Steganography and Image File Forensics

Module 13

Designed by Cyber Crime Investigators. Presented by Professionals.

C|HFI™
Computer Hacking Forensic Investigator

April 29, 2011

Researchers Hide Files on Unencrypted Disks

A process, devised by researchers in California and Pakistan, hides data on hard disks "in plain sight"

A new application can hide **sensitive data** on a hard drive without encrypting it or leaving any obvious signs that the data is present, according to the academic researchers who developed it.

The new software uses "**steganography**", the process of hiding information in plain sight, according to researchers from the University of Southern California and the National University of Science and Technology in Pakistan. The technique exploits the way the operating system normally splits up file data in numerous small chunks, called **clusters**, and writes them wherever there is free space on the hard drive.

The method employs a "covert channel" to encode sensitive information. Instead of the operating system writing small pieces of the file in random areas on the hard drive, the software chooses the positions according to a **secret code**. The person who wants to access the file needs to know the key to figure out where the fragments were written and re-assemble the clusters accordingly.

computing.co.uk

Cyber Crooks vs Digital Sleuths

12 May 2011

Speculation that the late Osama bin Laden and Al-Qaeda used steganography to distribute data to members of the terrorist group by hiding it in seemingly innocuous web sites may ultimately prove to be just that. But security and academic researchers have been looking at ways that steganography, along with digital watermarking and cryptography, can be used for a range of less malign purposes, including copyrighting and law enforcement activities.

Steganography involves embedding data, such as text or other images, into an image so that it is not immediately obvious to the human eye. It can be done in subtle ways, like slightly altering the color red green blue (RGB) or greyscale pixel depth, or replacing the eighth bit, or least significant bit (LSB), with something else.

Digital evidence pertaining to traffic enforcement and insurance claims is admissible in court, where forensics obtained from digital cameras and mobile phone cameras can be used to prevent fraud.

read more...

http://www.computing.co.uk

Scenario

A couple from Manchester was charged of plotting a terrorist act. Taxi driver Habib Ahmed of Elmfield Street, Cheetham Hill and his wife Mehreen Haji were held in police custody in London.

Ahmed, age 27, was accused of making computer records of possible terror targets and undergoing a course of weapons training at a Pakistani terror camp between April and June of 2006. 25-year-old Haji was accused of providing just under £4,000 to finance her husband's alleged terrorist activities.

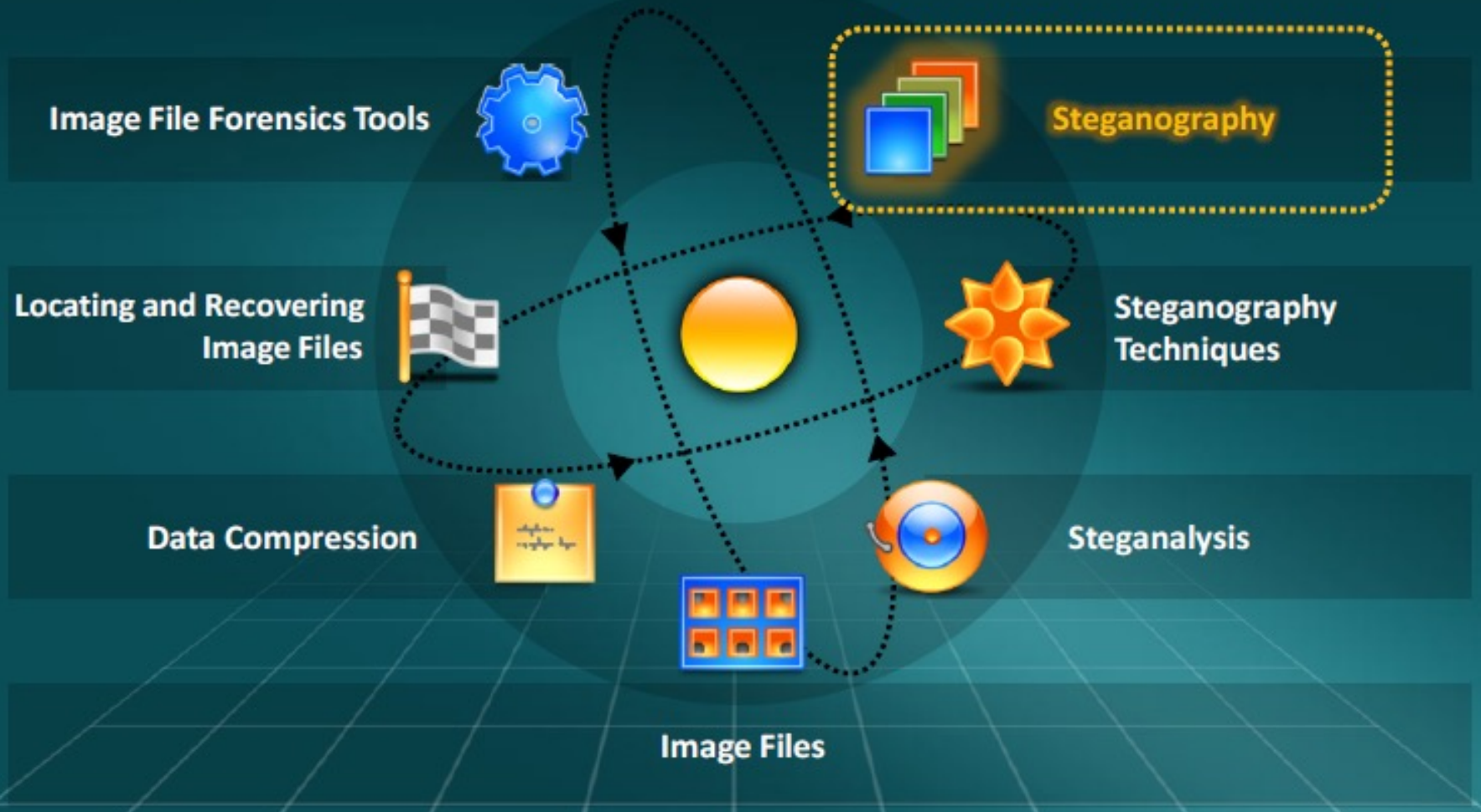
The police seized the computer from Ahmed and also recovered a significant amount of material from the computers.

Module Objectives

- What is Steganography?
- Application of Steganography
- Steganography Techniques
- Classification of Steganography
- Types of Steganography
- Steganalysis
- How to Detect Steganography
- Steganography Detection Tools

- Image Files
- Understanding Image File Formats
- Data Compression
- Forensic Image Processing Using MATLAB
- Locating and Recovering Image Files
- Identifying Unknown File Formats
- Picture Viewer Tools
- Image File Forensic Tools

Module Flow



What is Steganography?

- Steganography is a technique of **hiding a secret message** within an ordinary message and **extracting it at the destination** to maintain confidentiality of data
- **Utilizing a graphic image as a cover** is the most popular method of concealing data in files

Hiding Messages

List of the compromised servers

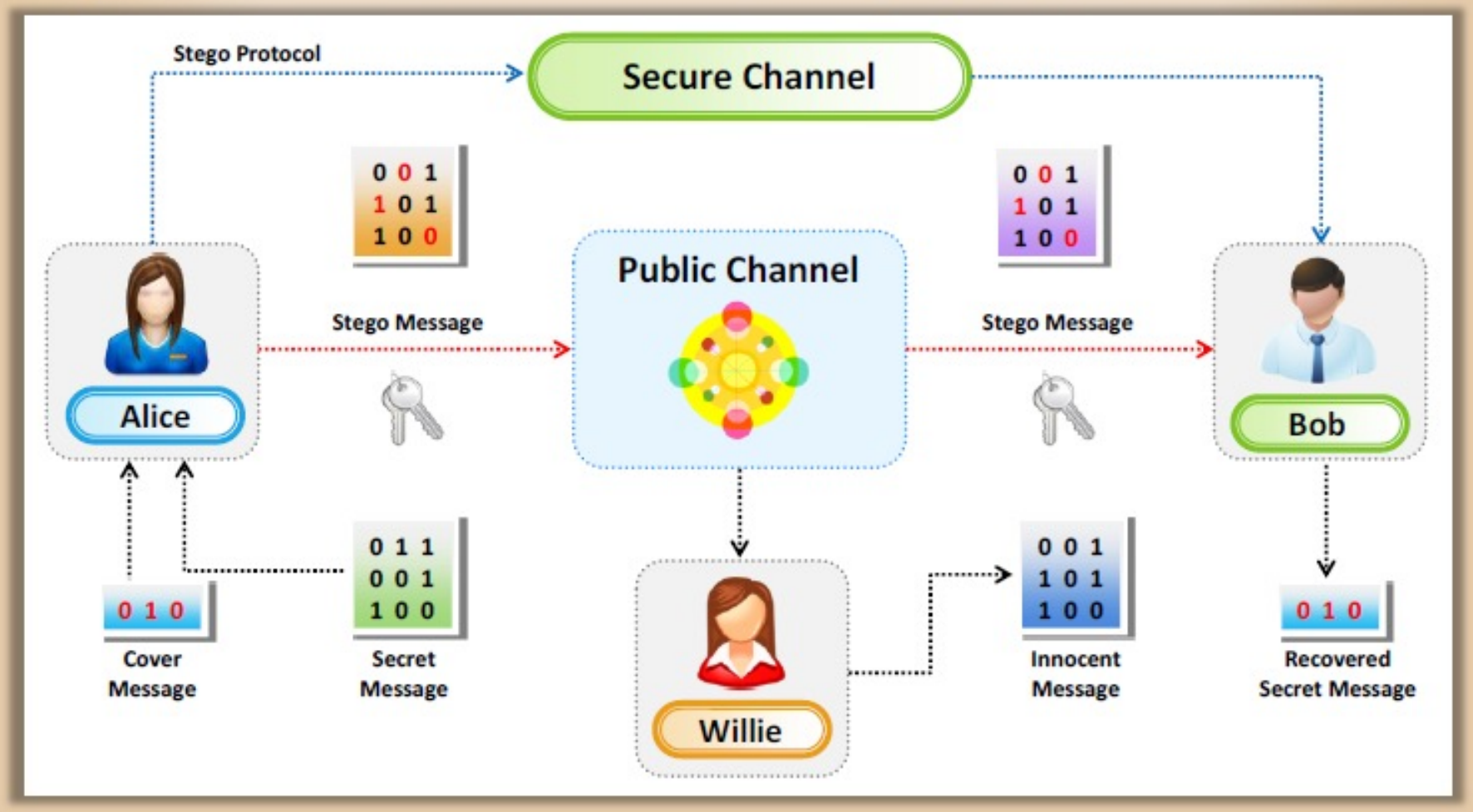
Source code for the hacking tool

Communication and coordination channel

Plans for future attacks

How Steganography Works

Stegosystem describes the process that is used in performing steganography



Legal Use of Steganography

Law enforcement agencies use steganography to:

Steganographically watermark intermediation materials after authorization and under public prosecutor control with predefined marks

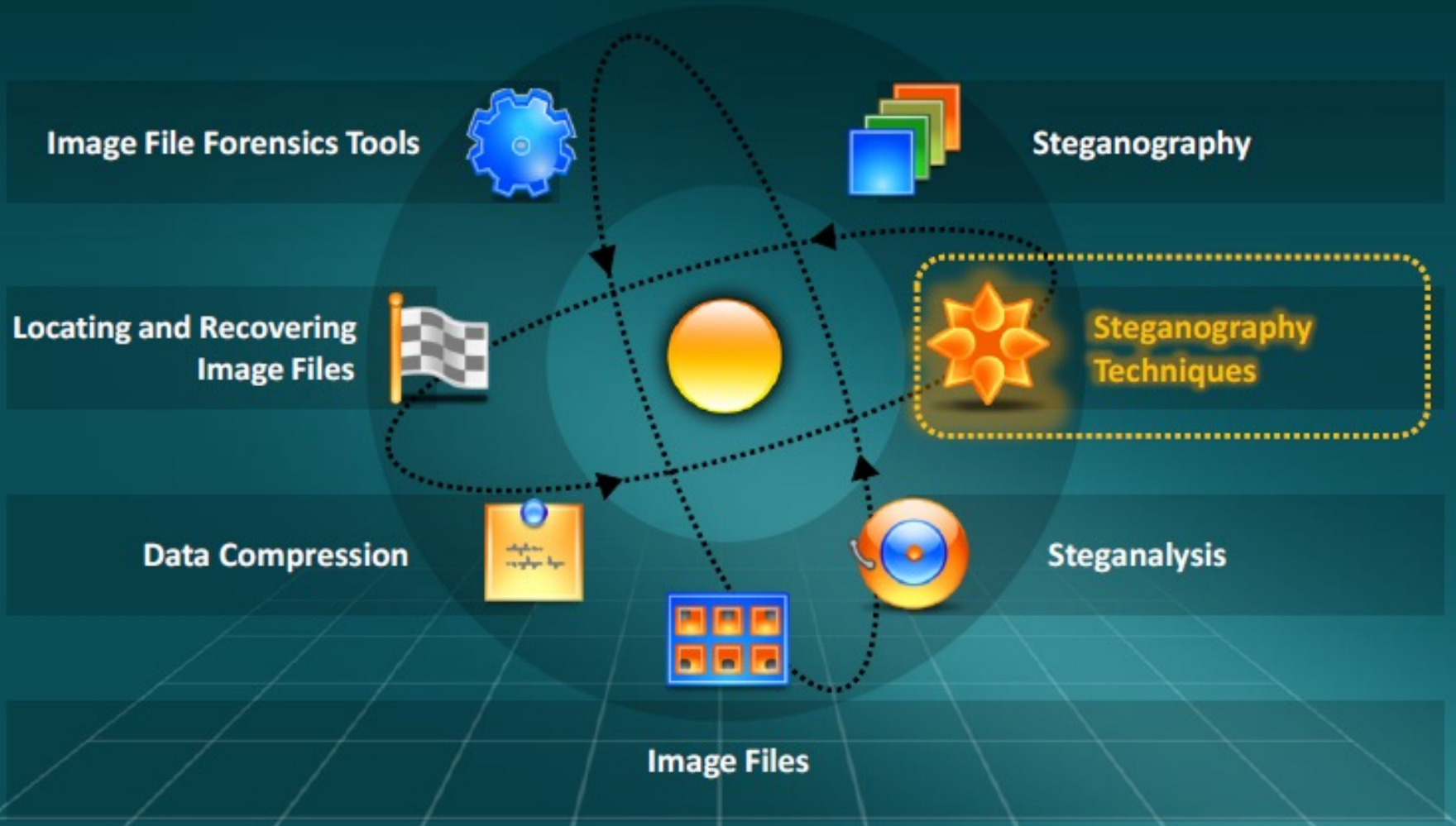
Trace trade materials

Build an international data bank to collect data on the trading controlled by investigative bodies

Provide network nodes where trade material is monitored

Unethical Use of Steganography

Module Flow



Steganography Techniques

Substitution Techniques

Substitute redundant part of the cover object with a secret message

Transform Domain Techniques

Embed secret message in a transform space of the signal (e.g. in the frequency domain)

Cover Generation Techniques

Encode information that ensures creation of cover for secret communication

Spread Spectrum Techniques

Adopt ideas from spread spectrum communication to embed secret messages

Distortion Techniques

Store information by signal distortion and in the extraction step measure the deviation from the original cover

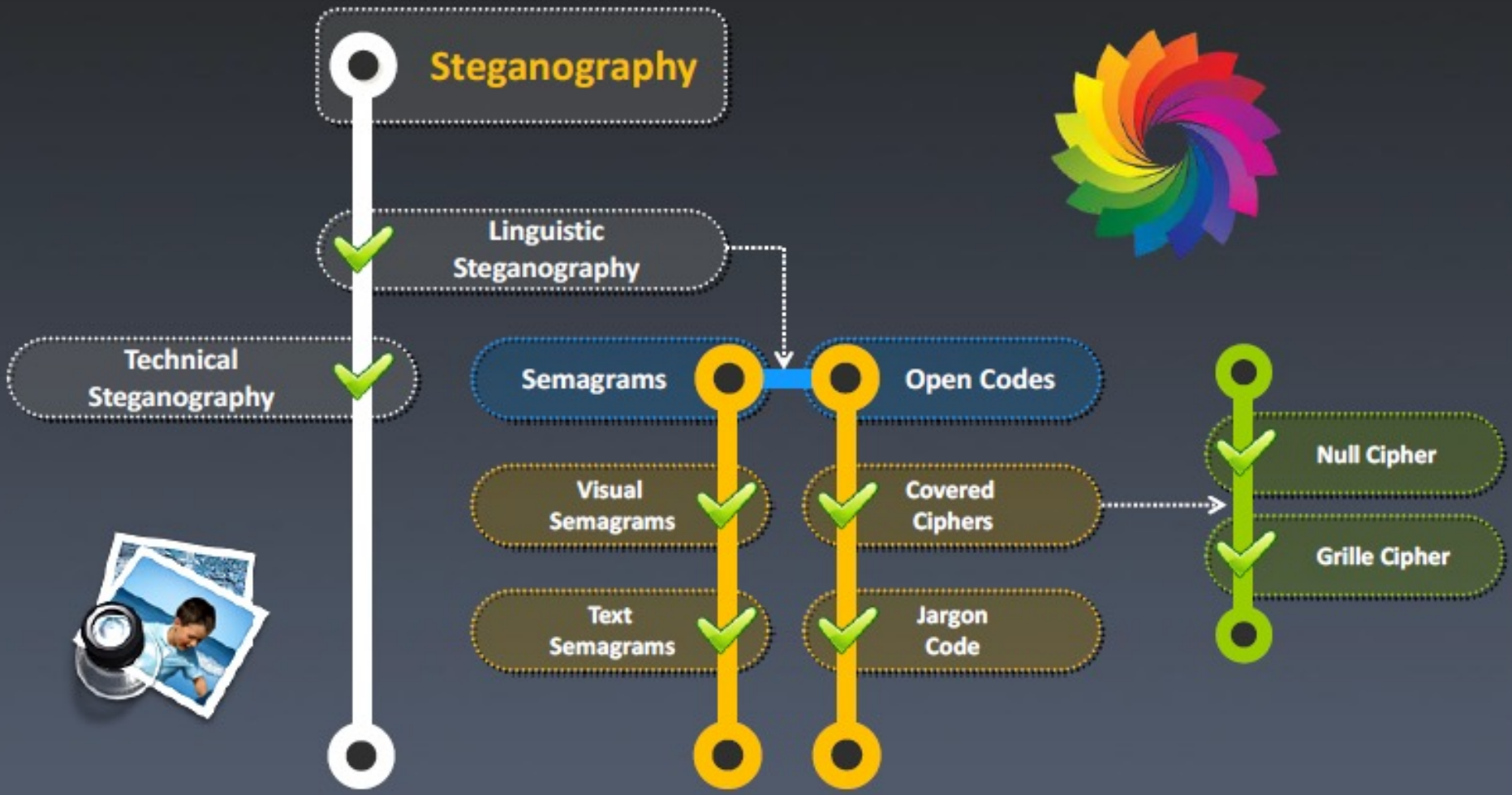
Statistical Techniques

Embed messages by altering statistical properties of the cover objects and use hypothesis methods for extraction

Application of Steganography

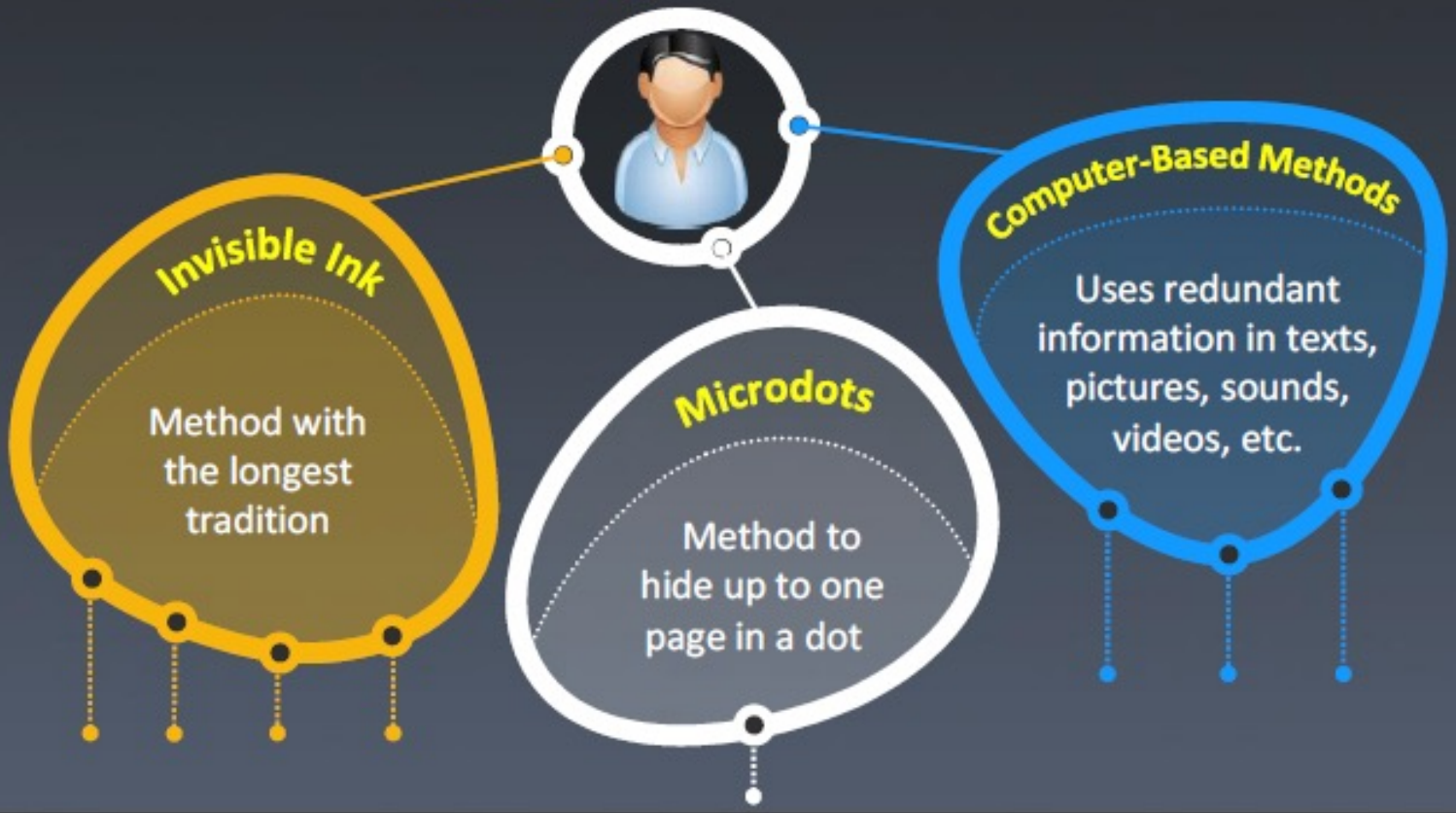
Steganography is applicable to the following areas:

Classification of Steganography



Technical Steganography

- Technical steganography uses physical or chemical means to **hide the existence of a message**
- Technical steganography uses tools, devices, or methods to **conceal messages**
- Some methods of **technical steganography** include:



Linguistic Steganography (Cont'd)

- Linguistic steganography utilizes written natural language to hide the message in the carrier in some non-obvious ways
- It is further categorized into Semagrams and Open codes
- Semagrams utilize visual symbols or signs to hide secret messages. They are classified into Visual and Text Semagrams

Visual Semagrams

Use innocent-looking or everyday physical objects to convey a message, such as doodles or the positioning of items on a desk or website

Text Semagrams

Hides a message by modifying the appearance of the carrier text, such as subtle changes in the font size or type, adding extra spaces, or different flourishes in letters or handwritten text

Linguistic Steganography

- Open code hides the secret message in a specifically designed pattern on the document that is unclear to the average reader
- Open code steganography is divided into:

1. Jargon Code
   - It is a language that a group of people can understand but is meaningless to others
2. Covered Ciphers
   - The message is hidden openly in the carrier medium so that anyone who knows the secret of how it was concealed can recover it

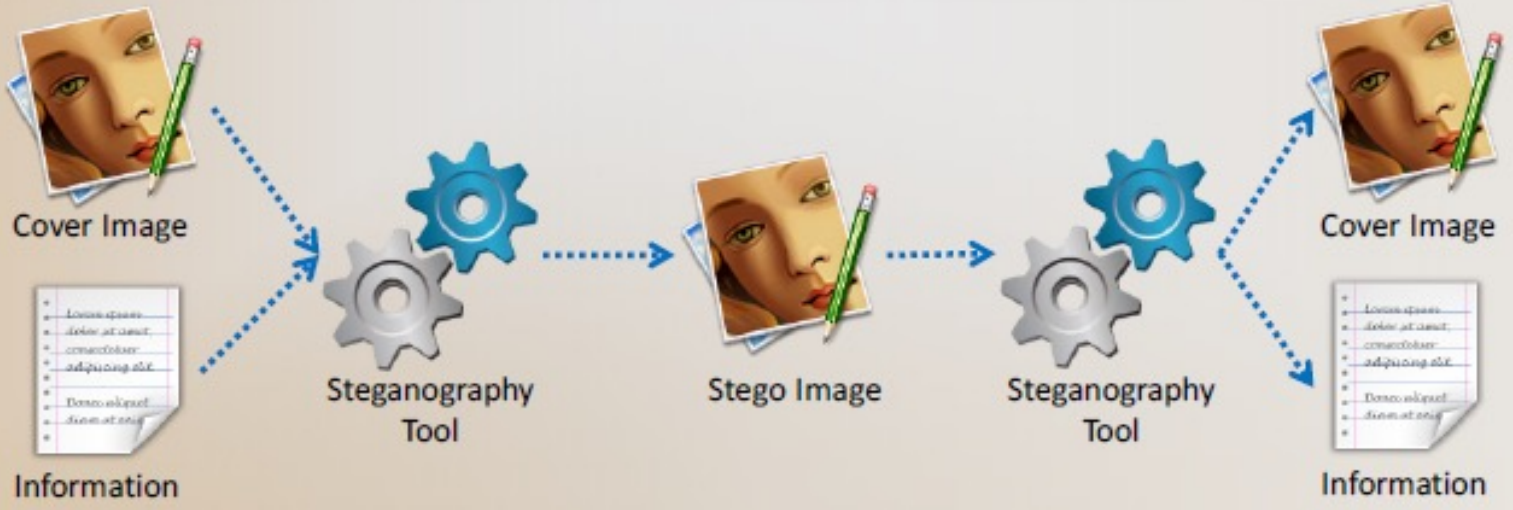
Covered Cipher is categorized into null ciphers and grille ciphers

1. Null Ciphers
   - A null cipher is an ancient form of encryption where the plaintext is mixed with a large amount of non-cipher material
   - It can also be used to hide ciphertext
2. Grille Ciphers
   - In this technique, a grille is created by cutting holes in a piece of paper
   - When the receiver places the grille over the text, the intended message can be retrieved

Types of Steganography

Image Steganography

1. In image steganography, the information is hidden in image files of different formats such as .PNG, .JPG, .BMP, etc.
2. Image steganography tools replace redundant bits of image data with the message in such a way that the effect cannot be detected by human eyes
3. Image file steganography techniques:
   - Least Significant Bit Insertion
   - Masking and Filtering
   - Algorithms and Transformation



Least Significant Bit Insertion

LSB

1. The right most bit of pixel is called the Least Significant Bit (LSB)
2. Using this method, the binary data of the hidden message is broken and then inserted into the LSB of each pixel in the image file in a deterministic sequence
3. Modifying the LSB does not result in a noticeable difference because the net change is minimal and can be indiscernible to the human eye

Example: Given a string of bytes

- (00100111 11101001 11001000) (00100111 11001000 11101001) (11001000 00100111 11101001)
- The letter "H" is represented by binary digits 01001000. To hide this "H" above stream can be changed as:
- (00100110 11101001 11001000) (00100110 11001001 11101000) (11001000 00100110 11101001)
- To retrieve the " H" combine all LSB bits 01001000

Masking and Filtering

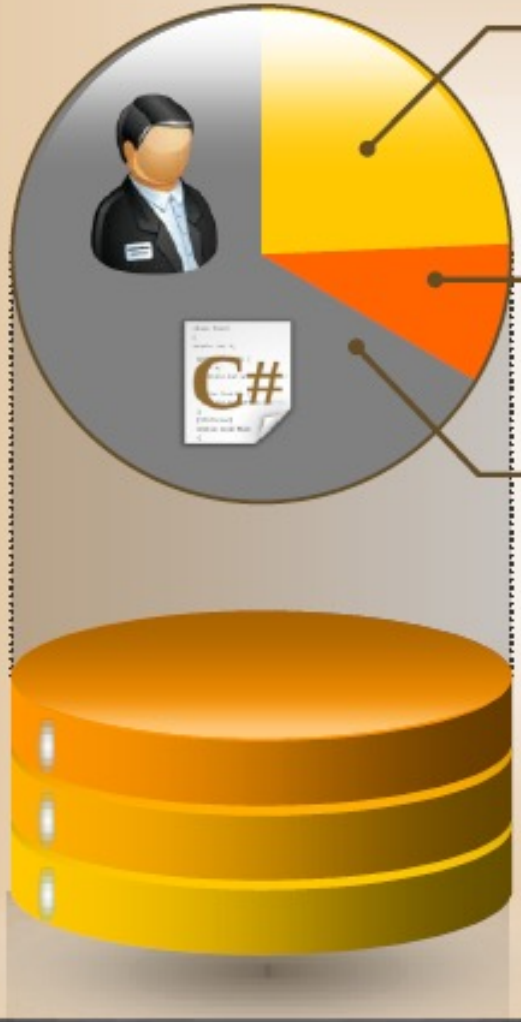
1. Masking and filtering techniques are mostly used on 24 bit and grayscale images

2. The masking technique hides data using a method similar to watermarks on actual paper, and it can be done by modifying the luminance of parts of the image

3. Masking techniques hide information in such a way that the hidden message is inside the visible part of the image

4. The information is not hidden at the "noise" level of the image

Algorithms and Transformation



Another steganography technique is to hide data in mathematical functions that are in compression algorithms

The data is embedded in the cover image by changing the coefficients of a transform of an image

JPEG images use the Discrete Cosine Transform (DCT) technique to achieve image compression

Types of transformation techniques

- Fast fourier transformation
- Discrete cosine transformation
- Wavelet transformation

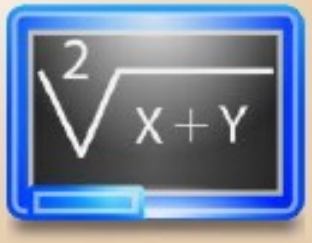
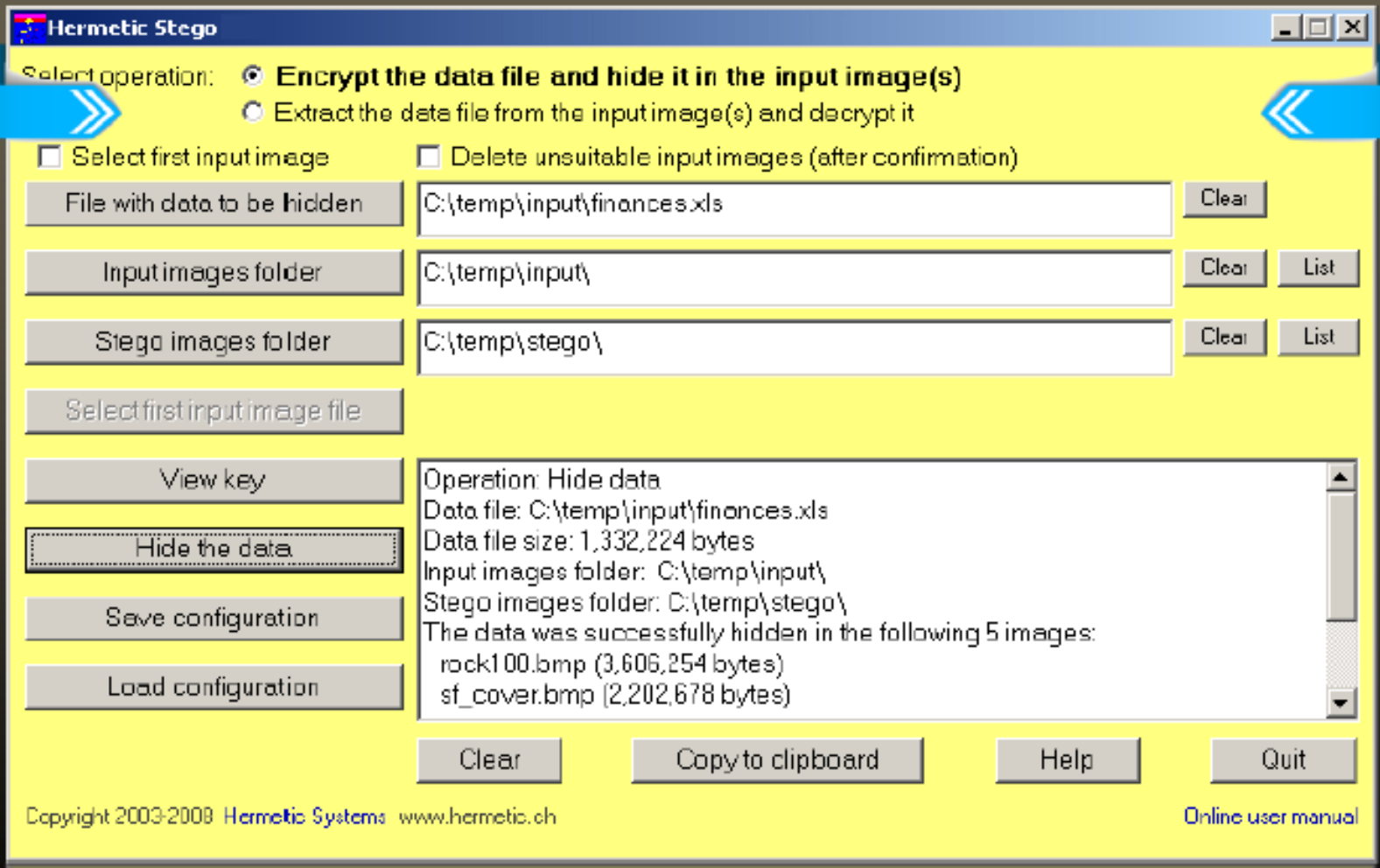


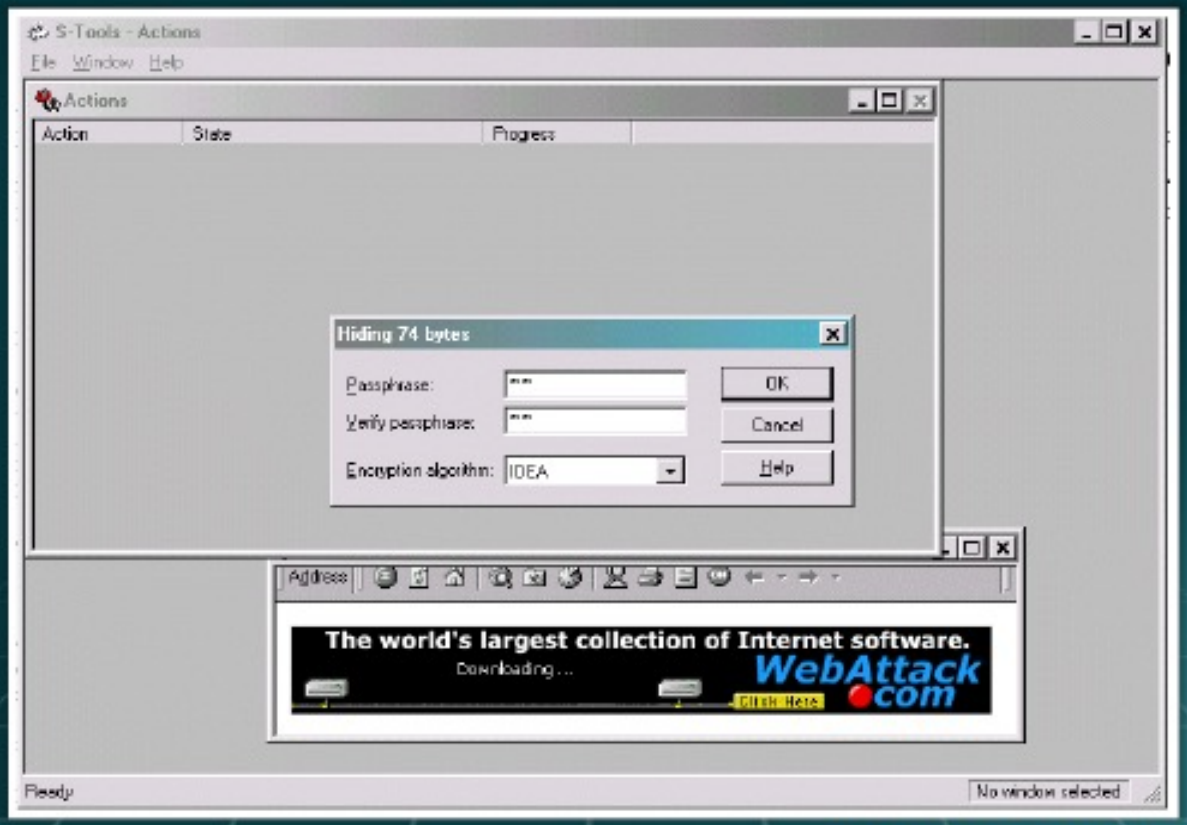
Image Steganography: Hermetic Stego



http://www.hermetic.ch

Steganography Tool: S- Tools

- S- Tools can hide multiple applications in a single object
- It hides files in BMP, GIF, and WAV files



http:/www.spychecker.com

Image Steganography Tools

ImageHide
http://www.dancemammal.com

Contraband
http://jthz.com

QuickStego
http://www.quickcrypto.com

Camera/Shy
http://sourceforge.net

gifshuffle
http://www.darkside.com.au

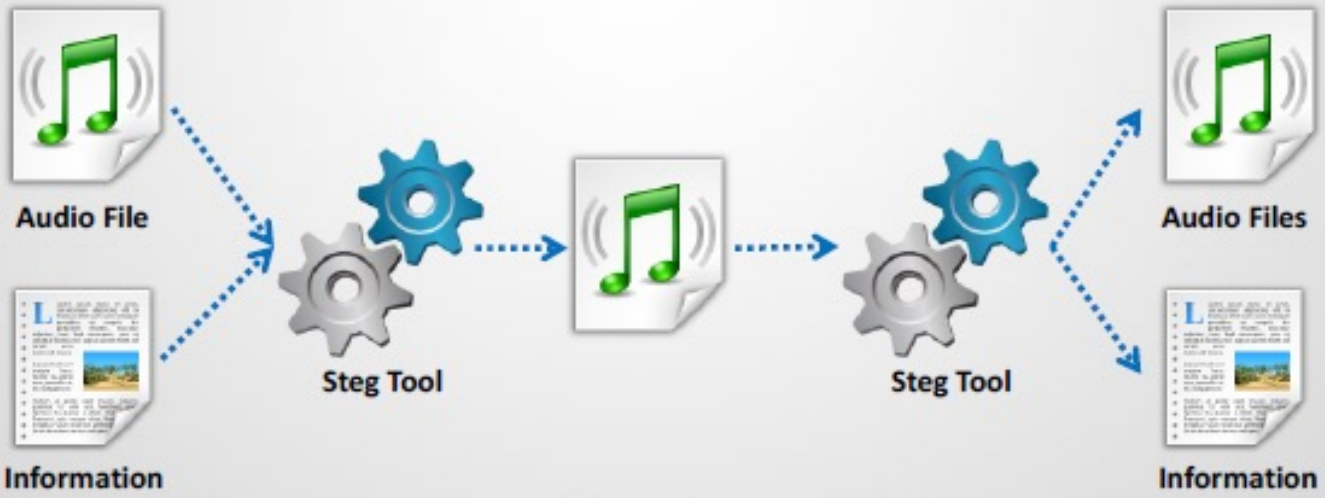
JPHIDE and JPSEEK
http://linux01.gwdg.de

OutGuess
http://www.outguess.org

StegaNote
http://www.planetsourcecode.com

Audio Steganography

- Audio steganography refers to **hiding secret information in audio files** such as .MP3, .RM, .WAV, etc.
- Information can be hidden in an audio file by using **LSB or by using frequencies** that are inaudible to the human ear (>20,000 Hz)



Audio Steganography Methods (Cont'd)

- **Echo Data Hiding**
  - In echo data hiding technique, the secret message is embedded into a cover audio signal as an echo
  - Parameters of the echo of the cover signal, namely amplitude, decay rate and offset from the original signal, are varied to represent an encoded secret binary message
  - An echo cannot be easily resolved because the parameters are set below levels audible to human

- **Spread Spectrum Method**
  - It encodes data as a binary sequence that sounds like noise but can be recognized by a receiver with the correct key
  - Two approaches are used in this technique, namely direct sequence spread spectrum (DSSS) and frequency hopping spread spectrum (FHSS)
  - In DSSS, the secret message is spread out by chip rate (constant) and then modulated with a pseudo-random signal that is then interleaved with the cover signal
  - In FHSS, the audio file's frequency spectrum is altered so that it hops rapidly between frequencies
  - Spread spectrum method plays a major role in secure communications – commercial and military

Audio Steganography Methods

LSB Coding

- The low bit encoding technique is similar to the least bit insertion method done through image files
- It replaces the LSB of information in each sampling point with a coded binary string

Tone Insertion

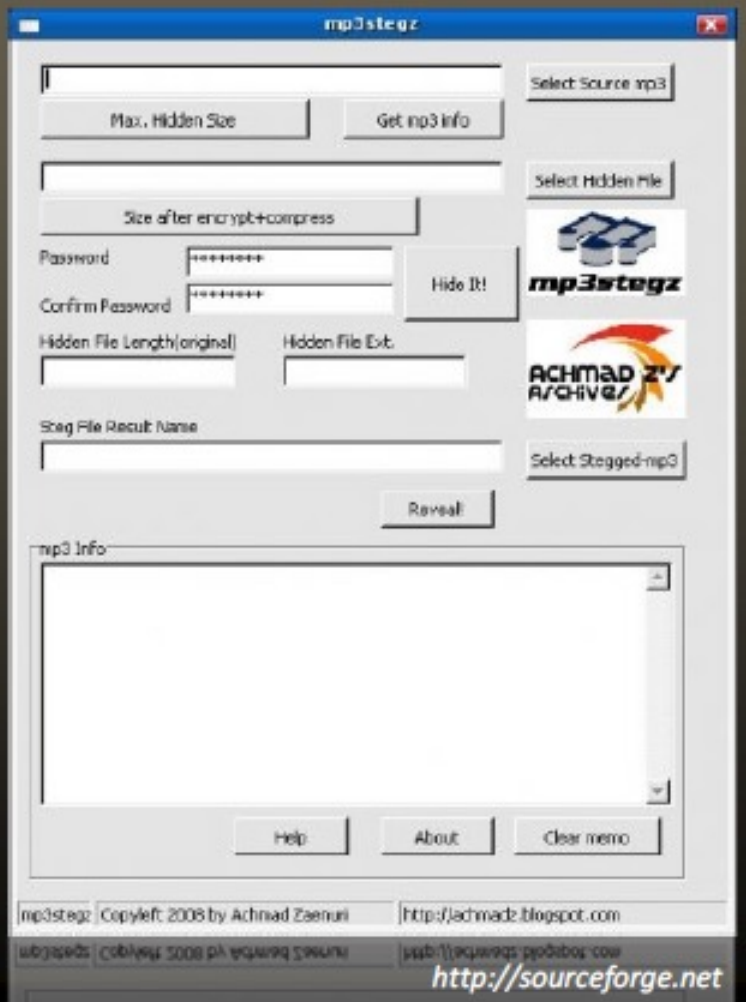
- It depends on the inaudibility of low power tones in the presence of significantly higher spectral components
- An indirect exploitation of the psychoacoustic masking phenomenon in the spectral domain

Phase Decoding

- Phase coding is described as the phase in which an initial audio segment is substituted by a reference phase that represents the data
- It encodes the secret message bits as phase shifts in the phase spectrum of a digital signal, achieving a soft encoding in terms of signal-to-noise ratio

Audio Steganography: Mp3stegz

- Mp3stegz is an application that applies a steganographic (steganography) algorithm in mp3 files
- It maintains the original mp3 file's size and sound quality
- The hidden message is compressed (zlib) and encrypted (Rijndael)



http://sourceforge.net

Audio Steganography Tools

MAXA Security Tools
http://www.maxa-tools.com

Stealth Files
http://www.froebis.com

Audiostegano
http://www.mathworks.com

BitCrypt
http://bitcrypt.moshe-szweizer.com

MP3Stego
http://www.petitcolas.net

Steghide
http://steghide.sourceforge.net

Hide4PGP
http://www.heinz-repp.onlinehome.de

CHAOS Universal
http://safechaos.com

Video Steganography

- Video Steganography refers to **hiding secret information** or any kind of files with any extension into a carrier video file
- In video steganography, the information is hidden in **video files** of different formats such as .AVI, .MPG4, .WMV, etc.
- **Discrete Cosine Transform (DCT)** manipulation is used to add secret data at the time of the transformation process of video
- The techniques used in audio and image files are used in video files, as video consists of audio and images
- A **large number of secret messages** can be hidden in video files since they are a moving stream of images and sound

