Cryptography, Sem: 2020-2, 3CM9, Laboratorio 1 (2da parte), 10/Febrero/2020

Laboratorio 1 (2da parte): Classical Cryptography

Miranda Mojica Erick
Torres Olivera Karla Paola

Escuela Superior de Cómputo
Instituto Politécnico Nacional, México
erimimo2@gmail.com
kpto997@gmail.com

Resumen: El presente trabajo consiste en la descripción e implementación del método que se llevo a cabo para poder decifrar los mensajes proporcionados por nuestros compañeros de clase.
Palabras Clave: Vigenère Cipher, Affine Cipher, criptoanálisis.

1 Introducción

La ciencia de la criptología se puede desglosar de dos maneras diferentes. Una forma de verla es que tiene que ver tanto con la creación de sistemas criptológicos (criptografía), así como con técnicas para descubrir el texto claro (criptoanálisis).

La criptografía es la ciencia que se encarga de estudiar las distintas técnicas empleadas para transformar ("cifrar") la información y hacerla irreconocible a todos aquellos usuarios que no tengan autorización de acceder a esta, de modo que solo los legítimos propietarios puedan recuperar ("decifrar") la información original. Mediante la criptografía es posible garantizar la confidencialidad, la integridad y la autenticidad de los mensajes y documentos.

El criptoanális se ocupa del estudio de las distintas técnicas y métodos que permiten "romper" los algoritmos de cifrado. En la práctica, el criptoanálisis se suele llevar a cabo estudiando distintos pares de "mensaje de texto original/ mensaje cifrado (criptograma)" generados utilizando la misma clave [1].

Tomando en cuenta lo anterior en esta práctica nosotros nos convertiremos en criptoanalistas con la finalidad de encontrar el texto en claro de los mensajes cifrados proporcionados. Para esto, se tomarán en cuenta únicamente los algoritmos de Vigenère y Affine vistos en la primera parte del laboratorio 1.

2 Conceptos Básicos

2.1 Criptografía clásica

La criptografía clásica consta de problemas y herramientas que incluyen cifrado, distribución de llaves, funciones unidireccionales, entre otras. En esta, existen dos tipos de sistemas criptográficos: los cifrados por sustitución y los cifrados por transposición. En estos últimos, las letras del texto claro se codifican (reordenan) sistemáticamente para que el texto se vuelva intangible. Por ejemplo, la palabra "software" puede codificarse para leerse como "fosawter", es decir, las letras se intercambian entre sí. Mientras que en los sistemas de sustitución las letras en el texto claro se reemplazan sistemáticamente por otras pertenecientes al alfabeto definido previamente por las personas que se desean comunicar [2].

2.1.1 Cifrados por sustitución

Como se mencionó anteriormente en los cifrados clásicos por sustitución las letras del texto claro se reemplazan por otras letras o números o símbolos. En [2] podemos encontrar algunos ejemplos de cifrados pertenecientes a este tipo, los cuales son:

1. Caesar Cipher
2. Mono-Alphabetic Cipher
3. Hill Cipher
4. Play-Fair Cipher
5. Vigenère Cipher
6. One-Time Pad

2.2 Criptoanálisis

El criptoanálisis consiste básicamente en la operación inversa de la criptografía, es decir, trata de convertir los criptogramas en textos en claro sin autorización.

2.2.1 Tipos de ataques

Los métodos de ataque pueden clasificarse en varios tipos generales según la información que conoce y desconoce el criptoanalista.

1. Ataques basados solo en el texto cifrado: solo se conoce el texto cifrado, aunque a menudo también se conoce el idioma del texto plano y el tipo de cifrado. El objetivo del criptoanalista es encontrar el texto plano y la clave. Este es el tipo de ataque más difícil. A veces, solo se tiene una serie de bits para trabajar [3].

2. Ataques basados en texto claro conocido: el criptoanalista dispone de varios textos cifrados y de los textos en claro de partida. Su objetivo es tratar de encontrar la llave utilizada para poder descifrar nuevos textos cifrados. Esta situación en la práctica es más frecuente de lo que se pudiera pensar, debido a que muchos mensajes que se van a cifrar pueden contener palabras o símbolos de inicio y finalización conocidos: cabeceras de mensajes de correo electrónico, determinados formatos de documentos o cabeceras de paquetes de datos en un determinado protocolo. Además, en algunas situaciones se cifran cadenas de texto predecibles, como ocurría en los ataques contra los criptosistemas alemanes y japoneses durante la Segunda Guerra Mundial [1].

3. Ataques basados en texto claro seleccionado: el criptoanalista no solo dispone de varios textos cifrados y textos en claro, si no que además ha podido seleccionar los que serán cifrados (aquellos que le podrían facilitar más información sobre las diversas transformaciones realizadas por el sistema criptográfico). Esta hazaña se puede lograr engañando al operador de la máquina de cifrado para que cifre un mensaje dado [1].

