Notas mínimas sobre Criptografía

Fernando Galindo Soria

9 de Septiembre del 2016

Método Base

Desde hace miles de años se han desarrollado técnicas para ocultar información, una de las mas simples consiste en tomar las letra de un texto, verlas como números y sumarles una constante a todos esos números y escribir el resultado.

Por ejemplo: podemos asociarle a cada letra números a partir del cero, entonces el *espacio en blanco* vale *0*, la *a* vale *1*, la *b* vale *2*, la *c* vale *3* y así sucesivamente hasta la *z*.
Entonces si tenemos el texto:
el perro ladra

se vería como la lista de números
5 12 0 16 5 18 18 15 0 12 1 4 18 1

Para criptografiar este texto le sumamos un numero, por ejemplo 4, quedando:
9 16 4 20 9 22 22 19 4 16 5 8 22 5

Quedando el texto:
ipdtivvsdpehve

Que como se ve no se sabe que dice.

Esta técnica es muy antigua, y es redescubierta cotidianamente por mucha gente, en particular yo la descubrí cuando era niño, pero dado que la uso Cesar para encriptar textos, se le conoce como el método Cesar.

Ahora bien es muy fácil decriptar los texto encriptados con esta técnica, en general el método de decriptación es muy conocido y consiste básicamente en tomar el texto encriptado y contar cuantas veces aparece cada letra en el texto, se ordenan las letras según su frecuencia de aparición y la que aparece mas veces se supone que corresponde a la letra mas frecuente en el idioma (por ejemplo el español), la segunda en aparecer corresponde a la segunda letra del idioma y así sucesivamente (es importante comentar que, entre mas grande es la muestra, mas probable es que correspondan las frecuencias).

Actualmente se tienen las tablas de frecuencias de muchísimos idiomas, como por ejemplo español, ingles, alemán, ruso, etc., y si no se tiene la tabla es muy fácil obtenerla, se toma un texto muy grande escrito en el idioma problema, se cuenta cuantas veces aparece cada letra y se ordenan por frecuencia de aparición.
Esta técnica de tener las tablas de frecuencias de los idiomas, sirve para otras cosas, por ejemplo para detectar el idioma en el que se encuentra el texto desconocido y poderlo traducir.

El problema de tomar un texto y lograr que no se entienda se conoce como criptografia, el problema de descubrir lo que dice el texto criptografiado se conoce como criptoanalisis y el área que engloba la criptografia y el criptoanalisis es la criptologia.

Metodo Polialfabetico
Dado que es tan fácil decriptar el texto se han desarrollado mucha otras técnicas basadas en esta idea, la siguiente se conoce como método polialfabetico y consiste básicamente en que a la primera letra se le suma un numero, a la segunda otro numero y así sucesivamente a la tercera y a las que siguen, De los métodos polialfabeticos mas usados se tiene el que usa cuatro alfabetos alternos (o números, para el caso es lo mismo), la primera letra se sustituye por su posición en el primer alfabeto, para la segunda letra se usa el segundo alfabeto, para la tercera el tercer alfabeto y para la cuarta se usa el cuarto alfabeto, para la quinta se usa nuevamente el primer alfabeto y así sucesivamente.

Yo les recomiendo que si quieren ocultar información usen el primer método, con eso tiene para que un lector casual no los pueda entender, o por mucho el polialfabetico de cuatro números o alfabetos, ya que la técnica es mas segura que la descrita inicialmente.

Teorema de Shannon
Pero aunque aumenten los alfabetos, si el texto es suficientemente grande se empiezan a encontrar patrones que permiten decriptarlo. En1949 Claude Shannon, demostró un teorema que básicamente dice que: si la cantidad de números (o alfabetos) que se usan para encriptar la información es al menos del tamaño del texto a encriptar entonces el código es indescifrable, la única condición que se pone es que la secuencia de números que se usa para encriptar sea aleatoria y no se pueda predecir su comportamiento (que no tengan un patrón de comportamiento predecible), o sea que se tiene un método polialfabetico, donde la cantidad de números que se usa para encriptar es al menos del tamaño del texto a encriptar, por lo que cada letra se encripta con un numero diferente.

El único problema es que es difícil recordar tantos números diferentes, por lo que la solución que se encontró fue la de usar un generador de números pseudoaleatorios.
Un generador de números aleatorios si esta bien hecho (existen libros completos sobre el tema) genera una secuencia de números que tarda mucho en repetirse, por lo que puede generar una secuencia de números suficientemente larga para encriptar textos, por otro lado tiene que ser pseudoaleatios, porque eso significa que el usuario puede lograr repetir la lista de números aleatorios (si no, no podría decriptar el texto, porque no sabría que números se usaron para la encriptación).

Este tipo de generadores necesitan un numero del cual partir, si les doy un numero inicial n, la maquina generara una secuencia de números aleatorios, ahora si le doy otro numero m la maquina genera otra secuencia, pero si vuelvo a dar como inicio n me da la secuencia original, a este numero que se da para generar la secuencia de números aleatorios se le conoce como semilla o clave o password.
Si se quiere una secuencia de números aleatorios y no nos interesa repetir esa secuencia, se le da la semilla en forma aleatoria (por ejemplo toma la hora del reloj de la maquina).