mathematical model of steganography is illustrated shedding the light on its basic parameters, elements, and processes. Finally, three detailed sections are devoted for the different types of steganography along with their related work and states-of-the-art including image, audio, and text steganography.

2. History of Steganography

The term Steganography is derived from the Greek words steganos, meaning "secret", and graphy, meaning "writing" [4]. In other terms, steganography simply means "secret writing". The usage of steganography is not new as it has been employed in various forms throughout history. The very first reference to steganography can be traced back to 440 B.C in ancient Greece. In one account, the Greek ruler Histiaeus shaved the head of one of his slaves, and ordered an artist to tattoo a secret message on the slave's skull. Histiaeus then waited until the slave's hair grew over the message. When the message was totally covered by the slave's hair, Histiaeus sent the slave to one of his allies warning him of a possible Persian attack. The ally, who received the slave, shaved the slave's head and recovered the secret message that was hidden underneath the slave's hair. Likewise, in the same era, Demaratus, the king of Sparta, carved a secret message on a wooden board, then covered it with wax. The unsuspicious wooden tablet was then delivered with the secret message hidden inside it [5, 6].

In the 15th century, in 1499, the German polymath and Christian abbot Johannes Trithemius, wrote a three-volume book known as "Steganographia" [7] which is considered to many as the first manuscript to discuss cryptography. In fact, "Steganographia" is principally concerned with the transport of secret messages by angels and spirits, which are grouped according to a certain hierarchy based on regions in the Earth they rule over. In order to send a secret message, the sender should first determine the appropriate group of spirits, and then write his message using a reference table containing the different names of these spirits. Finally, he should conjure these spirits. As a result, the final cover text would be a sort of a long prayer containing the names of the angels with the secret message appearing as a pattern of letters within the words. The cover message is then sent by courier to the receiver who then also has to conjure the appropriate spirits in order to retrieve the secret message from the angels. For example, using every other letter in every other word in the following cover text: "padiel a**p**o**r**s**y** mesarpon o**m**e**u**a**s** peludyn m**a**l**p**r**e**a**x**o", would reveal, when deciphered, the secret message "prymus apex" [8]. As this book appeared to deal with magic and spiritualism, it later influenced occultism.

In the 16th century, an Italian astronomer, mathematician, and gambler called Gerolamo Cardano invented what later bore his name, the Cardano grille. The Cardano grille is simply a cardboard having some rectangular-form holes cut in it. The secret message is written in the holes, and then a text is written around the secret letters to create as much as possible an innocent message. In order to recover the secret message out of the text, the recipient has to place the correct grille over the text and then line up the holes of the grille with the words in the larger message to produce the hidden message [9].

In the 17th century, Sir Francis Bacon, an English philosopher, scientist, and statesman, devised a steganography method, he called the Baconian cipher. In his approach, a secret plaintext message is encoded by converting every letter in it into another set of letters selected from a group of five letters mainly composed of only "A"s and "B"s. This conversion is done based on a predefined alphabet of the Baconian cipher. For instance, the word "car" is encoded using the Baconian cipher as "AAABA AAAAA BAAAA". In the long debate of proving William Shakespeare authorship, some scholars and conspiracy theorists suspected that Sir Francis Bacon was the true author of Shakespeare's literature work. They backed up their hypothesis with the mere proof that Bacon left his signature ciphered in Shakespeare's work. It was argued that the Latin word "honorificabilitudinitatibus", which was found in the "Love's Labour's Lost" play of Shakespeare, is actually an anagram, that when deciphered leads to the Latin phrase "Hi ludi F. Baconis nati tuiti orbi" which can be translated in English to "These plays, the offspring of F. Bacon, are preserved for the world" [10], or to "These games F. Bacon gave birth to the world" using Google translator (with "games" maybe referring to "plays").

During World War II, several steganography techniques were used to convey secret military data among fighting parties. One of these techniques was Microdots which refers to complete documents and images reduced to the size of a dot. Microdots needed to be embedded in a paper which is then sent to the recipient. The recipient, on the other side, had to enlarge the size of the printed microdot so that the hidden information gets revealed. Another technique is the invisible ink which refers to a chemical substance used for writing text that can be made invisible and then visible by some means such as heat or ultra violet light. Furthermore, null cipher is an old, yet a popular technique for steganography which constructs an unsuspicious plaintext having the secret message as part of its characters. One of null cipher basic examples is to take the first letter of each word of the secret message and spawn a set of new words out of it [11, 12]. For instance, taking the third letter in every word of the following sentence "Fishing freshwater bends and saltwater coasts rewards anyone feeling stressed. Resourceful anglers usually find masterful leapers fun and admit swordfish rank and overwhelming any day.", would reveal the following secret message "Send lawyers guns and money" [8].

Although all the above techniques are now obsolete and easy to be broken by recent technologies, they were the spark for the science of information hiding and the groundwork of modern steganography that took off with the advent of digital computers.