4. Ataques adaptativos basados en texto claro conocido: en este, además de poder seleccionar varios textos en claro y obtener sus correspondientes textos cifrados, el criptoanalista puede modificar su elección de los mensajes a cifrar teniendo en cuenta los resultados generados por cifrados previos. De este modo, el criptoanalista puede ir seleccionando bloques pequeños de texto en claro en etapas sucesivas para obtener información más precisa. Aunque el texto cifrado obtenido puede no ser un mensaje significativo, puede ayudar a encontrar la llave [1].

2.2.2 Técnicas de criptoanálisis

Algunas de las técnicas más utilizadas en las actividades de criptoanálisis presentadas en [1] son:

1. Criptoanálisis diferencial: trata de encontrar correlaciones entre el texto claro y el cifrado obtenido, partiendo del conocimiento de la existencia de ciertas diferencias entre varios textos claros que se han introducido en el sistema.

2. Criptoanálisis lineal: trata de encontrar correlaciones entre la clave, el texto claro y el cifrado basado en un cifrado en bloque.

3. Técnicas de análisis estadístico de frecuencias: los primeros métodos para "romper" los cifrados de sustitución polialfabética se basaban en el análisis estadístico de frecuencias, partiendo del estudio de las cadenas de texto repetidas en el mensaje cifrado para determinar la longitud de la clave y la correspondencia entre los caracteres cifrados y sin cifrar.

4. Interceptación de llaves: ataques de intermediación, mediante los cuales se pueden interceptar directamente las llaves sin despertar sospechas de los usuarios del sistema criptográfico y sin que sea necesario estudiar los textos cifrados.

2.3 Vigenère Cipher

El sistema de cifrado de Vigenère (en honor al criptógrafo fránces del mismo nombre) es un sistema polialfabético[1] o de sustitución múltiple. Este tipo de criptosistemas aparecieron para sustituir a los monoalfabéticos o de sustitución simple, basados en el Caesar, debido a que presentaban ciertas debilidades frente al ataque de los criptoanalistas en relación a la **frecuencia de aparición de elementos del alfabeto**.

La llave del cifrado Vigenère es una palabra de k letras, tal que, $k \geq 1$, del alfabeto $\mathbb{Z}_{26}$ ={A, B, C, ..., Z}; esta palabra es un elemento del producto cartesiano $\mathbb{Z}_{26} \cdot \mathbb{Z}_{26} \cdot \mathbb{Z}_{26}$ (k veces).

De esta forma, el mensaje a cifrar en texto claro se descompone en bloques de k elementos - letras - y aplica sucesivamente la llave a cada uno de estos.

Este método se consideraba invulnerable hasta que en el siglo XIX consiguieron decifrar algunos mensajes codificados con este sistema, mediante el estudio de la repetición de bloques de letras.Una mejora de este algoritmo fue introducida por el sistema de Vernam el cual, utiliza una llave aleatoria cuya longitud es igual que a la del mensaje [4].

Tomando en cuenta lo anterior una definición más formal de este algoritmo de cifrado se muestra en la Tabla 1.

Definición Vigenère Cipher
Sea
$P = p_0, p_1, ..., p_{n-1}$ *secuencia de letras del texto claro,*
$K = k_0, k_1, ..., k_{m-1}$ *una llave que consiste en una secuencia de letras tal que,* $m < n$

Cifrado: $C_i = (p_i + k_{i \ mod \ m}) \bmod 26$
Decifrado: $p_i = (C_i - k_{i \ mod \ m}) \bmod 26$

Tabla 1: Definición Vigenère Cipher

Con la finalidad de entender mejor cómo funciona, se realizó un ejercicio el cual se puede observar en la Tabla 2 especificando los pasos que se llevaron a cabo. Es importante recordar que el número asignado a cada una de las letras depende de su localización en el alfabeto (A-Z).

key:	*deceptive*
plaintext:	*wearediscovered*
cipherttext:	ZICVTWQNGRZGVTW

[1]Son aquellos que cifran letras en función de su posición en el texto claro.

M:	w	e	a	r	e	d	i	s	c	o	v	e	r	e	d
	22	4	0	17	4	3	8	18	2	14	21	4	17	4	3
K:	3	4	2	4	15	19	8	21	4	3	4	2	4	15	19
M + K:	25	8	2	21	19	22	16	39	6	17	25	6	21	19	22
(M + K) mod 26:	25	8	2	21	19	22	16	13	6	17	25	6	21	19	22
C:	Z	I	C	V	T	W	Q	N	G	R	Z	G	V	T	W

Tabla 2: Ejemplo Vigenère Cipher

2.4 Affine Cipher

Affine Cipher no comparte la propiedad con Shift Cipher que si se conoce la correspondencia entre una letra del texto claro y una del cifrado, el resto de las correspondencias siguen.

