Esteganografia

Roberto Gómez Cárdenas
ITESM-CEM
rogomez@itesm.mx

Lámina 1 Dr. Roberto Gómez C.

Esteganografia

- Area similar a la de criptología.
- Viene del griego stegos (ocultar).
- Conjunto de técnicas que nos permiten ocultar o camuflar cualquier tipo de datos, dentro de información considerada como válida.
- La información puede esconderse de cualquier forma
  - diferentes métodos se han ido desarrollando

Lámina 2 Dr. Roberto Gómez C.

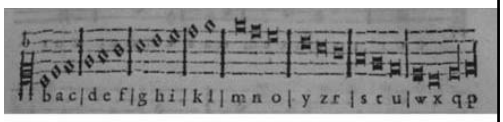
Algunos ejemplos históricos

- Herodoto:
  - 440 ac: Aristagoras de Milet usa esclavos calvos para la revuelta contra los persas
  - Demeratus envía mensaje (tablones cubiertos de cera) a Esparta para avisar de que Xerxes (rey de Persa) tenía intenciones de invadir Grecia.
- Tintas invisibles
  - naturales: jugo limón, leche, orina, sal de amoniaco
  - química: alumbre y vinagre, traspasar cáscara huevo duro
- Chinos: texto escrito sobre seda china

Lámina 3 Dr. Roberto Gómez C.

Algunos ejemplos históricos

- Siglo XVII: Schola Steganographica, Gaspar Schott partituras música
- Segunda Guerra mundial:
  - Microfilmes
  - prisioneros usan i, j, t, y f para ocultar mensaje en código morse
  - "Null Cipher"



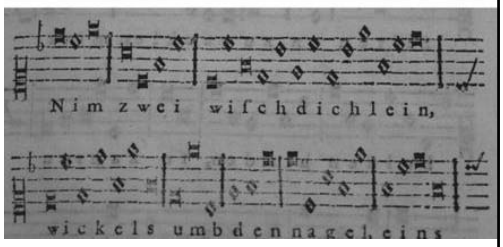


Lámina 4 Dr. Roberto Gómez C.

Un primer ejemplo de Null Cipher

Tomando la primera letra de cada palabra

News Eight Weather: Tonight increasing snow. Unexpected precipitation smothers eastern towns. Be extremely cautious and use snowtires especially heading east. The highways are knowingly slippery. Highway evacuation is suspected. Police report emergency situations in downtown ending near Tuesday.

Hidden Information !

Newt is upset because he thinks he is President.

Lámina 5 Dr. Roberto Gómez C.

Un ejemplo más de Null Cipher

Fishing freshwater bends and saltwater coasts rewards anyone feeling stressed. Resourceful anglers usually find masterful leapers fun and admit swordfish rank overwhelming anyday.

Fishing freshwater bends and saltwater coasts rewards anyone feeling stressed. Resourceful anglers usually find masterful leapers fun and admit swordfish rank overwhelming anyday.

Send Lawyers, Guns, and Money.

Lámina 6 Dr. Roberto Gómez C.

Un último ejemplo Null Cipher

Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on by products, ejecting suets and vegetable oils.

Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on by products, ejecting suets and vegetable oils.

Pershing sails from NYr June i

Lámina 7 Dr. Roberto Gómez C.

Usando imágenes digitales

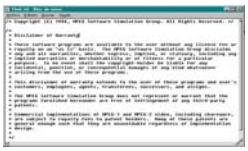


Lámina 8 Dr. Roberto Gómez C.

El proceso esteganografico

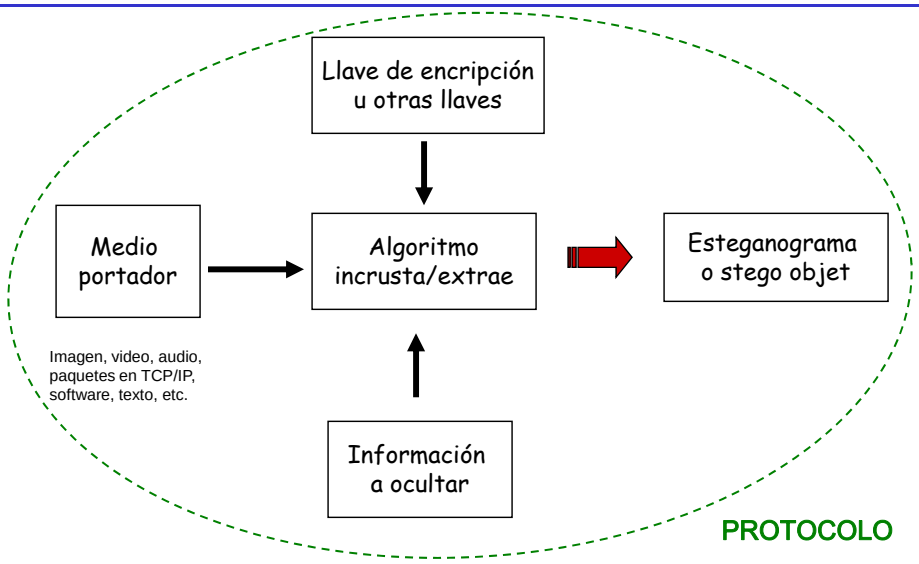


Lámina 9 Dr. Roberto Gómez C.

Técnicas esteganográficas

- Adición
  - se oculta el mensaje secreto en las secciones del medio portador que pueden ser ignoradas por la aplicación que lo procesa
- Generación
  - se crea el esteganograma a partir de la información secreta, sin contar con un medio portador previamente
- Susbtitución
  - se modifican ciertos datos del medio portador por los datos del mensaje secreto

http://www.wayner.org/books/discrypt2/bitlevel.php

Lámina 10 Dr. Roberto Gómez C.

Medios Portadores

- Archivos de imágenes, sonido, texto, video
- Archivos ejecutables
- Archivos de música y de películas
- Páginas Web
- Campos no usados de paquetes de redes (TCP/IP)
- Espacio no utilizado del disco: slack space
- Particiones escondidas
- HTML
- ...

Lámina 11 Dr. Roberto Gómez C.

Ejemplos adición

- Operaciones NOP en códigos ejecutables
- Archivos MP3
- HTML
- Paquetes TCP/IP
- Slack space

Lámina 12 Dr. Roberto Gómez C.

Archivos MP3

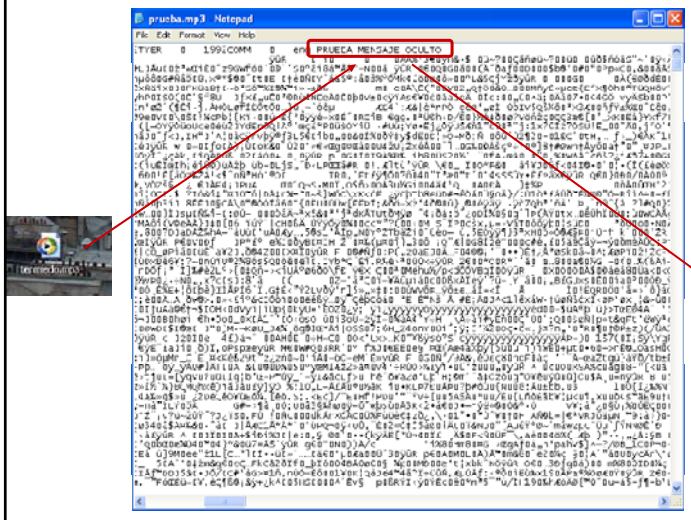
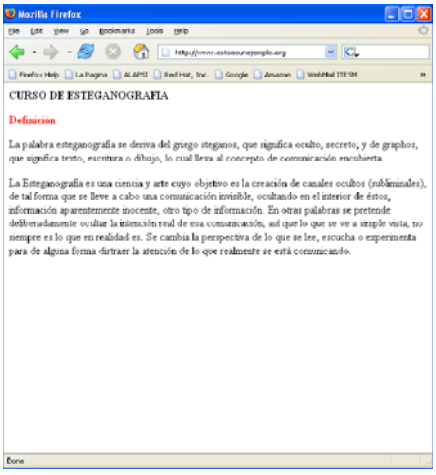
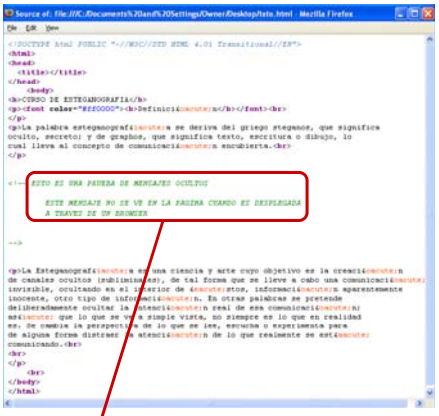


Lámina 13 Dr. Roberto Gómez C.

HTML





ESTO ES UNA PRUEBA DE MENSAJES OCULTOS

ESTE MENSAJE NO SE VE EN LA PAGINA CUANDO ES DESPLEGADA A TRAVES DE UN BROWSER

Lámina 14 Dr. Roberto Gómez C.

Paquetes redes

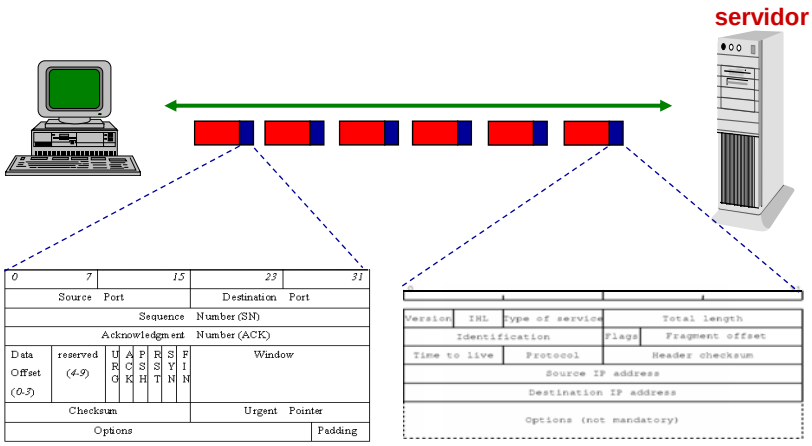
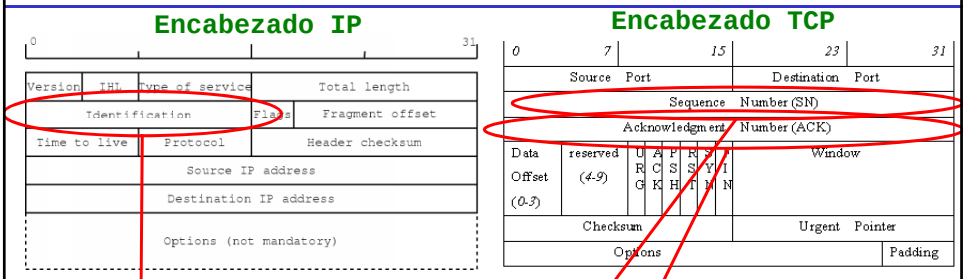


Lámina 15 Dr. Roberto Gómez C.

Esteganografía en paquetes redes



Opción 1:
IP Identification Field

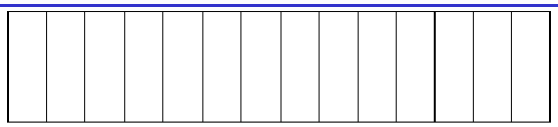
Opción 2:
Initial Sequence Number Field

Opción 3:
The TCP Acknowledge Sequence Number Field "Bounce"

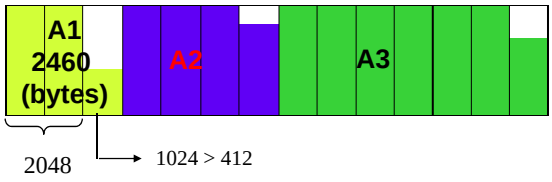
Lámina 16 Dr. Roberto Gómez C.

El slack space

14 clusters libres
c/cluster = 1024 bytes



Tres archivos:
A1, A2 y A3



Cluster = 512bytes

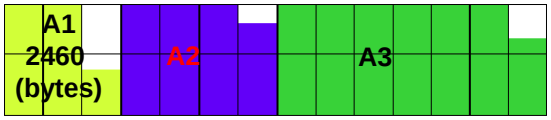


Lámina 17 Dr. Roberto Gómez C.

Ejemplo slack space: bmap

- Herramienta para ocultar información en el slack space producido por sistemas Linux
  - inserta, borra y despliega la información
- Ejemplo inserción
  echo "evil data is here" | bmap --mode putslack /etc/passwd
- Ejemplo despliegue
  # bmap --mode slack /etc/passwd
  getting from block 887048
  file size was: 9428
  slack size: 2860
  block size: 4096
  evil data is here
- Ejemplo borrado:
  bmap --mode wipeslack /etc/passwd

Lámina 18 Dr. Roberto Gómez C.

¿Y para Windows?

- Frag FS
  - desarrollado por Irby Thompson y Mathew Monroe
  - oculta información en sistema archivos NTFS
  - aprovecha espacio dejado por los atributos de las entradas MFT (Master File Table)
- Slacker
  - desarrollado por la gente anteforensics del Metasploit Project
  - manipula apuntadores de archivos físicos y lógicos para ocultar información en lugar de los ceros que debería de insertar el sistema de archivos

Lámina 19 Dr. Roberto Gómez C.

