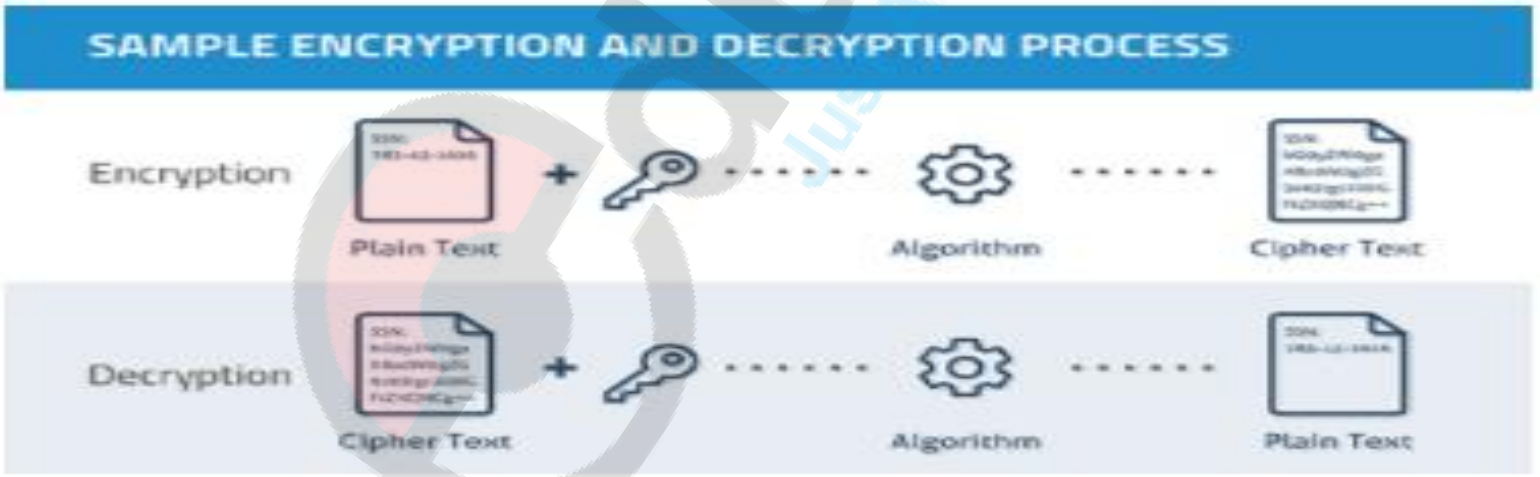
Chapter 2

DATA RECOVERY

Encryption and Decryption

Encryption is a method of turning meaningful information (known as the plaintext) into an obscured format (the ciphertext) by means of an algorithm (cipher). Encryption algorithms use a key to obscure the data (encryption) and to recover the plaintext (decryption).



Challenges in Digital Forensic

Digital forensics is a technique in the identification of computer based crimes. But digital forensics faces a few major challenges when it comes to conducting investigations.

Challenges of digital forensics can be categorized into three parts.

A) Technical challenges – e.g. differing media formats, encryption, steganography, anti-forensics, live acquisition and analysis.

B)Legal challenges – e.g. jurisdictional issues, privacy issues and a lack of standardized international legislation.

C)Resource challenges – e.g. volume of data, time taken to acquire and analyze forensic media.

Unlike many other sources of physical evidence, digital evidence is easy to modify, remove or hide, possibly without leaving tracks that might identify the criminal using advance Technology .