Este algoritmo de cifrado a diferencia de otros, cuenta con una llave compuesta por dos números. Cifra multiplicando el mensaje original por una parte de esta, seguido de la adición de la otra parte de la llave. La definición de Affine Cipher encontrada en [5] se muestra en la Tabla 3.

Definición Affine Cipher
Sea $x,\ y,\ a,\ b \in \mathbb{Z}_{26}$

Cifrado: $e_k(x) = y \equiv a \cdot x + b \bmod 26$
Decifrado: $d_k(y) = x \equiv a^{-1} \cdot (y - b) \bmod 26$

con llave: k=(a,b), con la restricción: gcd(a,26)=1

Tabla 3: Definición Affine Cipher

El decifrado se deriva fácilmente de la función de cifrado:

$$a \cdot x + b \equiv y \bmod 26$$
$$a \cdot x \equiv (y - b) \bmod 26$$
$$x \equiv a^{-1} \cdot (y - b) \bmod 26$$

La restricción $\gcd(a, 26) = 1$ proviene del hecho de que el parámetro a de la llave necesita tener inverso multiplicativo para el decifrado. Como se mencionó en la sección 2.2.1 para que esto se cumpla a y el módulo deben ser relativamente primos. Por lo tanto debe estar en el conjunto [5]:

$$a \in \{1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25\}$$

Con base en lo anterior se dará un ejemplo del funcionamiento de este algoritmo de cifrado, el cual se visualiza en la Tabla 4.

key:	$E_{(3,2)}(m_i)$
plaintext:	*coronavirus*
cipherttext:	ISBSPCNABKE

M:	C	O	R	O	N	A	V	I	R	U	S
	2	14	17	14	13	0	21	8	17	20	18
$a \cdot m_i + b$:	8	44	53	44	41	2	65	26	53	62	56
$(a \cdot m_i + b)$ mod 26:	8	18	1	18	15	2	13	0	1	10	4
C:	I	S	B	S	P	C	N	A	B	K	E

Tabla 4: Ejemplo Affine Cipher

Por otro lado, cuando se desee decifrar el mensaje se hará uso del algoritmo de Euclides extendido, explicado en la sección 2.2.2 para calcular el inverso de a. A su vez, este también nos ayuda a determinar si el valor a de una llave candidata es válido o no, es decir, si $\gcd(a, \mathbb{Z}_n) = 1$.

3 Experimentación y Resultados

3.1 Punto 1. Vigenère Cipher

Algunos puntos importantes a considerar antes de comenzar a explicar la forma en que se decifraron los mensajes son: se especifico que la llave del cifrado Vigenère debía tener una longitud de 5-15 caracteres. Por otro lado, mientras revisábamos los textos cifrados y los textos claros nos dimos cuenta que las mayúsculas y los signos de puntuación (, : ; “ . etc.) no estaban cifrados por lo que se diseño una función en python que nos ayudó a eliminarlos del texto cifrado, logrando que al momento de decifrarlo no se tomaran en consideración y se pudiera obtener el mensaje claro. Dicha función se muestra en el Listing 1.

```
import re
f = open ('ejemplo5.vig', encoding="utf8")
fd = open('Final.txt', 'w')
mensaje = f.read()
fd.write(re.sub('[\nA-Z.;,?   ]', '', mensaje))
f.close()
fd.close()
```

Listing 1: Función para eliminar caracteres

Dentro de los corchetes se pusieron los caracteres a eliminar. Más adelante se explicará que para dos mensajes se tomaron en consideración los números del 0-9, es decir, pertenecían al alfabeto mientras, que en los otros dos no.

3.1.1 Key 1

La técnica que se uso para decifrar el primer mensaje fue la de análisis estadístico de frecuencias (véase 2.2.2). Para esto vimos las cadenas que se repetían en el texto cifrado encontrándonos con que algunos $\_and\_$ pertenecientes al mensaje original se cifraban con la cadena $eg2mh$. Tomando esto en cuenta decidimos partir decifrando ese pequeño bloque.

Por otra parte, para determinar el alfabeto analizamos el mensaje cifrado observando que solo tomaba en cuenta los números y las letras minúsculas, por lo que, el alfabeto final se muestra en el Listing 2

```
abcdefghijklmnopqrstuvwxyz 1234567890
```

Listing 2: Primer alfabeto

Con estos datos pudimos hacer el decifrado. Los pasos a seguir fueron: buscar un número que sumado con m_i mod 37 nos diera como resultado las letras del texto cifrado. El resultado se muestra en la Tabla 5.

	a	n	d	
26	0	13	3	26
15	16	15	9	18
e	g	2	m	h
k:	p	g	p	j	s

Tabla 5: Posible clave

Sin embargo, al momento de decifrar el mensaje con esta llave únicamente salía correctamente una parte del mensaje original. Posteriormente, nos dimos cuenta que había exactamente 10 letras antes de una palabra descubierta exitosamente y debido a esto decidimos repetir el procedimiento con esas letras. El resultado final se muestra en la Tabla 6.