Ocultando información en archivos ejecutables

- Existen diferentes técnicas para esconder información en un archivo *.exe*.
- Una técnica común de inserción consiste en poner los datos después de la marca de fin de archivo.
  - El archivo se ejecutará, ya que la información se oculta después del fin de archivo.
  - Otra técnica consiste en substituir las operaciones NOP por la información a esconder.

Lámina 20 Dr. Roberto Gómez C.

Ejemplo (1/2)

```
/* hello.c */
main( )
{
   printf("Bonjour"n");
}
```

```
[raynal]$ gcc -o hello hello.c
[raynal]$ gdb -q hello
(gdb) disass main
Dump of assembler code for function main:
0x80483c8 <main>:       push %ebp
0x80483c9 <main+1>:     mov %esp,%ebp
0x80483cb <main+3>:     push $0x8048430
0x80483d0 <main+8>:     call 0x8048308 <printf>
0x80483d5 <main+13>:    add $0x4,%esp
0x80483d8 <main+16>:    leave
0x80483d9 <main+17>:    ret
0x80483da <main+18>:    nop
0x80483db <main+19>:    nop
0x80483dc <main+20>:    nop
0x80483dd <main+21>:    nop
0x80483de <main+22>:    nop
0x80483df  <main+23>:   nop
End of assembler dump.
```

Lugar para Ocular Información En este caso: **PRUEBA**

Lámina 21 Dr. Roberto Gómez C.

Ejemplo (2/2)

```
/* hello.c */
main( )
{
   printf("Bonjour"n");
}
```

```
[raynal]$ gcc -o hello hello.c
[raynal]$ gdb -q hello
(gdb) disass main
Dump of assembler code for function main:
0x80483c8 <main>:       push %ebp
0x80483c9 <main+1>:     mov %esp,%ebp
0x80483cb <main+3>:     push $0x8048430
0x80483d0 <main+8>:     call 0x8048308 <printf>
0x80483d5 <main+13>:    add $0x4,%esp
0x80483d8 <main+16>:    leave
0x80483d9 <main+17>:    ret
0x80483da <main+18>:    P
0x80483db <main+19>:    R
0x80483dc <main+20>:    U
0x80483dd <main+21>:    E
0x80483de <main+22>:    B
0x80483df  <main+23>:   A
End of assembler dump.
```

Lámina 22 Dr. Roberto Gómez C.

Otra técnica de esconder en ejecutables

- Archivos tipo PE (Protable Ejecutable)
  - Formato usados para ejecutables en versiones de 32 y 64 bits del sistema operativo Windows.
  - La información puede ocultar en
    - Slack spaces de cada sección del archivo ejecutable.
    - La sección .Text

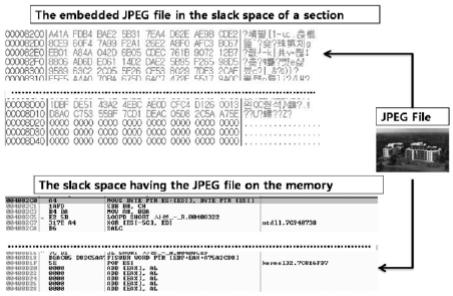


Lámina 23 Dr. Roberto Gómez C.

Generación

- Se genera el esteganograma a partir del texto que se desea enviar.
- Basado en lo que se conoce como funciones de imitación (mimic functions)
  - usadas para ocultar la identidad de un mensaje al cambiar su perfil estadístico de tal forma que corresponda al perfil de cualquier texto inocente
  - cambia un archivo A de tal forma que asume las características estadísticas de otro archivo B

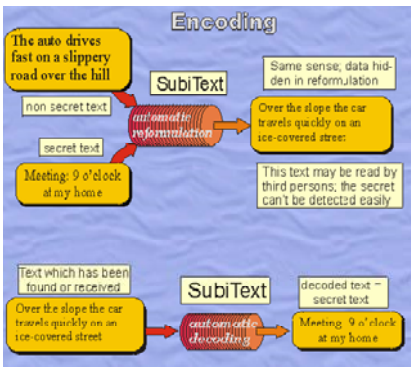


Lámina 24 Dr. Roberto Gómez C.

Técnicas para ocultar información dentro de texto

- Formato base del texto
  - ocultar texto o información en el formato físico del texto, es decir, introduce espacios o caracteres que no son desplegables, introduce faltas ortográficas en el texto, cambia en tamaño de la fuente en ciertas partes del texto, etc
- Generación aleatoria y estadística
  - secuencia de caracteres
  - secuencia de palabras
  - generación estadística de secuencias
- Métodos lingüísticos

Lámina 25 Dr. Roberto Gómez C.

Métodos lingüísticos

- Lenguaje regular
  - convierte datos dentro de cadenas regulares establecidas.
  - cadenas establecidas descritas en términos de funciones de compresión
- Lenguajes de contexto libre
  - convierte datos dentro de cadenas con un lenguaje ambiguo de contexto libre
- Lenguajes de numeración recursivos.
  - uso de este tipo de lenguajes para ocultar información

Peter Wayner, Mimic Functions, CRYPTOLOGIA Volume 16, Number 3, pp. 193-214

Lámina 26 Dr. Roberto Gómez C.

Un primer ejemplo

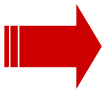
- Mensajes a esconder
  - Falla de sistema, soporte técnico satisfactorio
  - Falla de sistema, soporte técnico fallido
- Posibles mensajes a enviar
  - Los amigos son ángeles que nos ayudan a ponernos de pie cuando nuestras alas olvidan como volar".
  - Los amigos son ángeles que se entristecen cuando nuestras alas olvidan como volar".
- Relación
  - Los Amigos son ángeles – soporte técnico
  - Ayudan a ponernos de pie – satisfactorio
  - Entristecen - fallido
  - Nuestras alas olvidan como volar – falla de sistema

Lámina 27 Dr. Roberto Gómez C.

Un segundo ejemplo (usando GLC)

- Consideremos una gramática libre de contexto con sus producciones
- Información a esconder: "0110"

Inicio ⇨ sujeto verbo
Sujeto ⇨ persona// animal
verbo ⇨ comer//dormir
comer ⇨ sabroso// enojado
dormir ⇨ apacible//intranquilo
persona ⇨ paco//luis
animal ⇨ conejo//gato



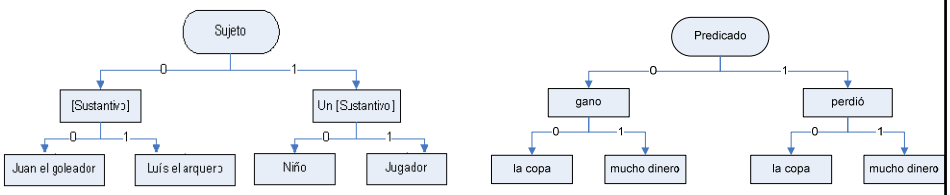
0 ⇨ persona
1 ⇨ luis
1 ⇨ dormir
0 ⇨ apacible

- El enunciado será: Luis duerme apacible

Lámina 28 Dr. Roberto Gómez C.

Un último ejemplo

- Arboles representando substitución de dos dígitos binarios para la formación del sujeto y del predicado dentro de una oración.



- Mensaje a ocultar
  - México ganara el próximo mundial en alemania

Lámina 29 Dr. Roberto Gómez C.

Continuación del ejemplo

- Tabla analogías

Palabra	Codificación en 1´s y 0´s
México	1111
ganara	1010
el	1100
próximo	0111
mundial	0000
en	0101
Alemania	0010

- Cadena a codificar
  - México ganara el próximo mundial en alemania
  - 11111010110001110000001010010
- Codificando los cuatro primeros bits 1111
  - Un jugador perdió mucho dinero

Lámina 30 Dr. Roberto Gómez C.

Terminando el ejemplo

- Siguiendo con los siguientes, se obtiene:

Un jugador perdió mucho dinero. Un niño perdió la copa. Un jugador gano la copa. Luís el portero perdió mucho dinero. Juan el goleador gano la copa. Luís el portero gano mucho dinero. Juan el goleador perdió la copa.

Lámina 31 Dr. Roberto Gómez C.

Implementaciones

- Tres ejemplos
  - Spam Mimic
    - http://www.spammimic.com/
  - Narrador de baseball
    - http://www.wayner.org/texts/mir
  - TextHide
    - http://www.texthide.com/

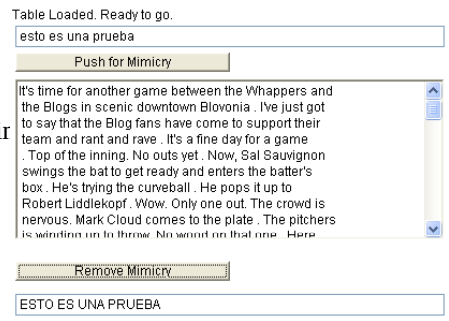


Lámina 32 Dr. Roberto Gómez C.

Spam: spammimic.com

- Dear Professional , This letter was specially selected to be sent to you . If you no longer wish to receive [illegible] publications simply reply with a Subject: of "REMOVE" and you will immediately be removed from our mailing list . This mail is being sent in compliance with Senate bill **1621** ; Title 1 , Section [illegible] ! This is not a get rich scheme . Why work for somebody else when you can become rich in 39 days ! Have you ever noticed how [illegible] the line-ups are at bank machines & how many people you know are on the Internet ! Well, now is your chance to capitalize on this ! We will help **you** decrease perceived waiting time by [illegible]% and deliver goods right to the customer's doorstep ! The best thing about our system is that it is absolutely risk free for you ! But don't believe us . **Mrs Simpson** of Maryland tried us and says "I was skeptical but it worked for me" . We assure you that we operate within all applicable laws ! We implore you - act now ! Sign up a friend and you get half off . Thanks .

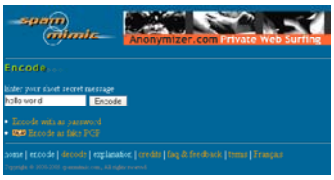
Lámina 33 Dr. Roberto Gómez C.

Spam: spammimic.com

- Dear Business person , We know you are interested in receiving amazing intelligence . This is a one time 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 [illegible] ; Title [illegible] , Section **301** ! This is not a get rich scheme ! Why work for somebody else when you can become rich as few as 98 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 !

Lámina 34 Dr. Roberto Gómez C.

La página de spam mimic



hello world

Dear Decision maker , We know you are interested in receiving amazing intelligence . This is a one time 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 33 days . Have you ever noticed people love convenience 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 190% and increase customer response by 150% . You can begin at absolutely no cost to you. But don't believe us . Ms Ames of Florida tried us and says "My only problem now is where to park all my cars" ! We are licensed to operate in all states! We implore you - act now ! Sign up a friend and you get half off . God Bless !

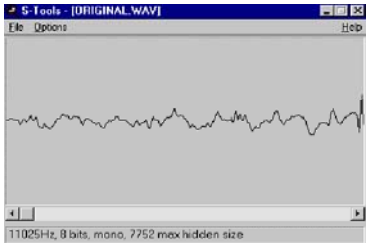
Lámina 35 Dr. Roberto Gómez C.

Ejemplos Susbtitución

- Principales métodos
  - LSB: Least-Significant Bit
  - La transformación matemática de la información
    - Transformación discreta del coseno (DCT)
    - Transformación discreta de Fourier
    - Transformación de Wavelet
- Posibles medios medios portadores (archivos digitales)
  - archivos de música
  - archivos de imagenes

Lámina 36 Dr. Roberto Gómez C.

Esteganografía en música



Información: 132 134 137 141 121 101 74 38

Binario: 10000100 **1000011**0 **1000100**1 **1000110**1 **0111100**1 **0110010**1
01001010 **0010011**0

Información a esconder: 11010101 (213)

Resultado: 133 135 136 141 120 101 74 39

Binario: 10000101 10000111 10001000 10001101 01111000 01100101
01001010 0010011

Lámina 37 Dr. Roberto Gómez C.

Imágenes

- Una imagen es una matriz de MxN Pixeles.
- Un Pixel es la unidad mínima de dibujo

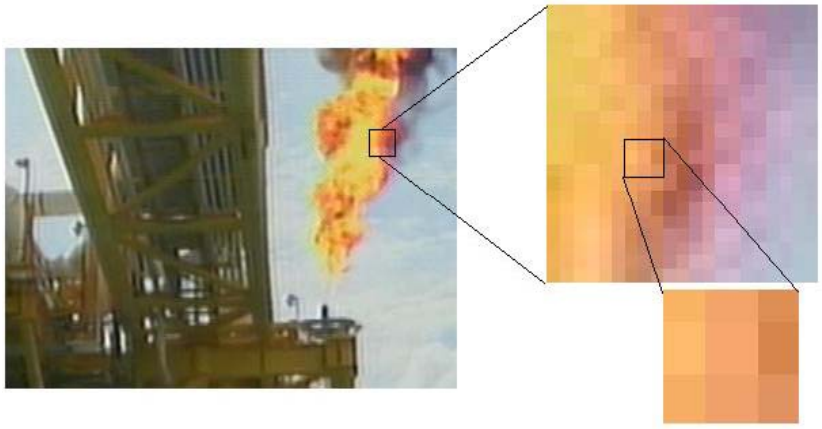
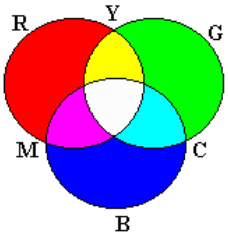
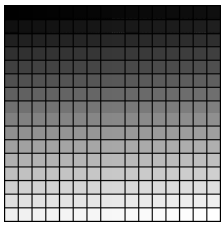


Lámina 38 C.