Video Steganography: MSU StegoVideo

MSU StegoVideo allows you to hide any file in a **video sequence**

Features:

- Small video distortions after hiding info
- It is possible to extract info after video compression
- Information is protected with passcode

Video Steganography Tools

Masker
http://www.softpuls.com

Max File Encryption
http://www.softeza.com

Xiao Steganography
http://xiao-steganography.en.softonic.com

RT Steganography
http://sourceforge.net

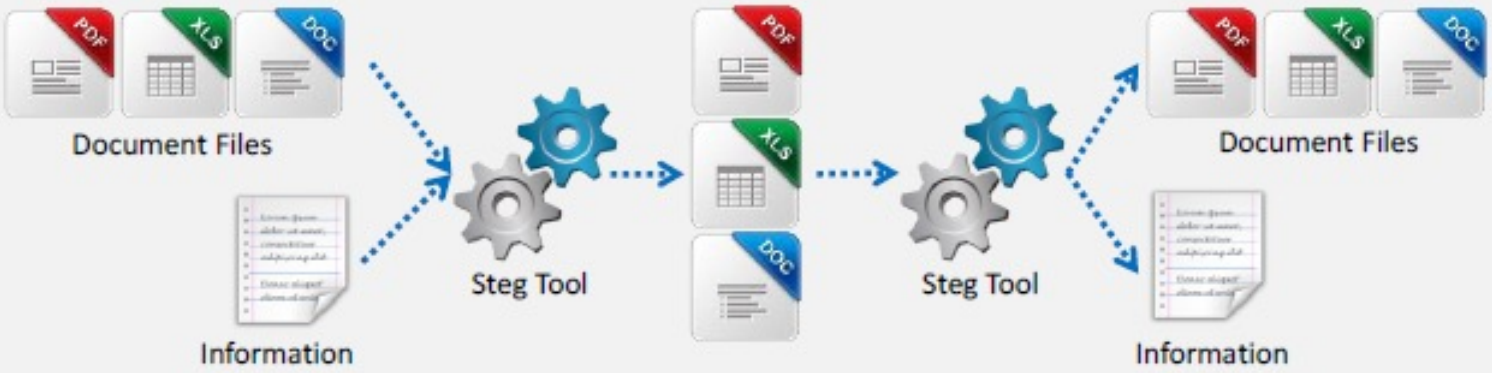
Our Secret
http://www.securekit.net

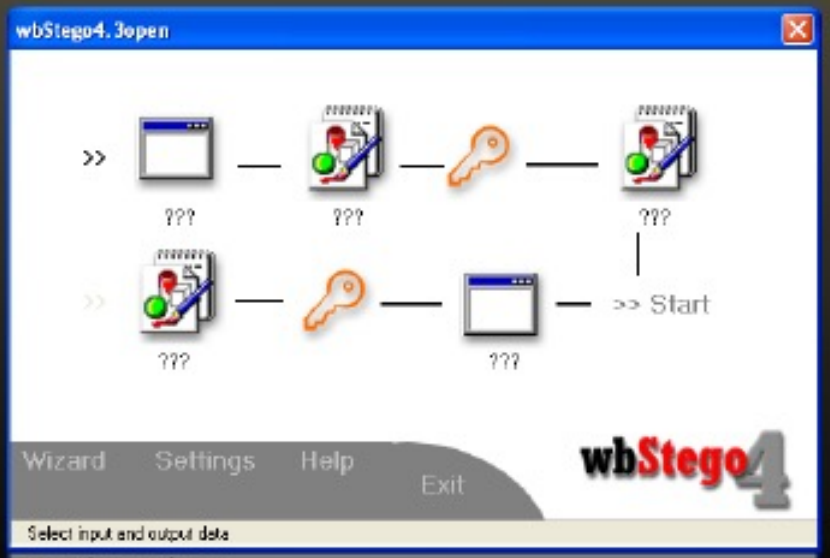
BDV DataHider
http://www.bdvnotepad.com

CHAOS Universal
http://safechaos.com

OmniHide PRO
http://omnihide.com

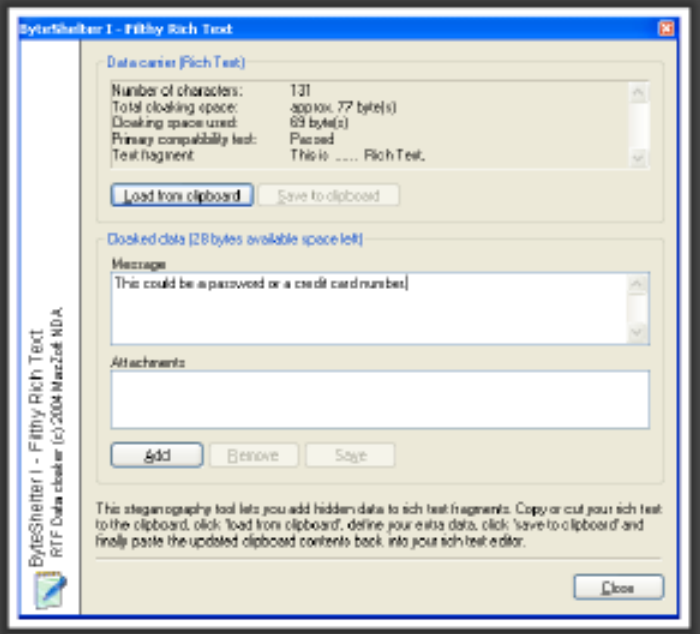
Document Steganography: wbStego





Byte Shelter I

- Byte Shelter I encrypts data and hides it in .doc files or email messages
- Byte Shelter can be used to hide files and/or text in rich text fragments



Document Steganography Tools

Merge Streams
http://www.ntkernel.com

FoxHole
http://foxhole.sourceforge.net

Office XML
http://www.irongeek.com

Xidie Security Suite
http://www.stegano.ro

CryptArkan
http://www.kuskov.com

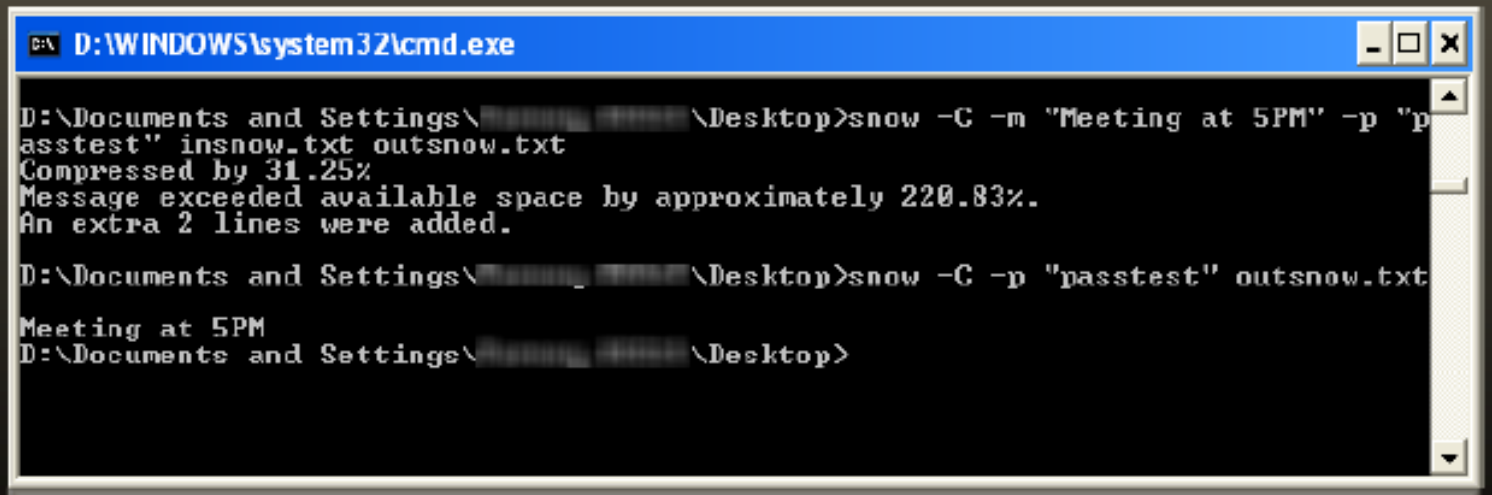
StegParty
http://www.fasterlight.com

Data Stash
http://www.skyjuicesoftware.com

Hydan
http://www.crazyboy.com

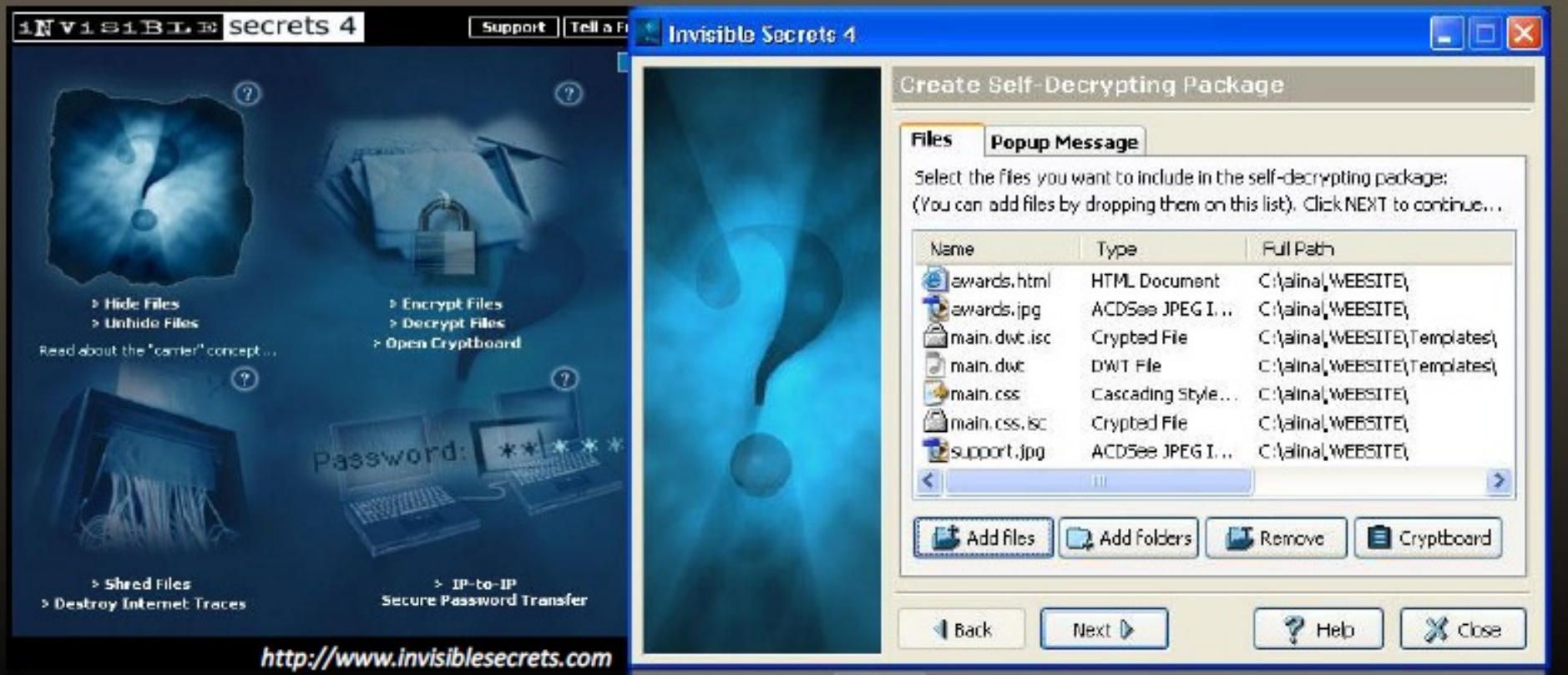
Whitespace Steganography Tool: SNOW

1. The program SNOW is used to conceal messages in ASCII text by appending white space to the end of lines
2. Because spaces and tabs are generally not visible in text viewers, the message is effectively hidden from casual observers
3. If the built-in encryption is used, the message cannot be read even if it is detected



Folder Steganography: Invisible Secrets 4

Folder steganography refers to hiding secret information in folders



Folder Steganography Tools

StegoStick
http://stegostick.sourceforge.net

QuickCrypto
http://www.quickcrypto.com

Max Folder Secure
http://www.maxfoldersecure.com

WinMend Folder Hidden
http://www.winmend.com

PSM Encryptor
http://www.powersoftmakers.com

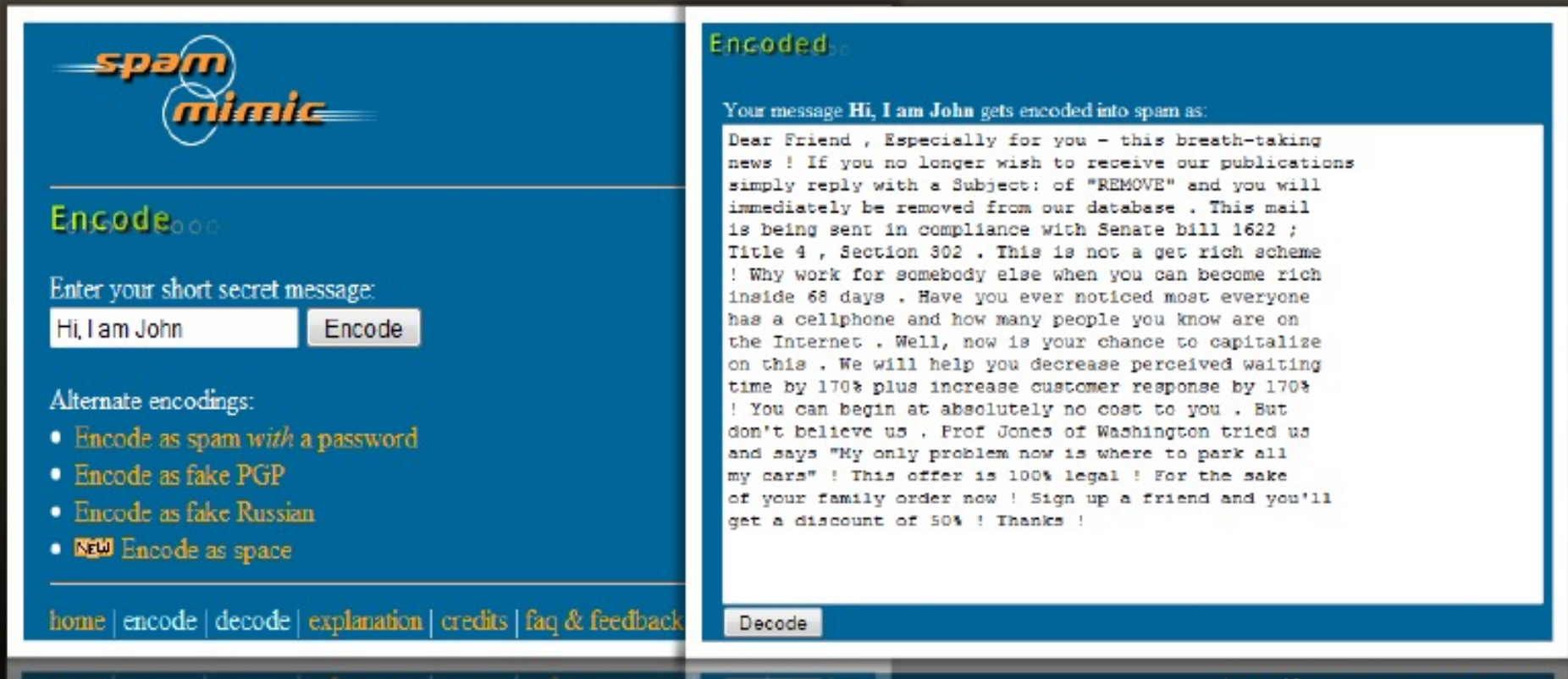
XPTools
http://www.xptools.net

Universal Shield
http://www.everstrike.com

Hide My Files
http://www.secretfilesoftware.com

Spam/Email Steganography: Spam Mimic

Spam steganography refers to hiding information in spam messages



http://www.spammimic.com

Steganographic File System

Steganographic file system is a method to store the files in a way that it encrypts and hides the data without the knowledge of others

It hides the user's data in other seemingly random files

It allows the user to give names and passwords for some files while keeping other files secret

Issues in Information Hiding

Levels of Visibility

- If the embedding process distorts the cover to the point that it is visually unnoticeable, meaning if the image is visibly distorted, then the carrier is insufficient for the payload. Likewise, if the image is not distorted, then the carrier is adequate

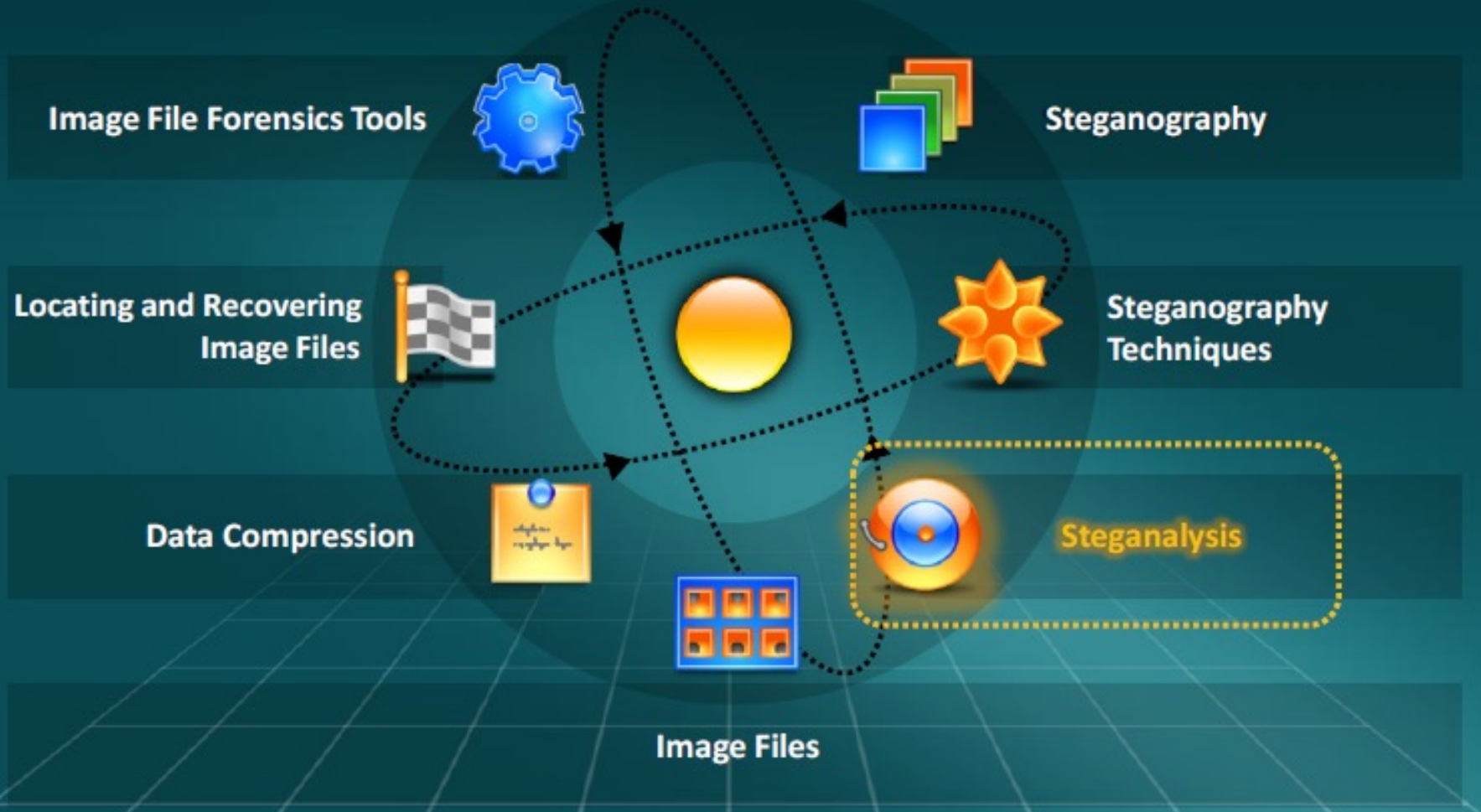
Robustness vs. Payload

- Redundancy is needed for a robust method of embedding the message, but it subsequently reduces the payload
- Robustness and payload are inversely related. Therefore, the smaller the payload, the more robust it will be

File Format Dependence

- Some image and sound files are lossy or lossless
- The conversion of lossless information to compressed lossy information destroys the hidden information in the cover

Module Flow



Steganalysis

Steganalysis is the art of discovering and rendering covert messages using steganography

Suspect information stream may or may not have encoded hidden data

Challenge of Steganalysis

Efficient and accurate detection of hidden content within digital images

Some of the suspect signals or files may have irrelevant data or noise encoded into them

Encrypts the hidden data before inserted into a file or signal

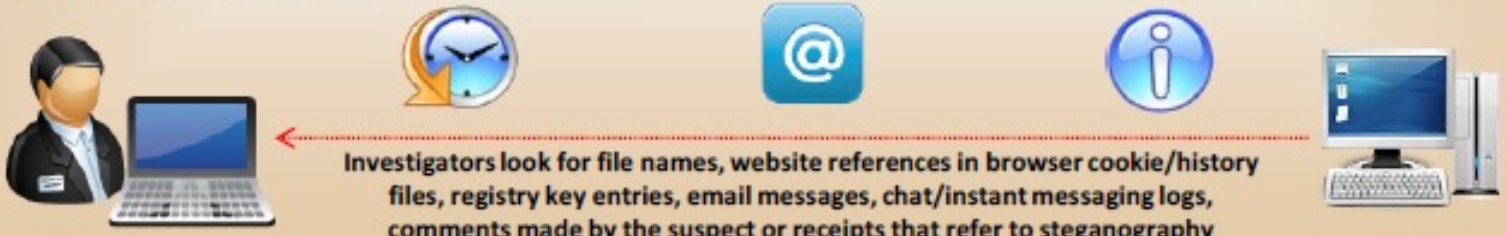
How to Detect Steganography (Cont'd)

Software Clues on the Computer

- Steganographic investigators need to be familiar with the names of the common steganographic software and related terminology, and even websites about steganography
- Investigators look for file names, website references in browser cookie/history files, registry key entries, email messages, chat/instant messaging logs, comments made by the suspect or receipts that refer to steganography
- These will provide hard clues for the investigator to look deeper

Other Program Files

- Non-steganographic software might offer clues that the suspect hides files inside other files
- Users with binary (hex) editors, disk wiping software, or specialized chat software might demonstrate an inclination to alter files and keep information secret



Investigators look for file names, website references in browser cookie/history files, registry key entries, email messages, chat/instant messaging logs, comments made by the suspect or receipts that refer to steganography