M:	e	n	o	c	i	d	e		i	s		a	n	d	
	4	13	14	2	8	3	4	26	8	18	26	0	13	3	26
	34	14	4	15	29	29	0	22	29	7	15	16	15	9	18
C:	b	1	s	r	a	6	e	l	a	z	e	g	2	m	h
K:	8	0	e	p	3	3	a	w	3	h	p	g	p	j	s

Tabla 6: Clave Final

Una vez teniendo la clave o llave final la ingresamos junto con el texto cifrado y el alfabeto al programa realizado en la parte 1 del laboratorio, obteniendo correctamente el mensaje claro como salida. En las Figuras 1 y 2 se muestra una parte de ambos textos.

Figura 1: Primer texto cifrado

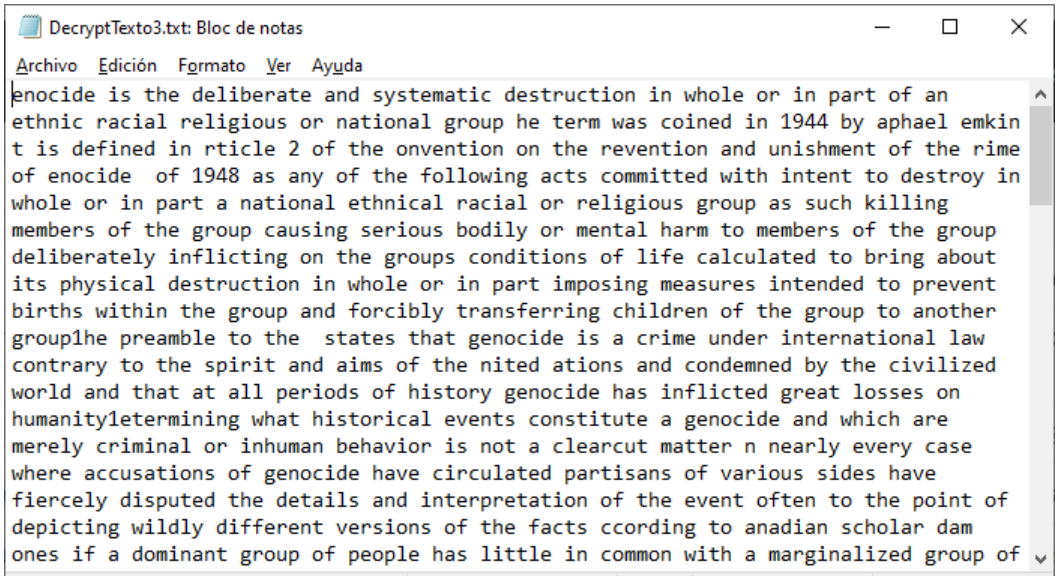

Figura 2: Primer texto decifrado

Haciendo uso de esta llave pudimos decifrar el segundo mensaje (veáse Figuras 3 y 4).