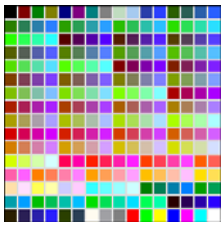
Los colores en las imágenes digitales



RGB



Paleta escala de grises



Paleta escala de colores

Lámina 39 Dr. Roberto Gómez C.

Modelo de Color RGB

- Emplea síntesis aditiva, es decir, suma colores para obtener nuevos colores.
  - el color de inicio es el negro y la suma de todos los colores da blanco.
- Los colores se representan con 24 bits
  - 8 para cada componente RGB.
- Cada componente 8 bits: 256 posibles niveles color
- Tres canales de color: Rojo (R), Verde (G), Azul (B)

10001100 10001100 10001100

Azul Verde Rojo

Lámina 40 Dr. Roberto Gómez C.

Ysi tomamos bits significativos

Bit más significativo

Segundo bit más significativo

Tercer bit más significativo

Cuarto bit más significativo

Quinto bit más significativo

Sexto bit más significativo

Lámina 41 Dr. Roberto Gómez C.

Usando LSB para insertar una "a"

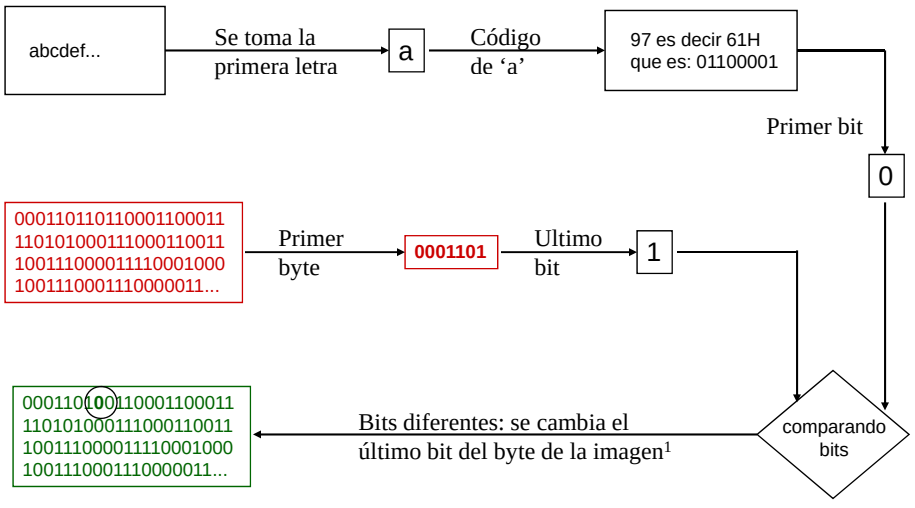


Lámina 42 [1]Bits iguales: no se cambia nada Dr. Roberto Gómez C.

Usando LSB para ocultar una imagen

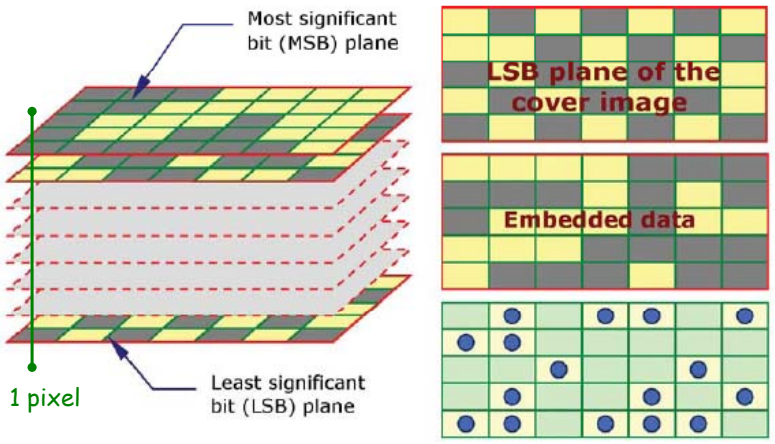


Lámina 43 Dr. Roberto Gómez C.

Algoritmo para ocultar información

1. Cargar imagen portadora e imagen a esconder.
2. Elegir número de bits a esconder
   - más bits usados en la imagen portadora, más se va a deteriorar esta imagen
   - incrementar el número de bits incrementa claridad de la imagen a ocultar
3. Crear nueva imagen combinando pixels dos imágenes
4. Para recuperar la imagen es necesario conocer el número de bits usados
   - recorrer imagen portadora, tomar los bits menos significativos y usarlos para crear una nueva imagen
   - solo un cambio: los bits extraídos se convierten en los bits más significativos

Lámina 44 Dr. Roberto Gómez C.

Ejemplo LSB con cuatro bits

- Si se decide usar 4 bits para esconder la imagen secreta, se tendrán 4 bits para la imagen portadora

Pixel medio portador: 10110001
Pixel info a ocultar: 00111111
Pixel stego objeto: 10110011

- Para recuperar la imagen es necesario saber cuantos bits se usaron para almacenar la imagen

Pixel stego objeto: 10110011
Bits usados: 4
Nueva imagen: 00110000

Lámina 45 Dr. Roberto Gómez C.

Variantes

- Utilizando una permutación de las ubicaciones de los pixeles para esconder los bits, generados tal vez pseudo aleatoriamente
- Poner los bits en solo ciertos lugares en la imagen donde los cambios en los valores de color no serán perceptibles.

Lámina 46 Dr. Roberto Gómez C.

Paleta de colores I

- Imagen de color de 8-bits: 256 colores diferentes que se indexan con los números 0,…,255
- Para incrusrtar información, por ejemplo, S-Tools reduce el número de colores de 256 a 32.
- Los 8 bits de colores después de la incrustación son muy parecidos visualmente a los mismos 8 de antes de la modificación, solo difieren en su representación a nivel de bits.

Lámina 47 Dr. Roberto Gómez C.

LSB y Paleta de colores

- Existen muchas herramientas para este tipo de esteganografía.
- Todas ellas sufren de una falta de robustez: un simple cambio en los valores LSB y la información se modifica.
- Los métodos basados en la paleta de colores generan patrones que se pueden analizar.

Lámina 48 Dr. Roberto Gómez C.

Ocultando información imágenes 8 bits

- Imágenes de 8 bits no permiten una buena manipulación del LSB debido a su limitación de colores.
- Cuando la información es insertada dentro de los LSBs de los datos a desplegar, los apuntadores a las entradas de la paleta son cambiados.
- Ejemplo
  - paleta de cuatro colores, con entradas
    - 0 00
    - 1 01
    - 2 10
    - 3 11

Lámina 49 Dr. Roberto Gómez C.

Ocultando información en el ejemplo

- Valores de cuatro pixeles a desplegar
  - 00 00 10 10
- Ocultar valor binario 1010 cambia los valores a
  - 01 00 11 10
- Estos cambios en la imagen son visibles y se pueda apreciar el efecto de usar imágenes de 8 bits

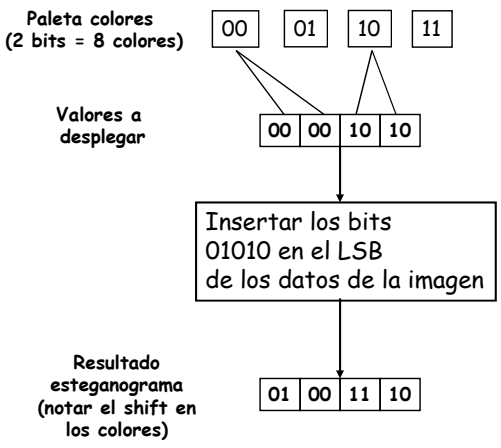
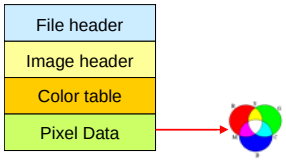


Lámina 50 Dr. Roberto Gómez C.

Archivos BMP



- Microsoft Bitmap Format
- Dos versiones
  - una para Windows
  - otra para OS/2
- Calidad imágenes abarca 1, 4, 8, 16, 24 y 32 bits por píxel,
- Principales virtudes son su simplicidad y su amplio soporte,
- Estructura general de cualquier archivo BMP está compuesta por
  - una cabecera del archivo (BITMAPFILEHEADER)
  - una cabecera de la imagen (BITMAPINFOHEADER)
  - una tabla de colores (RGBQUAD) y
  - los datos de los píxeles (BYTES)

Lámina 51 Dr. Roberto Gómez C.

Implementación en archivos BMP

La cabecera del archivo	14 bytes
La cabecera de la imagen	40 bytes
La tabla de colores	(4 bytes)(#colores)
Los datos de los píxeles	3 bytes por píxel

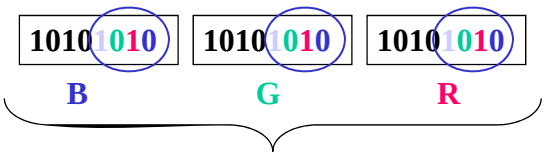


Lámina 52 Dr. Roberto Gómez C.

Ejemplo esteganografia en imágenes BMP

- Herramienta: Stools
- Archivos .BMP y .WAV
- Posible ocultar cualquier tipo de información.
- Aparte de ocultar cifra la información.

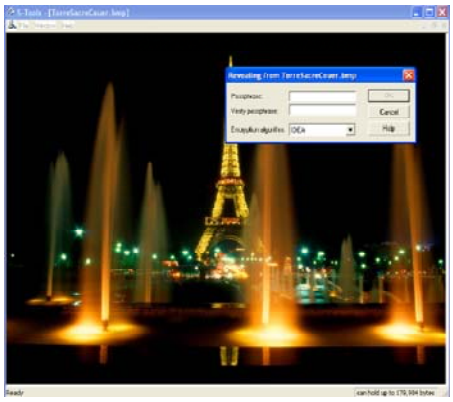


Lámina 53 Dr. Roberto Gómez C.

Implementación en archivos GIF

- Graphics Interchange Format
- Formato de mapas de bits
- Dos versiones del archivo
  - 87a
  - 89b
- Puede contener 2,4,8,16,32,64,128 o 256 colores que son almacenados en una paletad dentro de la imagen.
  - cada color descrito en base de colores RGB
- Principal diferencia con otros formatos
  - basado en streams, i.e. formado por bloques y extensiones

Lámina 54 Dr. Roberto Gómez C.

Bloques en GIF

- Los bloques de control
  - cabecera del archivo, el descriptor de la ventana lógica, la tabla de colores global, la extensión de control grafica, la extensión de control grafico y el de terminación del archivo
  - contienen información para el procesamiento de los datos del archivo grafico.
- Los bloques de despliegue grafico
  - la cabecera de la imagen, la extensión de texto
  - contienen los datos utilizados para desplegar la imagen.
- Los bloques de propósito especial
  - la extensión de comentarios y la extensión de aplicación
  - tienen como particularidad que no son tomados en cuenta por los decodificadores de este formato, ya que son prescindibles para su correcto despliegue.

Lámina 55 Dr. Roberto Gómez C.

¿Y donde oculto la información?

- Técnica de adición
  - información toma la forma de un bloque de aplicación que puede insertarse en cualquier lugar del archivo
  - desventaja: archivo incrementa su tamaño
- Técnica de susbtitución
  - reducir el tamaño de la tabla de colores en el bloque de imagen
  - si los colores son duplicados, un bit de cada uno de ellos puede ser usado para esconder información
  - desventaja: tamaño reducido

Lámina 56 Dr. Roberto Gómez C.

Formatos archivos GIF

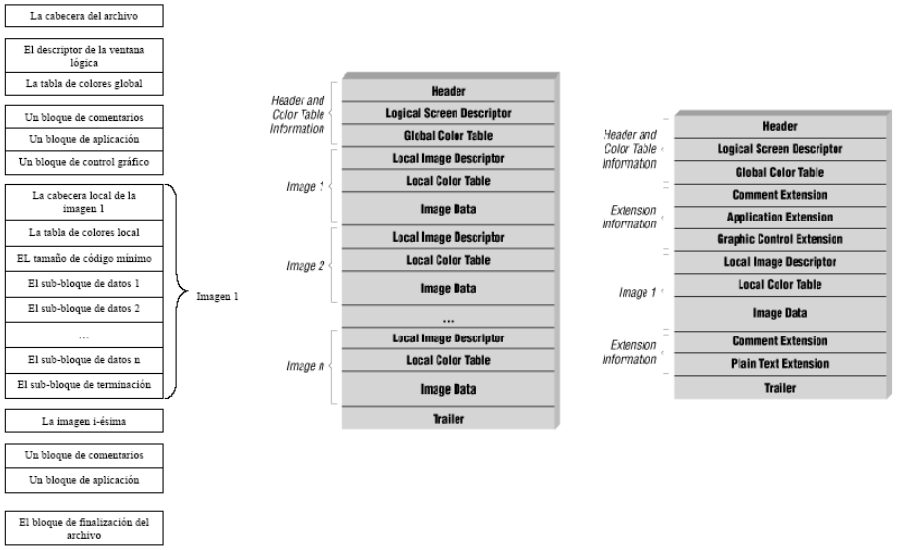


Lámina 57 Dr. Roberto Gómez C.