How to Detect Steganography

Multimedia Files

- Look for the **presence** of a large volume of the suitable **carrier files**
- A computer system with an especially large number of files that could be **steganographic carriers** are potentially suspected
- This is particularly true if there are a significant number of seemingly **duplicate "carrier" files**

Type of Crime

- The type of crime being investigated may also make an investigator think more about **steganography** than other types of crime
- Child pornographers, for example, might use steganography to **hide their wares** when posting pictures on a website or sending them through e-mail
- Crimes that involve **business-type records** are also good steganography candidates because the **perpetrator** can hide the files but still get access to them; consider accounting fraud, identity theft (lists of stolen credit cards), drugs, gambling, hacking, smuggling, terrorism, and more

Detecting Text, Image, Audio, and Video Steganography (Cont'd)

Text File

- For the text files, the alterations are made to the character positions for hiding the data
- The alterations are detected by looking for text patterns or disturbances, language used, and an unusual amount of blank spaces

Image File

- The hidden data in an image can be detected by determining changes in size, file format, the last modified timestamp, and the color palette pointing to the existence of the hidden data
- Statistical analysis method is used for image scanning

Detecting Text, Image, Audio, and Video Steganography

Investigator

Audio File

- Statistical analysis method can also be used for audio files since the LSB modifications are also used on audio
- The inaudible frequencies can be scanned for information
- The odd distortions and patterns show the existence of the secret data

Video File

- Detection of the secret data in video files includes a combination of methods used in image and audio files
- Special code signs and gestures can also be used for detecting secret data

Steganalysis Methods/Attacks on Steganography

Only the steganography medium is available for analysis	Stego-only	Reformat	The format of the file is changed. This works because different file formats store data in different ways
Original and stego-object are available and the steganography algorithm is known	Known-stego	Known-cover	The stego-object is compared with the original cover object to detect hidden information
The hidden message and the corresponding stego-image are known	Known-message	Chosen-message	The goal is to determine patterns in the stego-object that may point to the use of the specific steganography tools or algorithms
During the communication process, active attackers can change the cover	Disabling or Active	Chosen-stego	The stego-object and steganography algorithm are identified

Disabling or Active Attacks

Blur:

It softens the transitions and averages the adjacent pixels with significant color change

Noise:

- Random noise injects random colored pixels to an image
- Uniform noise inserts slightly similar pixels and colors of the original pixel

Noise Reduction:

It reduces the noise in the image by adjusting the colors and averaging the pixel values

Sharpen:

It increases the contrast between the adjacent pixels

Rotate:

It moves the image around a central point

Resample

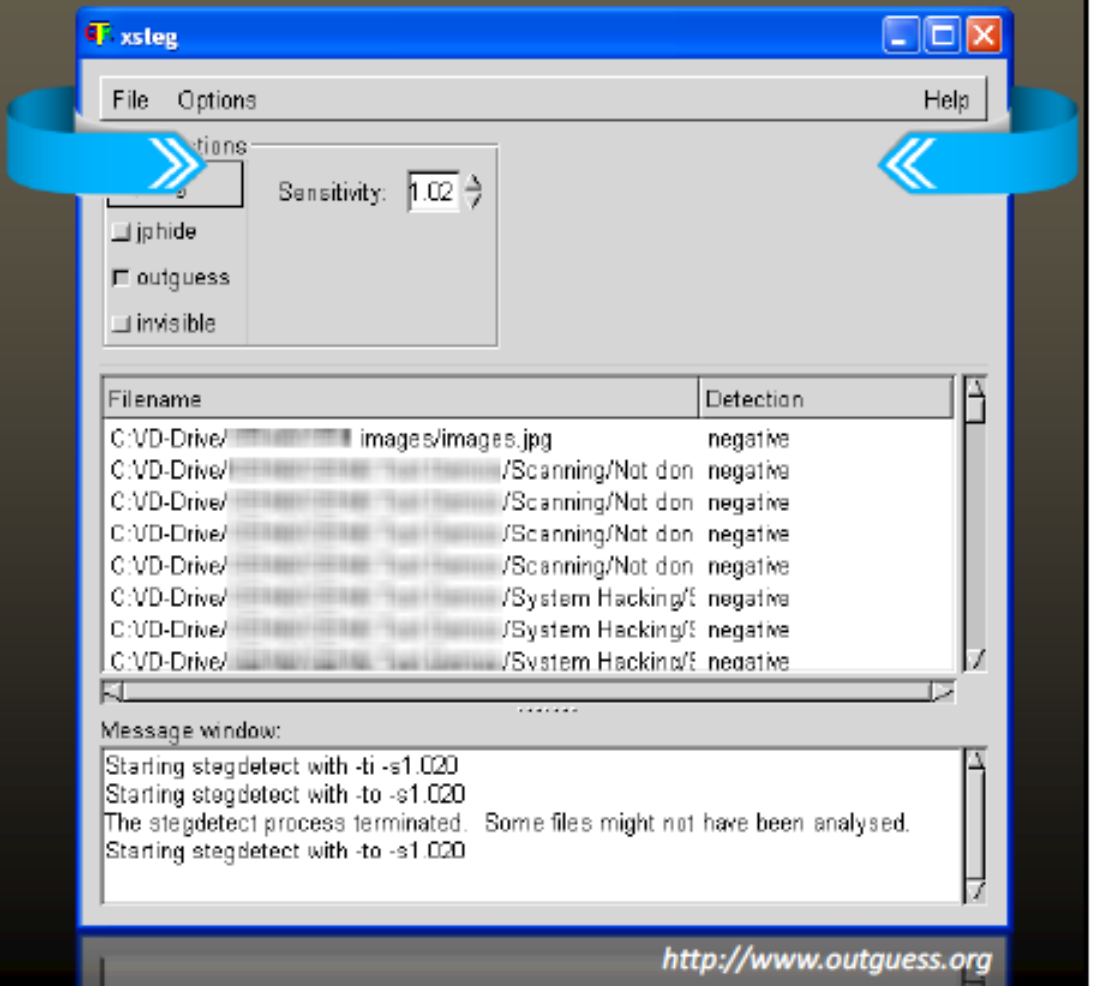
It is an interpolation process where the raggedness while expanding an image is reduced

Soften:

Soften is a uniform blur to an image that smooths the edges and decreases the contrasts

Steganography Detection Tool: Stegdetect

- It is an automated tool for detecting steganographic content in images
- It is capable of detecting several different steganographic methods to embed hidden information in JPEG images



http://www.outguess.org

Steganography Detection Tools

Xstegsecret
http://stegsecret.sourceforge.net

StegSpy
http://www.spy-hunter.com

Stego Watch
http://www.wetstonetech.com

Gargoyle Investigator™ Forensic Pro
http://www.wetstonetech.com

StegAlyzerAS
http://www.sarc-wv.com

StegAlyzerSS
http://www.sarc-wv.com

StegAlyzerRTS
http://www.sarc-wv.com

StegMark
http://www.datamark-tech.com

Module Flow

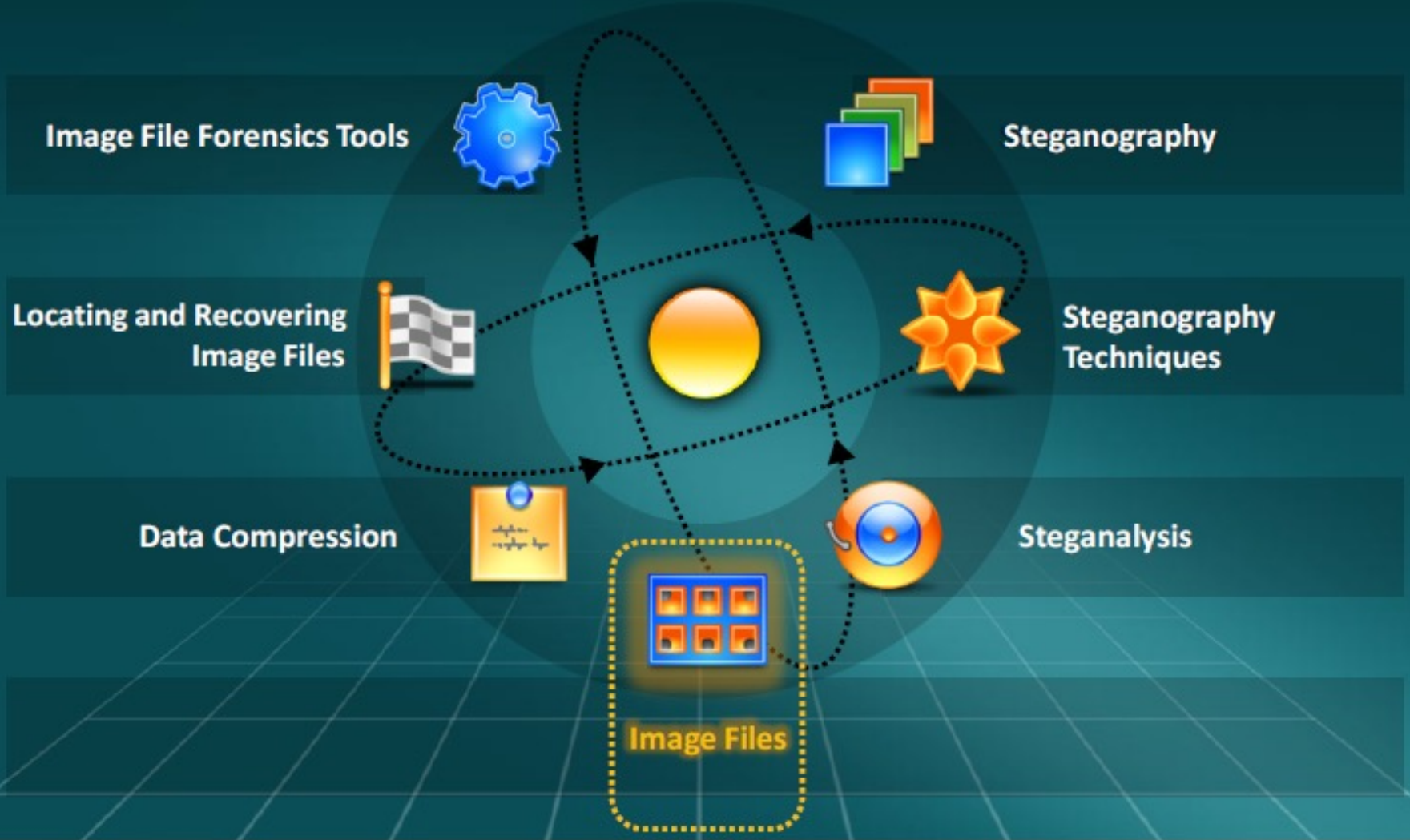


Image Files

Image may be black and white image, grayscale image, color image, indexed color image

An image is an artifact that reproduces the likeness of some subject

Images can be broadly categorized into:

- Vector images
- Raster images

These are produced by optical devices i.e. cameras, mirrors, lenses, telescopes, and microscopes

Common Terminologies

Pixel

- Pixel (Picture Element) is a single point in a graphic image
- Number of pixels combine together to form an image

Bit Depth

- Refers to the number of colors available for each pixel in an image

Resolution

- Refers to the sharpness and clarity of an image
- The term describes monitors, printers, and bit-mapped graphic images

File Formats

- Particular way to encode information for storage in a computer file

Image File Size

- Expressed as the number of bytes
- It increases with the bit depth of the pixel and number of pixels comprising an image

Compression

- Refers to the method of making image files smaller so that less disk space is used to store them

Understanding Vector Images

Vector graphics use **geometrical primitives** such as points, lines, curves, and polygons, which are all **based upon mathematical equations** to represent images in the computer

Advantages

Smaller file size and useful for representing the images of various shapes

Applications such as moving, scaling, and rotating do not affect the quality of the image

Can be indefinitely zoomed without loss in quality

Understanding Raster Images

A raster image is a data file or structure representing a generally rectangular grid of pixels, or points of color, on a computer monitor

Quality may be lost if raster graphics are scaled to a higher resolution

Color of each pixel is individually defined

A colored raster image has pixels with eight bits of information for each of the red, green, and blue components

Quality of a raster image is determined by the total number of pixels and the amount of information in each pixel

Metafile Graphics

Metafiles combine raster and vector graphics

Metafiles have similar features of both bitmap and vector images

When metafiles are enlarged, it results in a loss of resolution, giving the image a shaded appearance

Understanding Image File Formats

- A file format is a particular way to encode information for storage in a computer file
- All image formats differ in their ease of use, the size of the files they produce, and their image quality

Image File Formats

Standard Image File Formats

File Format	File Extension
Joint Photographic Experts Group (JPEG)	.jpg
JPEG 2000	.jp2
Graphics Interchange Format (GIF)	.gif
Tagged Image File Format (TIFF)	.tif
Windows Bitmap (BMP)	.bmp
Portable Network Graphics (PNG)	.png

Nonstandard Image File Formats

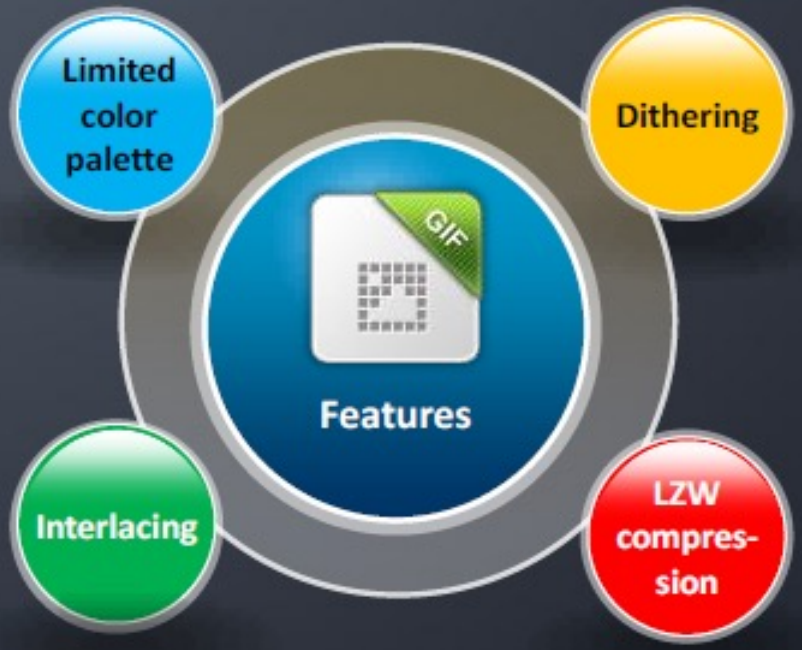
File Format	File Extension
Targa	.tga
Raster Transfer Language	.rtl
Photoshop	.psd
Illustrator	.ai
Freehand	.h9
Scalable Vector Graphics	.svg
Paintbrush	.pcx

GIF (Graphics Interchange Format) (Cont'd)

- GIF is an **8-bit RGB bitmap image format** for images with up to 256 distinct colors per frame

Each color in the GIF color table is described in **RGB values**, with each value having a range of 0 to 255

It is a mechanism that makes images appear **faster** on-screen by first displaying a low-res version of the image and gradually showing the full version



This method is used to create the **illusion** of greater color depth by blending a smaller number of colored "dots" together

GIF supports **LZW lossless** compression algorithms

GIF (Cont'd)

1

Each file begins with a Header and a Logical Screen Descriptor

2

A Global Color Table may optionally be displayed after the Logical Screen Descriptor

Each image stored in the file contains a Local Image Descriptor, an optional Local Color Table, and a block of the image data

3

The last field in every GIF file is a Terminator character, which indicates the end of the GIF data stream

4

GIF

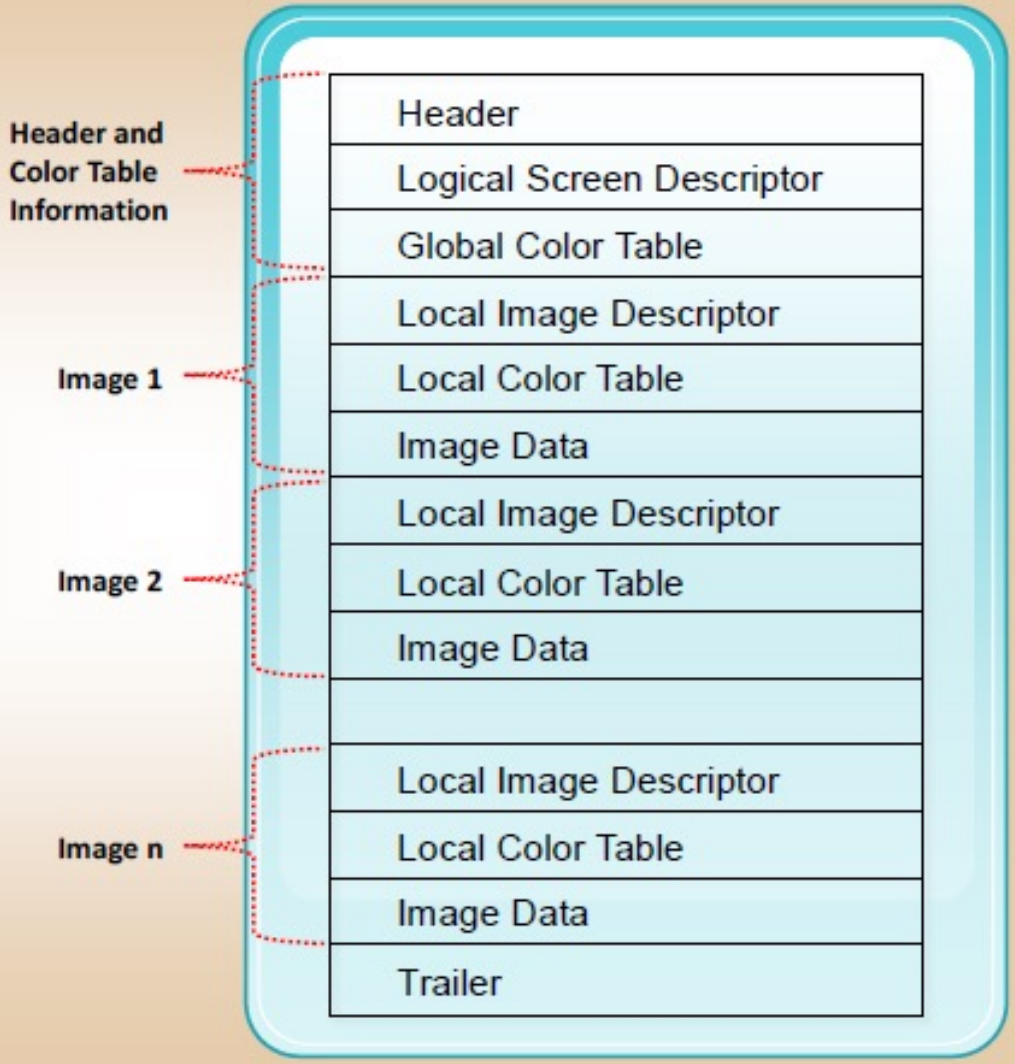
There are two versions of the GIF format:

GIF 87a

It supports LZW file compression, interlacing, 256-color palettes, and multiple image storage

GIF 89a

It supports properties such as background transparency, delay times, and image replacement parameters, which helps to store multiple images



JPEG (Joint Photographic Experts Group)

JPEG is a commonly used method for compression of photographic images

It performs the file compression in four phases:

1. The JPEG lossy algorithm divides the image in separate blocks of 8x8 pixels
2. The compression algorithm then starts applying a Discrete Cosine Transform (DCT) for the whole block
3. The compression algorithm checks the JPEG image quality requested by the user, and then computes the separate tables of quantization constants for luminance and chrominance
4. Any of the Huffman or arithmetic encoding schemes are employed to compress these coefficients

JPEG Lossy Algorithm → 8x8 Pixel Block → Discrete Cosine Transform → Quantizer → Binary Encoder → 011010 Output Data Stream

JPEG File Structure (Cont'd)

JPEG Image

Contents	Name	Description
0xFF 0xD8	SOI	Start of image
Segments		
0xFF 0xD9	EOI	End of image

JPEG Segments

Description
Segment marker (2 bytes)
Segment size (2 bytes) excl. marker
Segment data

JPEG File Structure

Some JPEG Segment Markers

Contents	Name	Description
0xFF 0xEo	APP0	Application marker (in every JPEG file)
0xFF 0xDB	DQT	Quantization table
0xFF 0xC0	SOF0	Start of frame
0xFF 0xC4	DHT	Define Huffman Table
0xFF 0xDA	SOS	Start of scan
0xFF 0xED	APP14	This is the marker where Photoshop stores its information

JPEG 2000

JPEG 2000 is the new version of JPEG compression

It produces as much as 20% improvement in compression efficiency over the current JPEG format

Its compression has been mainly developed for use on the Internet

It can handle RGB, LAB, and CMYK with up to 256 channels of information

BMP (Bitmap) File

- BMP is a standard file format for computers running the Windows operating system
- BMP images can range from black and white (1 bit per pixel) up to 24 bit color (16.7 million colors)
- Each bitmap file contains:

1. Header

Contains information about the type, size, and layout of a file

2. The RGBQUAD Array

Specifies the dimensions, compression type, and color format for the bitmap

3. Information Header

Color array contains a color table that is absent for bitmaps with 24 color bits because each pixel is represented by 24-bit red-green-blue (RGB) values in the actual bitmap

4. Image Data

These are the actual image data, represented by consecutive rows, or "scan lines," of the bitmap

BMP File Structure

Basic BMP File Format			
Name			Size
Header			14 bytes
	Signature		2 bytes
	File Size		4 bytes
	reserved		4 bytes
	DataOffset		4 bytes
Color Table			4 * NumColors bytes
		Red	1 byte
		Green	1 byte
		Blue	1 byte
		reserved	1 byte
	Repeated NumColors times		
Raster Data			Info ImageSize bytes

Basic BMP File Format (Cont'd)		
Name		Size
InfoHeader		40 bytes
	Size	4 bytes
	Width	4 bytes
	Height	4 bytes
	Planes	2 bytes
	BitCount	2 bytes
	Compression	4 bytes
	ImageSize	4 bytes
	XpixelsPerM	4 bytes
	YpixelsPerM	4 bytes
	ColorsUsed	4 bytes
	ColorsImportant	4 bytes

PNG (Portable Network Graphics)

PNG bitmap image format uses lossless data compression

PNG was created to improve upon and replace GIF (Graphics Interchange Format) as an image-file format

PNG supports:

- Indexed / Palette-based images (24-bit RGB or 32-bit RGBA colors)
- Grayscale images (with or without alpha channel)
- Transparency (both normal and alpha channel)

PNG Image

PNG File Structure

PNG Signature

- A PNG file starts with an 8-byte signature that contains decimal values (137 80 78 71 13 10 26 10)
- This signature shows that the remainder of the file contains a single PNG image
- This image consists of a series of chunks starting with an IHDR chunk and ending with an IEND chunk