3. Steganography in the 20th Century before the Computer Age

Steganography is not only related to digital computers as it has been seamlessly employed in various forms and in different domains before the rise of computers and the Internet. More specifically, it has been intensively used in military and espionage during warfare. Several techniques have been experimented; they include Null cipher [53], Jargon Code [13], Anamorphosis [14], Semagram [15], Invisible Ink [16], and Microdots [17].

3.1 Null Cipher

NULL cipher [53] is in essence based on a selection technique that constructs an unsuspicious plaintext having the secret message as part of its characters. For example, in World War I, the German embassy in United States sent a telegraph to Berlin stating "**P**resident's **E**mbargo **R**uling **S**hould **H**ave **I**mmediate **N**otice. **G**rave **S**ituation **A**ffecting **I**nternational **L**aw. **S**tatement **F**oreshadows **R**uin **O**f **M**any **N**eutrals. **Y**ellow **J**ournals **U**nifying **N**ational **E**xcitement **I**mmensely". It is by reading the first character of every word of this statement, that the secret message can be revealed as "Pershing Sails from NY June I" [54].

3.2 Jargon Code

Jargon code is used within a language that is understood by a group of individuals but is nonsense to others. Jargon codes are often manifested by using an innocuous dialog that mimics special meaning to those who understand the language. For example, a group of spies might devise their own jargon code to secretly communicate such as using the phrase "The bird is flying" to mean that "the target is moving" or "he is having dinner tonight at 6:00" to mean that "The meeting is to take place tonight at 6:00", etc.

3.3 Anamorphosis

Anamorphosis is a Greek word which means "change shape". It is a technique which distorts an image by changing its perspective. It is often used to cover a secret shape or image within a carrier image. To reconstitute the secret image, one must look at the carrier image from a particular angle or using a special device.

3.4 Semagram

Semagrams hides information by using symbols, signs, or visual objects. It is more like an indicator of a larger, previously agreed upon message. For example, Bob wants to tell Alice that the party will take place on Friday. A semagram could be a postcard with a picture of a Chevrolet car, which Bob and Alice have already agreed that a Chevrolet car means affirmative (the party will take place); while, a Ford car means negative (the party won't take place).

3.5 Invisible Ink

Invisible ink is a substance that requires heat, light, or a special liquid to become invisible. It is used for secret writing, which become invisible sometime after its application. Invisible ink has been used for years in espionage. The ink liquid can be organic such as milk, vinegar, lemon, juice, or even more advanced chemicals such as Chloride, Ammonia, and Copper sulfate.

3.6 Microdots

Microdots are large photographic image that has been reduced in size to that of a dot normally around 1mm in diameter. In order to create a microdot, first, a photograph of the secret message is taken, then reduced to the size of a postage stamp, and further shrunk down to 1 millimeter using reverse microscope. Next, the negative film is created, generating the final microdot image. Microdots were used during World War I and World War II by military troops to pass messages via insecure postal channels.

4. Digital Steganography in the Computer Age

Fundamentally, steganography refers to "secret writing" in Greek, and is the art and science of hiding information inside innocuous files such as images, audio files, and video files, in ways that avoid the detection of the hidden information [18]. The outcome of steganography is a covert channel of communication through which secret data can be transmitted in total secrecy avoiding drawing eavesdroppers' suspicions. In current practice, steganography is used to hide secret data such as text messages into a carrier files such as images while maintaining the size and quality of the carrier file.

As an algorithmic model, steganography can be thought as the "Prisoners' Problem" [19]. In this model, two prisoners put in jail, Alice and Bob, want to communicate about an escape plan. The challenge is that they can only communicate through the warden of the prison, Wendy, who prohibits both Alice and Bob to communicate in code using standard cryptography. As a result, Alice and Bob devise a new method for secret communication called steganography. It is about hiding the message that needs to be communicated in an innocent-looking image. This image is then handed in by Bob to the warden Wendy to pass it along to Alice. Wendy, looking at the image, would not notice anything suspicious and would subsequently pass it along, not knowing that the pixels of the image encode the secret message. Alice, after receiving the image, would recover the secret message of Bob as she knows how the trick works. Formally, the steganography model can be mathematically defined as follows [20]: the original file into which the secret message is to be concealed is denoted by A, the secret message to hide by M, and the carrier file by C. Actually, the carrier file C is visually or audibly identical to the original file A but with one difference is that it has the message M embedded inside it. The steganography encoding algorithm which is used to cover a secret message M into a file A, is denoted by S(A, M, Enc)=C where "Enc" denotes the encoding mode. On the other hand, the steganography decoding algorithm which is used to recover the secret message M out of the carrier file C, is denoted by S(C, Dec)=M where "Dec" denotes the decoding mode. Figure 1 depicts the basic mathematical model of steganography.