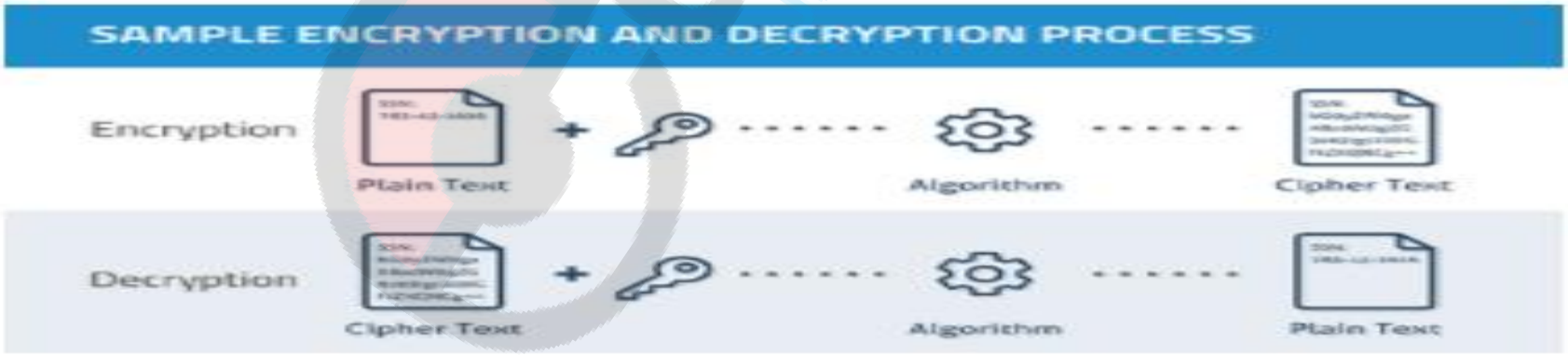
This kind of anti-forensics has become a major challenge for digital forensics

Anti-forensic techniques can be classified into following categories as :

1. Encryption
2. Steganography
3. Covert Channel
4. Data hiding in storage space
5. Residual Data Wiping
6. Tail Obfuscation
7. Attacking the tools
8. Attacking the investigators

1. **Encryption :** Encryption is process of scrambling information that can only be decoded and read by someone who has the correct decoding key. Encryption is used to hide or make the evidence unreadable on the compromised system.

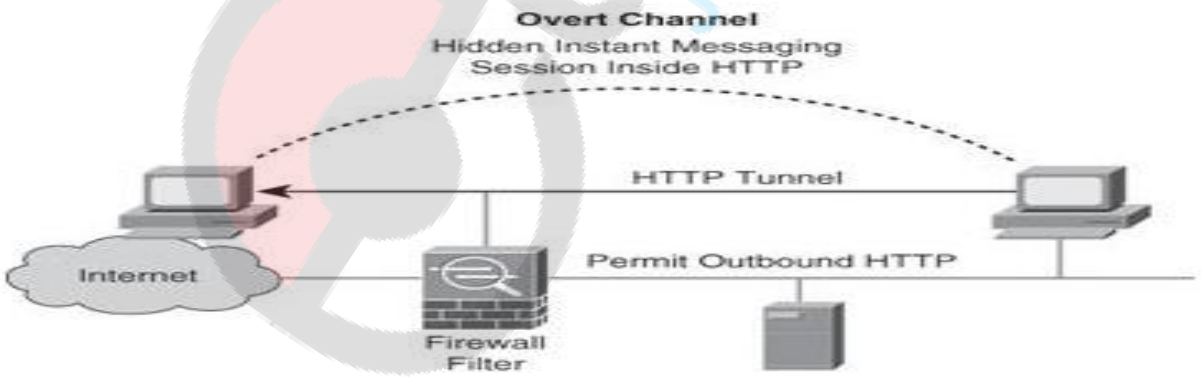
Attackers use many different encryption methods and in order to make the data usable, investigators have to decrypt the encrypted data. It is time consuming and sometimes the encrypted data cannot be decrypted.



2. **Steganography :** "Steganography is an encryption technique that can be used along with cryptography as an extra-secure method in which to protect data." (Janssen, 2014). Steganography is a technique that is used to hide any information inside a file carrier without modifying its outward appearance. Attackers use this steganography to hide their hidden data (payloads) inside the compromised system. When investigating computer crimes, the investigator has to identify these hidden data in order to reveal the information for further reference.

