Trabajo Práctico Especial

72.44 - Criptografía y Seguridad

2021 1Q

Instituto Tecnológico de Buenos Aires

TPI - ITBA

Grupo 3

Baader, Juan Martín. 58647

Bergagna, Federico Manuel. 58446

Rodríguez Brizi, Manuel. 58459

Recuperación del secreto

Recuperación exitosa

Recibimos 6 imágenes portadoras en las cuales se había escondido una imagen secreta. El valor de k para recuperarla es 6 al igual que n.

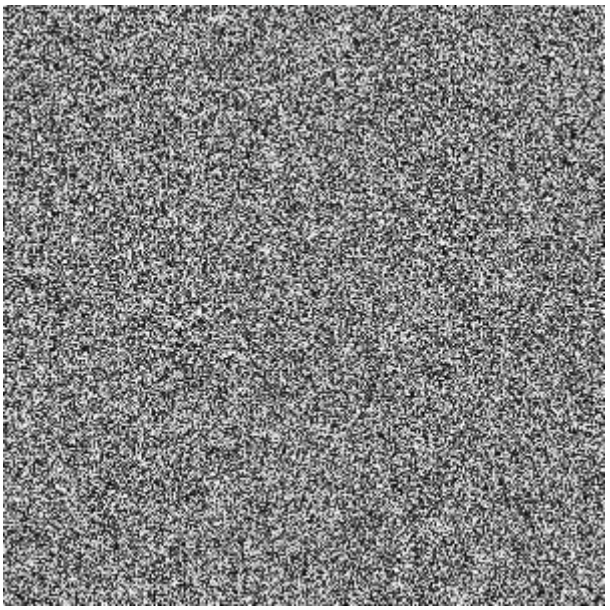
Para lograrlo, teniendo en cuenta que dichas imágenes se encuentran en el directorio *secret*, se ejecuta lo siguiente:

```
fede@fede-MS-7984:~/Desktop/cripto-tpe$ ls
includes  main.c  Makefile  README.md  secret  src
fede@fede-MS-7984:~/Desktop/cripto-tpe$ make all
gcc -c -Wall -Werror -o src/decrypt.o src/decrypt.c -I includes/
gcc -c -Wall -Werror -o src/encrypt.o src/encrypt.c -I includes/
gcc -c -Wall -Werror -o src/galois.o src/galois.c -I includes/
gcc -c -Wall -Werror -o src/lagrange.o src/lagrange.c -I includes/
gcc -c -Wall -Werror -o src/bmp_handling.o src/bmp_handling.c -I includes/
gcc  -Wall -Werror main.c src/bmp_handling.o src/galois.o src/lagrange.o src/dec
rypt.o src/encrypt.o -o ss -lm -I includes/
fede@fede-MS-7984:~/Desktop/cripto-tpe$ ./ss r secret.bmp 6 secret
Recuperacion terminada.
fede@fede-MS-7984:~/Desktop/cripto-tpe$
```

De esta forma, el secreto oculto en las portadoras queda guardado en la imagen *secret.bmp*:

Intento con menos imágenes

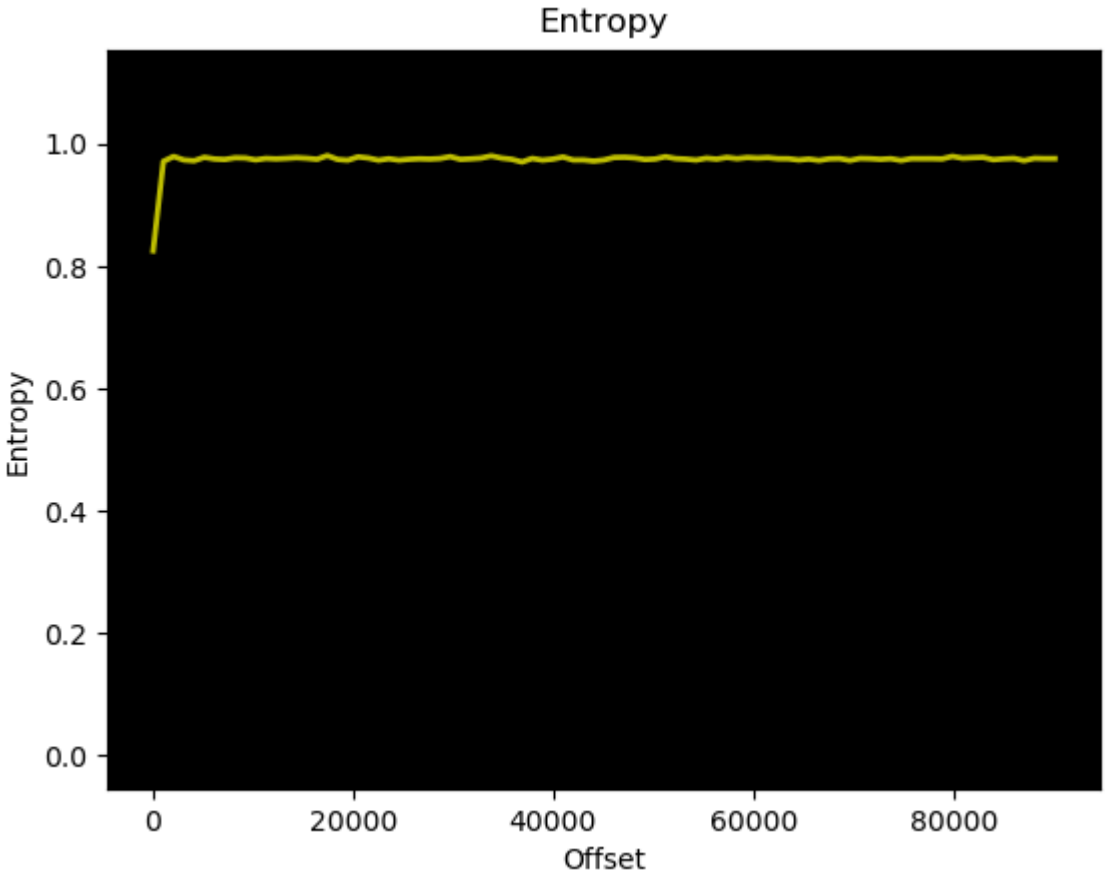
Supongamos que alguien encuentra 5 de las 6 portadoras necesarias para recuperar la imagen, al ejecutar el programa con k = 5 se obtiene lo siguiente:



Como se puede apreciar, si no se alcanza el umbral de recuperación, la imagen obtenida es ruido y no se puede inferir nada del contenido de la imagen original, cumpliendose el postulado de que no hay flujo de información en un esquema de Shamir cuando no se cuenta con la cantidad necesaria de portadoras.[1]

Si usamos el programa binwalk para analizar la entropía de esta imagen obtenemos el siguiente gráfico:

[1] C. L. F. Corniaux and H. Ghodosi, "An entropy-based demonstration of the security of Shamir's secret sharing scheme," *2014*



Donde se ve un primer momento en que la entropía vale 0.8 que corresponde al header de la imagen y luego sube hasta el valor máximo para mantenerse ahí por el resto del archivo, afirmando lo dicho anteriormente, sin contar con toda la información necesaria habría una gran incertidumbre al momento de determinar el verdadero valor del secreto.

A modo de comentario, se presenta una situación similar si al intentar recuperar el secreto se utiliza un polinomio generador distinto al indicado en el enunciado.

Cuestiones a analizar

Pregunta 1

Organización formal de Documento

La organización formal del documento es adecuada ya que lleva al lector a través del proceso completo, empieza por presentar el polinomio y el método de shamir. Luego muestra como el secreto se separa en bloques y como se encripta y esconde en las imágenes portadoras, y por último terminar muestra como se encripta.

Descripción de los algoritmos

A la hora de hablar de la descripción del algoritmo creemos que es poco clara a la hora de hablar de el algoritmo de desencriptación, en particular a la hora de explicar como obtener los S a partir de los pares (Xi,j, Yi,j), esto se da en parte por la confusa notación utilizada, pero también los autores fallan en explicar este proceso con palabras. Los autores no logran explicar claramente el cálculo de Y' o el proceso iterativo luego de utilizar el primer polinomio de Lagrange.

La notación utilizada

Como se mencionó anteriormente, la notación se vuelve confusa en ciertos puntos. Uno de los problemas principales, que asumimos es por problemas del formato del documento y no culpa de los autores, es la falta de símbolos de resta, que al tratar de entender el documento por primera vez, dificulta mucho la comprensión. Otro problemas es la notación utilizada para definir el polinomio (F(x)) y el polinomio evaluado (F(X)). Además, los autores utilizan 3 notaciones diferentes para referirse al polinomio evaluado a lo largo del documento (F(Xi,j) Ti,j Yi,j), lo cual complica el seguimiento de estas variables a lo largo de la lectura.

Por último, un problema de notación se generó a la hora de definir Y', ya que se utiliza el elemento S1,j lo que indicaría que para definir Y' se utiliza siempre el primer valor de S obtenido por lagrange en 0, mientras que el valor correcto debería ser Si-1,j, implicando así que es el valor anterior a cada iteración.

Pregunta 2

La carga útil o payload se refiere a la cantidad de información que pueden transportar las portadoras. En el documento, todo el tiempo se menciona la frase "optimizar la carga útil", y se debe a que el método que se desarrolla en este trabajo permite tener una carga útil más grande en comparación con otras propuestas anteriores. La máxima carga útil puede ser determinada por el valor de k y el tamaño de las imágenes portadoras.

Pregunta 3

La principal ventaja que provee trabajar en $GF(2^8)$ por sobre las operaciones con módulo es la de recuperar el secreto sin pérdidas, ya que evita el truncamiento del valor de los píxeles, y esto a la vez nos permite reconstruir de forma perfecta la imagen secreta y cualquier otro dato secreto. Una desventaja de $GF(2^8)$ por sobre las operaciones de módulo es que tiene que hacerse procesamiento adicional para hacer la suma, multiplicación y división de los valores, como calcular los inversos multiplicativos de los números (que pueden pre-calcularse para mayor eficiencia).