El dominio de las transformaciones

- Transformación Discreta del Coseno
- Transformación Discreta de Wavelet
- Transformación Discreta de Fourier
- Transformación de Mellin-Fourier
- Otras:
  - Descomposición de valores singulares
  - Descomposición de Eigenvalores

Lámina 58 Dr. Roberto Gómez C.

Ejemplo DCT: el estándar JPEG

- Creado por los grupos CCITT e ISO (1992)
- Objetivo: establecer un estándar de compresión internacional para imágenes en escalas de grises (8 bits) y a colores (24 bits)
- Incluye dos métodos de compresión básicos y cada uno con varios modos de operación
  - método con perdida basado en DCT
  - método sin perdida basado en un método de predicción
- Capaz de comprimir los datos de una imagen a menos del 10 por ciento de su tamaño original.

Lámina 59 Dr. Roberto Gómez C.

Fases procesamiento

- Una transformación matemática
  - la transformada discreta del coseno
  - trasladar una señal del domino espacial y se traducen al dominio de frecuencias
  - se cuenta con una transformación inversa
- Una cuantificación
  - perdida información relevante
- Codificación

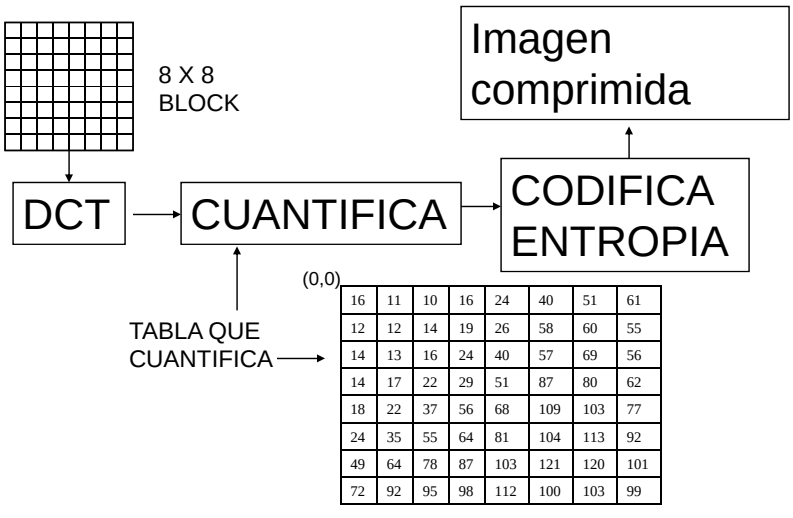
Lámina 60 Dr. Roberto Gómez C.

Transformación Discreta del Coseno (III)

- Idea básica de JPEG:
  1. Convierte una imagen a un espacio de colores.
  2. Divide la imagen en bloques de 8x8 pixeles.
  3. Se aplica DCT a cada bloque.
  4. Se cauntifican los valores con valores de cuantificación pre-establecidos (en una tabla)
  5. Los valores se redondean al entero más próximo.

Lámina 61 Dr. Roberto Gómez C.

Interfase de compresión de JPEG



16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Lámina 62 Dr. Roberto Gómez C.

Transformación del espacio de color

- Ojo más sensible a la intensidad de luz que a los colores
  - reducir tamaño: dar más importancia a la información de la luz que a de los colores
- Pixel codificado en rojo, verde y azul
  - tres variables Y (luz) Cr (colores) Cb (colores)
  - por ejemplo, para su convertirlo:

$$Y = 0.299 * rojo + 0.587* verde + 0.114 * azul$$
$$Cr = rojo - Y$$
$$Cb = azul - Y$$

  - para el regreso

$$rojo = Cr + Y$$
$$azul = Cb + Y$$
$$verde = Y * 1.7 - rojo * 0.509 - azul * 0.914$$

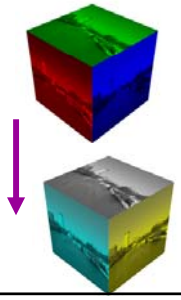
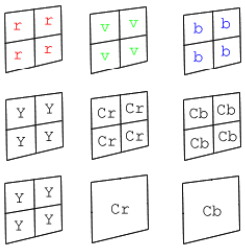


Lámina 63 rto Gómez C.

Ejemplo transformación del espacio de color



color compuesto por rojo, verde y azul

descomposición Y Cr Cb

supresión de Cr y Cb (media de las cuatro)

Imagen original

Y (intensidad)

Cb (azul/amarillo)

Cb (rojo/verde)

Lámina 64 http://www.stanford.edu/~esetton/table.htm Dr. Roberto Gómez C.

Transformación discreta de coseno o DCT

- Cada componente de la imagen se divide en pequeños bloques de 8×8 píxeles, que se procesan de forma casi independiente
- Transformación de los valores en frecuencias
  - describir números, no por su valor, pero sus coeficientes en la fórmula matemática
- Ecuación

$$b(u,v) = \frac{2}{N} C(u)C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} a(x,y) \cos\left(\frac{\pi u(2x+1)}{2N}\right) \cos\left(\frac{\pi v(2y+1)}{2N}\right)$$

- Ecuación inversa

$$a(x,y) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v)b(u,v) \cos\left(\frac{\pi u(2x+1)}{2N}\right) \cos\left(\frac{\pi v(2y+1)}{2N}\right)$$

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}} & si\, u = 0 \\ 1 & en.otro.caso \end{cases}$$

Ejemplo transformación discreta de coseno

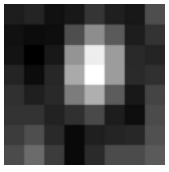
139	144	149	153	155	155	155	155
144	151	153	156	159	156	156	156
150	155	160	163	158	156	156	156
159	161	162	160	160	159	159	159
159	160	161	162	162	155	155	155
161	161	161	161	160	157	157	157
162	162	161	163	162	157	157	157
162	162	161	161	163	158	158	158

valores pixeles

1260	-1	-12	-5	2	-2	-3	1
-23	-17	-6	-3	-3	0	0	1
-11	-9	-2	2	0	-1	-1	0
-7	-2	0	1	1	0	0	0
-1	-1	1	2	0	-1	1	1
2	0	2	0	-1	1	1	-1
-1	0	0	-1	0	2	1	-1
-3	2	-4	-2	2	1	-1	0

coeficientes DCT

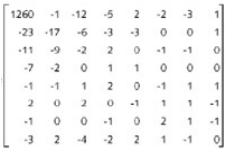
antes de, un bloque de 8x8



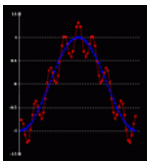
después de, un bloque de 8x8 (se notan errores respecto a la primera imagen)

Cuantificación

- Coeficientes son cuantificados
  - dividos uno por uno con una tabla de valores y redondeados
- Entre más grandes sean los valores de la tabla más detalles serán eliminados
- Tablas almacenadas en encabezado
  - parámetro a usar cuando se quiere almacenar un X% calidad JPG
- Proceso en el que se pierde la mayor parte de la información



coeficientes DCT



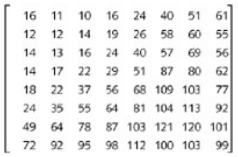
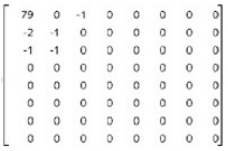


tabla cuantificación

lugar para esconder información



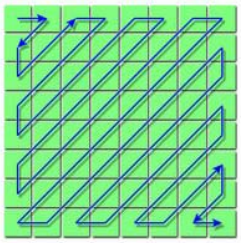
coeficientes DCT cuantificados

Lámina 67 Dr. Roberto Gómez C.

Ordenamiento zig zap

- Se cuenta con un bloque de 8x8 con algunos coeficientes vivos y varios ceros.
- Se reordenan los coeficientes en un orden zig-zag
- Objetivo: poner el máximo numero de ceros cercanos, para poder comprimirlos.
- En el ejemplo se obtendría la siguiente salida:

  79 0 -2 -1 -1 -1 0 0 -1 0 0 0 0 0……0



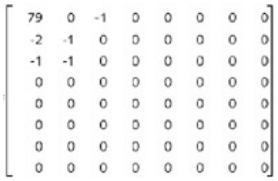


Lámina 68 Dr. Roberto Gómez C.

Compresión

- Dos algoritmos son utilizados
  - RLE
    - Run-Length Encoding
    - usado para comprimir los coeficientes de alta frecuencia (debido a la existencia de muchos ceros)
  - DPCM
    - Differential Pulse Code Modulation
    - muy similar al anterior
    - comprimir los primeros coeficientes de baja frecuencia
- Después se usa un algoritmo de Huffman para comprimir todo
  - árboles Huffman son almacenados en el encabezado del archivo

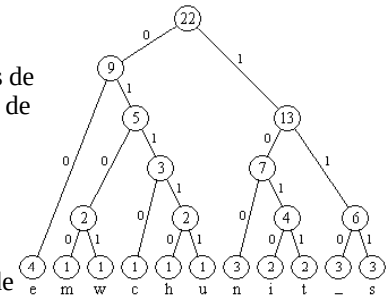


Lámina 69 Dr. Roberto Gómez C.

¿Donde oculto la información?

- No es tan simple como en Bitmap o GIF
- Algunas usan técnicas muy simples
  - añaden los datos a ocultar al final del archivo
  - utlizan el campo de comentario en el encabezado
- Verdadera esteganografia implica el mezclar el mensaje a ocultar con los pixels que conforman la imagen.
- Programas que llevan a cabo lo anterior:
  - F5
  - JPHide/JPSeek/JPHSWin
  - StegHide
  - JSteg

Lámina 70 Dr. Roberto Gómez C.

La herramienta JSteg

- Aparentemente la primer herramienta en implementar steganografia en archivos JPEG
  - no encripción
  - fuente disponible
  - interfaz gráfica: JStegShell
- Información se oculta en la matriz de coeficientes DCT cuantificados
  - aquellos con un valor de cero o 1 no son modificados
  - el resto son usados para ocultar secuencialmente un bit de la información a ocultar
  - se sobrescribe el LSB de la información a ocultar

Lámina 71 Dr. Roberto Gómez C.

Ocultando información en JSteg

- Formato información oculta:

A: cinco bits: expresa en bits la longitud (en bits) del campo B
B: número de bis, entre cero y 31, la longitud en bytes del archivo a ocultar
C: los bits del archivo a ocultar

- Para extraer la imagen:
  - Se leen los LSB de los primeros cinco coeficientes DCT, que contienen el tamaño del siguiente campo
  - Se extrae la información del campo B
  - Se leen el numero de bytes especificados en B, extrayendo el LSB de cada uno de ellos.

Lámina 72 Dr. Roberto Gómez C.

Ejemplo JSteg

original

original + poema if

Lámina 73 Dr. Roberto Gómez C.

Poema IF de Rudyard Kipling (1.5 Kb)

If you can keep your head when all about you
Are losing theirs and blaming it on you,
If you can trust yourself when all men doubt you,
But make allowance for their doubting too;
If you can wait and not be tired by waiting,
Or being lied about, don't deal in lies,
Or being hated don't give way to hating,
And yet don't look too good, nor talk too wise:

If you can dream and not make dreams your master;
If you can think and not make thoughts your aim;
If you can meet with Triumph and Disaster
And treat those two impostors just the same;
If you can bear to hear the truth you've spoken
Twisted by knaves to make a trap for fools,
Or watch the things you gave your life to, broken,
And stoop and build 'em up with worn-out tools:

If you can make one heap of all your winnings
And risk it on one turn of pitch-and-toss,
And lose, and start again at your beginnings
And never breathe a word about your loss;
If you can force your heart and nerve and sinew
To serve your turn long after they are gone,
And so hold on when there is nothing in you
Except the Will which says to them: "Hold on!"

If you can talk with crowds and keep your virtue,
Or walk with Kings nor lose the common touch,
If neither foes nor loving friends can hurt you,
If all men count with you, but none too much;
If you can fill the unforgiving minute
With sixty seconds' worth of distance run,
Yours is the Earth and everything that's in it,
And which is more you'll be a Man, my son!

Lámina 74 Dr. Roberto Gómez C.