3. **Covert Channel :** "Covert channel in communication protocols allows attackers to hide data over the network and possibly bypass intrusion detection techniques. Typically, a network protocol is chosen and its header is modified to leak messages between attackers, exploiting the fact that few fields of the header are modified during transmission."

Attackers use these covert channels in order to maintain a hidden connection between the attacker and the compromised system. It is less identifiable.



4. **Data hiding in storage space**

Attackers hide some data inside storage areas and make them invisible to the usual system commands and programs. It makes the investigation more complex and more time consuming and sometimes data can be corrupted too. Rootkit is one of the most popular techniques used to hide data in storage space.

5. **Residual Data wiping**

When the attacker uses a computer for his goal, a few hidden processes (e.g. temporary files, history of commands) are running without the knowledge of the attacker. But an intelligent attacker can avoid this risk by wiping out the tracks that were made by his process and making the system work as if it has not been used for such a purpose.

6. Tail Obfuscation – attacking the tools

It is the most common technique is the obfuscation of the source of the attack. Here, the attacker uses some false information in order to mislead the investigator (e.g. false email headers, changing file extensions). Therefore sometimes the investigator might miss some data that has forensic value.

7. Resource Challenges

Depending on the scenario, the volume of data involved in the case might be large. In that case the investigator has to go through all the collected data in order to gather evidence. It may take more time for the investigation. Since time is a limiting factor, it becomes another major challenge in the field of digital forensics.

Data sources which are damaged cannot be easily used in investigations. So it is a major issue when an investigator finds a valuable source that is not usable.

8. **Legal Challenges**

Privacy is also important to any organization or victim. In many cases it may be required that the computer forensics expert share the data or compromise privacy to get to the truth. A private company or an individual user might generate lots of private information in their day to day usage. So asking an investigator to examine their data might risk their privacy being revealed.

Recovery Deleted Files

The widespread use of computers means they may be used in the commission of a crime or be a target. Computers can be used in committing fraud, stealing an organization's formulas or property, damaging other computers, sexual harassment, cyber terrorism, etc. Criminals who misuse computers for such illegal purposes might clean out the computer before they depart by deleting everything or reformatting the hard drive. In dealing with such cases, to conduct a forensic investigation an investigator needs to use a variety of techniques and proprietary forensic applications to examine the hard drive copy, and search hidden folders and unallocated disk space for copies of encrypted, deleted, or damaged files.

File and partition recovery allows you to recover critically important documents and other files that have been lost by accidental deletion, intentional deletion to conceal the evidence, a system crash due to a virus, a software malfunction, or even sabotage. Forensic recovery of deleted files and partitions is achieved by using recovery tools that identify the contents of these lost files on the hard drive and allow for recovering and preserving the data forensically.

- Recovery of deleted or lost files emptied from the Recycle Bin
- Disk recovery after a hard disk crash
- Data recovery from a hard drive that has been reformatted or repartitioned
- Recovery of financial records and other documents

- Recovery from a USB drive, memory card, memory stick, camera card, zip disk, floppy disk, or other storage media
- Recovery of files with the original date and timestamp
- Finding partitions automatically, even if the boot sector or FAT has been erased or damaged

As we know, deleted files are actually not deleted , but merely marked for deletion. For ex. From the FAT system, when a file or directory is deleted, the first letter of its file name is set to sigma character **(Ó)** or in hex ,**0×E5.** The deleted file located on a hard disk will remain intact until a new file or data.

Special tools can find these “intact” deleted file and recover them for review as soon as possible. After a file has been marked for deletion, each hard drive I/O could overwrite the data you want to recover.

In attempting to better understand the inner-workings of a computer system at its core level, it is necessary to understand the difference between a file system and an operating system and the functionality of each. Examples of different file systems include the following: FAT (FAT12, FAT16, FAT32) and NTFS which are implemented on Windows operating systems.

HFS and HFS+ which are implemented on Macintosh operating systems;
And finally, EXT2 and EXT3 which are implemented on Linux operating systems.