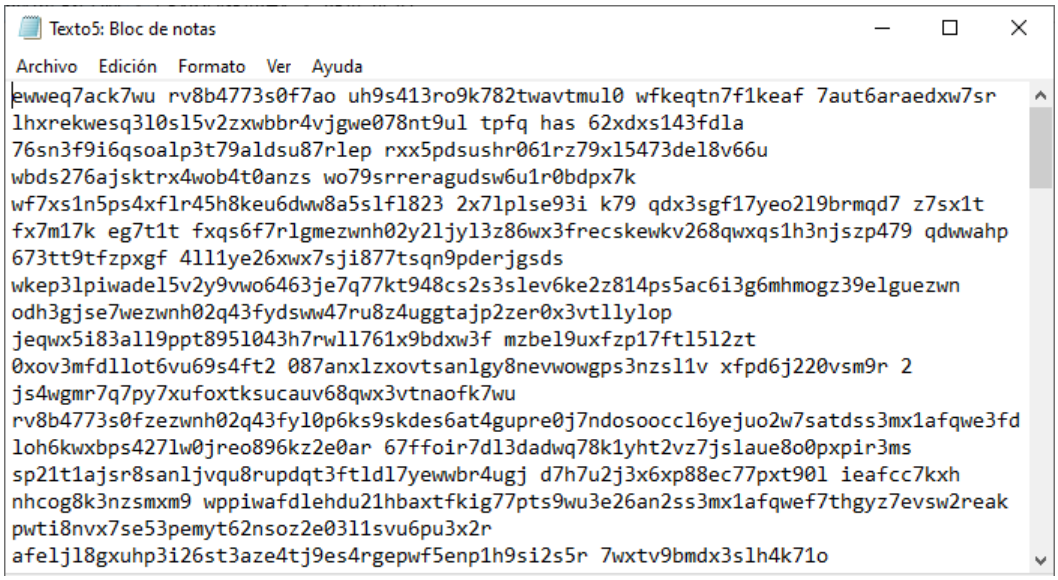

Figura 3: Segundo texto cifrado

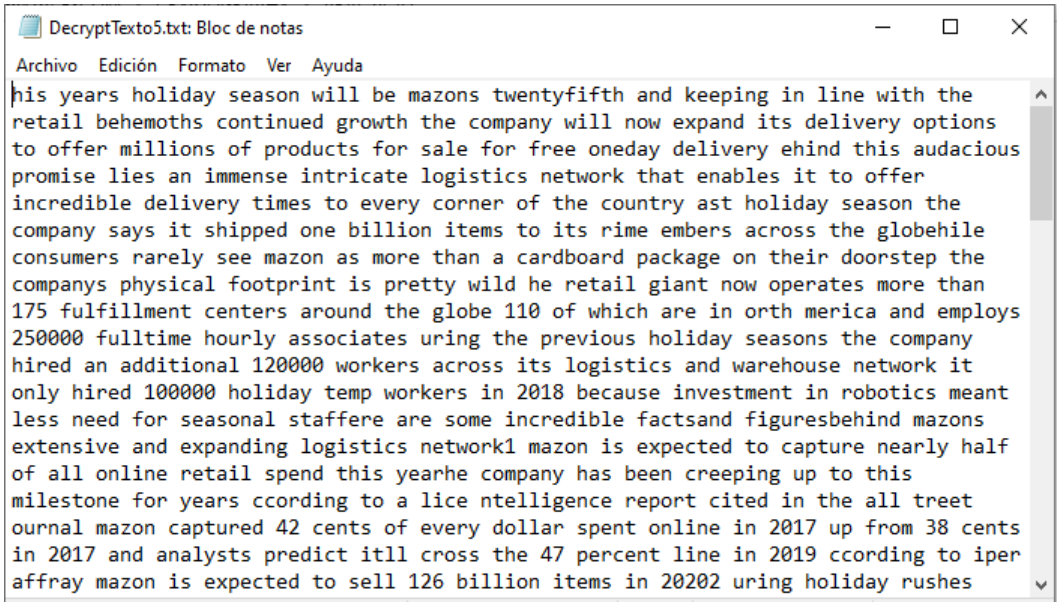

Figura 4: Segundo texto decifrado

3.1.2 Key 2

Para determinar la segunda llave perteneciente a los últimos dos textos cifrados decidimos empezar tomando en cuenta la longitud de esta (veáse 3.1). A partir de esto, tomamos los primeros 15 caracteres del texto cifrado y del texto claro proporcionados y realizamos el procedimiento visto con anterioridad. El resultado final se muestra en la Tabla 7.

M:	n		t	h	i	s		e	s	s	a	y		o	n
	13	26	19	7	8	18	26	4	18	8	0	24	26	14	13
	25	18	23	10	4	18	23	7	19	22	20	25	18	23	10
C:	l	r	p	r	m	j	w	l	k	n	u	w	r	k	x
K:	z	s	x	k	e	s	x	h	t	w	u	z	s	x	k

Tabla 7: Método para encontrar la llave

Podemos ver que a partir de la la letra y del mensaje se vuelven a repetir los caracteres, con lo cual, podemos concluir que la llave es: $zsxkesxhtwu$. Por otro lado, como se mencionó anteriormente en este alfabeto no fueron considerados los números por lo que el tamaño de este se redujo a 27 (veáse Listing 3).

```
1 abcdefghijklmnopqrstuvwxyz
```

Listing 3: Segundo alfabeto

El texto cifrado así como el texto claro se muestran en las Figuras 5 y 6 respectivamente.