Chunk Layout

- It is a series of chunks, each of which gives some information about the image
- Each chunk has a header specifying the size and type of chunk
- Each chunk consists of four parts: Length, Chunk Type, Chunk Data, and CRC (Cyclic Redundancy Check)

TIFF (Tagged Image File Format)

- Tagged Image File Format is a flexible and platform-independent image file format
- It supports numerous image processing applications

Extendibility

- This is the ability to add new image types without invalidating the older types

Portability

- TIFF is independent of the hardware platform and the operating system on which it executes

Revisability

- TIFF was designed to be an efficient medium for exchanging image information
- It is used as a native internal data format for image editing applications

TIFF File Structure (Cont'd)

TIFF files are made up of three unique data structures:

Image File Header (IFH)

- The IFH is an 8-byte structure located at offset zero in the file
- The IFH contains important information necessary to correctly interpret the remainder of the TIFF file

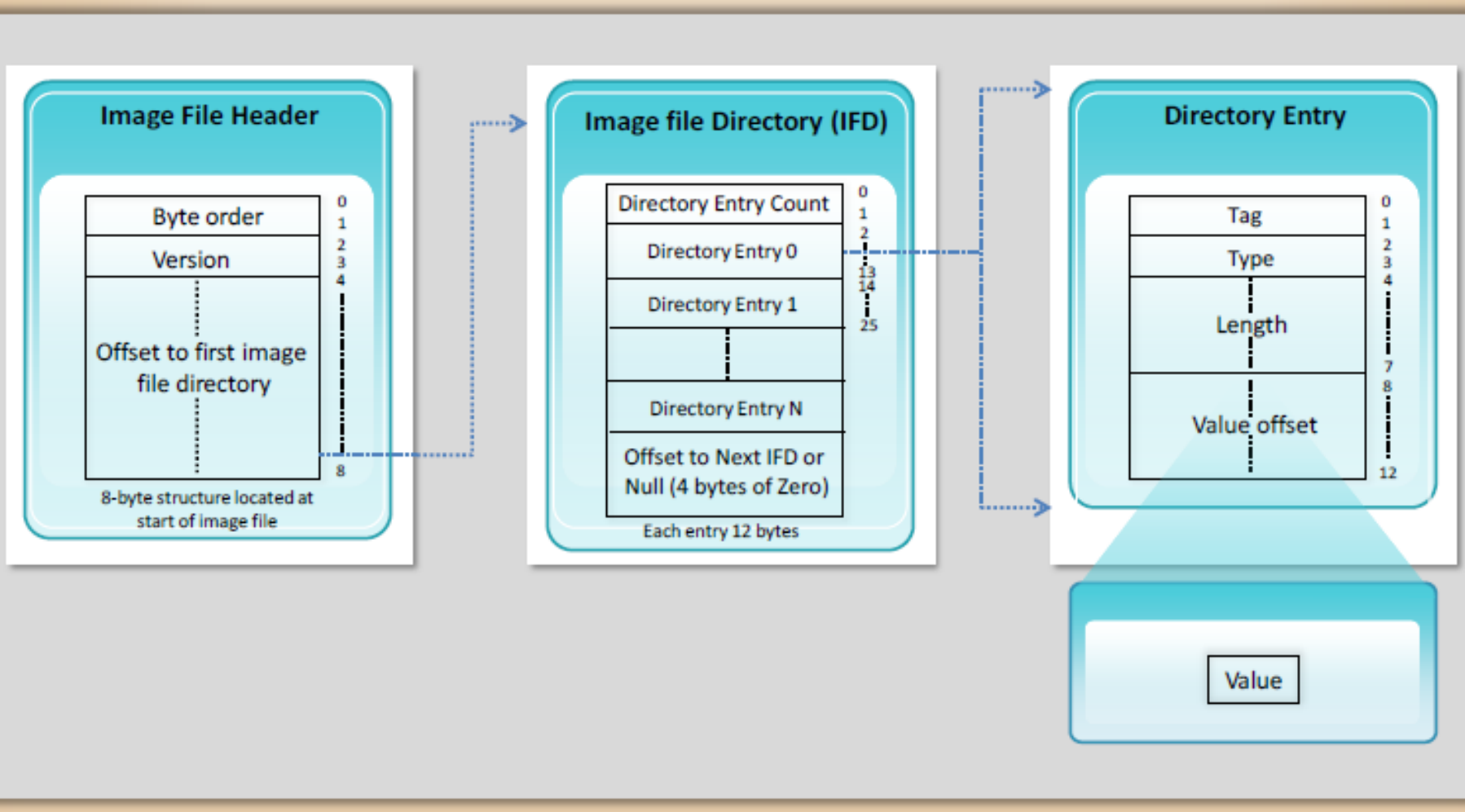
Image File Directory (IFD)

- An IFD consists of a count N, the number of directory entries
- Each entry is of 12 bytes
- If more than one IFD is present, the file contains more than one image

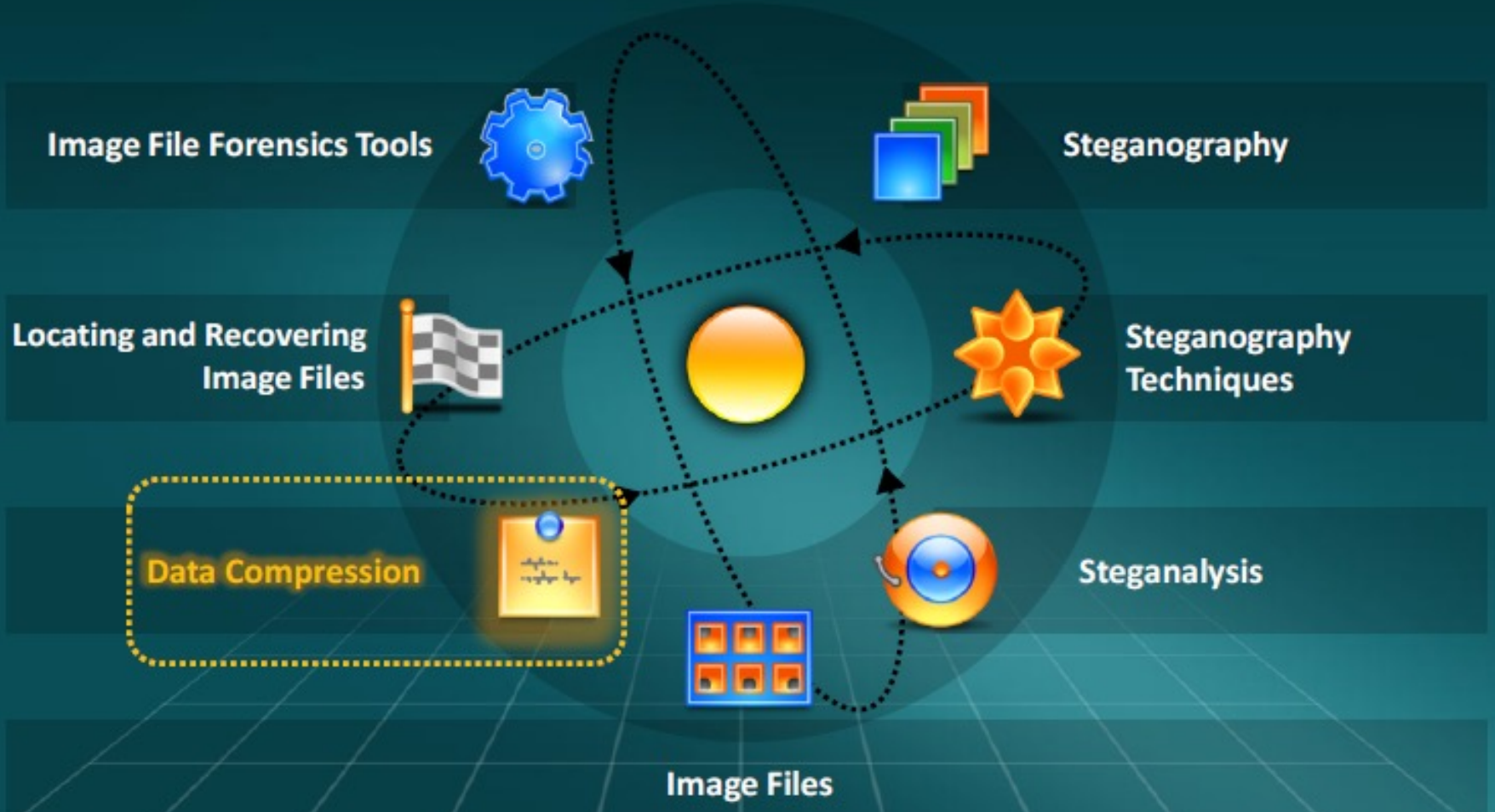
Directory Entry (DE)

- Each DE is exactly 12 bytes in length
- It is segmented into four fields

TIFF File Structure



Module Flow



Understanding Data Compression

Data compression means encoding the data to take up less storage space and less bandwidth for transmission

It helps in saving cost and high data manipulation in many business applications

There are two techniques of data compression:

Lossless Compression, which maintains the data integrity

Lossy Compression, which does not maintain the data integrity

How Does File Compression Work?

In John F. Kennedy's 1961 inaugural address, he delivered this famous line: *"Ask not what your country can do for you -- ask what you can do for your country"*

So, if this is your dictionary:

- ask
- what
- your
- country
- can
- do
- for
- you

The sentence now reads:

"1 not 2 3 4 5 6 7 8 -- 1 2 8 5 6 7 3 4"

When you go through Kennedy's famous words, pick out the words that are repeated and put them into the numbered index

Then, simply write the number instead of writing out the whole word

Lossless Compression

1

This technique uses data compression algorithms that allow the exact original data to be re-created from the compressed data

Original Data

Compressed Data

Restored Data

2

It maintains the data integrity and is used in many applications such as WinZip file format

3

Most image file formats such as PNG, GFI, and TIFF use a lossless compression method

4

Lossless data compression algorithms

- Huffman Coding Algorithm
- Lempel-Ziv Coding Algorithm

Huffman Coding Algorithm (Cont'd)

Huffman Coding Algorithm is a **fixed-to-variable length code**

A Huffman encoder takes a block of input characters with a **fixed length** and produces a block of output bits of **variable length**

The basic idea in Huffman coding is to assign a **short codeword** to those input blocks with **high probabilities** and a **long codeword** to those with **low probabilities**

A Huffman code is designed by **merging** the two least probable **characters** together

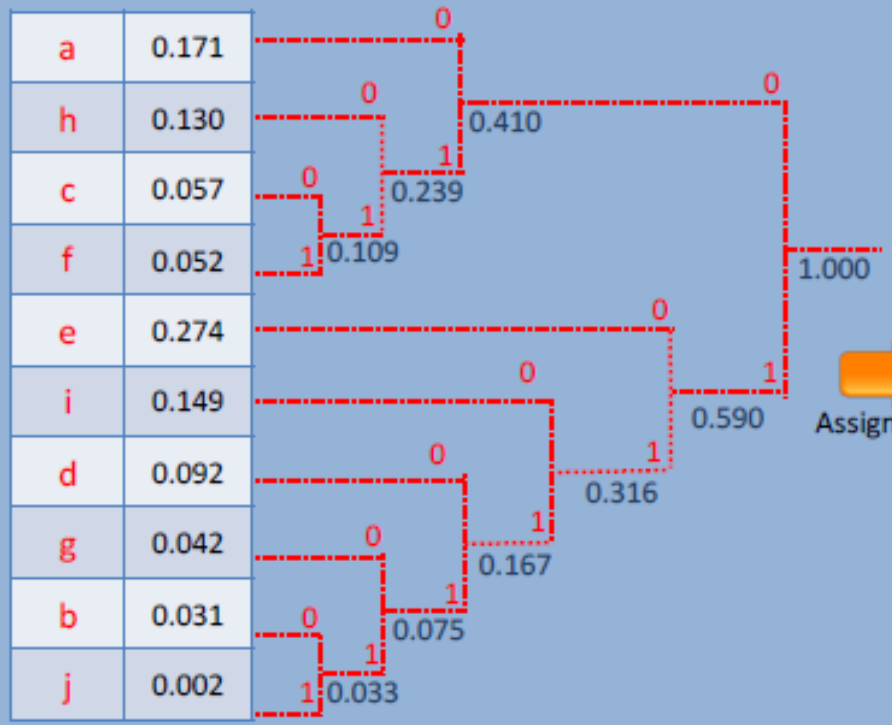
Repeat this merging **process** until there is only **one character remaining**

Huffman Coding Algorithm

Example shows how it works:

X	p(X)
a	0.171
b	0.031
c	0.057
d	0.092
e	0.274
f	0.052
g	0.042
h	0.130
i	0.149
j	0.002

Build Tree



Assign Code

X	Code
a	00
b	111110
c	0110
d	1110
e	10
f	0111
G	11110
H	010
I	110
J	111111

Lempel-Ziv Coding Algorithm (Cont'd)

Lempel-Ziv Coding Algorithm is a variable-to-fixed length code

The Lempel-Ziv code is not designed for any particular source but for a large class of sources

In this, the input sequence is parsed into non-overlapping blocks of different lengths

Dictionary of these blocks is constructed and the following algorithm is used

Encoding Algorithm

- Initialize the dictionary to contain all blocks of length one (D={a,b})
- Search for the longest block W in the dictionary
- Encode W by its index in the dictionary
- Add W followed by the first symbol of the next block to the dictionary
- Go to Step 2

Lempel-Ziv Coding Algorithm

An example of encoding:

Data	a	b	b	a	a b	b a	a b	a b b	a a	a a	b a a	b b	a
	0	1	1	0	2	4	2	6	5	5	7	3	0

Dictionary			
Index	Entry	Index	Entry
0	a	7	b a a
1	b	8	a b a
2	a b	9	a b b a
3	b b	10	a a a
4	b a	11	a a b
5	a a	12	b a a b
6	a b b	13	b b a

Lossy Compression

- Lossy methods provide a high degree of compression and small compressed files, but during decompression, a certain amount of data is lost
- It does not maintain data integrity
- It is never used for business data or text files

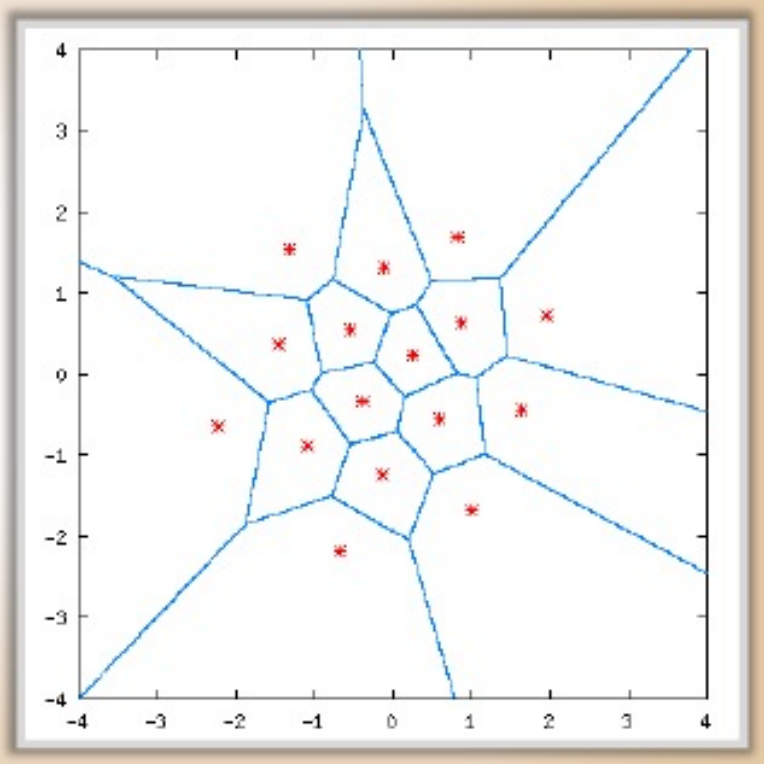
Original Data

Compressed Data

Restored Data

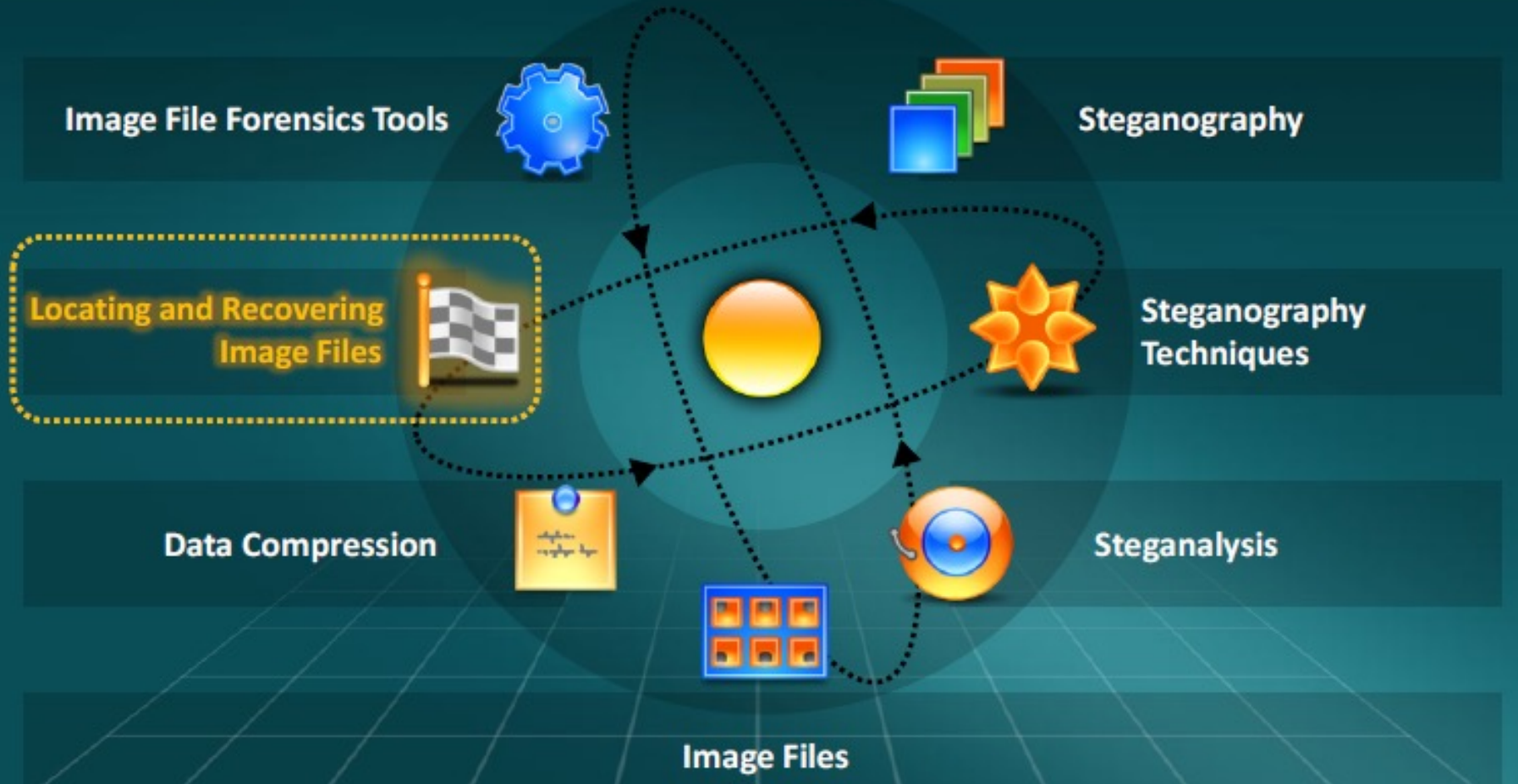
Vector Quantization

Vector quantization is a lossy data compression technique



This technique is based on the principle of block coding, which means it replaces a block of information with an approximate average value

Module Flow



Best Practices for Forensic Image Analysis

Methods of acquiring evidence should be forensically sound and verifiable

Forensic image(s) should be captured using hardware/software that is capable of capturing a "bit stream" image of the original media

Digital evidence submitted for examination should be maintained in such a way that the integrity of the data is preserved

Properly prepared media should be used when making forensic copies to ensure no commingling of data from different cases

Forensic image(s) should be archived to media and maintained in a manner consistent with departmental policy and applicable laws

Use write blockers to prevent the evidence from being modified

Prevent exposure to evidence that may be contaminated with dangerous substances or hazardous materials

Document the current condition of the evidence

Forensic Image Processing Using MATLAB

1. It ensures the image processing steps used are completely documented and hence can be replicated

2. The source code for all image processing functions is accessible for scrutiny and testing

3. It ensures that numerical precision is maintained all the way through the enhancement process

4. Advanced image processing algorithms are used

Advantages of MATLAB

Advantages

Recording of the processing used

- MATLAB is used to **process images** by writing function files or script files
- These files form a **formal record** of the processing used and ensure that the final results can be **tested and replicated**

Numerical accuracy

- It ensures maximal **numerical precision** in the final result
- An image can be read into memory and the data cast into double precision **floating point values**