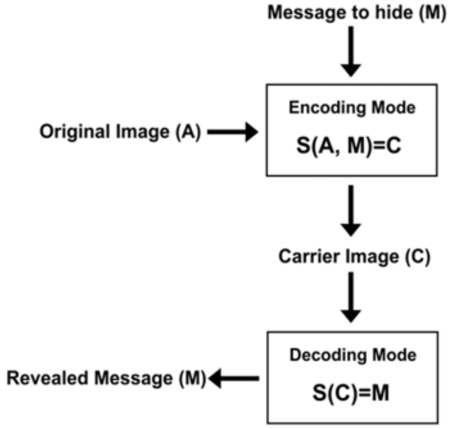

Figure 1. Mathematical Model of Steganography

5. Elements of Digital Steganography

Algorithmically, steganography has two processes, one for covering and one for uncovering secret data. The covering process is about hiding overt data into a cover medium, also known as stego or carrier file. In contrast, the uncovering process is just the reverse; it is about extracting the covert data from the carrier file and returning them back to their original state. Fundamentally, modern digital steganography is governed by five key elements. They are as follows [21]:

1. Covert Data: Often known as the payload and refers to the overt data that need to be covertly communicated or stored. The covert data can be anything convertible to binary format, from simple text messages to executable files.
2. Carrier Medium: It is basically a file into which the covert data are concealed. The carrier medium can be any computer-readable file such as image, audio, video, or text file.
3. Stego File: Sometimes called package, it is the resulting file which has the covert data embedded into it.
4. Carrier Channel: It denotes the file type of the carrier, for instance, BMP, JPG, MP3, PDF, etc.

5. Capacity: It denotes the amount of data the carrier file can hide without being distorted.

6. The Applications of Steganography

The applications of steganography are diverse; they include secret communication, digital watermarking, and data integrity [22, 23].

Secret Communication: Characteristically, transmitting a cryptographic message may raise unwanted suspicions as it travels overtly. Besides, cryptography is restricted by law in many situations. Contrariwise, steganography does not publicize secret communication; and hence, it evades the message from being detected and recovered by malicious parties.

Digital Watermarking: A secret copyright mark also known as digital watermark can be embedded inside an image to verify its authenticity. A digital watermark can be used to identify the owner of an intellectual property such as an image, audio, or video file. It also helps fight against copyright infringements as it detects the source of illegally copied materials.

Data Integrity: It refers to the assertion of data that they are accurate after they have been transmitted over the Internet and received by the recipient. Using steganography, a special mark can be embedded within an image file to determine that no variations, compromises, or damages have occurred to the image after it is received by the recipient. This form of steganography is referred to as Digitally Signing images so as to be able to confirm their reliability at any time.

7. Steganography Requirements

The effectiveness of a steganography system is determined by three requirements: The capacity of data that can hidden without distorting the carrier file; the imperceptibility of the carrier file after hiding the secret data into it; and the irrecoverability of the hidden data in case they were detected [24, 25]. Once either one of these requirements is compromised by third parties, the steganography system is defeated.

Hiding Capacity: It determines the number of bytes that can be covered within the carrier file without distorting or damaging it. For instance, in image steganography, it is important to hide as much as possible data inside the carrier image without increasing its brightness, without making it blurry, without pixelizing it, and without changing its size. This would be a key element in making the hidden data imperceptible and the carrier image innocent and unsuspicious.

Imperceptibility: It refers to the ability of the steganography algorithm to hide data in an undetectable way so much so that no one can see any visible artifacts or distortions in the carrier file. It therefore avoids drawing suspicions and obscures the fact that a secret communication is taking place.

Irrecoverability: It refers to how much an intercepted carrier file can be easily decoded and reverse-engineered so as to extract the data hidden inside it. An irrecoverable steganography algorithm makes it hard for eavesdroppers and unauthorized third parties to recover the hidden data from the carrier file despite knowing that steganography has been employed.

8. Methods of Hiding

Steganography can be achieved by means of three types of techniques: injection, substitution, and generation [26].

Injection: The injection technique implants the data to hide in the insignificant part of the carrier file, which is normally ignored by operating systems and software applications. For example, most computer files comprise what so called an end-of-file marker or EOF for short, which indicates that no more data can be read from a data source. Likewise, an executable file usually ends with an EOF marking the end of binary instructions. Another example is the PDF file which ends with an EOF indicating to the reader application that no more pages are to be fetched and that the file has ended. Steganography by injection exploits the EOF section and injects secret data after the EOF marker which eventually has no side effect on the carrier file and is often disregarded by the execution environment.

Substitution: The substitution technique substitutes the insignificant bits in the carrier file with the bits of the data to hide. Insignificant bits are those bits that can be modified without damaging the quality or destroying the integrity of the carrier file. For example, in audio files, every unit of sound is made up of a sequence of bits. If the least significant bit of this sequence is modified, its impact is minimal on the perceptible sound so much so that the human ear cannot tell the difference between the original version and the altered one. This technique takes advantage of the limited capabilities of the human auditory system (HAS) which cannot recognize two sounds that are slightly not alike.

Generation: The generation technique reads the data to hide and generates out of them a new set of data. It is a dynamic method of creating a carrier file based on the information contained in the data to hide. For example, one generation technique would take the message to hide and turn it into colored pixels that can eventually make up a real image, such as turning letter A into the green color, letter B into the yellow color, and so forth. The result is a dotted picture more like a fractal image exhibiting odd patterns and colored blots.