Analizando el ejemplo

A: siempre 5 bits a decodificar: 00 01 00 01 01
D6 (1101 0110) 69 (0110 1001) 12 (0001 0010) 05 (0000 0101) 03 (0000 0011)
= 01011 = 11_{10}

B: 11 bits a decodificar: 01 01 00 00 00 01 00 00 00 01
= 11000010001 = 1551310

C: 1553 x 8 bits, empezando en 00 01 00 00 01 00 00 01 or
= 01001001 in binary or 73 = I

D6 69 13 05 03 15 F2 EB	D6 69 12 05 03 15 F3 EA	00 01 00 01 01 01 01 00
FF 04 01 00 FA FB F9 FF	FE 04 01 00 FA FB F8 FE	00 00 01 00 00 01 00 00
06 02 FE FF 00 00 00 FF	06 03 FE FF 00 00 00 FE	00 01 00 01 00 00 00 00
01 03 02 01 01 FF 00 00	01 02 03 01 01 FE 00 00	01 00 01 01 01 00 00 00
01 00 00 00 00 00 00 00	01 00 00 00 00 00 00 00	01 00 00 00 00 00 00 00
00 FF FF 00 00 00 00 00	00 FE FF 00 00 00 00 00	00 00 01 00 00 00 00 00
00 00 00 00 00 01 00 00	00 00 00 00 00 01 00 00	00 00 00 00 00 01 00 00
00 00 00 00 00 00 00 00	00 00 00 00 00 00 00 00	00 00 00 00 00 00 00 00

Coeficientes DCT cuantificados imagen original

Coeficientes DCT cuantificados imagen con texto oculto

Coeficientes DCT cuantificados, solo el LSB de cada byte

Lámina 75 Dr. Roberto Gómez C.

Ejemplo esteganografia usando DCT

Original Image

Watermarked Image

JPEG compressed

Lámina 76 Dr. Roberto Gómez C.

Ejemplo herramienta JSteg

- Compuesta por dos programas
  - cjpeg
  - Djpeg

Lámina 77 Dr. Roberto Gómez C.

¿Qué información podemos "esconder" en una imagen?

- Dentro de una imagen podemos utilizar 1 ó 2 bits por cada canal de cada pixel.
- Esos bits pueden formar bytes
- Con Bytes podemos almacenar cualquier tipo de información: texto, archivos de sonido, programas e incluso otras imágenes.

Lámina 78 Dr. Roberto Gómez C.

Substitución en ejecutables: HYDAN

- Desarrollado por El-Khalil
- Presentado en 2003 en DEFCON
- Se basa en redundancias en el set de instrucciones del Intel x86
  - lugares donde dos diferentes instrucciones realizan lo mismo
  - la elección de entre dos opciones redundantes puede representar un bit de los datos a esconder
  - por ejemplo: sumar 50 a un valor es equivalente a substraer -50 al mismo valor.
  - Funciona en sistemas operativos Linux, Windows XP, NetBSD, FreeBSD, y OpenBSD

Lámina 79 Dr. Roberto Gómez C.

¿Cómo funciona?

- Selecciona, con mucho cuidado, ciertas variaciones de código ejecutable.
- Toma de entrada un código ejecutable proporcionado por el usuario, la información a esconder y una frase para cifrar lo anterior.
- Utiliza dos conjuntos de instrucciones que realizan las mismas funciones
  - El conjunto 0
  - El conjunto 1

Lámina 80 Dr. Roberto Gómez C.

Ejemplo codificación

- Por convención se decidio que todas las instrucciones de suma (add) representan un bit 0 y que las instrucciones de substracción representan un bit 1.

Original code

83 e8 30	sub	%eax, $0x30
83 f8 36	cmp	%eax, $0x36
77 e5	ja	$-27
83 c0 08	add	%eax, $0x8
89 04 24	mov	%eax, [%esp]

Encoding 00

83 c0 d0	add	%eax, $-0x30
83 f8 36	cmp	%eax, $0x36
77 e5	ja	$-27
83 c0 08	add	%eax, $0x8
89 04 24	mov	%eax, [%esp]

Encoding 01

83 c0 d0	add	%eax, $-0x30
83 f8 36	cmp	%eax, $0x36
77 e5	ja	$-27
83 e8 f8	sub	%eax, $-0x8
89 04 24	mov	%eax, [%esp]

Encoding 11

83 e8 30	sub	%eax, $0x30
83 f8 36	cmp	%eax, $0x36
77 e5	ja	$-27
83 e8 f8	sub	%eax, $-0x8
89 04 24	mov	%eax, [%esp]

Lámina 81 Dr. Roberto Gómez C.

Ejemplos HYDAN

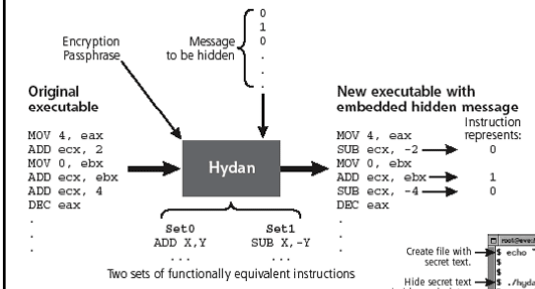


Lámina 82 Dr. Roberto Gómez C.

Un último ejemplo

- Escondiendo 0100

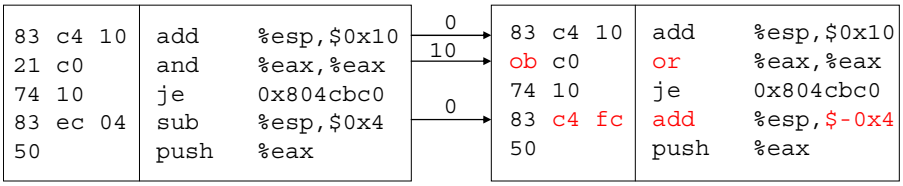


Lámina 83 Dr. Roberto Gómez C.

Otras técnicas

- Técnicas de texto
- Técnicas basadas en el sonido
- Técnicas basadas en el video
- Técnicas basadas en el DNA

Lámina 84 Dr. Roberto Gómez C.

Técnicas texto

- Puede ser formateo del texto o características de los caracteres.
- Se modifican estas características de tal forma que el ojo humano no lo detecte y que pueda ser decodificado por una computadoras.
- Técnicas usadas
  - Line Shift Coding Protocol

Lámina 85 Dr. Roberto Gómez C.

Corrimientos texto y palabras (i)

- Line Shift Coding Protocol
  - se recorren varias líneas dentro del documento hacia arriba o hacia abajo por una pequeña fracción (una 1/300th de pulgada)
  - líneas no detectadas por el ojo humano pero sí por una computadora
  - calculando si una línea fue recorrida hacia arriba o hacia abajo se puede representar un 0 o un 1
- Word Shitf Coding Protocol
  - mismo principio anterior, pero en lugar de medir corrimientos línea abajo/arriba, se recorren palabras a la izquierda o derecha
  - codebook indica al codificador cuales palabras recorrer y el sentido

Lámina 86 Dr. Roberto Gómez C.

Corrimientos texto y palabras (ii)

- Word Shitf Coding Protocol
  - decodificación: medir espacios entre cada palabra y un recorrido izquierdo puede representar un 0 y uno a la derecha un 1
  - ejemplo:

The quick brown fox jumps over the lazy dog.
The quick brown fox jumps over the lazy dog.

  - primera línea: usa espaciamiento normal
  - segunda línea: cada palabra esta recorrida a la izquierda o a la derecha 0.5 puntos, para codificar la secuencia 010000001 (valor 65 que es el ASCII de la letra A)

Lámina 87 Dr. Roberto Gómez C.

Técnicas texto (ii)

- Featuring Coding Protocol
  - documento es pasado a través de un parser
  - examina el documento y automaticamente construye un codebook específico al documento
  - tomará cada una de las características que puedan ser útiles para esconder dentro del documento
  - p.e. altura de ciertos caracteres, puntos letras i y j, líneas letras t y f, también corrimientos
- White Space Manipulation
  - manipular espacios blancos almacenar bits
  - añadir cierto espacio al final de las líneas
  - espacio añadido corresponde a un cierto valor
  - programa SNOW lleva a cabo lo anterior

Lámina 88 Dr. Roberto Gómez C.

Tecnicas texto (iii)

- Contenido del texto
  - esconder el mensaje en lo que parece ser texto discretro
  - la gramática dentro del texto puede ser usada para almacenar información
  - cambiar oraciones para almacenar información y mantener el significado original
  - ejemplo: The auto drives fast on a slippery road over the hill
  - cambio por: Over the slope the car travels quickly on an ice-covered street
  - uso de palabras al azar como un medio de codificar información
    - diferentes palabras pueden proporcionar diferentes valores
    - ejemplo: SpamMimic

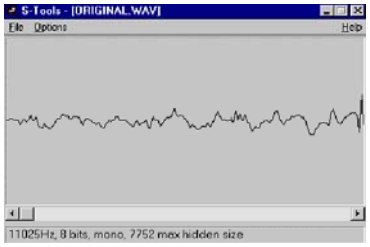
Lámina 89 Dr. Roberto Gómez C.

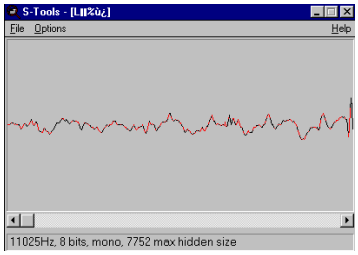
Técnicas basadas en sonido

- Spread Spectrum
  - Codifica los datos como una secuencia binaria q suene como ruido pero que puede ser reconocida por el receptor con una llave correcta.
- MIDI (Musical Instrument Digital Interface)
  - buenos archivos para ocultar información
  - MIDI no transmite señales de audio, sino datos de eventos y mensajes controladores que pueden interpretarse de forma arbitraria
  - mensaje PC (Program Change)
    - valores entre 0 y 127 representa los diferentes instrumentos
    - agrupar el número de mensajes que contiene el dato oculto

Lámina 90 Dr. Roberto Gómez C.

Ejemplo esteganografía en sonido: Stools





Información: 132 134 137 141 121 101 74 38

Binario: 10000100 10000110 10001001 10001101 01111001 01100101 01001010 00100110

Información a esconder: 11010101 (213)

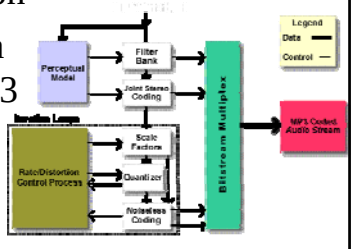
Resultado: 133 135 136 141 120 101 74 39

Binario: 10000101 10000111 10001000 10001101 01111000 01100101 01001010 0010011

Lámina 91 Dr. Roberto Gómez C.

Técnicas basadas en sonido: MP3 (i)

- Formato compresión más usado para archivos de música
- Muy bueno para ocultar información
- Muy pocos ejemplos prácticos para esconder información archivos MP3
  - programa MP3Stego
- Técnica similar a la de transformaciones de frecuencia



- Los datos se esconder conforme el archivo MP3 es creado, es decir durante la fase de compresión

Lámina 92 Dr. Roberto Gómez C.

Técnicas basadas en sonido: MP3 (ii)

- Durante nivel 3 del proceso de codificación, los datos a perder son seleccionados dependiendo del "bit rate" especificado por el usuario
- Los datos a esconder son codificados en el bit de paridad de esta información
- Archivos MP3 divididos en varios frames
  - cada uno con su propio bit de paridad,
  - da espacio para almacenar una cantidad importante de información
- Recuperación: descomprimir archivo MP3 y leer bits de paridad cuando el proceso se esta llevando a cabo.

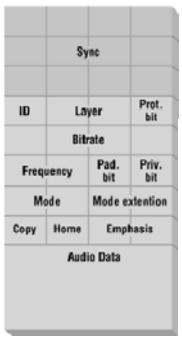


Lámina 93 Dr. Roberto Gómez C.

Técnicas basadas en el video

- Combinación de sonido e imagen
- Se basa en el hecho de que el video generalmente tiene archivos separados dentro de los archivos por el video (consistente de varias imágenes) y el sonido
- Técnicas pueden aplicarse en ambas áreas para esconder los datos.
  - Tarea: y en el protocolo de combinación
- Dada el tamaño de los archivos de video, la "ventana" para añadir grandes cantidades de datos es mayor y la probabilidad de encontrar los datos es menor.

Lámina 94 Dr. Roberto Gómez C.

Técnica basadas en el DNA (i)

- Area relativamente nueva
- Mensaje:
  - JUNE6_INVASION:NORMANDY}
  - escondido dentro algún DNA
- Técnica similar a algunas de texto
- DNA consiste de una cadena de moléculas llamadas bases
  - Adenina, Timina, Guanina y Citosina
- Tabla fue dibujada con tres combinaciones igualan palabra en el alfabeto junto con otras cosas

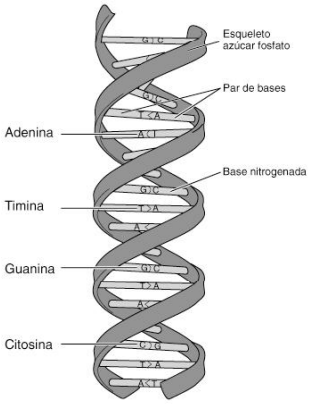


Lámina 95 Dr. Roberto Gómez C.

Técnica basada en el DNA (ii)

- DNA sintetizado de acuerdo a la tabla con las bases en el orden correcto
- Después es rodeada entro otras dos cadenas de DNA que actúan como marcadores
  - indican al emisor y al receptor del mensaje donde empieza y termina el mensaje
- Paso final:
  - se mezcla con algunas cadenas DNA para prevenir la detección del mensaje secreto
- DNA es increiblemente pequeño
  - puede ocultarse en un punto en un libro o una revista

Lámina 96 Dr. Roberto Gómez C.