When trying to understand the mechanics of how computers actually manage data, it is necessary to delve into the physics of how a computer's OS uses the file system architecture.

1. Using Windows-based tools to recover files on FAT file systems :

To recover the files on FAT file system, the EnCase and FTK tools. Both these tools have built-in capability to automatically recover any files

The purpose of the experiment was to prove or disprove that variations in file systems affect the recoverability of files that were deleted and then wiped. The difference between the trials is the procedural steps undertaken when the 32 bit entry File Allocation Table (FAT32) or Windows New Technology File System (NTFS) file system architecture is used locally to store and manipulate data on the computer. Microsoft OS Vista (and all of its flavors) was not considered for this experiment because Vista cannot be installed on a FAT32 partition (http://support.microsoft.com/kb/927520/en-us, 2008) and all new copies of the OS are defaulted to run on NTFS.

Again one disk was formatted with FAT32 and the other formatted with NTFS using the patterns described previously. The assessment, acquisition, and analysis phases were followed exactly as before and the results of all trials were the same. Specifically, the results were duplicated on each file system loaded on all three sizes of hard drives for a total of 12 duplicated experimental trials.

However, if you are interested, simply find the Oxe5 character , and use a hex editor and rebuild the cluster chain(Dir/FAT/raw cluster) by hand.

2. Using Linux tools to recover files on FAT file systems :

The following capabilities should be provided by an operating system to value to a computer forensic examiner :

a) Supports a wide variety of file systems, including FAT12,FAT16,FAT36,NTFS,HPFS,Macintosh,OS/2,EXT2,EXT3 and UFS.
b) Recovers file slack and not allocate space. The improved loopback kernel makes it easy to recognise slack and not allocate drive space.
c) Provides an efficient, effective and accurate undelete utility.

d) Delivers keywords search competences and performs all function in a read only state on the file system being processed. The NASA kernel also provides the read only option to setup.

e) Handles compressed drives (DriveSpace , Dblspace and DriveSpace 3).

f) Delivers widespread checking and cataloging of all forensic activities.

g) Delivers for data authentication and reliability.