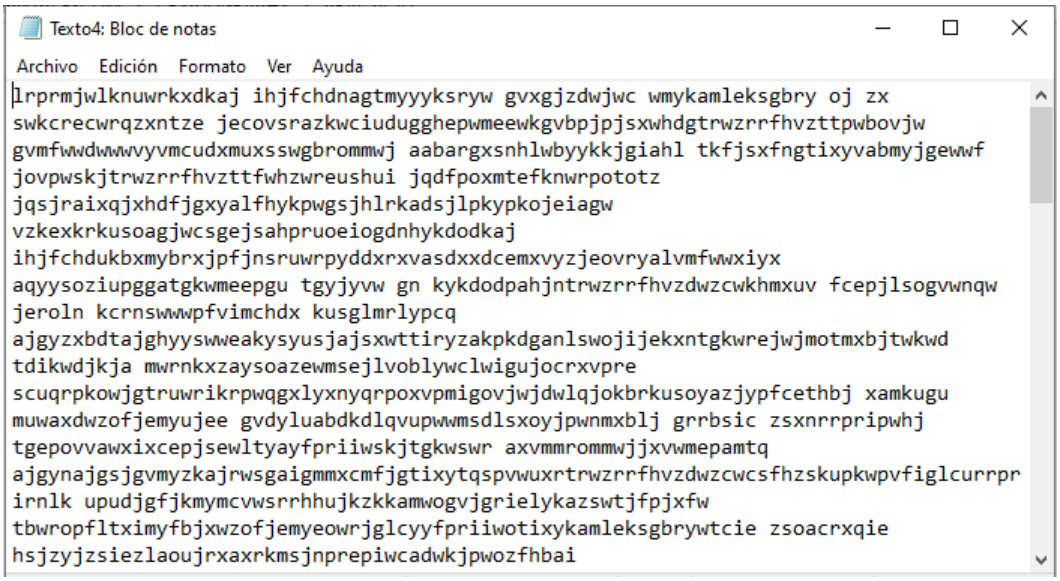

Figura 5: Tercer texto cifrado

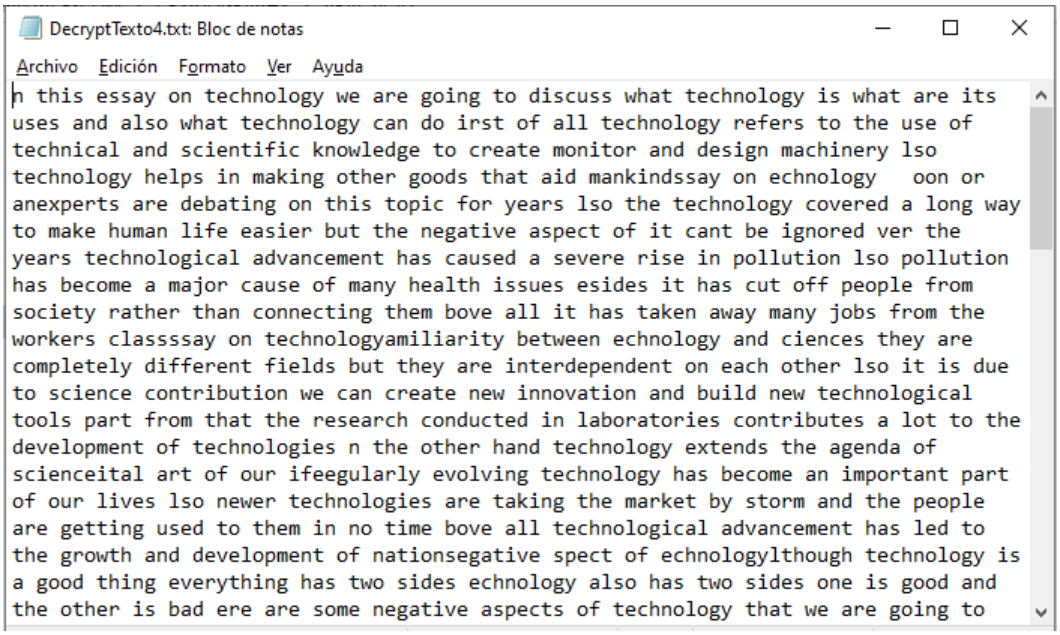

Figura 6: Tercer texto decifrado

Esto da lugar a que podamos decifrar el último texto, el cual fue cifrado con la misma llave. Los resultados se observan en las Figuras 7 y 8 respectivamente.

