Criptografía y Seguridad

Trabajo Práctico Especial: Secreto Compartido + Esteganografía

Grupo 10

Manganaro Bello, Santiago (56289)
Soracco, Tomás (56002)
Vazquez, Agustin (55345)

Introducción

En pos de introducirnos en el mundo de la esteganografía y sus aplicaciones, se implementó un programa ***stego*** en C que se encarga de distribuir y recuperar una imagen secreta de extensión ***.bmp*** en o desde otras imágenes del mismo formato. Este programa está basado en el algoritmo detallado en el documento *"Verifiable Image Secret Sharing Using Matrix Projection"* cuyas autoras son Nurfathiya Faradiena Azzahra y Kiki Ariyanti Sugeng de Universitas Indonesia, de Indonesia.

El enfoque de este análisis estuvo puesto en la efectividad de la esteganografía como herramienta para mantener esquemas de secreto compartido desde el punto de vista de la aplicación práctica; se busca dar un contexto más detallado de las decisiones tomadas a la hora de la realización del trabajo.

Cuestiones a analizar

1. Discutir los siguientes aspectos relativos al documento. a. Organización formal del documento. b. La descripción del algoritmo de distribución y la del algoritmo de recuperación. c. La notación utilizada, ¿es clara? ¿cambia a lo largo del documento? ¿hay algún error?

a) La estructura del documento es común a gran parte de la bibliografía provista. Está compuesto por un abstract (breve resumen del paper) y la introducción, poniendo en contexto la relevancia de la investigación y los avances al momento.
Posteriormente se pasa a nombrar los supuestos y documentos previos en los que se basará la investigación y se realiza la propuesta propiamente dicha, explicitando las variables, algoritmos y explicación del modelo matemático que lo sustenta, con un ejemplo de realización.

Finalmente se procede a demostrar resultados de la implementación y llevar a algunas conclusiones sobre el algoritmo propuesto.

b) Algoritmo de distribución:

- Determinar la matriz de la imagen secreta S
- Construir una matriz mxk aleatoria de rango k, donde m>2(k-1)-1
- Calcular la matriz $\mathbb{S}$ realizando *proj(A)* y la matriz resto R = S - $\mathbb{S}$
- Generar n vectores X_j de dimensión k
- Computar los vectores $V_j = (A\ x\ X_j)$
- Computar las matrices G_j utilizando los pixels de la imagen secreta, la matriz R y los vectores V
- Elegir la imagen W y computar la matriz Rw = (W-$\mathbb{S}$)
- Calcular los *shares* y distribuirlos $Sh_j = [V_j G_j]$

Algoritmo de recuperación

- Obtener k *shares* de los participantes
- Computar la matriz B haciendo un merge de los vectores V_j
- Calcular la matriz $\mathbb{S}$ realizando *proj(B)*
- Construir la matriz R usando las matrices G_j
- Calcular la matriz secreta S = ($\mathbb{S}$ + R)
- Verificar veracidad de la matriz utilizando Rw y el metodo PSNR

c) En primer lugar, se detectan cuestiones de semántica y sintáctica. Entre ellas, es posible resaltar numerosas omisiones de verbos, mezcla de plurales y singulares, unión de palabras y variables, entre otras.

Sin embargo, el mayor problema radica en la manera en que las cosas están definidas. Por un lado, a lo largo del documento se hace referencia a variables que no están explícitamente referenciadas paralela o previamente y su significado debe obtenerse por contexto. De hecho, a veces, hasta no analizar los

ejemplos (que aparecen mucho después), es literalmente imposible entender de qué se está hablando. En algunos casos, sucede que en una oración se presentan muchas variables a la vez y hasta en desorden, lo cual dificulta la lectura. Por otro lado, no se respetan las notaciones definidas. Dentro del documento, por ejemplo, las n imágenes que se generan a partir de la imagen secreta, por momentos son referidas como "*share images*" y por otros momentos como "*shadow images*" sin antes explicar el procedimiento. A su vez, se hace referencia a los *papers* de Li Bai y Thien/Lin, y en lugar de mantener las notaciones de estos autores, se crean nuevas y se deriva al lector la tarea de comprender y buscar la información particular a la que hace alusión.