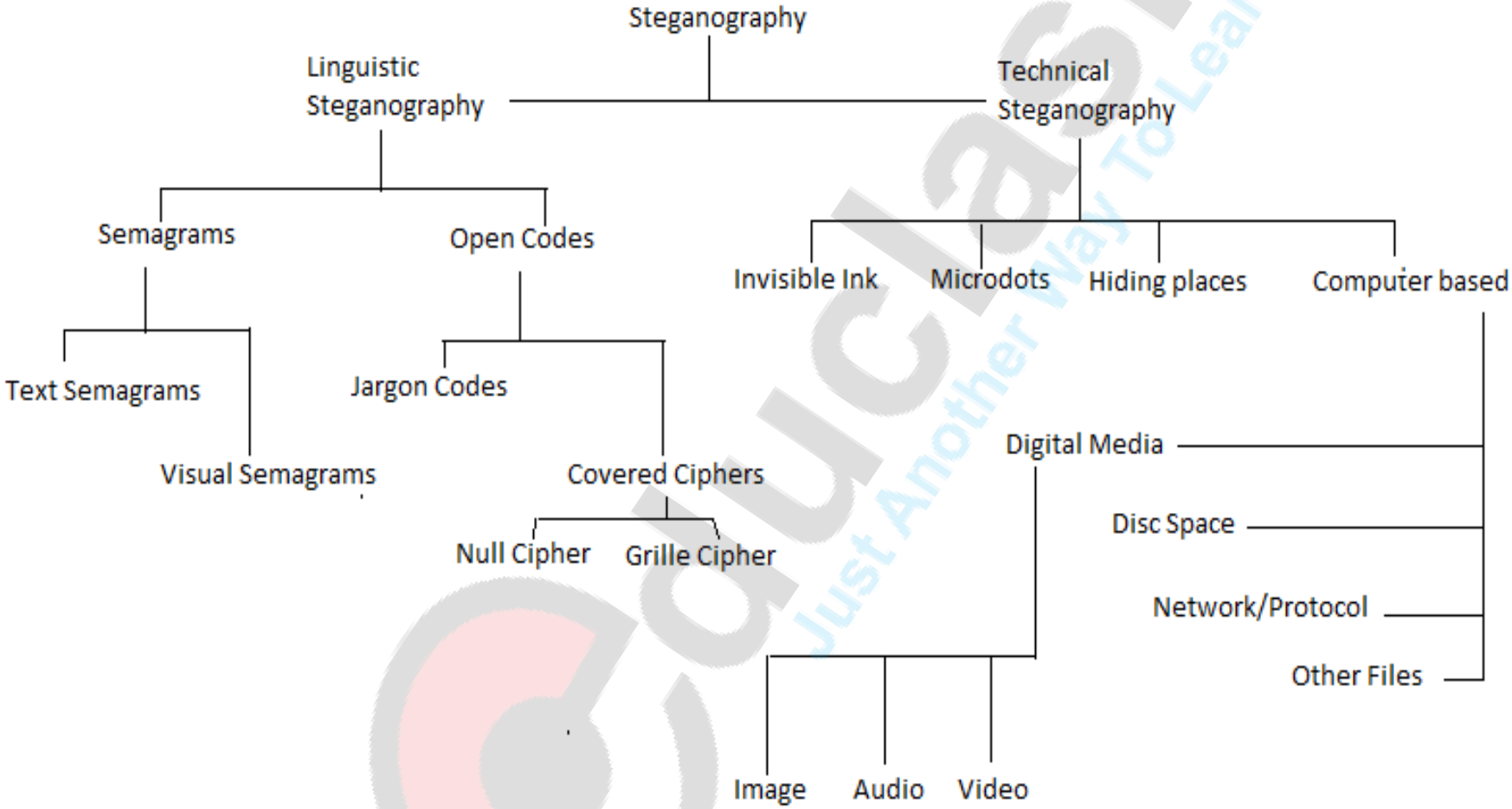
An Overview of Steganography

Steganography is the art of covered or hidden writing. The purpose of steganography is covert communication to hide a message from a third party. This differs from cryptography, the art of secret writing, which is intended to make a message unreadable by a third party but does not hide the existence of the secret communication. Although steganography is separate and distinct from cryptography, there are many analogies between the two, and some authors categorize steganography as a form of cryptography since hidden communication is a form of secret writing

Steganography hides the covert message but not the fact that two parties are communicating with each other. The steganography process generally involves placing a hidden message in some transport medium, called the carrier. The secret message is embedded in the carrier to form the steganography medium. The use of a steganography key may be employed for encryption of the hidden message and/or for randomization in the steganography scheme. In summary:

steganography medium = hidden message + carrier + steganography key

Fig.: steganography classification



Types of Steganography

1. Technical steganography uses scientific approaches to hide a message, such as the use of undetectable ink or microdots and other size-reduction approaches.
2. Linguistic steganography hides the message in the transporter in some non-obvious ways and is further categorized as Seagram's or open codes.

i) Semagrams hide information by the use of symbols or signs. A visual semagram uses innocent-looking or everyday physical objects to convey a message, such as doodles or the positioning of items on a desk or website.A textsemagram hides a message by amending the appearance of the transporter text, such as indirect changes in font size or type, adding extra spaces, or different flourishes in letters or handwritten text.

ii) Open codes hide a message in a legitimate carrier message in ways that are not obvious to an innocent observer. The transporter message is sometimes called the unconcealed communication whereas the hidden message is the concealed communication. This category is subdivided into jargon codes and covered ciphers.

-> Jargon code, as the name suggests, uses language that is understood by a group of people but is meaningless to others. Jargon codes include War chalking (symbols used to indicate the presence and type of wireless network signal), underground terminology, or an innocent conversation that conveys special meaning because of facts known only to the speakers. A subset of jargon codes is cue codes, where certain prearranged phrases convey meaning.