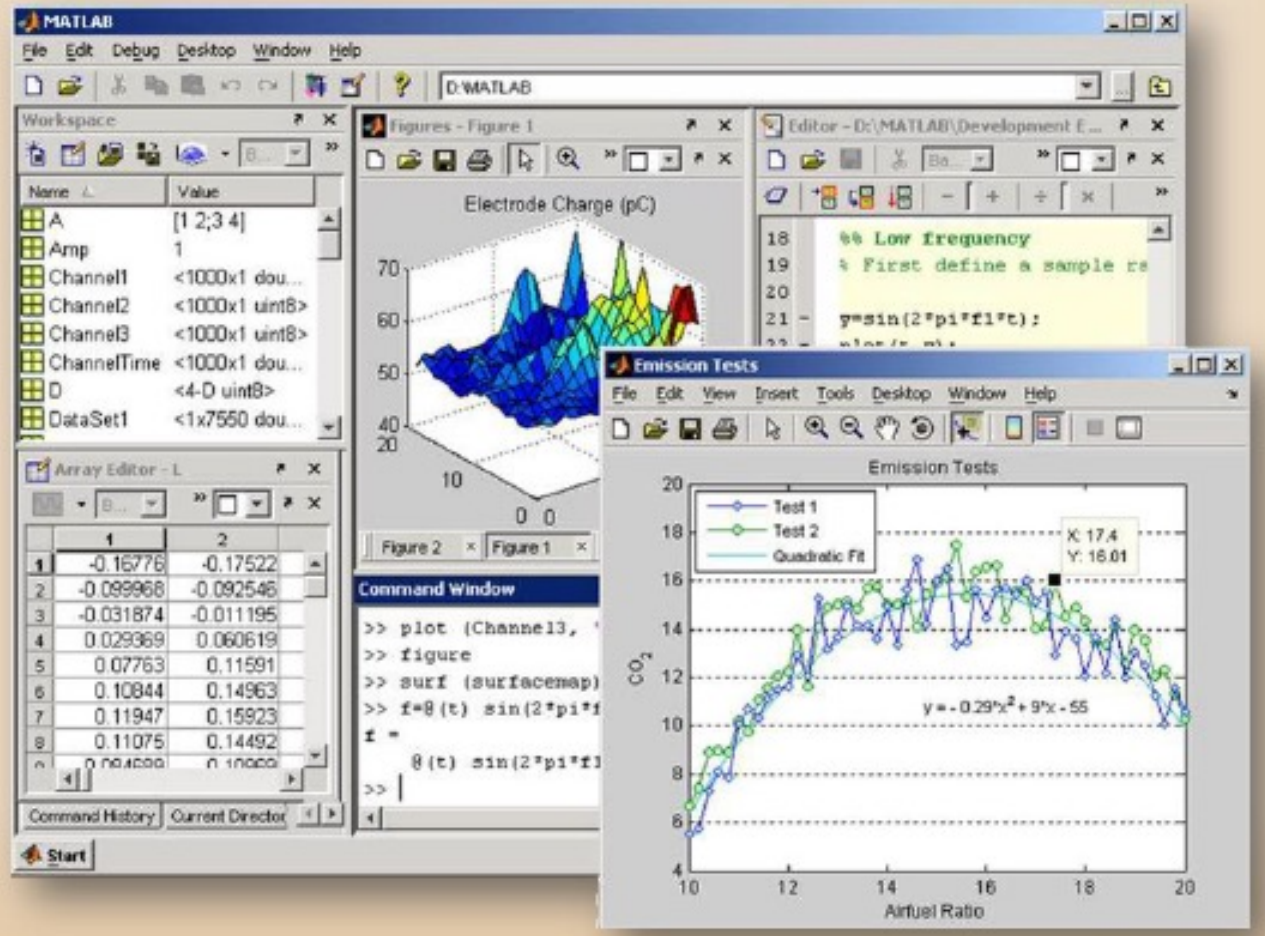
Access to implementation details

- Functions written in the MATLAB language are **publicly readable** as plain text files

Advanced algorithms

- It provides strong **mathematical and numerical support** for the implementation of advanced algorithms

MATLAB Screenshot



http://www.mathworks.com

Locating and Recovering Image Files

Salvaging

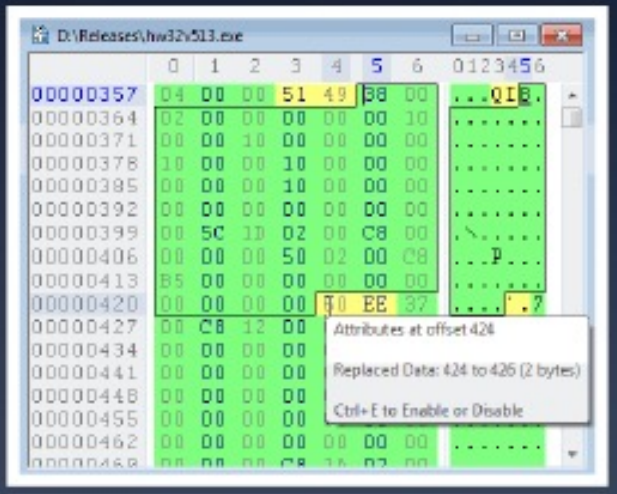
- Collecting and regenerating the image from pieces of an image file dispersed into many areas on the disk is known as salvaging

Carving

- It is the process of data recovery
- It uses the database of headers and footers (essential string of bytes) for a specific file type and recovers files from the raw disk image
- File carving also works if the file system metadata has been destroyed

Analyzing Image File Headers

- Investigators **analyze image file headers** when new file extensions are present that forensic tools cannot recognize
- File headers are accessed with the help of a **hexadecimal editor** such as the **Hex Workshop**
- **Hexadecimal values** present in the header can be used to **define a file type**



Repairing Damaged Headers (Cont'd)

JPEG files would include letters "JFIF" after hexadecimal values

Example: JPEG files have a hexadecimal value of: FF D8 FF E0 00 10

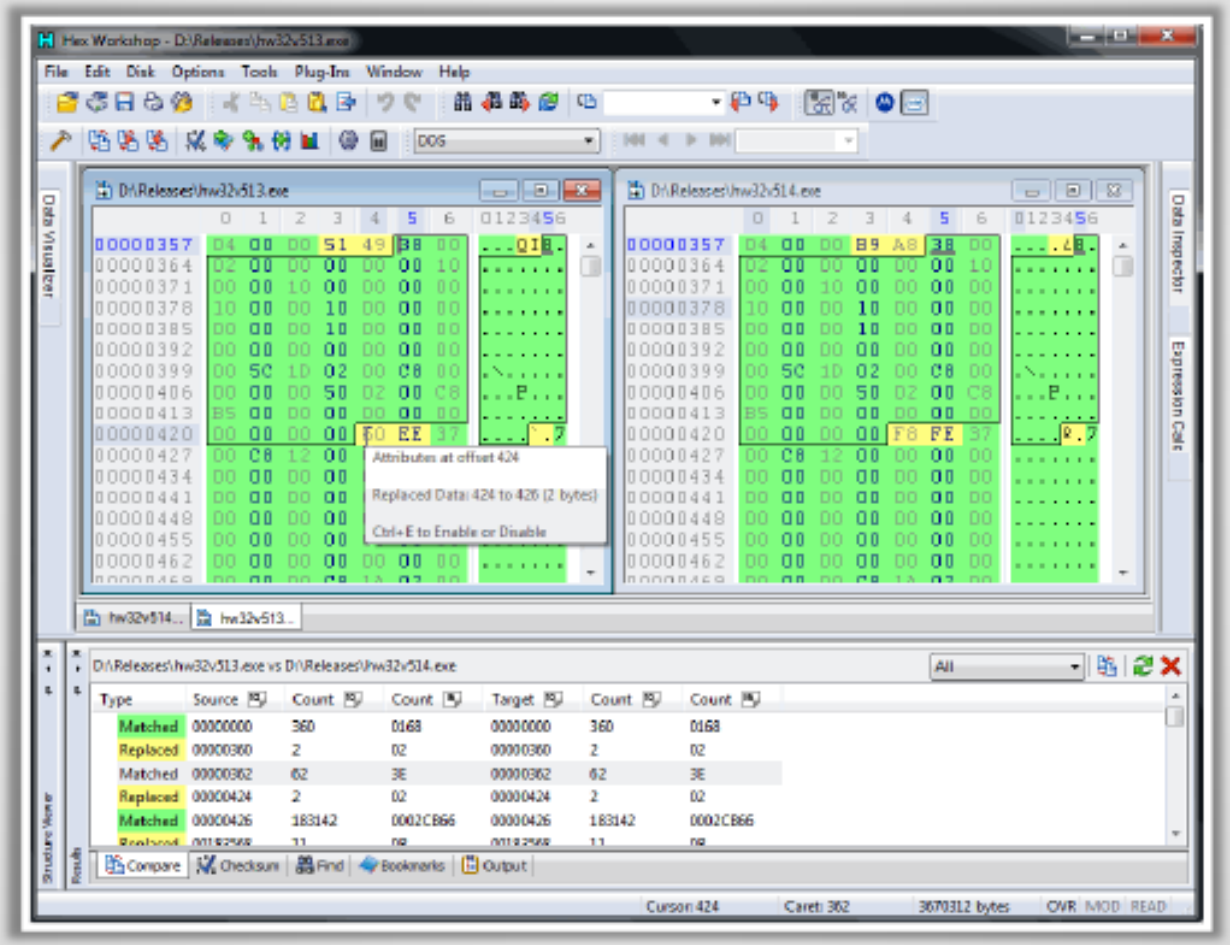
The HEX Workshop application can be used to repair the damaged headers by the process of comparison

Investigators recover data remnants from free space

This data would be similar to headers from common image files

Header data that is partly overwritten can be used to repair the damaged headers

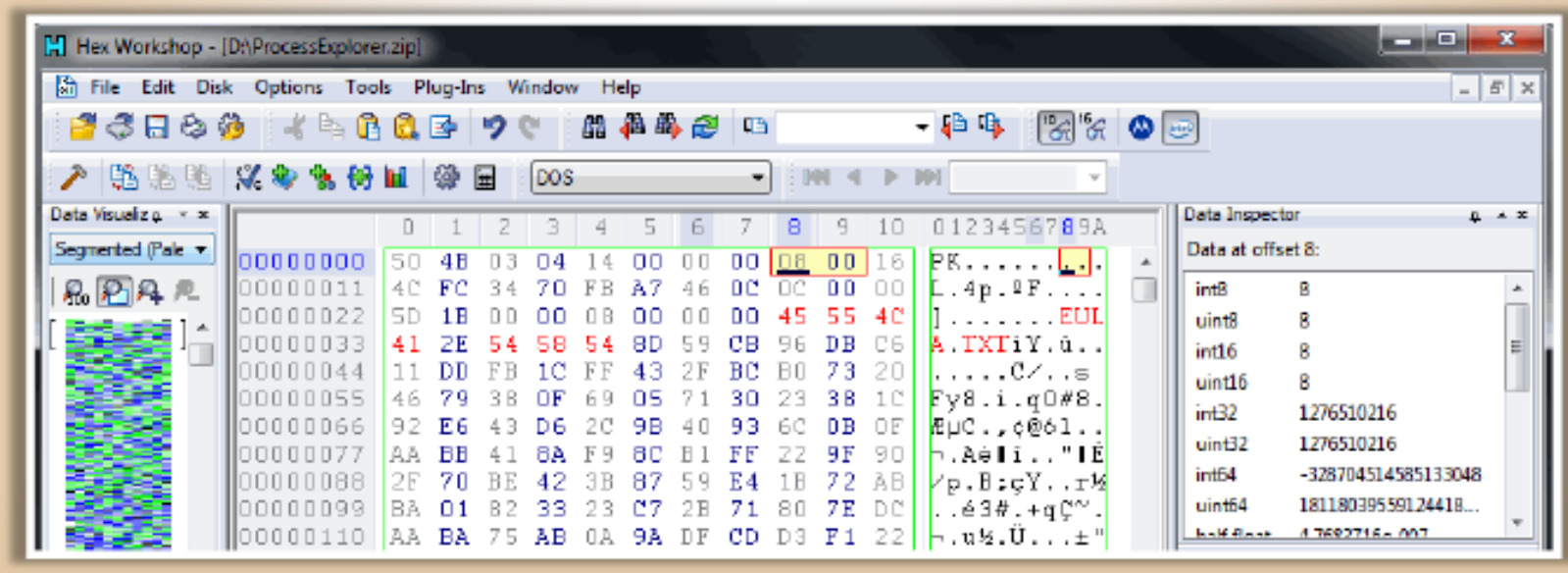
Repairing Damaged Headers



Reconstructing File Fragments

- Corruption of the data prevents investigators from reconstructing file fragments for image files
- Data corruption can be:
  - Accidental
  - Intentional

- File fragments can be reconstructed by examining a suspect disk with the help of the DriveSpy application
- Investigators can build the case based on the data reconstructed



Identifying Unknown File Formats

To understand unknown image file formats ,you should know about non-standard file formats:

- Targa (.tga)
- Raster Transfer Language (.rtl)
- Photoshop (.psd)
- Illustrator (.ai)
- Freehand (.h9)
- Scalable vector graphics (.svg)
- Paintbrush (.pcx)

Tools to identify the unknown file formats:

- IrfanView
- ACDSee Photo Manager 12
- Thumbsplus
- AD Picture Viewer Lite
- Max
- FastStone Image Viewer
- XnView

Identifying Image File Fragments

The first step in recovering the deleted data files is to **identify the image files' fragments**

Recover all the fragments to re-create the image if the image file is fragmented across different disk areas

Recovering a piece of file is called **salvaging** or **carving**

After recovering the parts of the fragmented image file, **restore the fragments** and continue the forensic investigation

Identifying Copyright Issues on Graphics

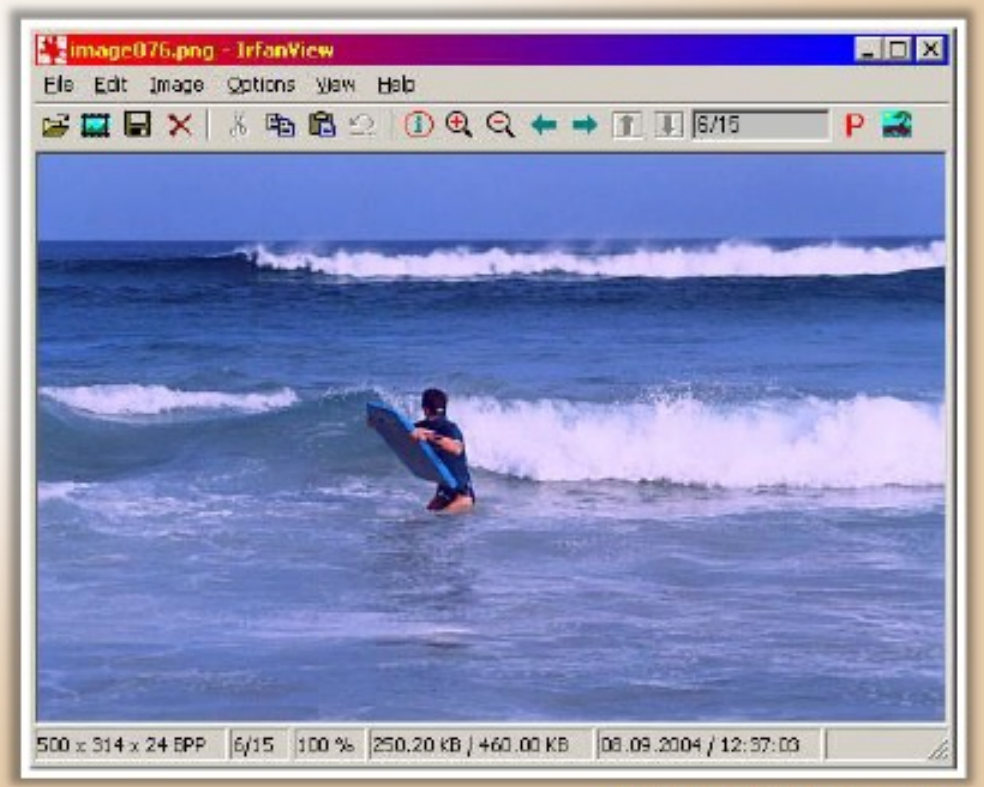
Section 106 of the 1976 Copyright Act:

The owner of copyright under this title has the exclusive rights to do and to authorize any of the following:

- To reproduce the copyrighted work in copies or phonorecords;
- To prepare derivative works based upon the copyrighted work;
- To distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending;
- In the case of literary, musical, dramatic, and choreographic works, pantomimes, and motion pictures and other audiovisual works, to perform the copyrighted work publicly;
- In the case of literary, musical, dramatic, and choreographic works, pantomimes, and pictorial, graphic, or sculptural works, including the individual images of a motion picture or other audiovisual work, to display the copyrighted work publicly;
- In the case of sound recordings, to perform the copyrighted work publicly by means of a digital audio transmission

Picture Viewer: IrfanView

- IrfanView is an image/graphics viewing program that supports many file formats, such as:
  - Targa (.tga)
  - Illustrator (.ai)
  - Scalable vector graphics (.svg)
  - FlashPix (fpx)
  - Portable network graphics (.png)
  - Joint photographic experts group (.jpg)

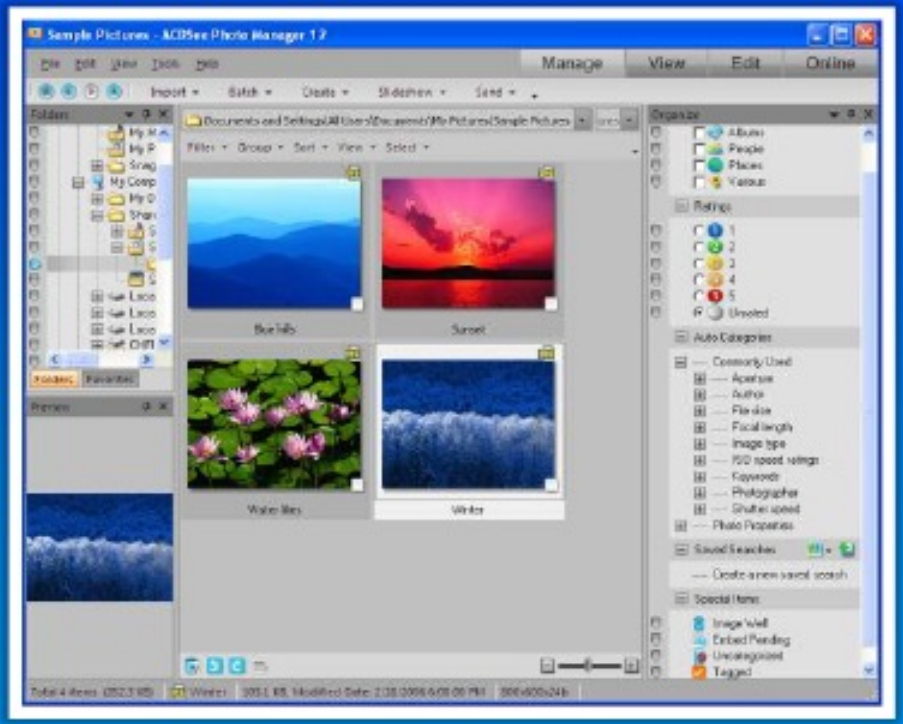


http://www.irfanview.com

Picture Viewer: ACDSee Photo Manager 12

ACDSee Photo Manager 12 is an image viewing program that enables investigators to:

01	Find images
02	View images
03	Manage image files on the drive
04	Search and view unknown file formats



http://store.acdsee.com

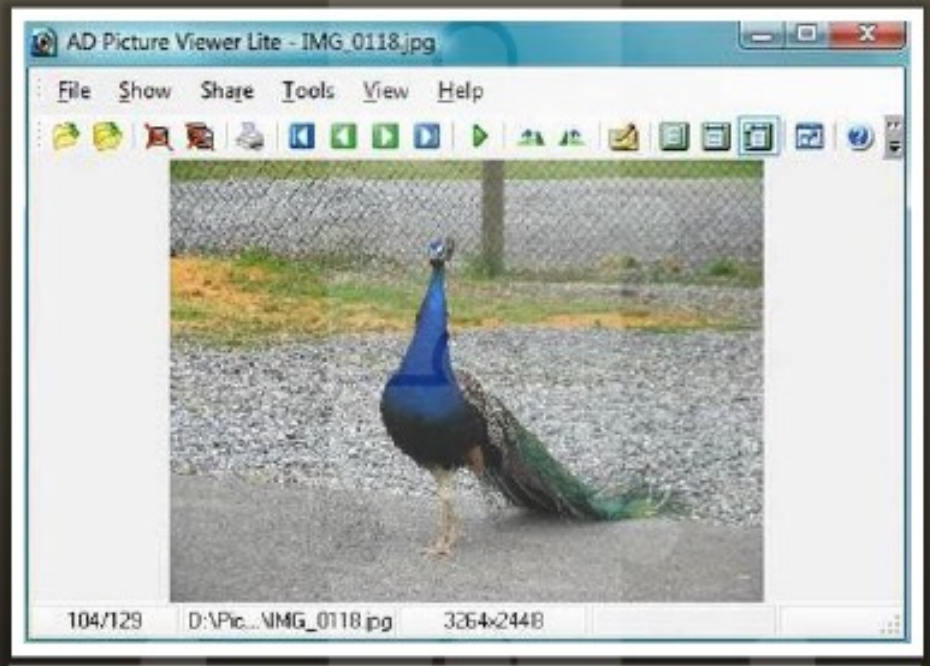
Picture Viewer: Thumbsplus

- ThumbsPlus is an image viewing program that enables investigators to:
  - View images from a drive database
  - View files other than images such as audio and multimedia files
  - Catalog image files for future reference
  - Save RAW files

http://www.cerious.com

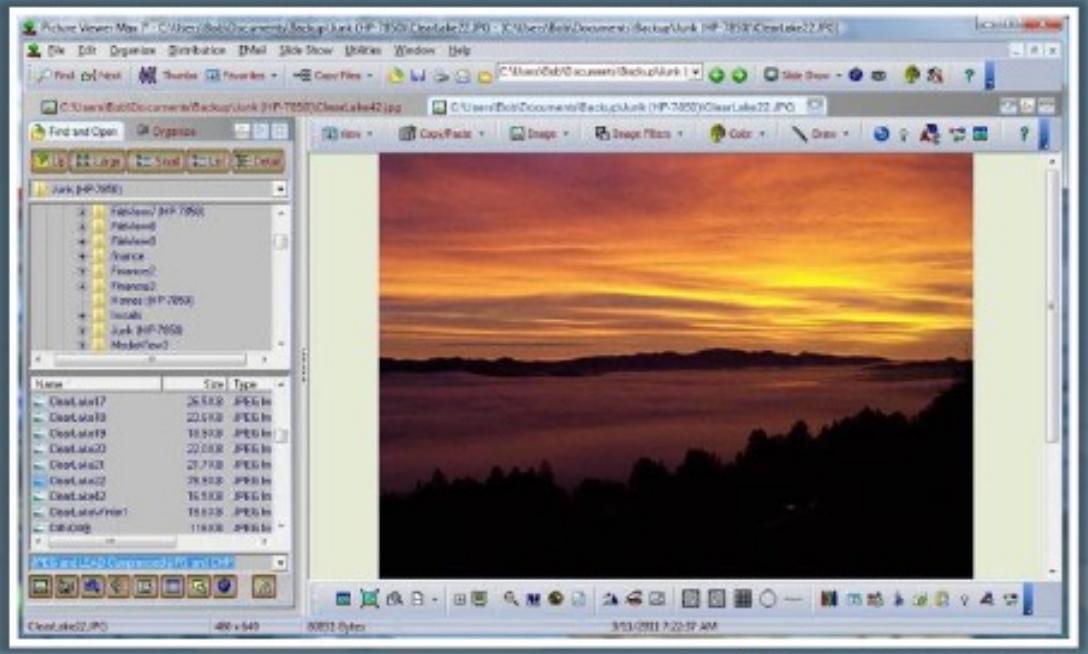
Picture Viewer: AD Picture Viewer Lite

- AD Picture Viewer Lite is an image viewer that support all popular **image file formats**
- It allows you to **view**, **print**, **organize**, and **catalog** the image
- It opens all images in a folder, including in subfolders, using a **single command**