¿Y cómo lo hicieron?

- Tabla (llave) DNA

A = CGA	K = AAG	U = CTG	0 = ACT
B = CCA	L = TGC	V = CCT	1 = ACC
C = GTT	M = TCC	W = CCG	2 = TAG
D = TTG	N = TCT	X = CTA	3 = GAC
E = CGC	O = GGA	Y = AAA	4 = GAG
F = GGT	P = GTG	Z = CTT	5 = AGA
G = TTT	Q = AAC	_ = ATA	6 = TTA
H = CGC	R = TCA	, = TCG	7 = ACA
I = ACG	S = ACG	. = GAT	8 = AGG
J = AGT	T = TTC	= GCT	9 = GCG

- Mensaje: JUNE6_INVASION:NORMANDY
- Mensaje codificado en secuencia de 69 bases
  - AGTCTGTCTGGCTTAATAATGTCTCCTCGAACGATGGGATCTGCTTCTGG ATCATCCCGATCTTTGAAA
- Marcando inicio y final del mensaje
  - TCCCTCTTCGTCGAGTAGCA -y el complemento de- TCTCATGTACGGCCGTGAAT

Lámina 97 Dr. Roberto Gómez C.

Herramientas esteganográficas

- Covert.tcp
- dc-Steganograph
- EzStego
- FFEncode
- Gif-it-Up V1.0
- Gifshuffle
- Gzsteg
- Hide4 PGP
- Hide and Seek
- jpeg-jsteg
- MandelSteg
  - and GIF Extract
- MP3 Stego
- Outguess
- Paranoid
- PGE
  - Pretty Good Envelope
- PGPn123
- Publimark
- Stools
- Scytale
- Snow
- Stealth
- Steganos
- Steghide
- Stego
  - John Walker
- Stego
  - Romana Machado
- Stegonosaurus
- StegonoWav
- Stegodos
- Stegtunnel
- Texto
- wbStego
  - Werner Bailer
- Wnstorm
  - WhiteNoise Storm

Fuentes: http://www.jjtc.com/Security/stegtools.htm
http://www.theargon.com/achilles/steganography/

Lámina 98 Dr. Roberto Gómez C.

Características Información Oculta

- La información oculta es muy sensible al encontrarse en los bits menos significativos
- La pueden destruir:
  - La aplicación de cualquier filtro
  - Cambios en brillo o contraste
  - Un cambio de tamaño en la imagen
  - Cualquier trazo o cambio en la imagen
  - Recortar la imagen

Lámina 99 Dr. Roberto Gómez C.

Las marcas de agua: Watermarking

- Misma características esteganografía
- Robustez en contra de posibles ataques
  - esteganografia esta relacionada con la detección de un mensaje oculto, mientras que watermarking involucra el borrado/duplicación de un pirata
- Watermarking no siempre necesita estar oculto
- Tipos
  - invisibles robustas
  - invisibles frágiles
  - visibles robustas
  - visibles frágiles

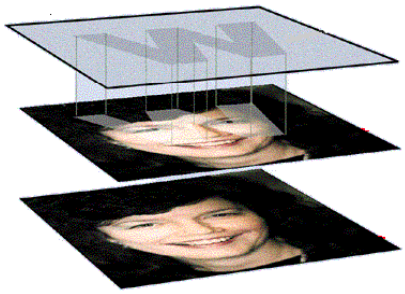


Lámina 100 Dr. Roberto Gómez C.

Marcas agua visibles vs invisibles

Imagen sin marca | Marca de agua | Imagen con marca

Imagen sin marca | Imagen con marca | Marca de agua

Lámina 101 Dr. Roberto Gómez C.

Medidas seguridad billetes

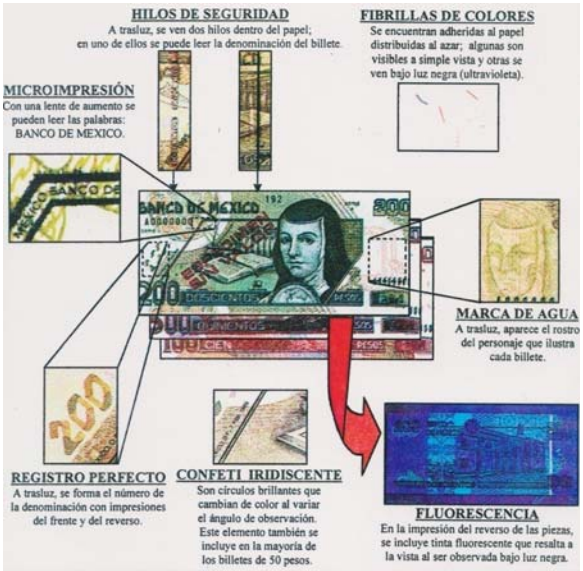


Lámina 102 Dr. Roberto Gómez C.

Robustas vs frágiles

- Marcas de agua robustas
  - soportan un cierto grado de modificación, dependiendo de las necesidades de la aplicación.
  - tienen que considerar los ataques a los que pueden ser sometidas las imágenes marcadas
- Marcas de agua frágiles
  - son diseñadas para destruirse o modificarse ante cualquier distorsión sobre la imagen que la contiene, verificando así la integridad de la imagen.
  - algunas marcas de agua permiten localizar las áreas en el espacio que han sido afectadas, e incluso caracterizar cierto tipo de distorsión

Lámina 103 Dr. Roberto Gómez C.

Requerimientos marca de agua

- Imperceptible para el ojo humano.
- No afectar a la calidad de la imagen.
- La marca recuperada debe identificar de forma unívoca al propietario de la misma.
- No debe ser detectada mediante pruebas estadísticas.
- Debe ser difícil, imposible, de eliminar excepto por el propietario.

Lámina 104 Dr. Roberto Gómez C.

Esteganografia vs Watermarking

- La información ocultada por un sistema de marca de agua, siempre se asocia al objeto digital a ser protegido.
- Comunicaciones esteganograficas son del tipo punto a punto, mientras que watermarking son del tipo punto-multipunto.
- Software
  - AiS Watermark Pictures Protector
  - Easy Watermark Creator
  - Alphatec Watermarking Suite 1.0
- Software de prueba
  - StirMark Benchmark 4 I
  - AudioStirMark
- Referencias:
  - http://www.elis.ugent.be/~banckaer/watermarking.html

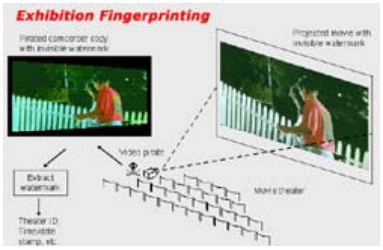


Lámina 105 Dr. Roberto Gómez C.

Características en común y diferencias

	Requerimientos	Watermarking		Steganografía
		Privado	Público	
Objetivo	Protección propiedad intelectual	++++		-
	Transmisión mensaje secreto sin despertar sospechas	-		++++
Especificación	Invisibilidad perceptual	++++		+++++
	Invisibilidad estadística o algorítmica	+		+++++
	Robustez contra borrado hostil, destrucción	+++++		-
	Resistencia contra un normal procesamiento de señales	++++		+
	Capaz sobrevivir códigos de compresión	++++		++
	Muy grande sobrecarga	++		++++
Detección/ extracción	Extracción/detección sin el host/objeto de cobertura	-	++++	++++
	Extracción con presencia del objeto/host de cobertura	++++	-	-
	Requerimiento de complejidad baja en extracción/detección	++		+++
	Capacidad opcional de bajado automático del objeto	+		++

Nota: Crucial +++++ Necesario: ++++ Importante +++ Deseable ++ Útil + Innecesario o irrelevante -

Lámina 106 Dr. Roberto Gómez C.

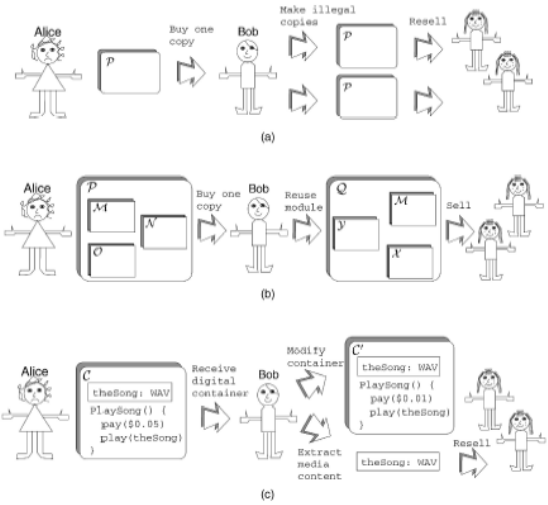
Marcas agua ejecutables

- Tan solo se trata de diferenciar un original de una copia.
- No son técnicas que se puedan usar para:
  - Evitar copias ilegales de software.
  - Contramedida de la ingeniería inversa.
- Basadas en el concepto de esteganografia.

Lámina 107 Dr. Roberto Gómez C.

Tipos de ataques

- Piratería software
- Ingeniería inversa maliciosa
- Software Tampering



Tomado sin permiso de: Watermarking, Tamper-Proofing, and Obfuscation Tools for Software Protection, Christian S. Collberg

Lámina 108 Dr. Roberto Gómez C.

Marcas estáticas de datos

- Usado para protección de software (copyrigtht)
- Incluir la marca como una cadena de caracteres dentro de inicialización de las variables.

```
char mark[] = “All your base…”
switch (a) {
        case 1: return “are”;
        case 2: return “belong”;
        case 3: return “to us”;
    …
}
```

```
{
int gonads, strife;

gonads = 1;
strife = 1;
printf (“weeeeee”);

}
```



```
{
int gonads, strife;

printf (“weeeeee”);
gonads = 1;
strife = 1;

}
```

- Si no hay datos o dependencias de control entre dos enunciados adyacentes son S1 y S2
  - marca puede insertarse dependiendo si S1 y S2 se encuentran en un orden lexicográfico o no.

Lámina 109 Dr. Roberto Gómez C.

Pros y contras

- Ventajas
  - fácil de implementar

```
> strings /usr/local/bin/netscape | \
     grep -i copyright
Copyright (C) 1995, Thomas G. Lane
```

- Desventajas
  - fácil de romper

Lámina 110 Dr. Roberto Gómez C.

Secuencia de las gráficas de control

- Un número de software puede ser codificado dentro del bloque de secuencia de un grafo del flujo del programa.

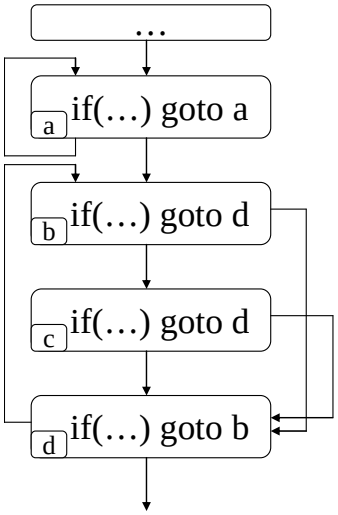


Lámina 111 Dr. Roberto Gómez C.

Marcas Dinámicas

- El usuario ejecuta el programa con un conjunto específico de entradas, después de los cuales el programa llega a un estado que representa la marca.
- Tipos de marcas
  - marca del “easter egg”
  - estructuras de datos
  - trazado de ejecución

Lámina 112 Dr. Roberto Gómez C.

Easter egg

- Parte de código que es activada dada una entrada inusual a la aplicación.
- Característica esencial del easter egg: lleva a cabo alguna acción que es inmediatamente perceptible por el usuario
  - i.e. desplegar el mensaje oculto
- Por ejemplo:
  - entrar al URL about:mozilla en Netscape 4.0 provocará que una imagen aparezca

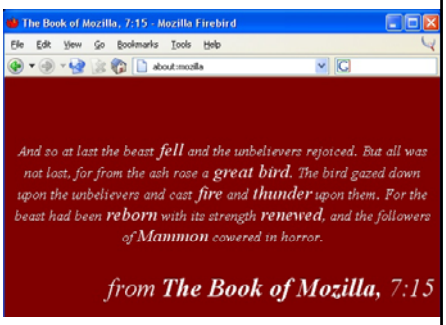
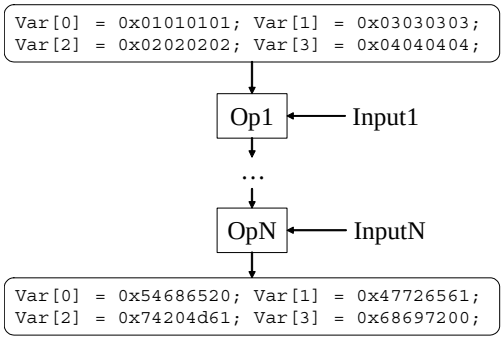


Lámina 113 Dr. Roberto Gómez C.

Ejemplo estructura datos dinámica

- El contenido de una estructura de datos cambia conforme el programa se ejecuta.
- El estado final de la estructura representa la marca almacenada.



"The Great Mahir"

Lámina 114 Dr. Roberto Gómez C.