2. ¿Por qué la propuesta de Azzahra y Sugeng supone una mejora a la propuesta de Li Bai?

En el documento de Li Bai se propone un esquema mediante el cual se podían generar y compartir los archivos que posteriormente se utilizarían para develar la imagen secreta (de ahora en más *shares*) . Esto tomaba como supuesto que el distribuidor de las imágenes era honesto y que los *shares* eran verídicos, es decir, realmente generaban la imagen secreta.

Para asegurarse de que así fuera, Azzhara y Sugeng propusieron un esquema en el cual interviene una imagen llamada *watermark* en el proceso de creación de los *shares*. Está imagen servira posteriormente para verificar que el *share* obtenido es el que corresponde a la imagen secreta que se intenta develar. Comprobando la veracidad del distribuidor.

3. ¿Qué dificultades se encuentran al elegir pares (k, n) distintos de los establecidos en este TP?

A la hora de elegir los pares (k, n), es necesario tener en cuenta algunos factores:

- El primero de ellos, radica en la definición de k y n, particularmente en la relación que existe entre ambas (2 <=k <= n).
- El segundo, hay que tener en cuenta que valores más grandes de k y de n implican un mayor computo y por lo tanto mayor tiempo de ejecución, con lo cual, no deberían ser excesivamente grandes.
- El tercero a tener en cuenta que el valor de k está, a su vez, estrictamente relacionado con el tamaño de las *share images*.

4. ¿Por qué es importante controlar el rango de A y el resultado de At .A?

El rango de A debe ser controlado ya que si su rango fuera menor a k, entonces al computar los vectores v_j podríamos obtener vectores que no sean linealmente independientes.

El resultado de AA^T debe ser una matriz invertible, ya que de lo contrario no se podrá computar la matriz $\mathbb{S}$, obtenida realizando *proj(A)*.

5. ¿Por qué es válida la forma de generar los Xi?

La forma de generar los Xi es válida porque son vectores de k, linealmente independientes, por lo tanto cuando se realice el producto por la matriz A (rango k) nos aseguramos de tener j vectores de dimensión k linealmente independientes.

6. La imagen RW que se obtiene es una imagen “con ruido”. ¿Sería necesario ocultarla mediante esteganografía? ¿Cómo podría hacerse?

Los bytes obtenidos de la matriz RW son de conocimiento público, dado que es el metodo de verificacion para la veracidad

del distribuidor, por lo cual ocultarlo mediante esteganografía no tendría mucho sentido. En caso de que se quiera hacer por cuestiones de privacidad, sería posible utilizando el método *LSB replacement* tal como se hizo para las sombras en el esquema (4,8). Los primeros 7 bits de cada byte podrían rellenarse de manera aleatoria o bien utilizando una imagen predeterminada, lo cual sería indistinto.

7. ¿Por qué siempre hay que indicar n, aún al recuperar?

En la generación, n se utiliza para determinar el número de vectores aleatorios que deben obtenerse, para luego poder obtener n imágenes diferentes. Sin embargo, el motivo por el cual siempre se indica el n es por que en base al par (k,n) se utiliza un método diferente de esteganografía para realizar el ocultamiento. Para (2,4) se utiliza LSB2 y para (4,8) se utiliza LSB. De no indicarlo, no podríamos determinar qué esquema se debe utilizar para recuperar las sombras de las imágenes portadoras.

8. ¿En qué otro lugar puede guardarse el número de sombra?

Como opción para guardar el número de sombra, podemos guardarlo en alguno de los bytes del header (7, 8 y 9) que están reservados. Otra opción, aunque un poco más rebuscada, sería incluirlo en el ocultamiento por esteganografía.

9. Discutir los siguientes aspectos relativos al algoritmo implementado:

a. Facilidad de implementación

La implementación del algoritmo toma su complejidad del manejo de matrices con aritmética modular y de realizar las verificaciones correspondientes al rango o validez de matrices.

Más allá de estos puntos, el algoritmo está separado en pasos suficientemente claros como para realizar su implementación sin muchas complicaciones.

b. Posibilidad de extender el algoritmo o modificarlo.

