Laboratorio 1 – Criptoanálisis de Criptosistemas Básicos

1. Introducción

El objetivo de este laboratorio es que los estudiantes se familiaricen con el uso de técnicas básicas de croptoanálisis. El estudiante deberá programar las herramientas que le permitan efectuar el "ataque" a los criptosistemas de Shift, Sustitucion, Afin y Vigenere. El objetivo del atacante será a partir de los textos cifrados averiguar las claves y los textos en claro.

2. Ambiente de Trabajo

En este laboratorio el estudiante deberá programar las herramientas en el lenguaje C. Se proporciona la herramienta "cripto" que le permite encriptar y desencriptar los textos en los criptosistemas básicos.

3. Especificación de la herramienta "cripto"

Se define el mapeo de los símbolos del alfabeto español a Z_{27} de la siguiente forma: $a \to 0, ..., z \to 25$, ñ $\to 26$.

El modo de uso de la herramienta es: *cripto tipo direccion entrada clave* donde:

- *tipo*
    1. shift
    2. sustitucion
    3. afin
    4. vigenere
    5. trasposicion
- *direccion*
    - e = encripta
    - d = desencripta
- *entrada* archivo con el texto a procesar
- *clave* depende del tipo:
    1. K, siendo K la clave / Y = X + K
    2. archivo con una permutación de: *a,...,z,ñ*
    3. A B, siendo (A,B) la clave / Y = A.X + B
    4. archivo con un texto con símbolos pertenecientes a *a,...,z,ñ*
    5. archivo con M seguido de una permutación de *0,...,M-1*, siendo M el largo de la permutación.

Observaciones:

- El texto a procesar se lee desde el archivo entrada. Empieza en el primer carácter del archivo y termina en el primer carácter que no pertenezca al alfabeto. Este último no se incluye.

- El texto procesado se envía a la salida estándar. Si desea almacenar el resultado en un archivo de texto, debe redireccionar el flujo de salida utilizando el símbolo mayor.
- El parámetro tipo establece el criptosistema a utilizar.
- El parámetro dirección establece si encripta o desencripta.
- La clave a utilizar depende del tipo.

Ejemplos de uso:

```
./cripto 1 e texto_plano.txt 3 > cifradoShift.txt
./cripto 2 e texto_plano.txt keySustitucion > cifradoSustitucion.txt
./cripto 3 e texto_plano.txt 7 8 > cifradoAfin.txt
./cripto 4 e texto_plano.txt keyVigenere.txt > cifradoVigenere.txt
./cripto 5 e texto_plano.txt keyTrasposicion.txt > \
                  cifradoTrasposicion.txt
```

4. Tarea 1: Programación de las herramientas de Criptoanálisis

El objetivo de este tarea es programar las distintas herramientas que permitirán efectuar el ataque de los desafíos cifrados. En la carpeta *Tarea 1* encontrará dos subcarpetas. En la carpeta *Codigo C* estarán los archivos fuentes que el estudiante deberá completar para el desarrollo de cada herramienta. Estos archivos pretenden servir de guía para facilitar el desarrollo. El estudiante es libre de modificar y ampliar dichos fuentes como prefiera. El objetivo de esta tarea es que implemente cada herramienta con la información pedida. En la carpeta *Ejemplos de ejecucion* encontrará para cada herramienta un ejemplo de entrada y salida con la información que se pretende mostrar en cada caso. El formato de salida no tiene porque ser estrictamente idéntico, alcanza con brindar **al menos** los mismos datos que se muestran. Si el estudiante cree que debe mostrar más información de la pedida, puede hacerlo, dado que el objetivo es utilizar dichas herramientas para el criptoanálisis.

4.1. frecuencia entrada.txt

Esta herramienta cuenta las ocurrencias de los símbolos, digramas y trigramas en el texto e imprime en la salida estándar.

Ejemplo de salida:

```
Simbolo    Ocurrencias
b      1
c      1
d      1
...
Simbolo    Frecuencia
k   0.203390
t   0.118644
f   0.101695
...
Digrama    Ocurrencias
bo       1
co       1
dq       1
...
Digrama    Frecuencia
kf    0.051724
kq    0.051724
tk    0.051724
...
Trigrama   Ocurrencias
bol      1
cor      1
dqf      1
...
Trigrama   Frecuencia
fgt   0.035088
gtl   0.035088
...
```

4.2. coincidencia entrada.txt

La herramienta calcula el índice de coincidencia de una cadena de caracteres y lo imprime en la salida estándar. El índice de coincidencia del string x de n símbolos del alfabeto: $x = x_1x_2...x_n$ nos indica la probabilidad de que agarremos dos elementos de x y estos sean iguales.

4.3. coincidenciaBloques entrada.txt largoClave

Esta herramienta sirve para saber si, dado un texto cifrado por Vigenere, el supuesto largo de la clave es un buen indicio. Se basa en que el índice de coincidencia del idioma español es aproximadamente: $0{,}073$ y el de un texto aleatorio es $0{,}038$. Entonces, si se parte el texto en M bloques, y el supuesto largo de la clave es "correcto" el índice de coincidencia de cada bloque se aproxima al del idioma español. De lo contrario, estará próximo al del texto aleatorio.

