Instituto Tecnológico de Buenos Aires

Trabajo Práctico

Bendayan, Alberto (62786)

Boullosa Gutierrez, Juan Cruz (63414)

Gonzalez Rouco, Lucas (63366)

Pérez de Gracia, Mateo (63401)

Criptografía y seguridad - 72.44

Segundo cuatrimestre 2024 - Grupo 10

Índice

Introducción

La estenografía es una ciencia que se ocupa de ocultar un mensaje dentro de otro objeto. A lo largo de la historia, se usó mucho la estenografía para transmitir información entre partes sin correr el riesgo que sea interceptada y usada en su contra.

En lo que respecta a la informática, una de las formas más comunes de estenografía es la utilización de imágenes donde se modifican algunos de los bits menos significativos para que estos contengan el mensaje en cuestión.

Este trabajo práctico busca desarrollar y evaluar un sistema de esteganografía digital mediante la inserción de información en los bits menos significativos de imágenes BMP. También se incluyen técnicas de cifrado para proteger la información antes de ocultarla, ofreciendo así una capa extra de seguridad. El propósito es asegurar que la información sea imperceptible y, si llegara a ser descubierta, permanezca inaccesible sin la clave correcta.

Diseño del sistema

El sistema cuenta con una arquitectura modularizada en la cual se pueden reconocer y entender la funcionalidad de los componentes fácilmente.

- *Main:* Es el punto de entrada de nuestro sistema donde se llama al parser de los elementos y luego a las acciones pertinentes de acuerdo a la entrada provista.
- *arguments:* Este módulo cuenta con el parser utilizado y un enum con las acciones que pueden ser ejecutadas.
- *utils:* Diferentes clases que no están ligadas completamente con el funcionamiento del sistema.
- *encryption:* Contiene los algoritmos de encriptación y encadenamiento.
- *stenography:* Contiene los algoritmos de estenografía.

Estegoanálisis de los archivos provistos

Se tienen 4 archivos .bmp y se sabe que hay 1 con cada algoritmo de estenografia (LSB1, LSB4 y LSBI) y también se sabe que el otro archivo tiene información oculta. Entonces mediante prueba y error, se obtuvo un .png de un buscaminas cuando se hizo un extract al archivo montevideo.bmp con LSB1. Luego se obtuvo un .pdf cuando se hizo un extract al archivo avatar.bmp con LSBI. En dicho archivo .pdf se encontraba un mensaje indicando que la password es: metadata. A continuación se hizo un hexdump al archivo madrid.bmp y se encontró que en el final del archivo había un mensaje indicando que se debía cambiar la extensión del archivo .png obtenido, por .zip y descomprimirlo. El .zip contiene un archivo .txt con el siguiente texto:

"cada mina es un 1.
cada fila forma una letra.
Los ascii de las letras empiezan todos en 01.
Asi encontraras el algoritmo que tiene clave de 256 bits y el modo
La password esta en otro archivo
Con algoritmo, modo y password hay un .wmv encriptado y oculto."

Vale la pena mencionar que en la imagen hay minas que no están marcadas.

Traduciendo el buscaminas a binario, se obtiene la siguiente secuencia:

01000001
01100101
01110011
01001111
01100110
01100010

Que se traduce a "AesOfb"

Recopilando todo se tiene la siguiente información:

- Queda por utilizar LSB4.
- Queda por extraer información del archivo secreto1.bmp
- Está encriptado con AES256 modo OFB.
- La clave es metadata

Finalmente se extrae un archivo .wmv.

Cuestiones a analizar

I. Discutir los siguientes aspectos relativos al documento.

a) Organización formal del documento:

En el encabezado, se encuentra la siguiente información: título de la revista, fecha de publicación, volumen y número, derechos de autor, ISSN y el sitio web.

Luego se encuentra el título, nombre de los autores, afiliación (por ejemplo Universidad) e información de contacto en este caso un email y un abstract.

Finalmente, en el cuerpo del documento se encuentra la introducción, una explicación sobre el funcionamiento de los métodos tradicionales y la mejora que brinda el nuevo algoritmo, luego la conclusión y por último las referencias.

b) La descripción del algoritmo:

Entrada:

- Imagen de portada: La imagen en la que se ocultará el mensaje.
- Mensaje secreto: La secuencia de bits del mensaje que se desea ocultar en la imagen.

Proceso de incrustación:

- Patrones de bits: El algoritmo empieza identificando y clasificando los patrones de bits en la imagen de portada (específicamente el segundo y tercer bit menos significativo) en cuatro combinaciones posibles: 01, 10, 00 y 11.
- Inversión de bits: Para cada patrón de los bits identificados en la imagen de portada, se emplea el algoritmo estándar de LSB para crear una imagen "stego" inicial. Luego, se compara esta imagen "stego" con la imagen de portada para ver cuántos píxeles han cambiado. Si el número de píxeles modificados en un patrón

específico es mayor que el número de píxeles sin cambio, se invierte el último bit de esos píxeles para reducir la alteración visual.

Almacenamiento de patrones invertidos:

- El algoritmo guarda un registro de los patrones en los que se invirtieron bits, de modo que estos puedan ser utilizados en la recuperación del mensaje oculto.

Extracción del mensaje secreto:

- Clasificación de patrones: Para extraer el mensaje, el algoritmo vuelve a clasificar los patrones de la imagen "stego" en función de los bits invertidos almacenados previamente.
- Recuperación de bits: Utilizando los patrones guardados, el algoritmo re-invierte los bits y recupera los bits originales del mensaje oculto.

c) La notación utilizada, ¿es clara? ¿Hay algún error o contradicción?