9. Image Steganography

To a computer, an image file is an array of numbers that represent color intensities each of which denotes a pixel, short for picture element. In 16-bit images, every sequence of 16 bits makes up a pixel. In 24-bit images, every sequence of 24 bits makes up a pixel, and so forth. A 24-bit pixel has more bits than its 16-bit counterpart and thereby it can represent more colors. An image size is often designated by the term resolution denoted by the number of horizontal and vertical pixels that compose the image. For instance, a typical image of size 800x600 pixels and of color depth 16-bit (or 2^{16}=65553 different color intensities per pixel) contains about 800*600*16 = 7680000 bits = 960000 bytes = 937 kilobytes. This is only true for uncompressed image formats such as BMP and TIFF. In contrast, the size of compressed image formats such as JPG or PNG can be relatively much smaller as they use a compression algorithm that eliminates the color redundancies in the image; and thus, reducing its total size. All color intensities are essentially derived from three basic color components called channels, they are the red, green, and blue color channels often abbreviated as RGB. In 24-bit images, three 8-bit color channels are used to represent single pixel color intensity, such as, 8 bits for the red channel, 8 bits for the green channel, and 8 bits for the blue channel. For instance, a blue pixel is represented as p={R=0; G=0; B=255} (2^8=255 which means 255 different blue color combination of intensities). A green pixel is represented as p={ R=0; G=255; B=0}, and a red pixel is represented as p={R=255; G=0; B=0}. Now, in order to represent the spectrum of colors, different color values are set simultaneously for the channels R, G, and B. For example, p={R=224; G=176; B=255} represents the color Mauve, p={R=255; G=215; B=0} represents the color Golden, p={R=0; G=255; B=255} represents the color Cyan, and so on and so forth. Characteristically, a small fluctuation in these color values is often undetectable by the Human Visual System (HVS) as it cannot differentiate between two images whose color intensities in the high frequency spectrum are marginally unalike. For instance, the colors p={R=255; G=0; B=0}, p={R=254; G=0; B=0}, and p={R=253; G=0; B=0} technically represent different color intensities; however, to the HVS, they are equally red. The goal of image steganography is to exploit this weakness of the HVS for the sake of information hiding [27].

So far, massive research work has been conducted in the development of steganography for digital images. One of the earliest techniques is the LSB technique which obscures data communication by inserting the secret data into the insignificant parts of the pixels of an image file, more particularly, into the least significant bits (LSB) [28]. The modified version of the image, which is called carrier file or stego file, is then sent to the receiver through a public channel. The foremost requirement of the LSB technique is that it should not exhibit any visual signs in the carrier image so as to not give any indications that secret data are being communicated covertly. Basically, the LSB technique is an insertion-based image steganography method that embeds secret data into uncompressed computer image files such as BMP and TIFF. In this technique, the data to hide are first converted into a series of bytes, then into a series of smaller chunks each of which is of size *n* bits. Then, *n* LSBs of the pixels of the carrier image are replaced by each of the chunks of the original data to hide. The ultimate result of this operation is a carrier image carrying the secret data into the LSBs of its pixels. As the color values that are determined by LSBs are insignificant to the naked eye, it is hard to tell the difference between the original image and the tampered one, taking into consideration that no more than a certain number of LSBs were used to conceal the secret data; Otherwise, visual artifacts and damages would be produced in the carrier image which would in turn draw suspicions and raise attention about something unusual in the carrier image. For instance, in 24-bit True Color BMP images, using more than three LSBs per color component to hide data may result in perceptible artifacts in the carrier image [29]. As an illustration for the LSB technique, let's say that the letter H needs to be hidden into an 8-bit grayscale bitmap image. The ASCII representation for letter H is 72 in decimal or 01001000 in binary. Assuming that the letter H is divided into four chunks each of 2 bits, then four pixels are needed to totally hide the letter H. Moreover, assuming that four consecutive pixels are selected from the original image whose grayscale values are denoted by P_1=11011000, P_2=00110110, P_3=11001111, and P_4=10100011, then substituting every two LSBs in every of these four pixels by a 2-bit chunk of the letter H, would result in a new set of pixels denoted by P_1= 110110**01**, P_2=001101**00**, P_3=110011**10**, and P_4=101000**00**. Despite changing the actual grayscale values of the pixels, this has little impact on the visual appearance of the carrier image because characteristically, the Human Visual System (HVS) cannot differentiate between two images whose color values in the high frequency spectrum are marginally unalike [30].

On the other hand, other steganography techniques and algorithms for digital images have been proposed and researched both in spatial and frequency domains. They include masking and filtering [31], encrypt and scatter [32], transformation [33], and BPCS [34] techniques.