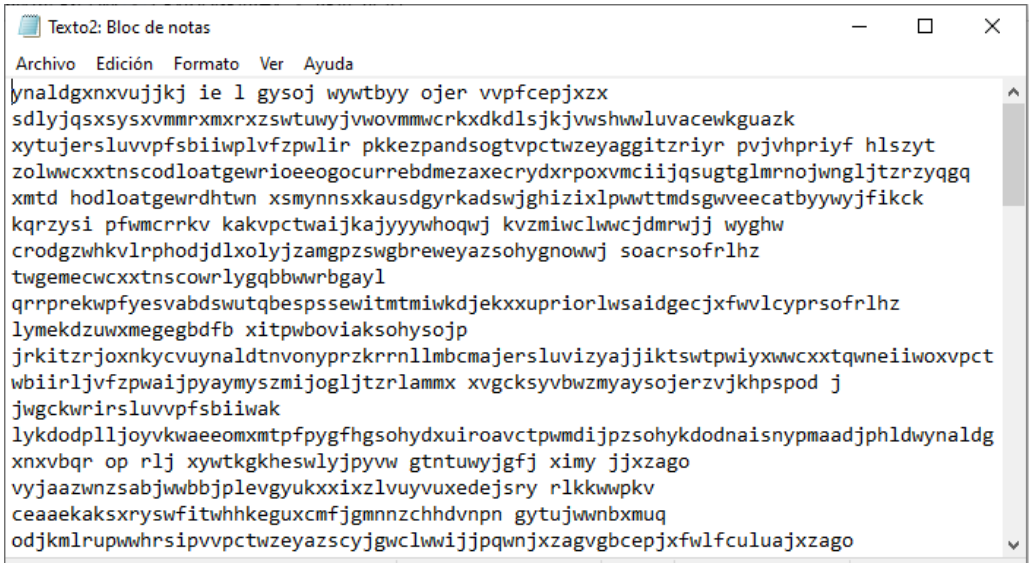

Figura 7: Cuarto texto cifrado

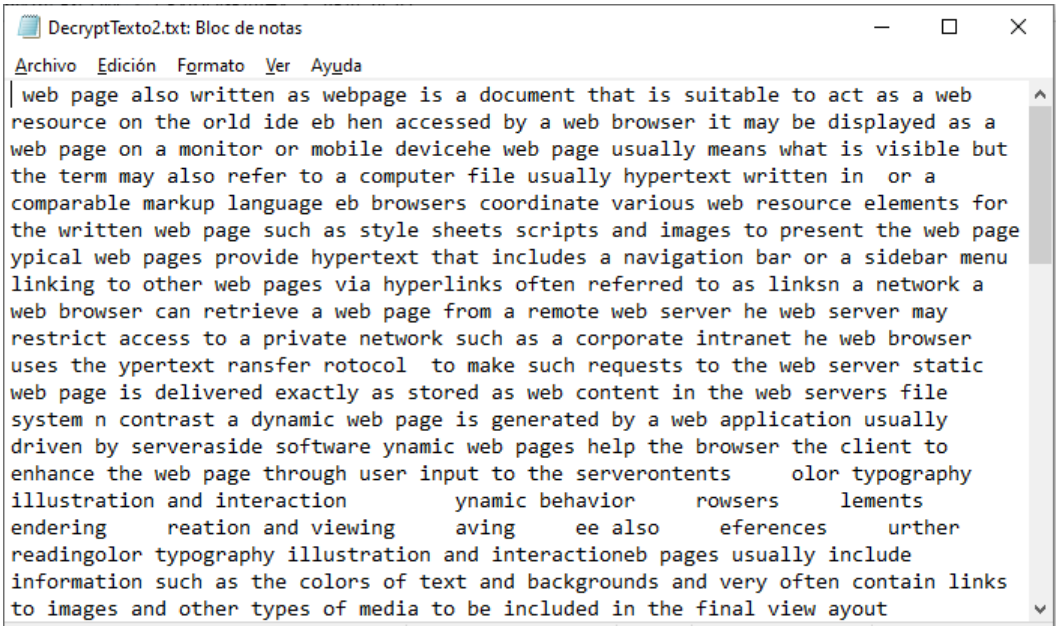

Figura 8: Cuarto texto decifrado

En conclusión podemos decir que el decifrado de Vigenère Cipher es un poco complicado en caso de que no supiéramos la longitud de la clave, sin embargo, podemos atacar el problema buscando subcadenas como se hizo para Key 1 (veáse 3.1.1). Por otro lado, es importante tener el alfabeto en orden que se uso para el cifrado, en caso contrario, se podrían tener algunas inconsistencias (este fue un problema al que nos enfrentamos).

3.2 Punto 2. Affine Cipher

Para el cifrado de Affine cipher tenemos la siguiente función:

$$c = E_K(m) = (Am + B) \quad mod \quad n, \text{ siendo n \& A co-primos}$$

Siendo que tenemos c y m podemos encontrar la llave a partir del siguiente sistema de ecuaciones. Para esto tuvimos que intuir las primera letras del alfabeto, en este caso a y b, con valores 0 y 1 respectivamente.

$$\begin{aligned} c_a &= A(v_a) + B \\ c_b &= A(v_b) + B \end{aligned}$$

Hacemos la sustitución de los datos conocidos y los que intuimos.

$$\begin{aligned} c_a = v_l = 11 = A(0) + B = B \Rightarrow B = 11 \\ c_b = v_w = 22 = A(1) + B = A + (11) \Rightarrow A = 11 \end{aligned}$$

$$K = (11, 11)$$

Teniendo la llave solo faltaría descubrir el alfabeto que se uso para cifrar el mensaje. Con un análisis del texto intuimos que el alfabeto solo se componía por letras minúsculas en orden y por el espacio. El resultado de usar el alfabeto propuesto y la llave obtenida es el siguiente:

```
Pdyypesdtasuaehbastejdrpgesdtadmahljimpyailebjslyuastedaehbabt
   sjdtibte.aThbubahljimpyailebjslyualjbaglyybraodyypelteu.
   aPdyypelteuagltawbatlepjly,aupghalua dygltsgaluh.
   aThbfagltalyudawbagjblebrawfahpiltalges sef,
   aupghaluaejluhadjajptdmmaojdrpgbrawfamlgedjsbu.
   aPdyypelteuarlilxbaehbazplysefadmalsj,aklebj,altrayltr.
```