La herramienta recibe un texto y la longitud de largo de clave e imprime el índice de coincidencia para cada bloque en la salida estándar.

Ejemplo de salida:

```
El indice de coincidencia del bloque 1 es: 0.091248
El indice de coincidencia del bloque 2 es: 0.070205
El indice de coincidencia del bloque 3 es: 0.071322
El indice de coincidencia del bloque 4 es: 0.070019
El indice de coincidencia del bloque 5 es: 0.078585
El indice de coincidencia del bloque 6 es: 0.065922
El indice de coincidencia del bloque 7 es: 0.065425
```

4.4. kasiski entrada.txt

El test de Kasiski se basa en la observación de que porciones idénticas de texto plano, cifradas por Vigenere con las mismas porciones de la clave, producen porciones idénticas de texto cifrado.
La distancia entre repeticiones de texto cifrado no sólo nos informa de la distancia entre repeticiones correspondientes de texto plano, sino que nos indica que la clave en esas posiciones se encontraba en fase, esto es, se encontraban en la misma posición relativa al comienzo de la clave. Esta distancia es necesariamente, por tanto, múltiplo de la longitud de la clave, ya que de otra forma, no encontraríamos la clave en la misma posición relativa.
La herramienta recibe un texto e imprime la longitud candidata de la clave en la salida estándar. Además imprime cual es el trigrama que ocurre más veces y las distancias relativas entre la primer ocurrencia del trigrama y las restantes. El largo candidato de la clave es el múltiplo común divisor entre estas distancias.

Notar que en este test solo se pide analizar y mostrar el **primer** trigrama que ocurre más veces. Por ejemplo, en el siguiente ejemplo, no aporta mucha información el test.

Ejemplo de salida:

```
El trigrama mas repetido es: mbi y ocurre 4 veces
Distancia de la nro ocurr 0 es 174
Distancia de la nro ocurr 1 es 223
Distancia de la nro ocurr 2 es 258
El largo de la clave es: 1
```

Entonces, con la herramienta antes dicha y esta, se tiene dos formas de verificar el supuesto largo de la clave de un texto cifrado por Vigenere.

4.5. mutua cifrado.txt largoClave

Muestra para cada bloque, el índice de coincidencia mutua de cada letra del alfabeto e indica cual es la letra para la cual ocurre el máximo. El parámetro largoClave es el supuesto largo de la clave con que se cifró el texto.

En este ejemplo de salida (que se encuentra en la carpeta) se puede obervar como la clave candidata es *gabriel*.

```
Bloque 1
Simbolo Indice
0 0.030075
1 0.025719
...
Maximo en el simbolo: g

Bloque 2
Simbolo Indice
0 0.070945
1 0.039545
2 0.033357
...
Maximo en el simbolo: a

Bloque 3
Simbolo Indice
0 0.039010
1 0.071331
2 0.036088
...
Maximo en el simbolo: b
...
```

4.6. sustitucionClaveParcial entrada.txt clave.txt

Esta herramienta permite descifrar un texto por Sustitución con una clave parcialmente completa. Cuando un carácter de la clave no se conoce se pone un guión y en el texto descifrado el carácter correspondiente aparece entre paréntesis.

Por ejemplo, si se quiere probar una conjetura sobre la clave conociendo solamente algunas posiciones de la permutación, en el archivo de la clave se debe escribir:

```
t-own---p--j-------e-------
```

Ejemplo de salida

```
(t)l(v)umord(t)lci(t)lo(v)a(q)ia(t)mp(t)zadoad
....
```

5. Tarea 2: Criptoanálisis de los desafíos cifrados

El objetivo de esta tarea es utilizar las herramientas programadas en la parte anterior para efectuar el criptoanálisis de los cifrados de Shift, Sustitucion, Afin y Vigenere. En la carpeta *Tarea 2-Desafios Cifrados* encontrará los textos cifrados para los cuales deberá obtener la clave y el texto claro. Es importante que se explique y fundamente el análisis efectuado en cada texto cifrado. En el informe debe poner todos los pasos que efectúa junto con las salidas correspondientes de las herramientas que van dando soporte a las conjeturas que hace.

6. Evaluación

El estudiante en la *Tarea 1* debe realizar el desarrollo de las herramientas solicitadas. Además, debe entregar un archivo makefile que permita compilar todos los archivos fuentes entregados. En la *Tarea 2* debe entregar un informe en **formato pdf** describiendo lo que ha hecho y observado en cada criptoanálisis explicando los resultados. Es importante que se explique claramente cual fue el razonamiento seguido paso a paso y se haga referencia a cada par de texto claro y clave que formen pasos intermedios de las conjeturas. Además, es deseable que se incluyan no solamente los resultados finales sino que se ilustre el proceso de aprendizaje explicando los errores cometidos en etapas anteriores.