Masking and Filtering Technique: This technique is based on digital watermarking but instead of increasing too much the luminance of the masked area to create the digital watermark, a small increase of luminance is applied to the masked area making it unnoticeable and undetected by the naked eye. As a result, the lesser the luminance alteration, little the chance the secret message can be detected. Masking and filtering technique embeds data in significant areas of the image so that the concealed message is more integral to the carrier file.

Encrypt and Scatter Technique: This technique attempts to emulate what is known by White Noise Storm which is a combination of spread spectrum and frequency hopping practices. Its principle is so simple; it scatters the message to hide over an image within a random number defined by a window size and several data channels. It uses eight channels each of which represents 1 bit; and consequently, each image window can hold 1 byte of data and a set of other useless bits. These channels can perform bit permutation using rotation and swapping operations such as rotating 1 bit to the left or swapping the bit in position 3 with the bit in position 6. The niche of this approach is that even if the bits are extracted, they will look garbage unless the permutation algorithm is first discovered. Additionally, the encrypt and scatter technique employs DES encryption to cipher the message before being scattered and hidden in the carrier file.

Transformation Technique: This technique is often used in the lossy compression domain, for instance, with JPG digital images. In fact, JPG images use the discrete cosine transform (DCT) to perform compression. As the cosine values cannot be calculated accurately, the DCT yields to a lossy compression. The transformation-based steganography algorithms first compress the secret message to hide using DCT and then integrate it within the JPG image. That way, the secret message would be integral to the image and would be hard to be decoded unless the image is first decompressed and the location of the hidden message is recovered.

BPCS Technique: This technique which stands for Bit-Plane Complexity Segmentation Steganography, is based on a special characteristic of the Human Visual System (HVS). Basically, the HVS cannot perceive a too complicated visual pattern as a coherent shape. For example, on a flat homogenous wooden pavement, all floor tiles look the same. They visually just appear as a paved wooden surface, without any indication of shape. However, if someone looks closely, every collection of tiles exhibits different shapes due to the particles that make up the wooden tile. Such types of images are called vessel images. BPCS Steganography makes use of this characteristic by substituting complex regions on the bit-planes of a particular vessel image with data patterns from the secret data.

10. Audio Steganography

Fundamentally, audio steganography is about hiding digital data such as text messages, documents, and binary files into audio files such as WAV, MP3, and RM files. The output audio file is called the carrier file and is the only intermediate to be sent to the receiver. Audio steganography takes advantage of the Human Auditory System (HAS) which cannot hear the slight variation of audio frequencies at the high frequency side of the audible spectrum; and thus, audio steganography can exploit and use this type of frequencies to hide secret data without damaging the quality of the audio file or changing its size [36].

The popularity and the abundance of audio files make them qualified to convey secret information. As a result, many researchers started to investigate how audio signals and audio properties can be used in the domain of information hiding [37]. Several approaches were conceived, the most popular ones are Least Significant Bit [38], Echo hiding [39], Hiding in Silence Interval [40], Phase Coding [41], Amplitude Coding [42], Spread Spectrum [43], and Discrete Wave Transform [44]:

Least Significant Bit (LSB): Basically, the LSB technique is based on embedding each bit from the data to hide into the rightmost bits of every audio sample of the carrier audio file. The LSB technique takes advantage of the HAS which cannot hear the slight variation of audio frequencies at the high frequency side of the audible spectrum. The LSB technique allows high embedding rate without degrading the quality of the audio file. Furthermore, it is relatively effective and easy to implement. However, it main drawback is that the secret data are concealed in a very predictable way, making them easy to be recovered by attackers.

Echo hiding: In this technique, the secret data are embedded into the audio signals as a short acoustic echo. In fact, an echo is a replication of sound, however, received by the listener some time after the original sound. As the echo is audible, its amplitude must be decreased so that it becomes imperceptible. In order to hide data, bits whose values are 0 are represented by an echo delayed 1ms; bits whose values are 0 are represented by an echo delayed 2ms. The limitation of echo hiding technique is the low hiding capacity as it would be computationally intensive to insert echo for every bit to hide.

Hiding in Silence Interval: This technique inserts a silence interval in the original audio signal to embed the secret data. The values that represent the length of the silence intervals are decreased by some value such as $0<value<2^n$ where n is the number of bits required to represent an element from the data to hide. The carrier audio file is then sent to the receiver having the new lengths for its silence intervals. Recovering the data is done via *mod(altered_length; 2*^{n}*)*.

Phase Coding: This technique substitutes the phase of an audio sample with a reference phase that expresses the secret data. The remaining samples are attuned so as to preserve the relative level between different audio samples. Algorithmically, the audio signal is divided into smaller samples whose size is equal to the size of the message to hide. Then, DFT (Discrete Fourier Transform) is applied to generate a matrix of phases. Then, the phase alteration between the contagious samples is calculated. Afterwards, the absolute phases are changed so that to embed the secret data in phase vector of the first audio sample. Finally, the audio signal is rebuilt by computing the DFT using the new generated phase matrix and the original matrix. Consequently, the sound samples are grouped together yielding to a carrier audio signal that encodes the secret data into it.