Esto depende mucho de qué tipo de extensiones o modificaciones se busquen realizar. Algunos ejemplos podrían ser:

- Agregar soporte para otros tipos de bitmap.
- Poder distribuir información de y hacia otro tipo archivos además de bitmaps, como por ejemplo, audios, videos, documentos.
- Soportar imágenes a color.
- Agregar soporte para otros esquemas, es decir, valores de k y n, con las consideraciones necesarias.
- Agregar un mayor número de métodos de esteganografía y no vincularlos uno a uno con los esquemas; y como tal, permitir al usuario elegirlo mediante un parámetro extra.

10. ¿En qué situaciones aplicarían este tipo de algoritmos?

El concepto de secreto compartido, surge como una manera de proteger claves. Debido al rápido crecimiento de la informática, la transmisión segura y la protección de información secreta se ha convertido en un tema importante. Se han desarrollado numerosos métodos, como la criptografía y la esteganografía para proteger los datos.

Pero ambos métodos son del tipo *Single Point of Failure (SPOF)*, ya que utilizan un mecanismo de almacenamiento único y, por lo tanto, no son robustos contra la pérdida o la manipulación. Los métodos de intercambio secreto que distribuyen un contenido secreto entre un conjunto de participantes podrían ser una de las posibles soluciones.

La protección efectiva y segura para mensajes importantes es una preocupación principal en aplicaciones comerciales y militares. Debido a su característica, hay muchos escenarios de aplicación, como la votación electrónica, comunicaciones en canales públicos no confiables, sistema de almacenamiento distribuido y control de acceso, etc.

La esteganografía y la marca de agua utilizan la cobertura multimedia para ocultar los datos secretos. La primera puede ser utilizada para transmitir un gran volumen de información encubierto en un archivo mediante técnicas de ocultación de información. La segunda puede usarse a partir de una marca en el archivo multimedia a partir de la cual sea posible detectar la localización de modificaciones no autorizadas subsiguientes y por lo tanto, puede servir como herramienta potencial para la autenticación de datos.

Recuperación del Secreto

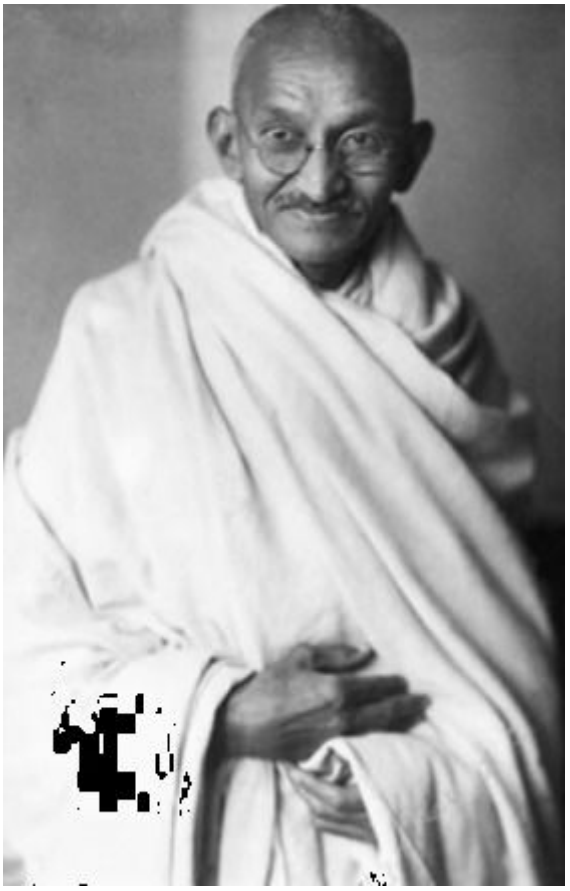
Esquema (2,4)

Secreto

Watermark

Esquema (4,8)

Secreto



Watermark

Como se puede ver en las imágenes, hay zonas que salen en negro y que, en teoría, desde un análisis real de la foto, deberían ser blancas. La explicación que le encontramos a esto es que a lo largo del trabajo se opera con módulo 251 y el rango negro-blanco va de 0 a 255. Por lo tanto, cuanto se detecta un blanco muy intenso, el módulo hace que se transformen en un negro muy intenso.