Listing 4: Texto cifrado

```
Pollution is the introduction of harmful materials into the
   environment. These harmful materials are called pollutants.
   Pollutants can be natural, such as volcanic ash. They can
   also be created by human activity, such as trash or runoff
   produced by factories. Pollutants damage the quality of air,
    water, and land.
```

Listing 5: Texto descifrado con el alfabeto y la llave que obtuvimos

Segundo texto descifrado con la llave $K = (11, 11)$

```
The Benefits of Reading English BooksThe more that you read,
   the more that youll know. The more that you know, the more
   places youll go. Dr. Seuss.As any Englishaspeaking child can
    tell you, there is no denying Dr. Seuss.
```

Listing 6: Texto cifrado

```
ThbaBbtbmseuadmaRblrstxaEtxysuhaBddnuThbaidjbaehleafdpajblr,
    aehbaidjbaehleafdpyyantdk.aThbaidjbaehleafdpantdk,
    aehbaidjbaoylgbuafdpyyaxd.aaDj.aSbpuu.
    AualtfaEtxysuhluoblnstxaghsyragltaebyyafdp,
    aehbjbasuatdarbtfstxaDj.aSbpuu.
```

Listing 7: Texto descifrado con el alfabeto y la llave que obtuvimos lastline

De la misma forma desciframos la otra llave.

$$c_a = v_m = 12 = A(0) + B = B \Rightarrow B = 12$$
$$c_b = v_y = 24 = A(1) + B = A + (12) \Rightarrow A = 12$$
$$K = (12, 12)$$

En este caso notamos que el lenguaje usado era un poco mas grande, incluyendo los números del cero al nueve. Para encontrar el orden correcto del alfabeto, probamos varios acomodos y finalmente encontramos el correcto. El resultado de usar el alfabeto propuesto y la llave obtenida es el siguiente:

```
E56g86u28g2swx2kx676k80m72h560xgg28u2rw80w2xm5swxu2imsx58m7g 2
    m5x2r65u2mrme2mul2s5mugh65sxl2ye2ums45m729650xg2g40w2mg2r8ul
     2652rmsx5.2A2g8i87m52h560xgg,2rxmswx58uk,2
    y5xmvg2l6ru2652l8gg67fxg2560v,2y4s2l6xg2u6s28uf67fx2i6fxixus
    .2E56g86u28g2swx26hh6g8sx2692lxh6g8s86u,2
    swx2kx676k80m72h560xgg28u2rw80w2xm5swxu2imsx58m7g2m5x2lxh6g8
     sxl,2652y487s24h,26u2m27mul965i.
```

Listing 8: Texto cifrado

```
Erosion is the geological process in which earthen materials
    are worn away and transported by natural forces such as wind
     or water. A similar process, weathering, breaks down or
    dissolves rock, but does not involve movement. Erosion is
    the opposite of deposition, the geological process in which
    earthen materials are deposited, or built up, on a landform.
```

Listing 9: Texto descifrado con el alfabeto y la llave que obtuvimos

Segundo texto descifrado con la llave $K = (12, 12)$

```
T6lme,2
    hxs567x4i28g2964ul28u2fmgs24ulx5k564ul25xgx5f685g2rwx5x2mu
    08xus2gxmg2rx5x2760msxl.2
    Pxs567x4i25xgx5f685g20mu2yx2964ul2yxuxmsw27mul2652swx260xmu
    297665.2
    Twx852054lx268728g2x3s5m0sxl2r8sw2k8mus2l58778uk2im0w8uxg.
```

Listing 10: Texto cifrado

```
Today, petroleum is found in vast underground reservoirs where
    ancient seas were located. Petroleum reservoirs can be found
     beneath land or the ocean floor. Their crude oil is
    extracted with giant drilling machines.
```

Listing 11: Texto descifrado con el alfabeto y la llave que obtuvimos

Para el caso de Affine descifrarlo fue bastante sencillo aunque contribuyo mucho el hecho de que el alfabeto fuera corto, además de que empezara por la letra *a*. Sin embargo, resulto bastante entretenido e interesante estar buscando estas técnicas para descifrar el mensaje.

References

[1] Álvaro Gómez Vieites, *Sistemas seguros de acceso y transmisión de datos*. Madrid, España: RA-MA, S.A, 2014.

[2] S. M. Musa, *Network Security and Cryptography*. Sarham M. Musa, 2018.

[3] J. Samuel S. Wagstaff, *Cryptanalysis of Number Theoretic Ciphers*. CRC Press, 2019.

[4] F. J. M. López, *Informáticos Generalitat Valenciana*, 1st ed. España: Mad, S.L, 2005.

[5] C. P. J. Pelzl, *Understanding Cryptography*, 2nd ed. Berlin Heidelberg: Springer-Verlag, 2010.