Amplitude Coding: This technique conceals secret data in the magnitude speech spectrum while not distorting the carrier audio signal. It is based on searching for safe spectral regions in the signal whose magnitude speech spectrum is below a certain value. Besides, the carrier locations are selected based on how much they can badly affect the audio signal.

Spread Spectrum: This technique scatters the secret data over the frequency spectrum of the audio file using a specific code independent of the actual signal. Basically, secret data are multiplied by a code known to the communicating parties only, and then embedded in the carrier audio file. The advantage of Spread Spectrum method is its speed in covering data; however, its drawback is that it introduces noise and distortions to the audio file.

Discrete Wave Transform: In this technique secret data are embedded in the least significant bits of the wavelet coefficients of the audio signals. Often, secret data are chosen to be hidden in the integer wavelet coefficients and not in silent sections of the audio signal so as to promote the imperceptibility of the audio file. The disadvantage of Discrete Wave Transform is that secret data can be lost during the recovering process as this technique is not that accurate.

11. Text Steganography

Hiding information in plain text can be done in many different ways. Some techniques consist of changing the outline of the carrier text such as adding whitespaces or altering the case of certain characters so as to represent secret text [45]. Others, consist of relating the characters to hide with the characters of the carrier text, creating a reference dictionary that maps words from the secret text with words from the carrier text [46]. This section sheds the light on the various techniques used in text steganography including hiding by selection [47], hiding in HTML [48], line and word shifting [49], hiding using whitespace [50], semantic-based hiding [51], and abbreviation-based hiding [52] techniques.

Hiding by Selection: The selection technique selects certain characters in the carrier text to convey the characters of the secret message such as selecting the first character of every word in the carrier text or the second character of every other word. Now, in order to recover the concealed secret message, all first characters of the words of the carrier text are extracted and concatenated together, producing the exact original message. A variation of this technique can be performed by selecting the first character from the first word, the second character from the second word, the third character from the third word, and so forth, until the characters of the message to hide are exhausted. The drawback of this method is that it requires a huge volume of text to hide a small message of few words.

HTML Documents: Secret text can be easily concealed within HTML documents because HTML tags are case insensitive. For instance, the tags <a title="clients">, <a TITLE="clients">, and <a TitlE ="Center">, are all the same and have the same effect on the rendering of the document. Text steganography applied in HTML documents can be performed by changing the case of the letters that make up the HTML tags. In particular, the secret message is represented by the capital version of the tags' letters, or vice versa depending on the algorithm being used. As for the recovering process, all the capital letters from the HTML document have to be captured and concatenated together in order to produce the original covert message.

Line and Word Shifting: In this technique, text lines are shifted vertically and words are shifted horizontally by a fixed space of *n* inches. That way, the distance between lines and words would convey the hidden characters. For instance, letter A can be encoded as a 0.01 inch space between two text lines. Similarly, letter B can be encoded as a 0.02 inch space between two other text lines. In effect, this technique is more suitable for printed text than for digital text, since printed spaces can be physically measured unlike their digital counterparts.

Hiding using Whitespace: Its concept is very straightforward. A message to hide is first converted into a binary format. Then, every bit whose value is 1 is represented by an extra whitespace between a particular set of two words in the carrier text; whereas, every bit whose value is 0 leaves the original single whitespace between the next particular set of two words. For example, "the boy went to school today " can be deciphered as "101001". In fact, two spaces exist between "the" and "boy", between "went" and "to", and between "today" and the end of the sentence. This results in a bit of value 1 in positions 0, 2, and 5 respectively. In contrast, only a single space exists between "boy" and "went", between "to" and "school", and between "school" and "today". This results in a bit of value 0 in positions 1, 3, and 4 respectively. Basically, the whitespace technique is very suspicious as a normal reader would right away notice the existence of some extra whitespaces in the text. Additionally, this method cannot encode too much information especially in small text.

Semantic-Based Hiding: The semantic-based text steganography technique uses synonyms of words to hide the secret information in the carrier text. For instance, the secret message "the boy went to school today" can be encoded using the semantic approach as "the child went to college today".

Abbreviation-Based Hiding: This technique uses a lexical dictionary containing words along with their abbreviations. These abbreviations are either labeled 0 or 1. While performing steganography, if a word in the carrier text is found in the dictionary, it is substituted by its abbreviation based on the current bit to hide. Different values of bits have different corresponding abbreviations. Some examples of these abbreviations can be "ASAP" for "As Soon As Possible", "CU" for "see you", "gr8" for "Great" etc. A variation of this method is the one proposed by [46], which consists of changing the spelling of words based on their American and British spellings. For example, "Favorite" is designated by 1 while "Favourite" is designated by 0, and "Center" is designated by 1 while "Centre" is designated by 0.

SpamMimic: SpamMimic [55] is a generation-type steganography technique that revolves around generating a natural English text from the message to hide. It matches letters from the secret text with words from a precompiled dictionary, and then uses some stored grammatical rules to make the generated text grammatically correct. Below is a paragraph generated using SpamMimic by encoding the world "hello". Decoding back the same paragraph would result in the original secret word "hello".