En general es bastante clara la notación, sobretodo porque utiliza ejemplos, de todas formas tiene algunos errores como por ejemplo:

“Calculate the pattern occurrences of these two bits on the cover-image, which are either 00, 10, 10, or 11.”

Donde repite 10, 10 en vez de escribir 10, 01.

II. *Esteganografiar un mismo archivo en un .bmp con cada uno de los tres algoritmos, y comparar los resultados obtenidos. Hacer un cuadro comparativo de los tres algoritmos estableciendo ventajas y desventajas.*

LSB1:

LSB4:

LSBI:

Sobre estas imágenes en particular, en ninguno de los tres casos se puede percibir a simple vista que hay información oculta ya que no se perciben diferencias visuales.

Método	Ventaja	Desventaja

LSB1	- Fácil de implementar - Baja alteración visual (al cambiar el bit menos significativo es imperceptible en la mayoría de los casos)	- Fácilmente identificable - Poco resistente a ataques de comprensión
LSB4	- Fácil de implementar - Alta capacidad de almacenamiento	- Fácilmente identificable - Poco resistente a ataques de comprensión - Alteración visual moderada
LSBI	- Mayor resistencia frente a ataques - Buena calidad de imagen	- Más complejo de implementar - Menor capacidad de almacenamiento comparado con LSB4

III. *Explicar detalladamente el procedimiento realizado para descubrir qué se había ocultado en cada archivo y de qué modo. Indicar qué se encontró en cada archivo.*

Esta información se encuentra en la sección “Estegoanálisis de los archivos provistos”.

IV. *Algunos mensajes ocultos tenían, a su vez, otros mensajes ocultos. Indica cuál era ese mensaje y cómo se había ocultado.*

En Montevideo.bmp se encuentra oculto un archivo .png dicho archivo tenía otro mensaje oculto que se revela cambiando la extensión del archivo de .png a .zip y luego descomprimiendolo obteniendo el archivo sol14.txt. Por ende un mensaje oculto era la imagen del buscaminas y el otro era “sol14.txt” que tenía el siguiente mensaje:

"cada mina es un 1.
cada fila forma una letra.
Los ascii de las letras empiezan todos en 01.
Asi encontraras el algoritmo que tiene clave de 256 bits y el modo
La password esta en otro archivo
Con algoritmo, modo y password hay un .wmv encriptado y oculto."

V. *Uno de los archivos ocultos era una porción de un video de una película, donde se ve ejemplificado una manera de ocultar información ¿qué se ocultaba y sobre qué portador?*

En el video se habla de que la persona que buscaba ocultar información terminó una relación por email, luego se habla de que hay un archivo que es más grande de lo que debería, y sacan la conclusión de que hay un error de compresión o bien su agente oculto información en dicho archivo.

Como allí finaliza el video no es claro que son los datos ocultos. También se puede suponer que el archivo del que hablan es dicho mail, aunque por lo que se ve en el video no parece serlo.

VI. *¿De qué se trató el método de esteganografiado que no era LSB1 ni LSB4 ni LSBI? ¿Es un método eficaz? ¿Por qué?*

El método de esteganografiado era simplemente insertar un mensaje de texto al final del archivo. No es muy eficaz, porque solo basta con hacer un hexdump -C para poder ver el mensaje. O usar otra herramienta que permite ver el contenido del archivo en hexadecimal para encontrar el mensaje pasando los valores hexadecimales a caracteres ascii.

VII. *¿Por qué la propuesta del documento de Majeed y Sulaiman es realmente una mejora respecto de LSB común?*

Porque le añade más seguridad, por un lado saltea los bits del color rojo, modificando únicamente el verde y el azul. Luego en base a los patrones de segundo y tercer bit menos significativo realiza una inversión de bits. Por ende dificulta más la extracción de la información para los atacantes, ya que

deben tener en cuenta que los bits del color rojo no deben usarse y en qué casos deben invertir los bits.

VIII. *En la implementación se optó por guardar los patrones invertidos antes del mensaje ¿de qué otra manera o en qué otro lugar podría guardarse el registro de los patrones invertidos?*

Es más que claro, que el mejor lugar para guardar tal información es este, sin embargo, también es el más obvio. El registro de los patrones invertidos se podría guardar en cualquier lugar, lo importante es que la persona que debería tener acceso a extraer el mensaje, sepa dónde está. Podrían ser los últimos 4 bytes, inclusive 4 bytes a partir del décimo, pero como se dijo anteriormente, lo importante es que se mantenga esa información entre las partes.

IX. *¿Qué dificultades encontraron en la implementación del algoritmo del paper?*

No se encontraron mayores dificultades a la hora de implementar el algoritmo ya que primero se implementó LSB1 y LSB4, por lo que ya se tenía un entendimiento acerca de cómo operar con bits. Lo único que se tuvo que tener en cuenta es que tantos bits se cambian según el patrón, y saltear los píxeles rojos. Sin embargo, se podría decir que la mayor dificultad, que no fue muy grande, fue decidir donde guardar la información acerca de los patrones ya que el paper no especificaba dónde realizar esto. Se decidieron utilizar los primeros 4 bytes de la data para guardar la información, tal como especificó la cátedra.

X. *¿Qué mejoras o futuras extensiones harías al programa stegobmp?*

La mejora que se puede plantear es que simplemente se pase un portador y en caso de encriptación una contraseña y automáticamente pruebe con todos los modos de estenografía y todos los algoritmos y modos de encriptación hasta que encuentre información oculta o hayas probado con todas las combinaciones posibles y no haya encontrado nada.