Pregunta 4

Si, sería posible utilizar otro polinomio generador. A niveles prácticos, podría utilizarse cualquier polinomio siempre que sea un polinomio generador del campo. Dicho polinomio podría guardarse como clave, pero no aportaría mucha seguridad dado que no hay muchos polinomios generadores.

Pregunta 5

Con el método propuesto, se pueden esconder todo tipo de archivos, entre los que se encuentran imágenes, ejecutables, documentos, audio, datos encriptados, etc. Lo que permite esconder todos estos archivos tan distintos es debido a que se trabaja con operaciones en $GF(2^8)$, una de las mayores ventajas de trabajar con esto. Estas contribuyen a una recuperación sin pérdida de los datos secretos, permitiendo así manejar secretos de todos los tipos mencionados anteriormente. Si la recuperación sufriera muchas pérdidas, no sería posible esconder gran parte de esos archivos.

Pregunta 6

Como se mencionó en el punto anterior, el esquema es capaz de encriptar cualquier tipo de archivo, con lo cual la modificación del programa sería bastante simple. El cambio consistirá en remover los métodos de obtención de píxeles del secreto, y simplemente se pasaría en plano para que sea guardado completamente. Hay que contemplar si el tamaño de las portadoras es suficiente para guardar todo el archivo por completo.

En cuanto a la desencriptación, el proceso es similar. En lugar de reemplazar los píxeles de una imagen y reutilizar su encabezado para escribir la imagen, el resultado de la descripción se guardará directamente. En el caso de que la modificación sea únicamente para archivos .bmp se agregara el formato al final, en otro caso se dejará sin formato y quien lo desencripte deberá asignarle el correspondiente.

Pregunta 7

Si utilizaramos imágenes a color, tendríamos que utilizar 24 bits en lugar de 8 para cada pixel. A nivel algorítmico no cambiaría más que la lectura de archivos, y las estructuras de datos utilizadas para contener y modificar las imágenes. Por otra parte, después del ocultamiento, resultaría que la calidad de las imágenes portadoras a color es superior a la calidad de las imágenes portadoras en escalas grises, y esto se debe principalmente a la capacidad de ocultamiento que cuentan las primeras.

Pregunta 8

Si bien se podría ampliar las matrices de 2x2, el estudio realizado es sobre este tamaño y también lo son las formas de esconder el secreto. Si mantenemos la idea de utilizar el bloque X para evaluar el polinomio y esconder los bits de dicha evaluación en el resto de los píxeles, cambiar la matriz de tamaño cambiaría la forma en la que ocultamos los bits de la evaluación. Esto se debe a que la cantidad de bits a ocultar permanece constante pero la cantidad de bytes disponibles para ocultar se aumentaría.

Sería necesario evaluar también si tomando bloques de mayor tamaño efectivamente podemos ocultar la totalidad del secreto. Si tomamos matrices muy grandes, es posible que nos quedemos sin lugar en la portadora antes de ocultar todo el secreto.

Pregunta 9

Facilidad de implementación

En cuanto al algoritmo, la complejidad de implementarlo provino principalmente de los problemas previamente mencionados en el documento y la complejidad de debuggear problemas ya que es un proceso de dos partes donde ambas pueden fallar.

En particular, el problema de notación fue resuelto con una implementación de ejemplo de shamir para verificar que los conceptos estaban bien aplicados a baja escala y que era posible recuperar el secreto si este era pequeño.

Por otro lado, tuvimos un problema que ocurría al momento de encriptar la imagen cuando 2 índices coinciden, lo que producía que únicamente un grupo de píxeles de la imagen resultante terminaran incorrectos. Este problema fue más difícil de ubicar, pero analizando los píxeles problemáticos detalladamente y observando como este desplazamiento ocurría pudimos resolverlo.

Posibilidad de extensión o modificación

Durante la implementación del trabajo práctico se tomaron en cuenta las buenas prácticas, como la modularización, lo que permite que el código sea modificado o extendido con facilidad. Además, el código se encuentra comentado para la fácil interpretación a futuro proveyendo maintainability.

En concreto, se pueden las modificaciones/extensiones que se podrían aplicar al algoritmo es encripción y desencripción paralelizada, matrices de diferentes tamaños como se menciona aquí, para reducir el efecto sobre los pixeles de las imágenes portadoras, o una versión con portadoras multicolor u en otro formato. Otra mejora posible es encontrar un valor estimado en caso de que se deba descartar un pixel por que no coincida el bit de paridad.

Pregunta 10

Las situaciones para implementar este tipo de algoritmo son aquellas donde el secreto debe ser revelado únicamente cuando un número específico de claves se encuentra presente. Existen muchas situaciones donde queremos que el secreto o recurso pueda obtenerse dadas una cantidad determinada de claves, la mayoría de ellos en ambientes cooperativos. Un caso muy claro es, por ejemplo, el lanzamiento de un misil, que necesita la confirmación de dos oficiales (dado por sus llaves).