Picture Viewer Max

Picture Viewer Max is an image and multimedia viewer for Windows 7 that helps to locate, view, edit, print, organize, and send/receive picture/image files over the Internet



http://accessoryware.com

Picture Viewer: FastStone Image Viewer

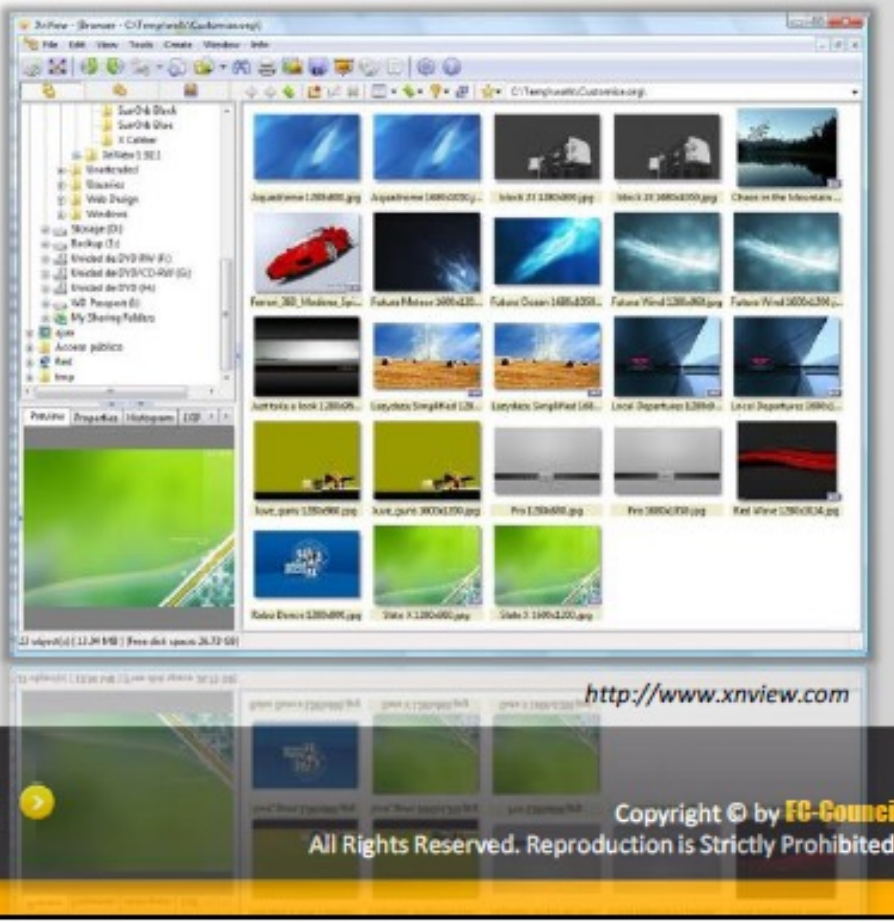
- FastStone Image Viewer is an **image browser**, **converter**, and **editor**
- It has many features that include image viewing, management, comparison, red-eye removal, emailing, resizing, cropping, retouching and color adjustments
- **Features:**
  - Supports **common image formats**, loading and saving of JPEG, JPEG2000, GIF, BMP, PNG, PCX, TIFF, WMF, ICO, CUR, and TGA
  - Supports **zoom - full screen viewer**
  - Crystal **clear** and **customizable** magnifier
  - Image **EXIF metadata** support

Picture Viewer: XnView

- XnView is software for viewing and converting graphic files
- It exists for Windows, MacOS X, Linux x86, Linux ppc, FreeBSD x86, OpenBSD x86, NetBSD x86, Solaris sparc, Solaris x86, Irix mips, HP-UX, and AIX

Features

- Imports about 400 graphic file formats
- Exports about 50 graphic file formats
- Supports multipage TIFF, animated GIF, and animated ICO
- Supports Image IPTC, EXIF metadata
- Supports lossless rotate and crop (jpeg)
- Creates web pages



http://www.xnview.com

Faces - Sketch Software

http://www.iqbiometrix.com

FACES contains a **data bank** of over 3,850 facial features, along with tools and accessories that allow the user to rapidly put a composite image together

Generally used by **law enforcement** agencies in identifying suspects

Digital Camera Data Discovery Software: File Hound

File Hound is a software package that helps to deal with crimes involving digital pictures

It searches images based on file signature

It can preview/print reports based on search results

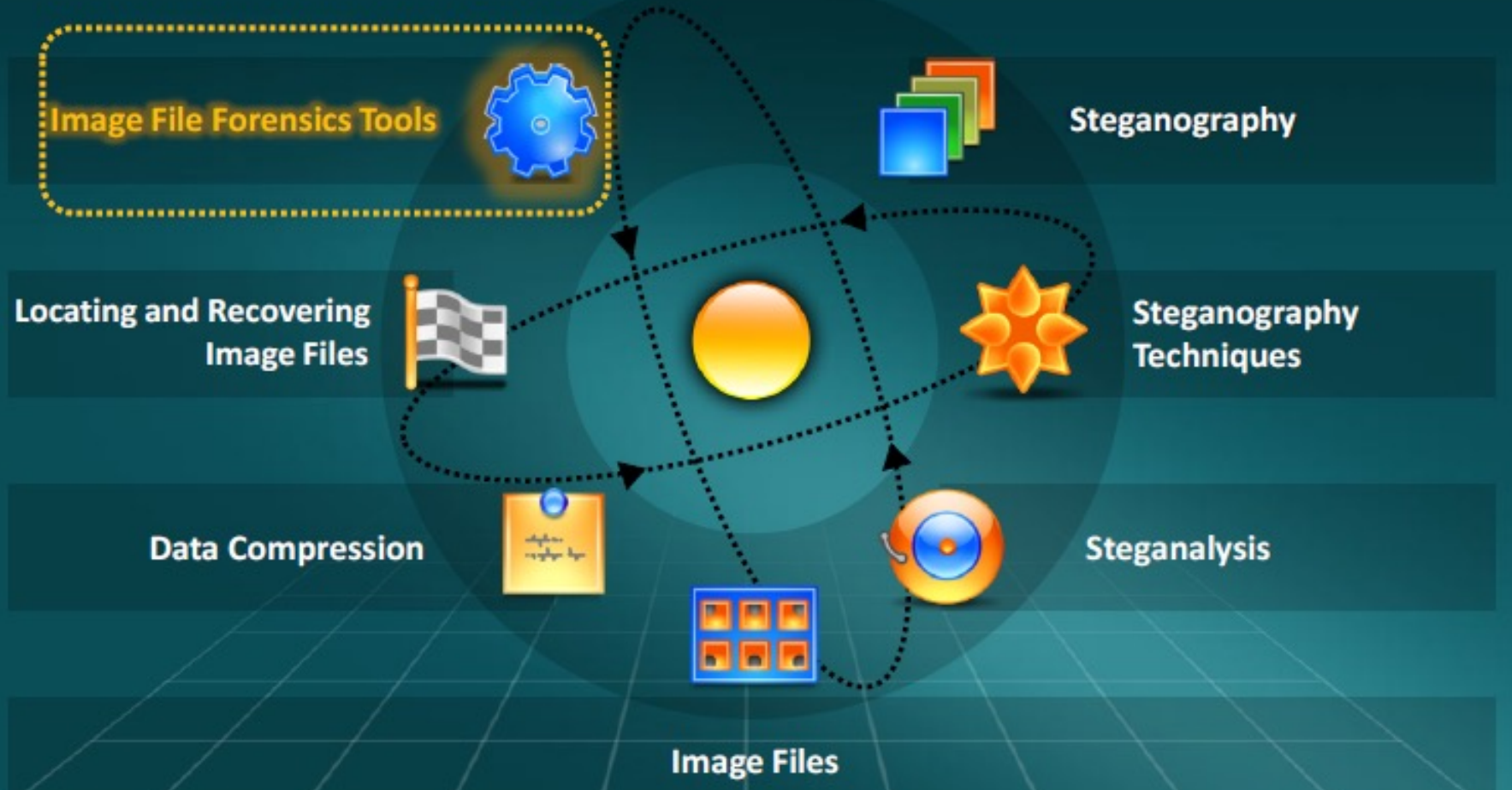
It distinguishes PNG, GIF, JPG, TIF, WMF, BMP, ICO

It previews thumbnail images and file information

It searches for files based on filenames

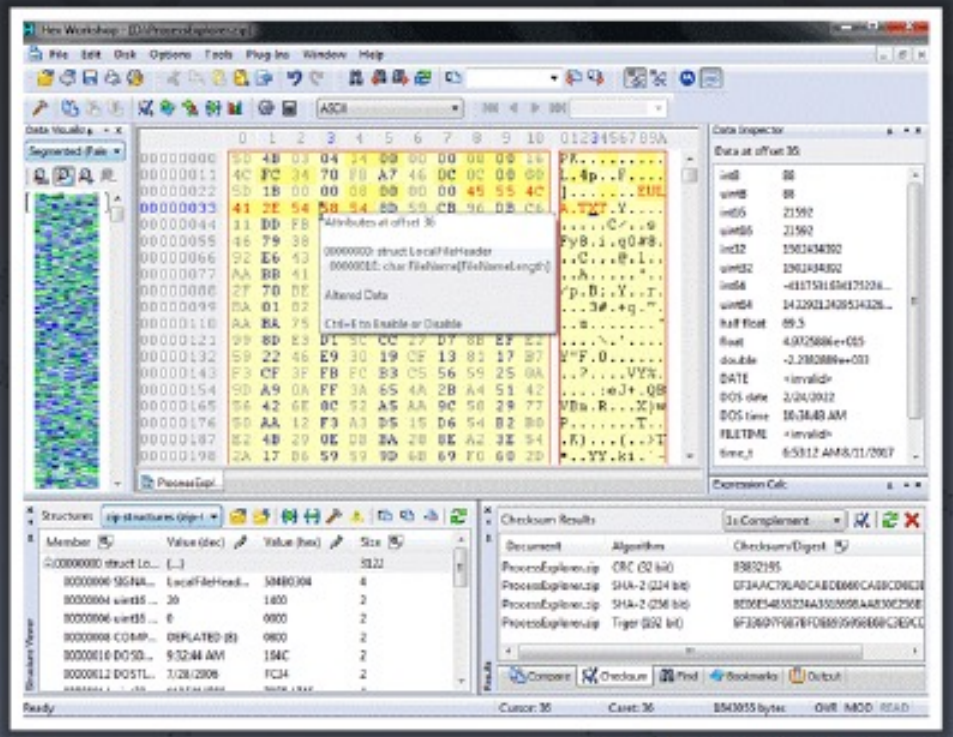
http://filehound.cerias.purdue.edu

Module Flow



Hex Workshop

- The Hex Workshop application can **detect** and **write messages** onto a file
- Investigators use the Hex Workshop tool to **reconstruct** the damaged file headers



GFE Stealth™ - Forensics Graphics File Extractor

- GFE Stealth™ is used to identify and capture graphics files (pictures) stored on computer hard disk drives
- It is a computer forensics tool that helps identify whether or not a targeted computer system contains evidence in the form of graphics files

Used effectively in computer investigations involving the distribution of child pornography and identity theft

Used in the recovery of deleted graphics files from computer hard disk drives and/or digital flash memory chips

Used to find evidence in corporate, civil and criminal investigations that involve computer graphics files

Used with other computer forensic software to quickly reconstruct and view previously deleted BMP, GIF and JPEG graphics files stored on computer storage media

http://www.forensics-intl.com

Ilook

- ILook is a multi-threaded, Unicode compliant, fast, and efficient forensic analysis tool designed to analyze images taken from seized computer systems and other digital media
- It can be used to examine images obtained from other forensic imaging tools that produce a raw bit stream image

Features

1	2	3	4
Supports FAT12, FAT16, FAT32, FAT32x, VFAT, NTFS, HFS, HFS+, Ext2FS, Ext3FS, SysV AFS, SysV EAFS, SysV HTFS, CDFS, Netware NWFS, Reiser FS, and ISO9660 file systems	Granular extraction facilities that allow all or part of a file system to be extracted from an image	It runs on Windows XP / Server platforms, both 32 and 64 bit versions 	It has file salvage (carve) facilities 

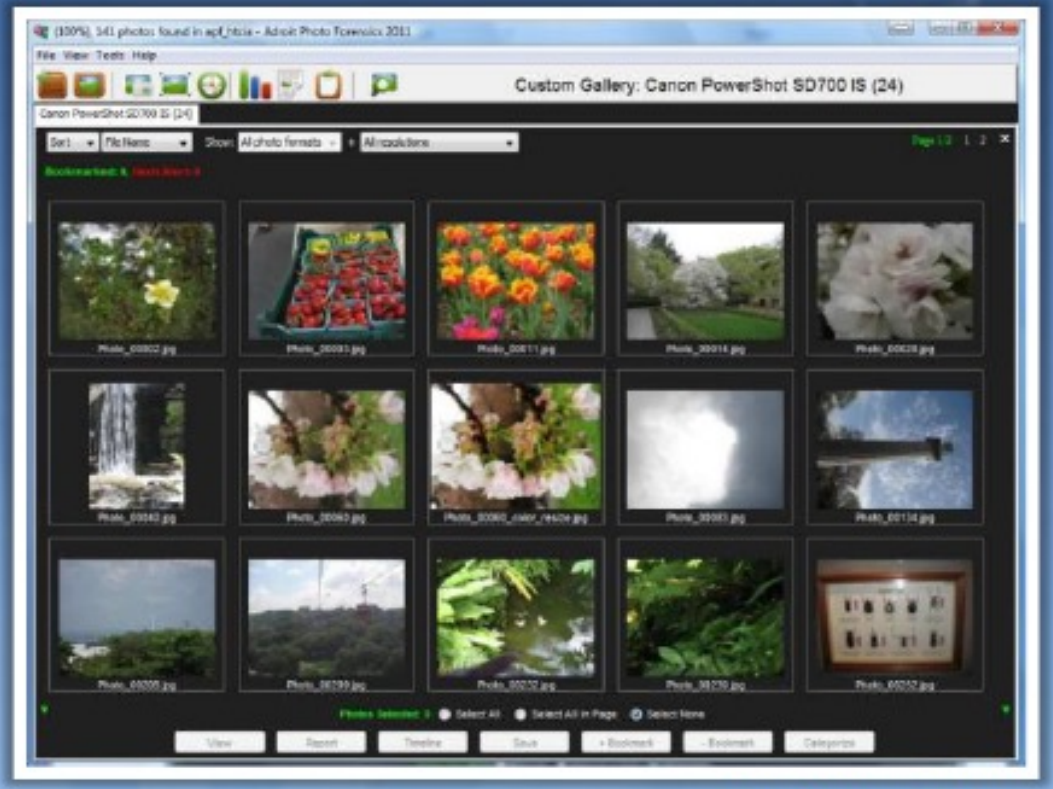
http://www.perlustro.com

Adroit Photo Forensics 2011

Adroit Photo Forensics 2011 is designed for professional evidence recovery and uses **SmartCarving and GuidedCarving** technologies for complete recovery of photo evidence

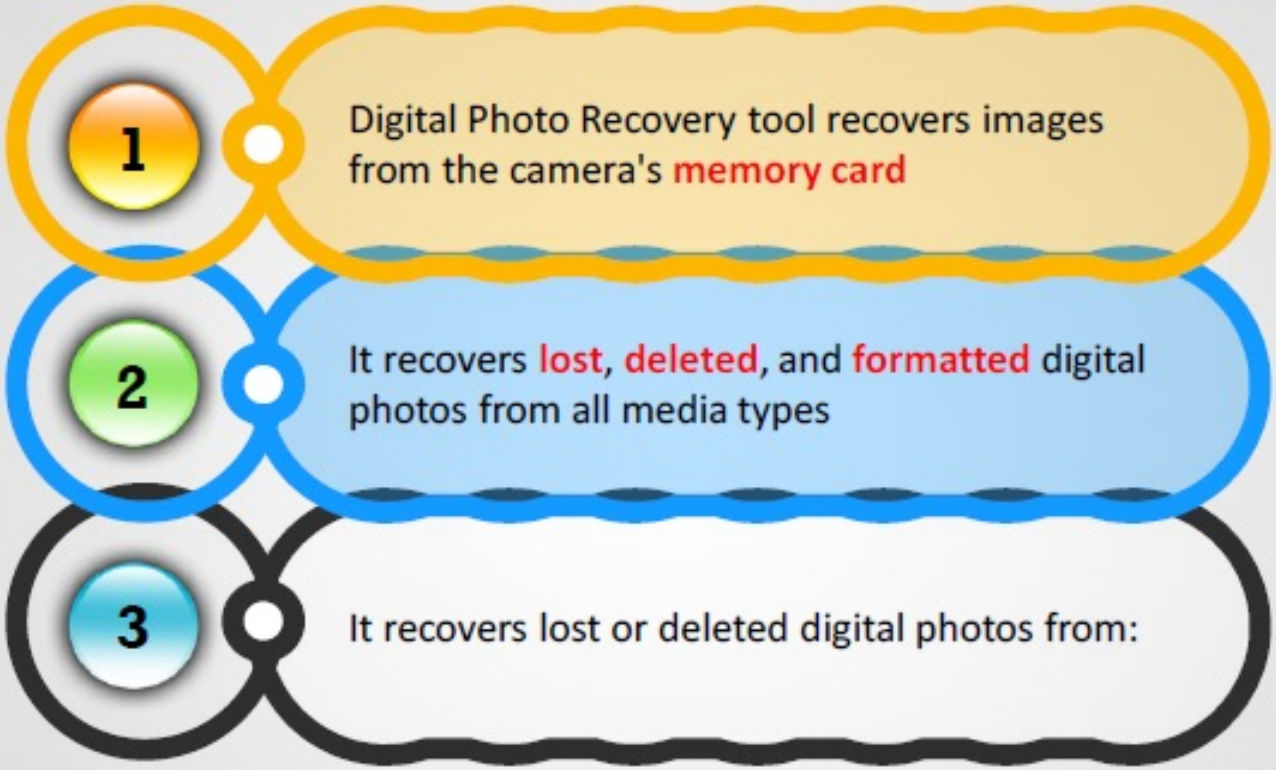
Features:

- Automatically generate case details based on selected evidence
- Supports Encase disk images (single and split), RAW/DD images (single and split), and logical/physical drives
- Generate MD5 and SHA1/SHA256 hashes of photos recovered
- Photo specific grouping based on camera, EXIF Date stamps etc.
- Carving support for unallocated space and slack space between partitions



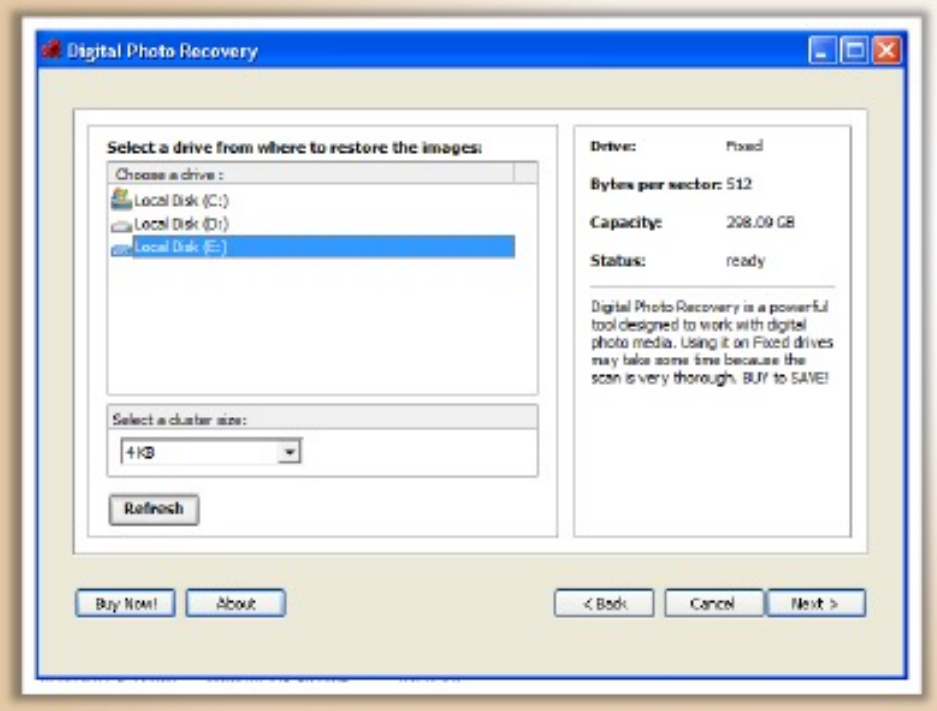
http://digital-assembly.com

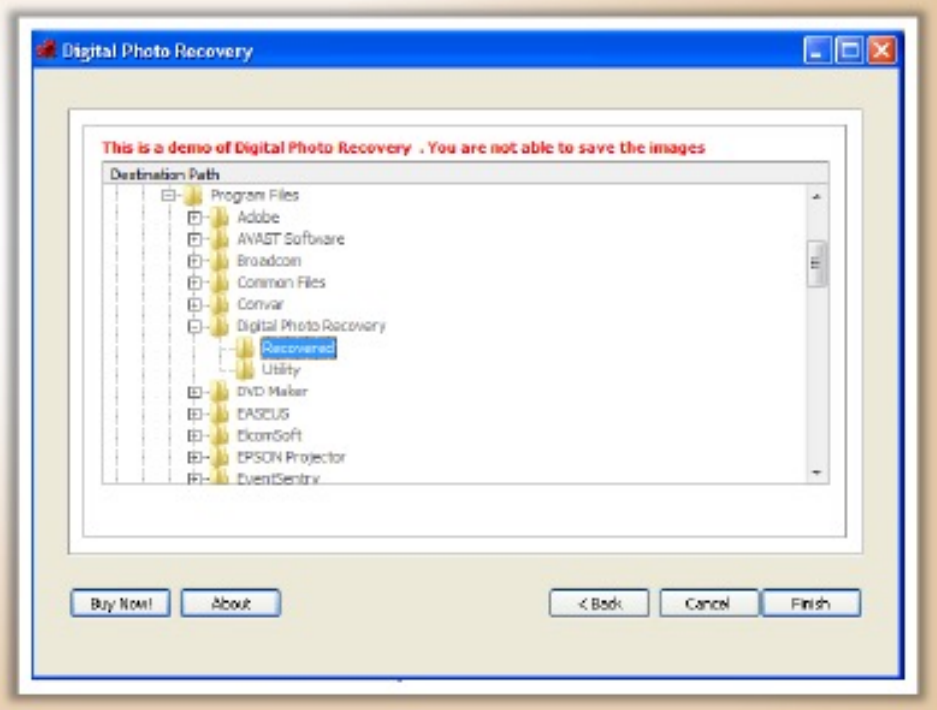
Digital Photo Recovery



- Hard drive
- Secure Digital Card (SD card)
- MultiMediaCard (MMC)
- CompactFlash (CF)
- Floppy Disk
- USB Memory Card
- Digital Cell Phones
- DVD+R DVD-R DVD +RW DVD-RW
- PDA PC Card

Digital Photo Recovery Screenshots





http://www.photosrecovery.com

Stellar Phoenix Photo Recovery Software

Stellar Phoenix Photo Recovery Software recovers lost/deleted/formatted photos, music, and video files from Windows based systems

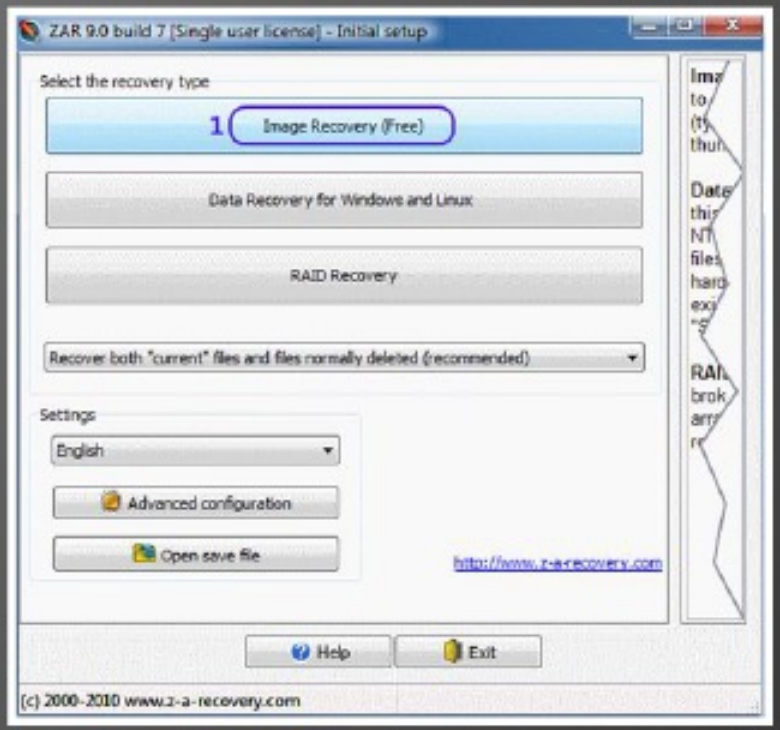
http://www.stellarinfo.com

It recovers photos lost due to deletion, corruption or formatting of SD cards and hard drives

It recovers photos from digital cameras, external hard drives, laptops and desktops, CDs, DVDs, USB Drives, and memory cards

Zero Assumption Recovery (ZAR)

- Zero Assumption Recovery (ZAR) software recovers digital images from memory cards of digital cameras
- It supports various file formats such as GIF, JPEG, TIFF, CRW (Canon RAW data), MOV, AVI movie, WAV, etc.