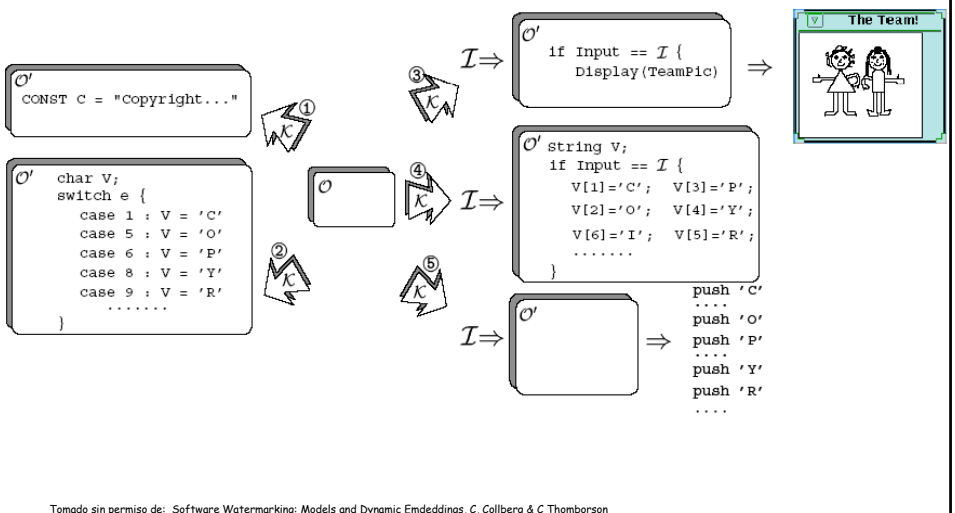
Ejemplo trazado ejecución dinámico

- Similar al de la estructura de datos.
- La información se oculta dentro del trazo (ya sea instrucciones o direcciones, o ambos) del programas conforme va corriendo de acuerdo a una entrada particular.
- La información se extrae con el monitoreo de algunas (tal vez estadísticas) propiedades del trazo de direcciones y/o de la secuencia de operadores utilizados.

```
80480d3:    85 db                   test   %ebx,%ebx
80480d5:    7e 29                   jle    0x8048100
80480d7:    83 7d 08 00             cmpl   $0x0,0x8(%ebp)
80480db:    74 23                   je     0x8048105
80480dd:    8b 45 08                mov    0x8(%ebp),%eax
80480e0:    a3 40 bc 08 08          mov    %eax,0x808bc40
80480e5:    80 38 00                cmpb   $0x0,(%eax)
                                    …
                                    …
8048100:    b8 00 00 00 00          mov    $0x0,%eax
8048105:    85 c0                   test   %eax,%eax
8048107:    74 0c                   je     0x8048115
8048109:    83 c4 f4                add    $0xfffffff4,%esp
```

Lámina 115 Dr. Roberto Gómez C.

Comparación entre las diferentes técnicas



Tomado sin permiso de: Software Watermarking: Models and Dynamic Embeddings, C, Collberg & C Thomborson

Lámina 116 Dr. Roberto Gómez C.

Stegoanálisis

- Arte de descubrir y convertir los mensajes en no útiles.
- Ataques y análisis de información oculta pueden tomar diferentes formas:
  - detección: solo detectar contenido esteganográfico
  - extracción: quitar la información
  - confusión: alteración, introducción, dejar inservible la información almacenada
  - deshabilitación de la información oculta
- Muchos casos requieren contar con porciones del objeto encubierto (stego-object) y posibles porciones del mensaje.
  - resultado: el stego-object

Lámina 117 Dr. Roberto Gómez C.

Métodos detección Steganografía

- Detección Visual
  - JPEG, BMP, GIF, etc.
- Detección Auditiva
  - WAV, MPEG, etc.
- Detección estadística o análisis de histogramas
  - cambios en lo patrones de lo pixeles o LSB
  - histograma: resumen gráfico de la variación de un conjunto de datos
- Detección estructural: verificar propiedades/contenidos de archivos
  - diferencia en el tamaño del archivo
  - diferencias en tiempo y fecha
  - modificaciones del contenido
  - checksum

Lámina 118 Dr. Roberto Gómez C.

Detección estructural

- Comparar las propiedades de los archivos
- Propiedades:
  - 04/04/2003 05:25p 240,759 helmetprototype.jpg
  - 04/04/2003 05:26p 235,750 helmetprototype.jpg
- Checksum
  - C:\GNUTools>cksum a:\before\helmetprototype.jpg
    3241690497 240759 a:\before\helmetprototype.jpg
  - C:\GNUTools>cksum a:\after\helmetprototype.jpg
    3749290633 235750 a:\after\helmetprototype.jpg

Lámina 119 Dr. Roberto Gómez C.

Detección visual

- Intenta revelar la presencia de comunicación secreta a través de la inspección, ya sea con el ojo o con la asistencia de una computadora.
- La computadora sirve para descomponer la imagen en planos de bits.
- Cualquier apariencia en el despliegue del plano LSB puede indicar la existencia de información secreta.
- Esta inspeccción puede tener éxito cuando los datos son insertados en areas suaves con valores de pixels cerca de la saturación.

Lámina 120 Dr. Roberto Gómez C.

Ejemplo detección visual

- Herramienta: www.guillermito2.net
- Elimina todos los bits de alto nivel para cada pixel excepto el último LSB
  - de tal forma que todos los bytes serán 0 o 1
- Problema
  - valores entre 0 y 1 en un rango de 256 valores do produce un color visible
- Solución
  - resaltar el valor de LSB
  - 0 se queda como y 1 toma un valor de 255

Lámina 121 Dr. Roberto Gómez C.

Ejemplo imagenes

Imagen original

LSB resaltados imagen pura

LSB resaltados con 1KB de datos aleatorios

LSB resaltados con 5KB de datos aleatorios

LSB resaltados con poema "if" (1.5 Kb)

Lámina 122 Dr. Roberto Gómez C.

Detección estadística

- Más poderoso ya que revela pequeña alteraciones en el comportamiento estadístico de las imagenes, provocado por la esteganografía
- Introducida por Westfeld y Piftzmann.
- Observarón que la inserción de datos cambia el histograma de la frecuencia de colores

Lámina 123 Dr. Roberto Gómez C.

Los pares de valores

- Ataque basado en pares de valores.
- Imagen definida con un solo color (p.e. negro)
  - todos los LSB's tendran valor de cero
  - si se oculta una imagen el valor de LSB's con valor de cero será diferente
  - si la información oculta es del mismo tamaño que la imagen el número de LSBs con valor de cero y de uno será el mismo.
  - tendrán una distribución de 50/50
  - a esto se le llama un par de valores: el número de 1 y de ceros

Lámina 124 Dr. Roberto Gómez C.

Detección estadística

- Considerando dos bits

(0**0**, 0**1**)	Primer par de valores. En caso de que un mensaje se encuentre oculto, los bits 00 se deben encontrar con la misma distribución que los bits 01.
(1**0**,1**1**)	Segundo par de valores. En caso de que un mensaje se encuentre oculto, los valores 10 se deben encontrar con la misma distribución que los bits 11.

- Considerando tres bits

(00**0**, 00**1**)	Primer par de valores originales de la imagen. En caso de que un mensaje se encuentre oculto, los bits 000 se deben encontrar con la misma distribución que los bits 001.
(01**0**, 01**1**)	Segundo par de valores. En caso de que un mensaje se encuentre oculto, los bits 010 se deben encontrar con la misma distribución que los bits 011.
(10**0**, 10**1**)	Tercer par de valores. En caso de que un mensaje se encuentre oculto, los bits 100 se deben encontrar con la misma distribución que los bits 101.
(11**0**, 11**1**)	Cuarto par de valores. En caso de que un mensaje se encuentre oculto, los bits 110 se deben encontrar con la misma distribución que los bits 111.

Lámina 125 Dr. Roberto Gómez C.

Detección estadística

- Se calcula la frecuencia real de valores de cada par de valores y se almacena en una tabla.
- Al mismo tiempo se calcula la frecuencia teórica de cada para de valores si un mensaje se encontrará oculto dentro de la imagen.
- Se comparan estas tablas con una prueba estadística
  - En caso positivo, distribución LSBs no es aleatoría:
    - la probabilidad de que la imagen no cuente con un mensaje oculto es alta.
  - Por otro lado, si no son muy diferentes,
    - la distribución de los LSBs es cercana a lo aleatorio y la probabilidad de encontrar un mensaje aleatorio oculto en los LSBs es alta.

Lámina 126 Dr. Roberto Gómez C.

Ejemplo detección estadística

Figura sin nada oculto

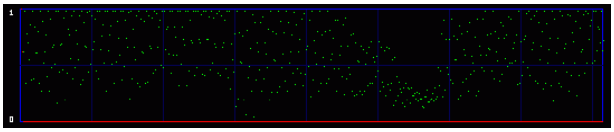


Figura con 1Kb de datos aleatorios

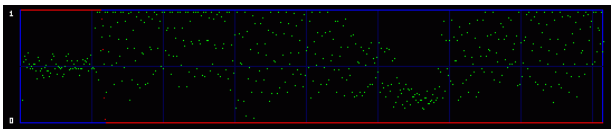


Figura con 5Kb de datos aleatorios

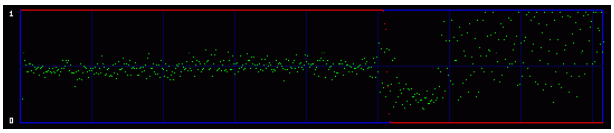


Figura con poema "if" (1.5 Kb)

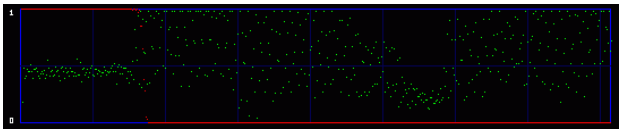
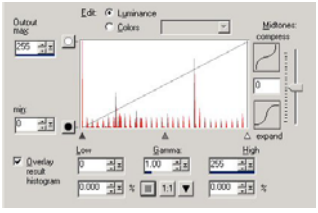


Lámina 127 Dr. Roberto Gómez C.

Otros métodos estadísticos

- Kurtosis
  - el grado de "plano" o "picos" de una curva describiendo una frecuencia de distribución
- Histogramas
  - posible detectar archivos ocultos
  - algunos presentan una tendencia repetitiva



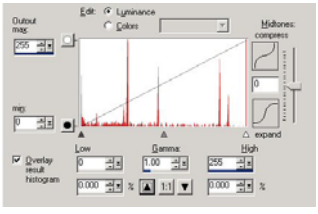


Lámina 128 Dr. Roberto Gómez C.

Otros metodos steganoanalisis.

Métodos steganográficos	Descripción	Técnicas steganográficas atacadas
RS Steganalisis	Sensibilidad de estadística dual en correlación espacial de pixeles con respecto a LSB aleatorio debido a la inserción esteganográfica usada	Varios técnicas de modificación del LSB
Po V basada en la prueba Chi-square	Una prueba Chi-square verifica cuando la ocurrencia de cada par de tendencia de valores tiende a ser la misma, indicando que algún dato ha sido insertado	Steganografía basada en intercambio de valores de pixeles o de coeficientes DCT
Verificación de paletas	Peculiaridades en el ordenamiento de la paleta es signo de una sistemática modificación.	Steganografía basada en paletas de imágenes
Método RQP	Método basado en analizar el incremento en el número de pares de colores cercanos, causado por la incrustación de información	LSB empotrado en imágenes de color verdadero
Verificar compatibilidad JPEG	Método detecta inusual inicio de la firma JPEG inherente en imágenes inicialmente almacenadas en formato JPEG	Steganografía del dominio del espacio usando imágenes inicialmente almacenada en formato JPEG
Análisis de histogramas	Método revela discrecionalidad o periodicidad en coeficientes particulares debido a la modificación relacionada con la cantidad	QIM u otros métodos cuantitativos de inserción
Detección ciega universal	Cantidades estadísticas construidas usando estadística de alto orden y un modelo de detección establecido con el umbral obtenido en un proceso de entrenamiento	Varias técnicas steganográficas

Lámina 129 Dr. Roberto Gómez C.

Firmas de archivos

Firma Hexadecimal	Extensión Archivo	Firma ASCII
FF D8 FF E0 xx xx 4A 46 49 46 00	JPEG (JPEG, JFIF, JPE, JPG)	ÿØÿà..JFIF.
47 49 46 38 37 61 47 49 46 38 39 61	GIF	GIF87a GIF89a
42 4D	BMP	BM

- Para una lista completa:
  www.garykessler.net/library/file_sigs.html

Lámina 130 Dr. Roberto Gómez C.

Analizando contenido archivos

- Si tiene una copia de un archivo original (virgen), este puede ser comparado con el archivo sospechoso
- Muchas herramientas pueden ser usadas para ver y comparar el contenido de un archivo oculto
- Se puede usar desde Notepad hasta un editor hexadecimal para identificar inconsistencia y patrones
- El verificar varios archivos puede identificar un patrón de firmas relacionado con el programa esteganográfico

Lámina 131 Dr. Roberto Gómez C.

La herramienta WinHex

- WinHex
  - www.winhex.com
  - Permite conversiones entre ASCII y Hex
  - Permite comparar archivos
    - guarda comparación como un reporte
    - guarda diferencia o bytes iguales
  - Posee capacidades de marcado
  - Permite búsqueda de strings
    - en ASCII y Hex
  - Otras características

Lámina 132 Dr. Roberto Gómez C.

