Ataques sobre métodos polialfabéticos.

Uso del método de Kasiski y el índice de coincidencia

- Antonio Miguel Pozo Cámara
- Javier Bolivar Valverde

22 de noviembre de 2015

Introducción

Programa en python para realizar ataques sobre métodos polialfabéticos haciendo uso del método de Kasiski y el Índice de Coincidencia.

El método de Kasiski fue introducido en 1863 por el militar Friedrich W. Kasiski. Este método analiza repeticiones en el texto cifrado para determinar el periodo que se usó para cifrarlo y así conseguir la longitud de la clave. Análogamente el índice de coincidencia es un método desarrollado por William Friendman, en 1920. La idea se fundamenta en analizar la variación de las frecuencias relativas de cada letra, respecto a una distribución uniforme. En un texto cifrado, no se cuenta con información suficiente para hallar tal variación. Sin embargo, se puede obtener por medio del IC. Al hacerlo, será posible aproximar el periodo de la clave.

Procedimiento

Buscamos en el texto cadenas de caracteres de 3 o más letras que se repitan en el texto. Una vez hecho esto, buscamos la separación que hay entre las dichas cadenas repetidas. El máximo común divisor de estas longitudes entre separaciones nos dará la posible longitud de la clave pero no las letras que la forman.

Para encontrar la clave, dividimos el texto en “n” cadenas, siendo “n” la longitud de la clave anteriormente descrita. Como cada una de esas cadenas será el resultado de una cifra monoalfabética con un desplazamiento dado por la letra clave, contabilizamos en una tabla las veces que aparecen las letras del alfabeto en cada una de las cadenas y anotamos las de mayor ocurrencia. Estas letras deberían corresponder a la clave buscada.

El Índice de Coincidencia de un texto dado se define como la probabilidad de que, al escoger al azar dos letras cualesquiera del texto, éstas sean idénticas.

Según el libro 'The Codebreakers' de David Kahn, el Índice de Coincidencia del castellano es aproximadamente 0,07750.

Lo que nosotros hacemos es comparar el índice de coincidencia de cada cadena n (todas sus letras están cifradas con la misma clave) con el índice de coincidencia del idioma en cuestión, en nuestro caso, el castellano. Si el índice de coincidencia que nos sale se aproxima al del castellano, es bastante probable que la letra que nos ha salido sea la correcta para la clave.

Explicación del código

El programa se encuentra en la siguiente url:

https://github.com/javibolibic/SPSI/blob/master/kasiskiAttack.py

findusbs(text, l):

"text" es el texto en el que vamos a buscar las subcadenas.

"l" es el tamaño de la cadena a buscar.

Modifica la variable global "lista_ocurrencias", que almacena los caracteres de separación entre subcadenas iguales.

filter(text):

Excluye los caracteres del texto pasado como argumento que no estén en el alfabeto "alphabet".

imprime(cadena):

Imprime la cadena pasada como argumento.

ktest(text):

Función principal del programa. Desarrolla los conceptos a tratar en el trabajo: Aplica el método de Kasiski y hace uso del Índice de Coincidencia.

- Filtramos el texto pasado como argumento [línea 53].
- Buscamos la distancia entre cadenas repetidas [líneas 59-60]
- Calculamos el máximo común divisor [línea 62]
- Dividir el texto en "x" cadenas, siendo "x" el máximo común divisor calculado en el paso anterior [líneas 72-83].
- Contar las letras que más se repiten en cada una de las subcadenascredas en el paso anterior [líneas 95-138].
- Calcular el índice de coincidencia de cada cadena [líneas 143-150]
- El programa llama a esta función principal, que imprimirá por pantalla todo el contenido que se muestra en la captura 1.

Ejemplo de ejecución

```
MacBook-Pro-de-Antonio:SPSI repo trabajo Apozo$ python kasiskiAttack.py
Introducir el texto cifrado:
PPQCAXQVEKGYBNKMAZUYBNGBALJONITSZMJYIMVRAGVOHTVRAUCTKSGDDWUOXITLAZUVAVVRAZCVKBQPIWPOU
YBN          8
AZU         48
VRA          8
VRA         24
Posible longitud de la clave = 8

cadena[0]== PEAAZAADAA
cadena[1]== PKZLMGUWZZ
cadena[2]== QGUJJVCUUC
cadena[3]== CYYOYOTOVV
cadena[4]== ABBNIHKXAK
cadena[5]== XNNIMTSIVB
cadena[6]== QKGTVVGTVQ
cadena[7]== VMBSRRDLRP

1º caracter mas repetido de cadena[0] = A -> 6 veces
2º caracter mas repetido de cadena[0] = D -> 1 veces
3º caracter mas repetido de cadena[0] = E -> 1 veces
Indice de coincidencia de cadena[0] 0.333 (0.333333333333)

1º caracter mas repetido de cadena[1] = Z -> 3 veces
2º caracter mas repetido de cadena[1] = K -> 1 veces
3º caracter mas repetido de cadena[1] = L -> 1 veces
Indice de coincidencia de cadena[1] 0.067 (0.0666666666667)

1º caracter mas repetido de cadena[2] = U -> 3 veces
2º caracter mas repetido de cadena[2] = J -> 2 veces
3º caracter mas repetido de cadena[2] = G -> 1 veces
Indice de coincidencia de cadena[2] 0.111 (0.111111111111)

1º caracter mas repetido de cadena[3] = O -> 3 veces
2º caracter mas repetido de cadena[3] = Y -> 3 veces
3º caracter mas repetido de cadena[3] = V -> 2 veces
Indice de coincidencia de cadena[3] 0.156 (0.155555555556)

1º caracter mas repetido de cadena[4] = A -> 2 veces
2º caracter mas repetido de cadena[4] = B -> 2 veces
3º caracter mas repetido de cadena[4] = K -> 2 veces
Indice de coincidencia de cadena[4] 0.067 (0.0666666666667)

1º caracter mas repetido de cadena[5] = I -> 2 veces
2º caracter mas repetido de cadena[5] = N -> 2 veces
3º caracter mas repetido de cadena[5] = M -> 1 veces
Indice de coincidencia de cadena[5] 0.044 (0.0444444444444)

1º caracter mas repetido de cadena[6] = V -> 3 veces
2º caracter mas repetido de cadena[6] = Q -> 2 veces
3º caracter mas repetido de cadena[6] = T -> 2 veces
Indice de coincidencia de cadena[6] 0.133 (0.133333333333)

1º caracter mas repetido de cadena[7] = R -> 3 veces
2º caracter mas repetido de cadena[7] = D -> 1 veces
3º caracter mas repetido de cadena[7] = L -> 1 veces
Indice de coincidencia de cadena[7] 0.067 (0.0666666666667)

CLAVE 1 = AZUOAIVR
CLAVE 2 = DKJYBNQD
CLAVE 3 = ELGVKMTL
```

Referencias

- http://www.robota.net/index.rsws?seccion=5&submenu=1&articulo=744
- http://mikelgarcialarragan.blogspot.com.es/2015/03/criptografia-i.html
- https://es.wikipedia.org/wiki/M%C3%A9todo_Kasiski
- https://www.youtube.com/watch?v=A7p2ydEPg1k