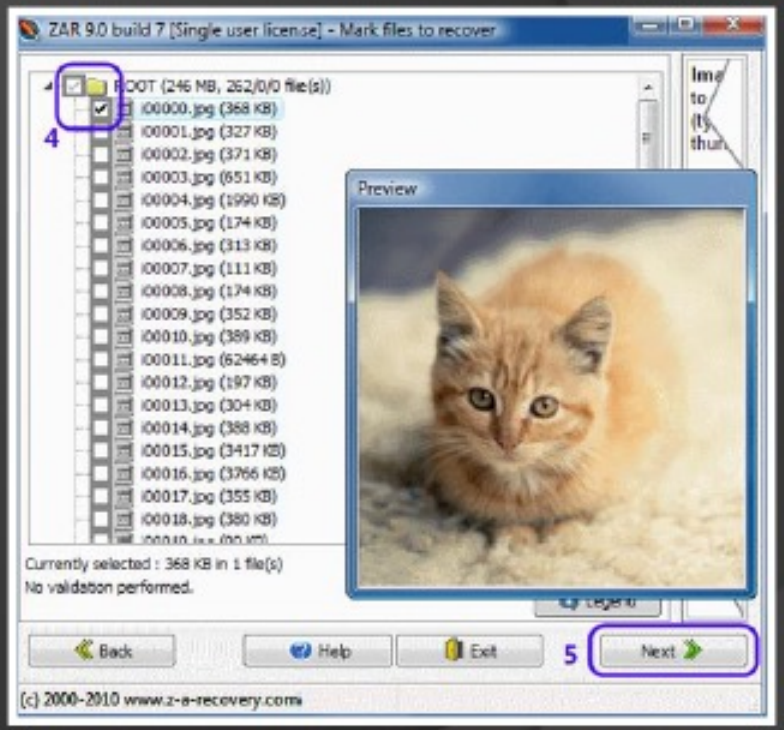
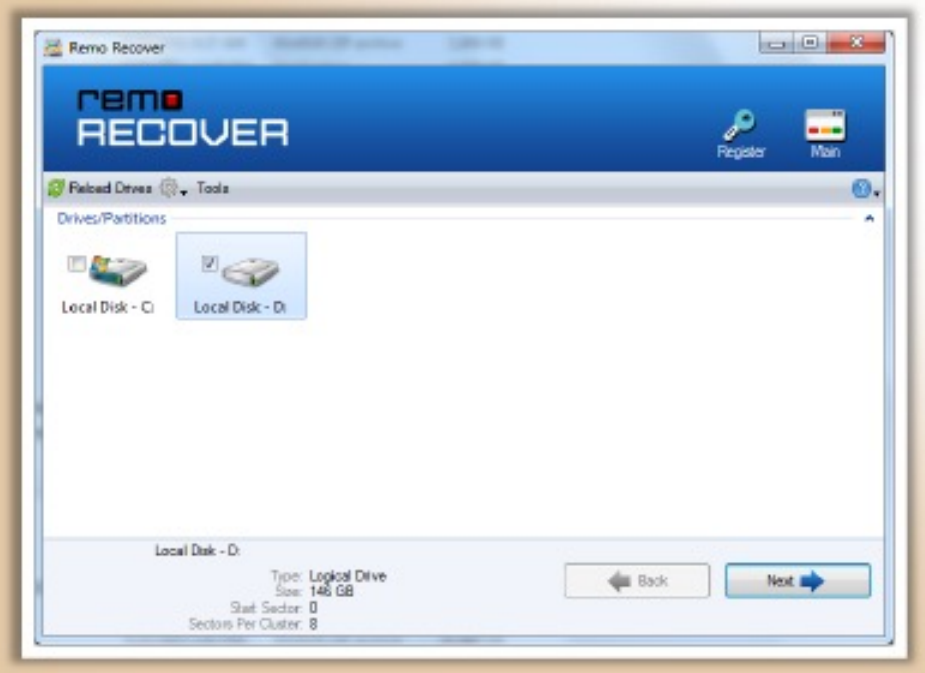
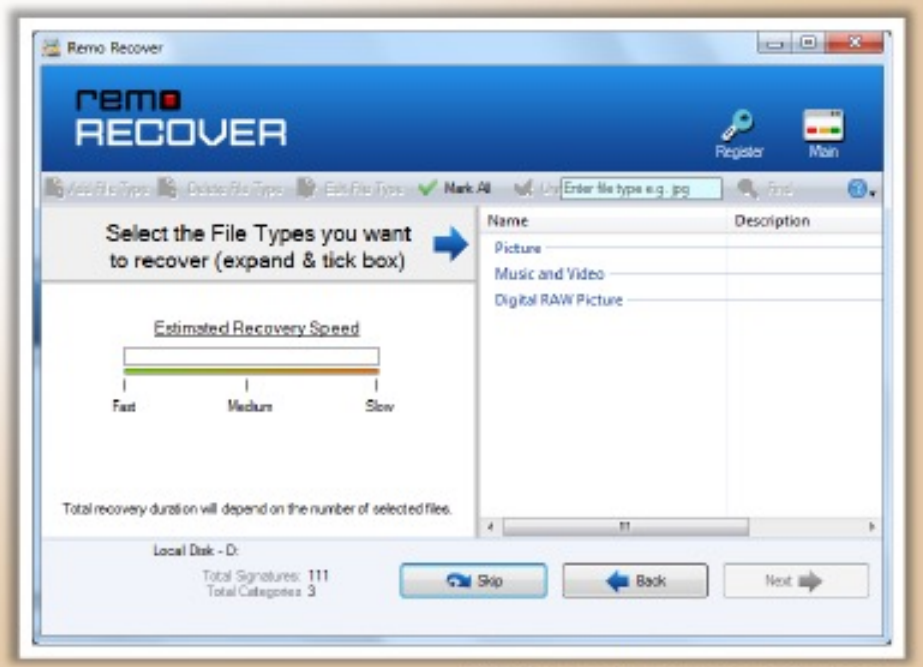


Photo Recovery Software

- Photo Recovery Software recovers digital photos from USB drives, flash cards, SD cards, CF cards, iPods, hard drives and FireWire drives
- It supports recovery of digital photos from all Windows accessible partitions





http://www.photorecoverysoftware.org

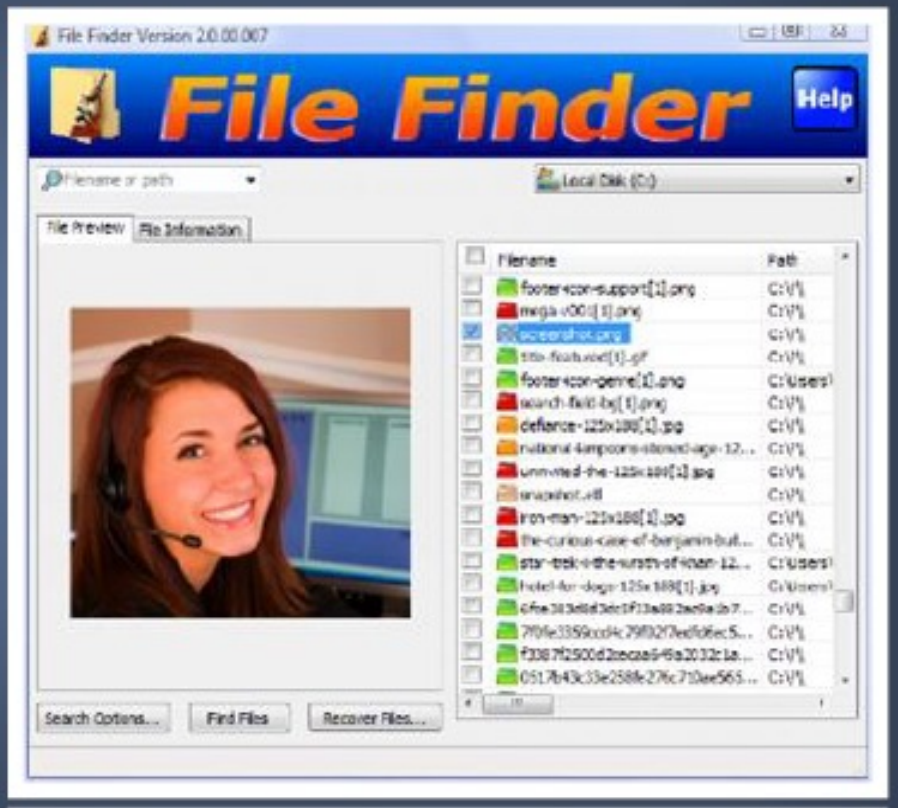
Forensic Image Viewer

- Forensic Image Viewer (FIV) is an in-development tool for the **processing** and **reporting** of still images (JPG's, PNG's, Gif's etc.)

http://www.sandersonforensics.com

File Finder

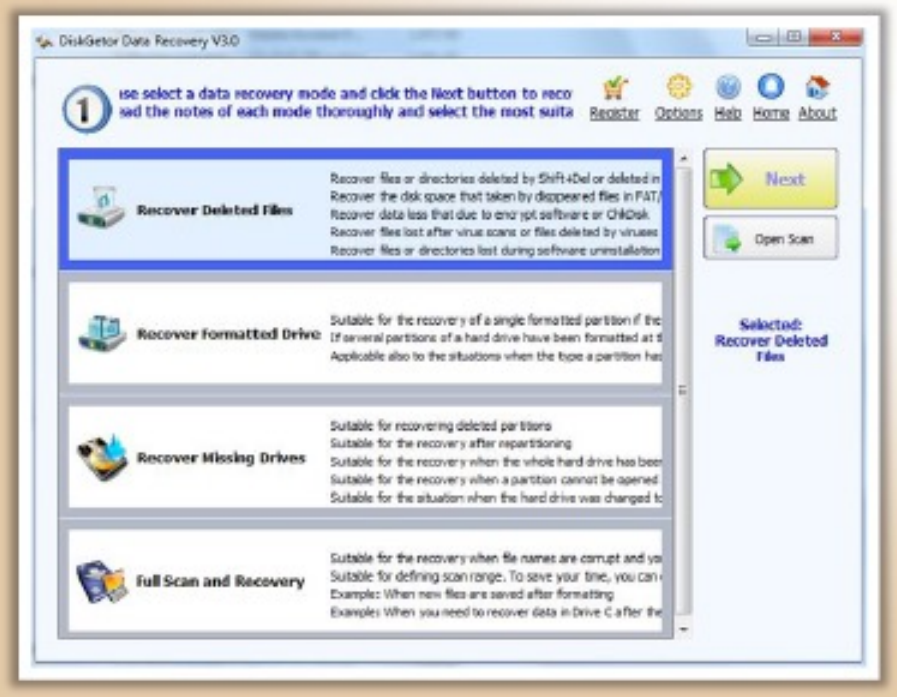
- File Finder recovers lost or deleted files on any computer or digital media
- It recovers image data deleted or lost on PCs and memory cards
- It recovers photos and pictures from MMC cards, SD cards and other photo memory cards

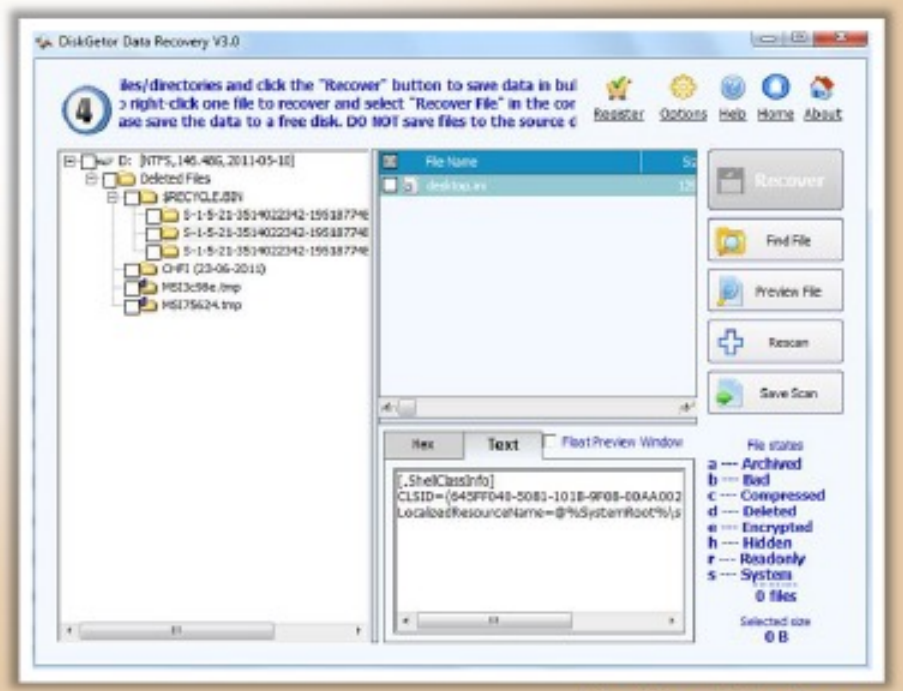


http://recover-lost-files.com

DiskGetor Data Recovery

- DiskGetor Data Recovery is hard drive data recovery software that recovers formatted data, deleted data, lost data, damaged data from hard drives
- It recovers image and photo files from the hard drive

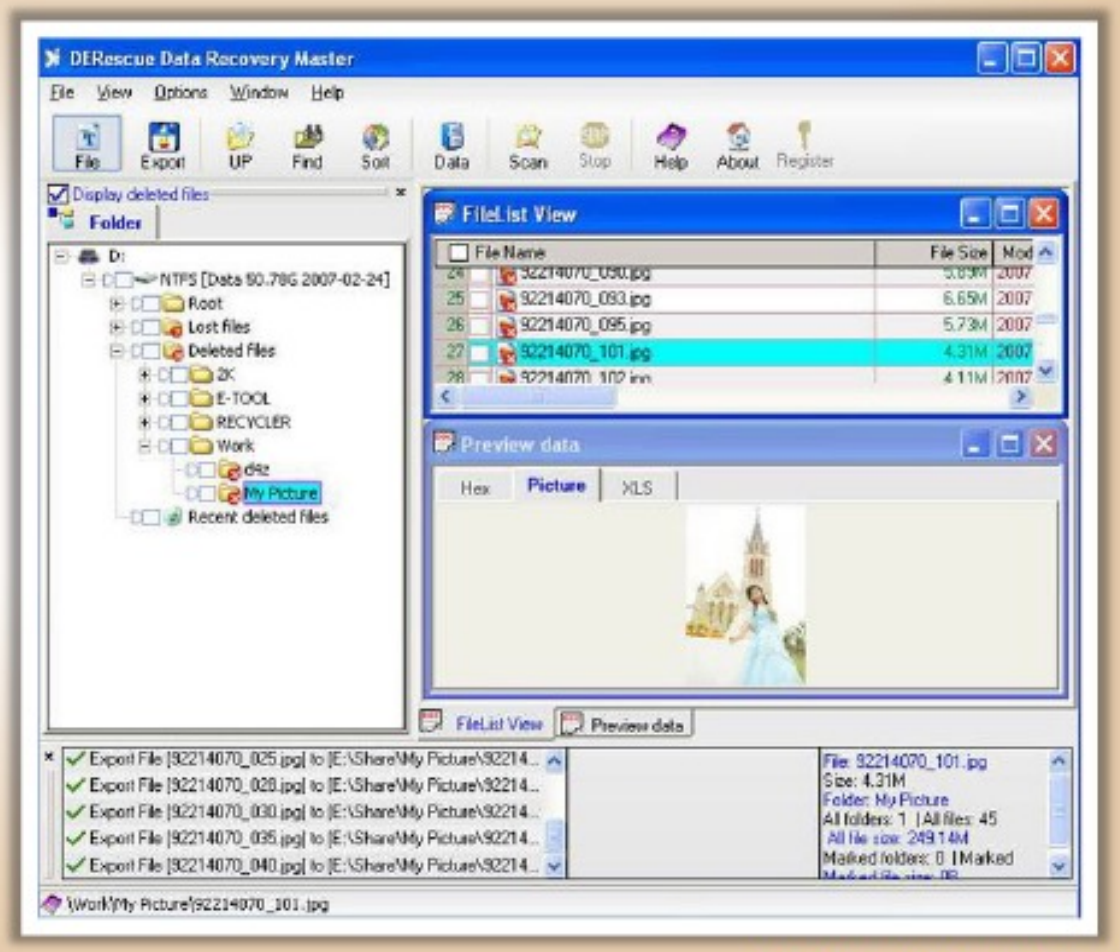




http://www.diskgetor.com

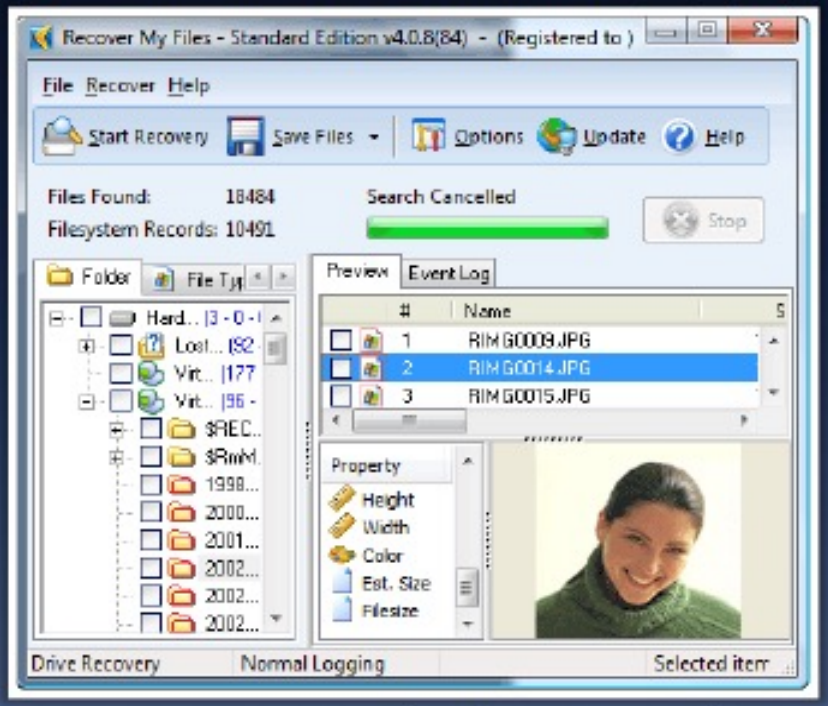
DERescue Data Recovery Master

DERescue Data Recovery Master recovers deleted, formatted and lost digital photos, images and pictures from any types of media cards used by digital cameras or computers



http://www.derescue.com

Recover My Files



http://www.recovermyfiles.com

- Recover My Files data recovery software recovers deleted files emptied from the **Windows Recycle Bin**, or lost due to the format or corruption of a hard drive, virus or Trojan infection, unexpected system **shutdown** or **software failure**
- It recovers **deleted** or **lost picture files** (.jpeg,.jpg,.gif and more) from hard drives, camera cards, USB drives, Zip disks, floppy disks or other removable drives and disks

Universal Viewer

- Universal Viewer is an advanced file viewer for a wide range of formats
- Supported file formats are images, multimedia, Word, Excel, PDF, RTF, Internet, text, etc.

http://www.uvviewsoft.com

Module Summary

- Steganography is the method of hiding information by embedding messages within other, seemingly harmless messages
- Steganalysis is the technology that attempts to defeat steganography—by detecting the hidden information and extracting it or destroying it
- Steganography is used to secure secret communications
- An image is an artifact that reproduces the likeness of some subject
- A file format is a "particular way to encode information for storage in a computer file"
- The standard image file formats include JPEG, GIF, BMP, TAG, and EPS
- Data compression means encoding the data to take up less storage space and less bandwidth for transmission
- Data is compressed by using a complex algorithm to reduce the size of a file
- Lossy compression compresses data permanently by removing information contained in the file

"Is there a file compression program that will help me squeeze 12 hours of work into an 8 hour schedule?"

"Someone sent me an e-mail attachment. From the size, I'm guessing it's either furniture or some sort of farm animal!"