Iii)Covered or concealment ciphers hide a message openly in the carrier medium so that it can be recovered by anyone who knows the secret for how it was concealed. A grille cipher employs a template that is used to cover the carrier message. The words that appear in the openings of the template are the hidden message. A null cipher hides the message according to some prearranged set of rules, such as "read every fifth word" or "look at the third character in every word."

Steganography process

During the steganography process, the secret message will first be embedded into a cover-object with an embedded algorithm and stego-key to generate a stego-object. This stego object can then be transported via OSN (Online Social Network), email, website, blog, etc. To the intended receiver. The receiver then extracts the secret message using the extraction algorithm and stego-key.

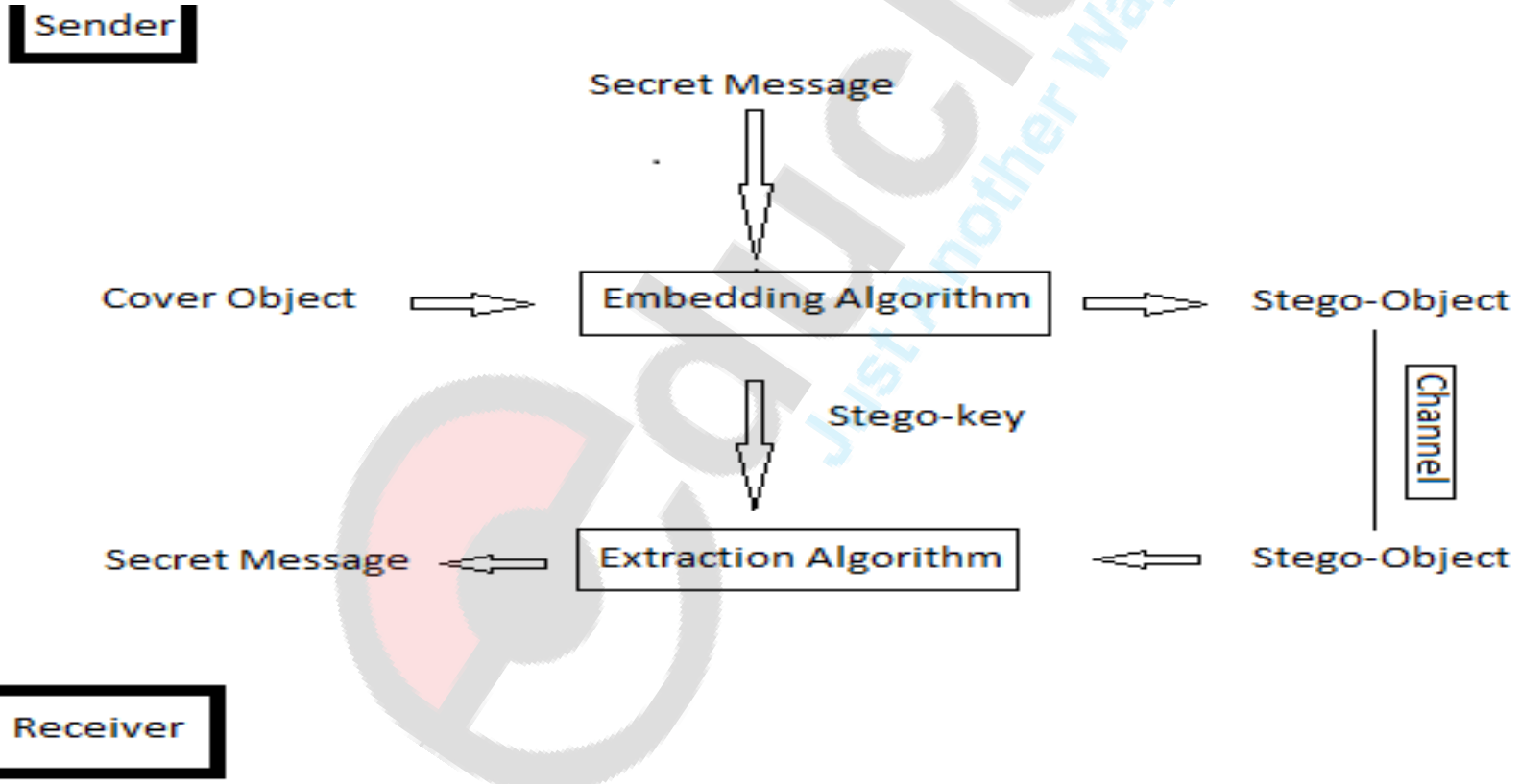
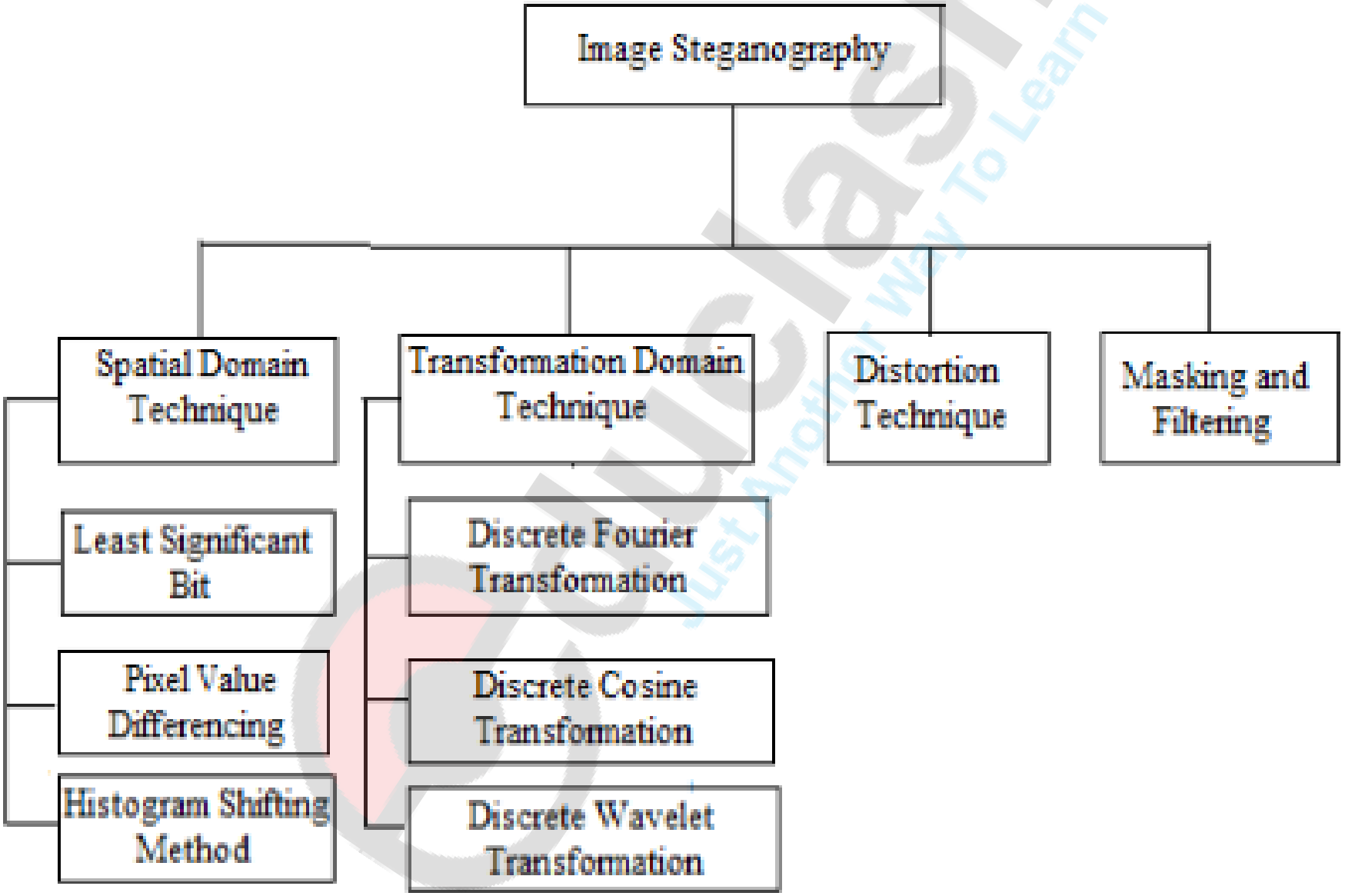