"Dear Decision maker, we know you are interested in receiving amazing intelligence. This is a onetime mailing there is no need to request removal if you won't want any more. This mail is being sent in compliance with Senate bill 1625; Title 4; Section 302. THIS IS NOT MULTI-LEVEL MARKETING! Why work for somebody else when you can become rich as few as 55 DAYS. Have you ever noticed most everyone has a cellphone and more people than ever are surfing the web. Well, now is your chance to capitalize on this! WE will help YOU decrease perceived waiting time by 200% and turn your business into an E-BUSINESS! The best thing about our system is that it is absolutely risk free for you! But don't believe us. Mr Ames of Massachusetts tried us and says "My only problem now is where to park all my cars"! We are licensed to operate in all states! We beseech you - act now. Sign up a friend and your friend will be rich too! Thank-you for your serious consideration of our offer!"

12. Conclusions

This paper presented a comprehensive overview on steganography and on its different techniques that have been proposed in the literature during the last decades.

Nevertheless, the million dollar question remains: how easy is to break a steganography system and reverse-engineer its algorithm which would allow malicious third parties to recover secret data and expose the communicating parties. Imperceptibility and Irrecoverability are the true dark horses for an uncompromisable steganography system. Imperceptibility makes the secret data innocuous and undetectable; while, irrecoverability makes the secret data irretrievable in case they were detected. These two properties must be carefully addressed when crafting a steganography system.

Acknowledgments

This research was funded by the Lebanese Association for Computational Sciences (LACSC), Beirut, Lebanon, under the "Stealthy Steganography Research Project – SSRP2012".

References

[1] Peter Wayner, "Disappearing Cryptography", 3rd edition, Morgan Kaufmann, 2008.

[2] Johnson, N. F. and Jajodia, S., "Exploring steganography: Seeing the unseen", Computer Journal, vol. 31, no.2, pp.26–34, 1998.

[3] Katzenbeisser, S. and Petitcolas, F.A.P., "Information Hiding: Techniques for Steganography and Watermarking", Artech House, 2000.

[4] Abdelrahman Desoky, "Noiseless Steganography: The Key to Covert Communications", Auerbach Publications, 2012.

[5] J.C.Judge, "Steganography: past, present, future", SANS Institute publication, 2001.

[6] P. Moulin, R. Koetter, "Data-hiding codes", Proceedings of the IEEE 93, vol.12, pp. 2083-2126, 2005.

[7] Johannes Trithemius, Steganographia, Frankfurt: Johannes Berner, 1606.

[8] Greg Kipper, Investigator's Guide to Steganography, Auerbach Publications, 2004.

[9] David Kahn, "The Codebreakers - The Comprehensive History of Secret Communication from Ancient Times to the Internet", 1996.

[10] K. K. Ruthven, "Faking Literature", Cambridge University Press, 2011.

[11] S. Lyu, H. Farid, "Steganalysis using higher-order image statistics", IEEE Transactions on Information Forensics and Security vol. 1, no.1, pp.111–119, 2006.

[12] Simon Singh, "The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography", Anchor Reprint edition, 2000.

[13] Kahn, D., "The Codebreakers", Macmillan, New York, 1967.

[14] Kent, P., Art of Anamorphosis, http://www.anamorphosis.com/.

[15] Zim, H.S., "Codes and Secret Writing", William Morrow, New York, 1948.

[16] Chaum, David, Richard Carback, Jeremy Clark, Aleksander Essex, Stefan Popoveniuc, Ronald L. Rivest, Peter Y. A. Ryan, Emily Shen, Alan T. Sherman, "Scantegrity II: End-to-End Verifiability for Optical Scan Election Systems using Invisible Ink Confirmation Codes", Proceedings of USENIX/ACCURATE EVT, 2008.

[17] White, William, "The Microdot: History and Application", Williamstown, NJ: Phillips Publications, 1992.

[18] R.J. Anderson, F.A.P. Petitcolas, "On the Limits of Steganography", IEEE Journal Selected Areas Communnication, 1998.

[19] G. J. Simmons, "The prisoners' problem and the subliminal channel," in Advances in

Cryptology: Proceedings of Crypto 83, pp. 51–67, 1984.

[20] K. Bailey, K. Curran, "Steganography The Art of Hiding Information", BookSurge Publishing, 2005.

[21] Jessica Fridrich, "Steganography in Digital Media: Principles, Algorithms, and Applications", Cambridge University Press, 2009.

[22] Rainer Böhme, "Advanced Statistical Steganalysis", Springer, 2012.

[23] Lin, E.T. and Delp, E.J., "A Review of Data Hiding in Digital Images", Purdue University.

[24] B. Pfitzmann, "Information hiding terminology", in Information Hiding, First International Workshop, vol. 1174, pp. 347–350, Springer, 1996.

[25] Eric Cole, "Hiding in Plain Sight: Steganography and the Art of Covert Communication", Wiley Publishing, 2003.

