Simmulated Anneling aplicado a Criptografía

Basados en los estudios de Carter and Magoc [1] y de Dimovski and Gligoroski [2],en este proyecto se busca estudiar la efectividad del método Simulated Anneling a la hora de buscar la clave con la que un texto fue cifrado. En particular, nos centraremos en 2 cifrados, el de substitucion monoalfabetica, que consiste en una biyeccion entre el alfabeto y si mismo, y en el cifrado de transposición, el cual consiste en separar el texto en bloques de tamaño P, y aplicar a cada estos una permutacion de sus elementos.

Para el análisis del primer tipo de cifrado, en el cual la clave consiste en la biyeccion hecha, se aplico simulated anneling, utilizando como energía, una variación de la formula fitness, encontrada en [1], la cual consiste en:

$$\text{fitness}(texto) = 1 - \left(\frac{\sum_i |F_{\text{ingles}}(i) - F_{texto}(i)| + \sum_{i,j} |F_{\text{ingles}}(i,j) - F_{texto}(i,j)|}{4} \right)^8$$

Esta función es una medida de cuanto se acerca la estadística del texto a la estadística del idioma que se está usando para el mismo, en este caso inglés. La variacion utilizada para esto, es la expresión: $1 - \text{fitness}$. En este caso se hace un análisis estadístico sobre la frecuencia relativa de aparición de un digrama en el texto con respecto al idioma, que se obtuvo desde la recopilación de información hecha por [3].

Mientras que para el otro tipo de cifrado, en el cual la clave es la permutación hecha a los bloques, se replico el código utilizado en el paper de Dimowski y Gligoroski [2], y se comparo la eficacia del mismo, para 2 funciones de energía, las cuales fueron:

$$\sum_{i,j} |F_{\text{ingles}}(i,j) - F_{texto}(i,j)| \text{ y } \sum_{i,j} \frac{|F_{\text{ingles}}(i,j) - F_{texto}(i,j)|}{F_{\text{ingles}}(i,j)}$$

Además, se comparo la eficacia al utilizar la frecuencia relativa de los digramas obtenidas de [3], contra la frecuencia relativa obtenida desde un texto de contexto similar al mensaje encriptado.

Conclusiones

Entre las conclusiones que se pudieron obtener, en el cifrado transposición, vemos que en claves de largos hasta 6, se logro encontrar parámetros que permiten obtener una efectividad cercana al 100 % del algoritmo, acercándose así, a los resultados del paper, pero para claves mayores no se logra alcanzar tal nivel de efectividad, distanciándose asi, de los resultados del paper, y sobre esto, se ahondara en la presentación

Por su parte se logra romper de forma total el cifrado monoalfabético sin requerir necesariamente usar métodos genéticos en el mismo. Observando ciertas invarianzas en el mismo, se logra un método de optimizar el cálculo de energía de un paso a otro y con ello realizar en menos de un minuto, decifrados de textos completos.

Referencias

[1] Brian A. Carter and Tanja Magoc. Classical ciphers and cryptanalysis. 2007.

[2] Aleksandar S. Dimovski and Danilo Gligoroski. Attacks on the transposition ciphers using optimization heuristics. 2003.

[3] Peter Norvig. English bigram and letter pair frequencies from the google corpus data in json format, Aug 2015.