Image steganography



a)Spatial Domain Technique, for hiding the data some bits are directly changed into the image pixel values bitwise also include, the intensity of pixels and noise manipulation. There are many ways to perform embedding file in Spatial Domain, the easiest is Least Significant Bit (LSB)

b) Transformation Domain Technique, this technique has a threat like an image processing operations (compression, docking and enhancement) to this technique, because of that transformation domain is hides the secret message in the significant area in the carrier image.In this the first thing to do is convert the image from spatial domain into transformation domain and then the secret message is embedded into carrier image.These techniques hide data by mathematical

c) Distortion Technique, the information is embedded and store in signal distortion. This technique requires knowledge and carefulness in looking different between the original carrier image and stego image after information embeds during process of decoding.

d) Masking and Filtering, the image with 24bit of size or greyscale type is usually applying masking and filtering technique and using different applications to hide a message. Hiding information by marking the image, this technique is similar to paper watermark, this technique imparts information in a more significant image area than just hiding in the noise level.

Audio Steganography

Human Auditory System(HAS) is more sensitive than Human Visual System(HVS), that is one of the reasons that makes embedding message in audio file in any different method is more difficult than other formats. In Audio steganography, there is some format that can use as a cover media for embedding the file such as MP3, WAV, MIDI etc.

The following method that uses for embedding process in audio file are as:

a. LSB Coding
b. Parity Coding
c. Echo Data Hiding

a. LSB Coding: The least significant byte of carrier file is replaced with the bytes of the secret message. In this method rightmost bit is chosen for the replacement, because considered as the LSB as it has the least impact on the quality of file.

b. Parity Coding: The parity bit of the cover file is checked for similarity, if similarity exists then no action will be done and if the dissimilarity exists then any bit LSB will be slightly changed (cover file or secret message) to make parity equal .

c. Echo Data Hiding: The information is inserted by adding an echo sound to the cover file. Embedding data is expressed in terms of decay rate, initial amplitude and delay.

i) The initial amplitude is used to determine the original data sound.

ii) Decay rate is useful for determination of echo function to be made.

iii) The Offset function is used to determine the distance between the original speech signals with the echo that has been made.

Video Steganography

It is hiding or embedding message in the video is like an art of hiding information because the sender is not only hiding but how that message is prevented open by anyone except receiver. Hiding message in the video is part of the art of hiding information, that avert the revealing of hiding messages. Video-based steganography techniques are same like image based, its classified into spatial domain and frequency domain based methods

There are various techniques of video steganography

a. Video Steganography by LSB substitution using different polynomial equations, Least significant bit (LSB) operates on LSB bit from media files to embed into a carrier file

b. Video Steganography using 32 * 32 vector quantization of DCT, One method that is capable of operating 32 * 32 is the quantization of DCT vectors. The first step is that all videos are sliced differently in the number of images. After all the sliced images are passed to the 32 * 32-pixel management procedure followed by the through quantitative LSB methods.

c. A high-capacity video steganography based on integer wavelet transform, The proposed system uses integer wavelet transforms on the cover image to obtain stego-images . The proposed algorithm capacity is further enhanced by a set of confidential imagery is considered .

d. Video Steganography using dynamic cover generation.

A new steganographic system where the cover media itself is produced by the system instead of using the existing cover and some data is the cover itself and the rest is embedded in the cover.