[26] Marvel, L.M., Boncelet, C.G., Jr., Retter, C.T., "Reliable Blind Information Hiding for Images", International Workshop on Information Hiding, 1998.

[27] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", 3rd edition, Prentice Hall, 2007.

[28] J. R. Smith and B. O. Comisky, "Modulation and information hiding in images," in information hiding, first international workshop, Germany: Springer-Verlag, vol. 1174, pp. 207–226, 1996.

[29] Fabien A. P. Petitcolas, Ross J. Anderson and Markus G.Kuhn, "Information Hiding - A Survey", Proceedings of the IEEE, special issue on protection of multimedia content, vol. 87, no.7, pp.1062-1078, 1999.

[30] Tovée, Martin J., "An introduction to the visual system", Cambridge University Press, 2008.

[31] Gruhl, D. and Bender, W., "Information Hiding to Foil the Casual Counterfeiter", Information Hiding Workshop, 1998.

[32] Frank Shih, "Digital Watermarking and Steganography: Fundamentals and Techniques", CRC Press, 2007.

[33] T. Zhang and X. Ping, "A Fast and Effective Steganalytic Technique Against JSteg-like Algorithms", Proceedings of the 8th ACM Symposium, Applied Computing, ACM Press, 2003.

[34] Eiji Kawaguchi and Richard O. Eason, "Principle and applications of BPCS-Steganography", Proceedings of SPIE: Multimedia Systems and Applications, vol.35, no.28, pp.464-473, 1998.

[35] Kandel ER, Schwartz JH, Jessell TM, "Principles of Neural Science", 4th edition, McGraw-Hill, 2000.

[36] Nedeljko Cvejic, Tapio Seppben, "Increasing the capacity of LSB-based audio steganography", FIN-90014, Finland, 2002.

[37] F.Djebbar, B. Ayad, K. Abed-Meraim and H. Hamam, "A view on latest audio steganography", 7th IEEE International Conference on Innovations in Information Technology, 2011.

[38] K. Gopalan, "Audio steganography using bit modification", Proceedings of International Conference on Multimedia, vol. 1, pp.629-632, 2003.

[39] D. Gruhl and W. Bender, "Echo hiding", Proceeding of Information Hiding Workshop, pp. 295-315, 1996.

[40] S. Shirali-Shahreza, M. Shirali-Shahreza, ”Steganography in Silence Intervals of Speech", proceedings of the Fourth IEEE International Conference on Intelligent Information Hiding and Multimedia Signal, pp. 605-607, 2008,

[41] Yin-Cheng Qi, Liang Ye, Chong Liu, "Wavelet Domain Audio Steganalysis for Multiplicative Embedding Model", Proceedings of the 2009 International Conference on Wavelet Analysis and Pattern Recognition, 2009.

[42] F. Djebbar, B. Ayad, K. Abed-Meraim, H. Habib, "Unified phase and magnitude speech spectra data hiding algorithm", Journal of Security and Communication Networks, John Wiley and Sons, 2012.

[43] Khan, K., "Cryptology and the origins of spread spectrum", IEEE Spectrum, vol. 21, pp. 70-80, 1984.

[44] N. Cvejic, T. Seppanen, "A wavelet domain LSB insertion algorithm for high capacity audio steganography", Proc. 10th IEEE Digital Signal Processing Workshop and 2nd Signal Processing Education Workshop, pp. 5355, 2002.

[45] Mohammad Shahreza, "Text Steganography by Changing Words Spelling", ICACT, 2008.

[46] C. Zhi-li, H. Liu-sheng, Y. Zhen-shan, Z. Xin-xin, Z. Xue-ling, "Effective Linguistic Steganography Detection", IEEE 8th International Conference on Computer and Information Technology Workshops, 2008.

[47] Adnan Gutub and Manal Fattani, "A Novel Arabic Text Steganography Method Using Letter Points and Extensions", World Academy of Science, Engineering and Technology, Vol. 27, 2007.

[48] Sudeep Ghosh, "StegHTML: A message hiding mechanism in HTML tags", 2007.

[49] S. Low, N. Maxemchuk, J. Brassil, L. O'Gorman, "Document marking and identification using both line and word shifting", Proceedings of the 14th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 95, 1995.

[50] W. Bender, D. Gruhl, N. Morimoto, A. Lu, “Techniques for data hiding”, IBM Systems Journal, vol. 35, no 3, pp. 313-336, 1996.

[51] M. H. Shirali-Shahreza, M. Shirali-Shahreza, “A New Synonym Text Steganography”, IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008.

[52] M. Shirali-Shahreza, M. H. Shirali-Shahreza, “Text Steganography in Chat”, 3rd IEEE/IFIP International Conference in Central Asia on Internet, 2007.

[53] Chun-Shien Lu, "Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property", Idea Group Publishing, 2005.

[54] Mark Stamp, “Information Security-Principles and Practice”, Wiley Student Edition, 2006.

[55] Peter Wayner, SpamMimic Tool, URL: http://www.spammimic.com, retrieved the 01-Nov-2012.