Herramienta analizada: Hiderman

- Analizando Hiderman
  - herramienta de tipo shareware

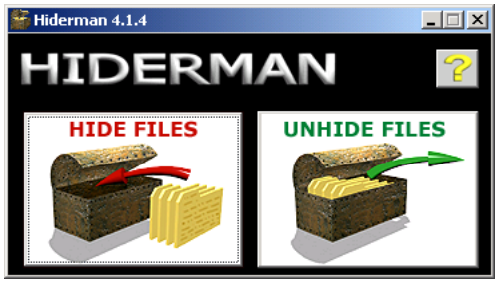


Lámina 133 Dr. Roberto Gómez C.

Empieza el análisis

- Primero se analiza la información de encabezado (principio archivo)
  - se puede ver que es un archivo tipo Bitmap, tal y como se ve por la firma BM

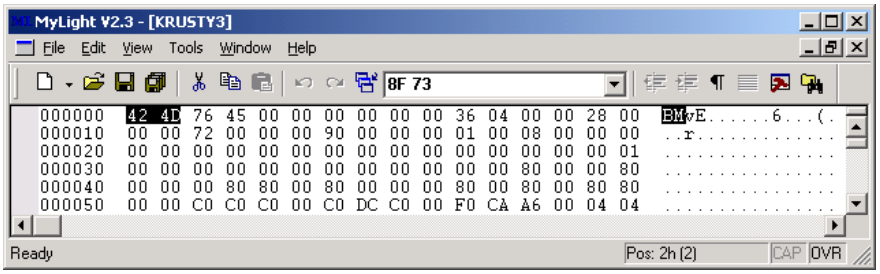


Lámina 134 Dr. Roberto Gómez C.

Siguiente paso

- Después se analiza el final del archivo, comparando el archivo virgen con el archivo portador
- A notar los datos añadidos al final del archivo

Lámina 135 Dr. Roberto Gómez C.

Mas información encontrada

- Aparte, hay que notar los tres últimos caracteres, “CDN” que es 43 44 4E en hexadecimal

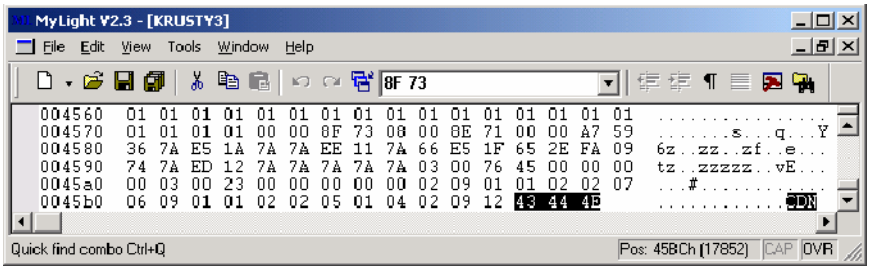
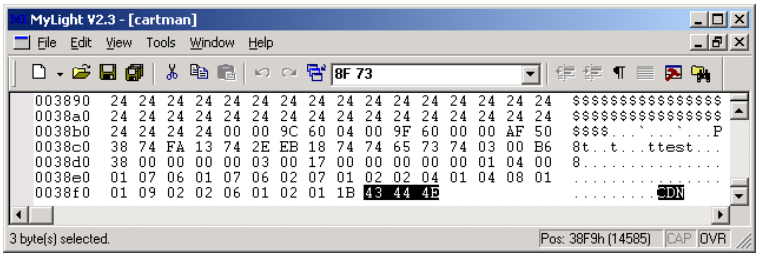


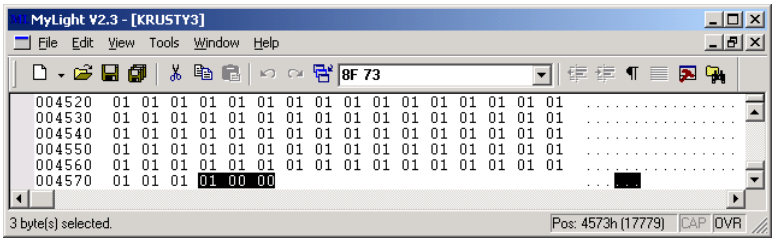
Lámina 136 Dr. Roberto Gómez C.

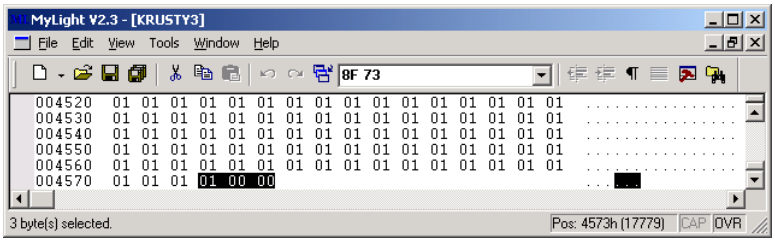
Encontrando la firma

- Escondiendo diferentes mensajes en diferentes archivos con diferentes passwords, se puede apreciar que los tres mismos caracteres (“CDN”) se añaden al final del archivo
- Se encontró la firma



Viendo final del archivo





Stegspy V2.1

- Programa de identificación de firmas
- Busca por firmas esteganográficas y determina el programa usado para ocultar el mensaje
- Identifica diferentes programas esteganográficos
- Identifica la ubicación del mensaje oculto
- Disponible en:
  - www.spy-hunter.com

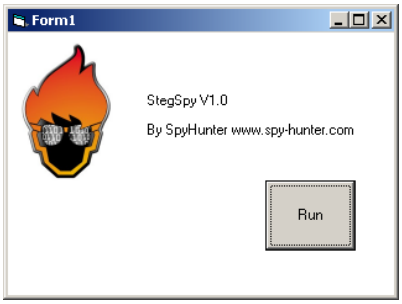


Lámina 139 Dr. Roberto Gómez C.

Programas detectables

- Hiderman
- JPHideandSeek
- Masker
- JPegX
- Invisible Secrets

Lámina 140 Dr. Roberto Gómez C.

http://www.guillermito2.net

Fecha	Programa	Precio	Metodo	Resultado	Herramienta extraer dato
16-09-02	Camouflage	Freeware	Fuse	Broken	Yes
18-09-02	JpegX	Freeware	Fuse	Broken	Yes
21-09-02	In Plainview	$10	LSB	Detectable	Yes
23-09-02	InThePicture	$25	LSB	Broken	Yes
29-09-02	Invisible Secrets 2002	$35	LSB	Detectable	Yes
04-12-03	Safe&Quick Hide Files 2002	$20	Fuse	Broken	No need
06-12-03	ImageHide	Freeware	LSB	Broken	Yes
03-01-04	Steganography 1.50 and 1.60	$25	Fuse	Broken	No need
18-02-04	JSteg	Open Source	LSB	Nothing to break	Yes
24-02-04	Cloak and DataStealth	Both $35	Fuse	Broken	No need
24-02-04	FortKnox	$45	LSB	Broken	Yes
27-02-04	Data Stash	$20	Fuse	Broken	No need

Lámina 141 Dr. Roberto Gómez C.

La herramienta Stegdetect

- Herramienta automática para la detección de contenido esteganográfico en imagenes.
- Capaz de detectar diferentes métodos esteganográficos
  - jsteg,
  - jphide (unix and windows),
  - invisible secrets,
  - outguess 01.3b,
  - F5 (header analysis),
  - appendX and camouflag
- Cuenta con un modulo, stegbreak, usado para lanzar ataques de tipo diccionario contra JSteg-Shell, JPHide y OutGuess 0.13b.

Lámina 142 Dr. Roberto Gómez C.

Detección automática de nuevos métodos esteganográficos

- Soporta análisis discriminante lineal.
- Dado un conjunto de imágenes normales y un conjunto de imágenes que contienen contenido oculto por una nueva aplicación esteganográfica, la herramienta puede determinar una función de detección lineal que puede usarse para detectar las nuevas imágenes.

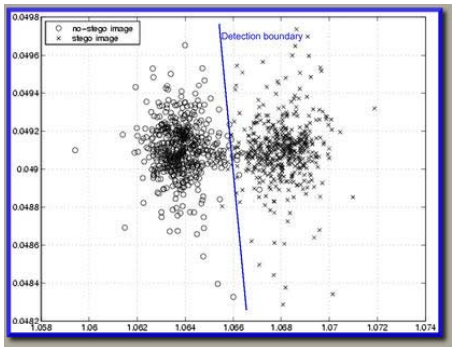


Lámina 143 Dr. Roberto Gómez C.

HTTP::StegTest

- Modulo diseñado en Perl para automatizar la recolección, detección y reporte de imágenes que han sido potencialmente alteradas por herramientas esteganográficas.
  - lleva a cabo comparaciones de imágenes
- Requiere del uso de otros programas para la prueba de imágenes.
  - todas las pruebas de imágenes se llevan a cabo con stegdetect 0.4
  - necesario definir path de la utileria unix cmp
- Disponible en http://www.duncanlamb.com/stegtest/

Lámina 144 Dr. Roberto Gómez C.

Detección software esteganográfico

- Necesario saber si en la computadora existe software estaganografico y cual es este.
- Una vez detectado se puede proceder a un análisis más dirigido de los archivos sospechosos.
- A tomar en cuenta
  - Sofrware esteganográfico en un medio de almacenamiento portable.

Lámina 145 Dr. Roberto Gómez C.

Gargoyle (StegoDetect)

- Detección de software esteganográfico en base a un conjunto de datos (hash set) propietario de los archivos de software esteganográfico.
- También puede ser usado para detectar la presencia de otro tipo de software
  - Criptografia, SMS, cracks

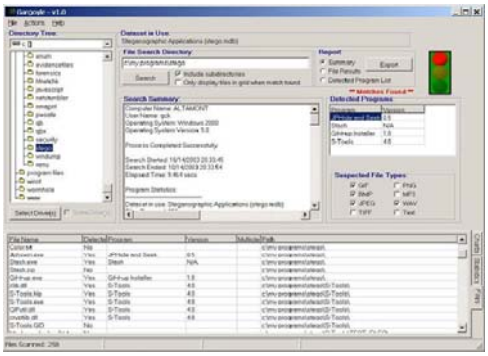


Lámina 146 Dr. Roberto Gómez C.

Forensic Toolkit y EnCase

- Detección de software esteganográfico
  - Pueden usar el HashKeeper, Maresware, y National Software Reference Library.
- A tomar en cuenta
  - Tamaño software esteganográfico en comparación con capacidad medios de almacenamiento temporal.

Lámina 147 Dr. Roberto Gómez C.

Stego Suite

- Conjunto de herramientas para investigación forense
- Herramientas que se incluyen
  - Stego Hunter
  - Stego Watch
  - Stego Analyst
  - Stego Break
- Producido por WetStone Technologies
  - https://www.wetstonetech.com/
- Precio (enero 2009): 1,495 USD

Lámina 148 Dr. Roberto Gómez C.

Stego Hunter

- Búsqueda de software esteganográfico

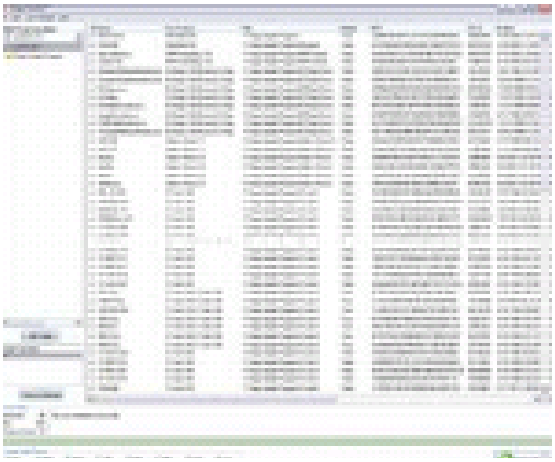


Lámina 149 Dr. Roberto Gómez C.

Stego Watch

- Detectar la presencia de mensajes ocultos.
- Posible revisar todo el sistema de archivos y presentar una lista de los archivos sospechosos.

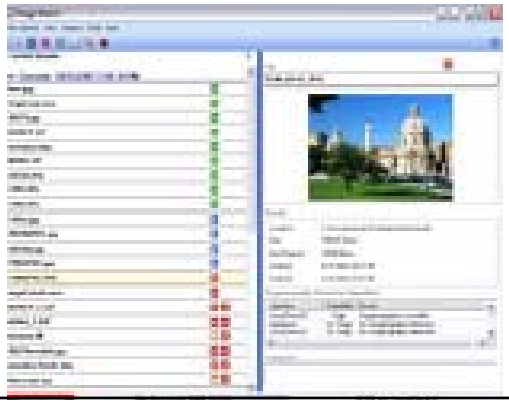


Lámina 150 Dr. Roberto Gómez C.

Stego Analyst

- Analizar las características de una imagen en búsqueda de elementos esteganográficos dentro de la imagen.

Lámina 151 Dr. Roberto Gómez C.

Stego Break

- Herramienta para obtener la contraseña usada para ocultar información dentro de un archivo.
- Incluye diccionarios y es posible añadir otros.

Lámina 152 Dr. Roberto Gómez C.