TIPE:chiffre de Vigenère

- Marie Sengler
- Numéro candidat:38088

Sommaire

Dans quelle mesure le code de Vigenère est-il fiable pour protéger ses données?

I) chiffrage de Vigenère

Le chiffrage de Vigenère s'obtient en substituant les lettres du texte originel par d'autres en utilisant une clé.
décalage : a=0 (place du a dans l'alphabet)

→	b	o	n	j	o	u	r
	1	14	13	9	14	20	17
→	a	i	u	a	i	u	a
	0	8	20	0	8	20	0
→	b	w	h	j	w	o	r
	1	22	33	9	22	14	17
			7				

Selon la place de la lettre dans l'alphabet, on décale de tant et tant.

II) cryptanalyse du chiffrage de Vigenère

Indice de coïncidence :

$$I_c = \sum_{k=1}^{26} \frac{n_k \times (n_k - 1)}{n \times (n-1)}$$

nk le nombre d'occurrences de la k ième lettre
n le nombre total de lettres
A:"on a deux fois la lettre A dans le texte"

Ic est la probabilité que deux lettres choisies aléatoirement soient identiques.
En français, il est de **0.0778**

$$P(A) = \frac{C_2^{nA}}{C_2^n} = \frac{nA(nA-1)}{n(n-1)}$$

Ic=P(AUB…. UZ) = P(A) +P(B) +.... +P(Z)
événements disjoints

II) cryptanalyse du chiffrage de Vigenère

Longueur de la clé :

-l'indice de coïncidence ne change pas
si on décale les lettres par un même écart.

$$I_c = \sum_{k=1}^{26} \frac{n_k \times (n_k - 1)}{n \times (n-1)}$$

-On subdivise k fois le texte ainsi
jkqsdluibnaeo
→ jsuno / kdia / qlbe

-On calcule l'indice de coïncidence
pour chaque subdivision,
si **Ic est environ 0,0778**,
alors la **longueur de clé est k**.

Dans un texte aléatoire,
Ic=26*(1/26)²=0,385.

II) cryptanalyse du chiffrage de Vigenère

Autre méthode : test de Kasiski

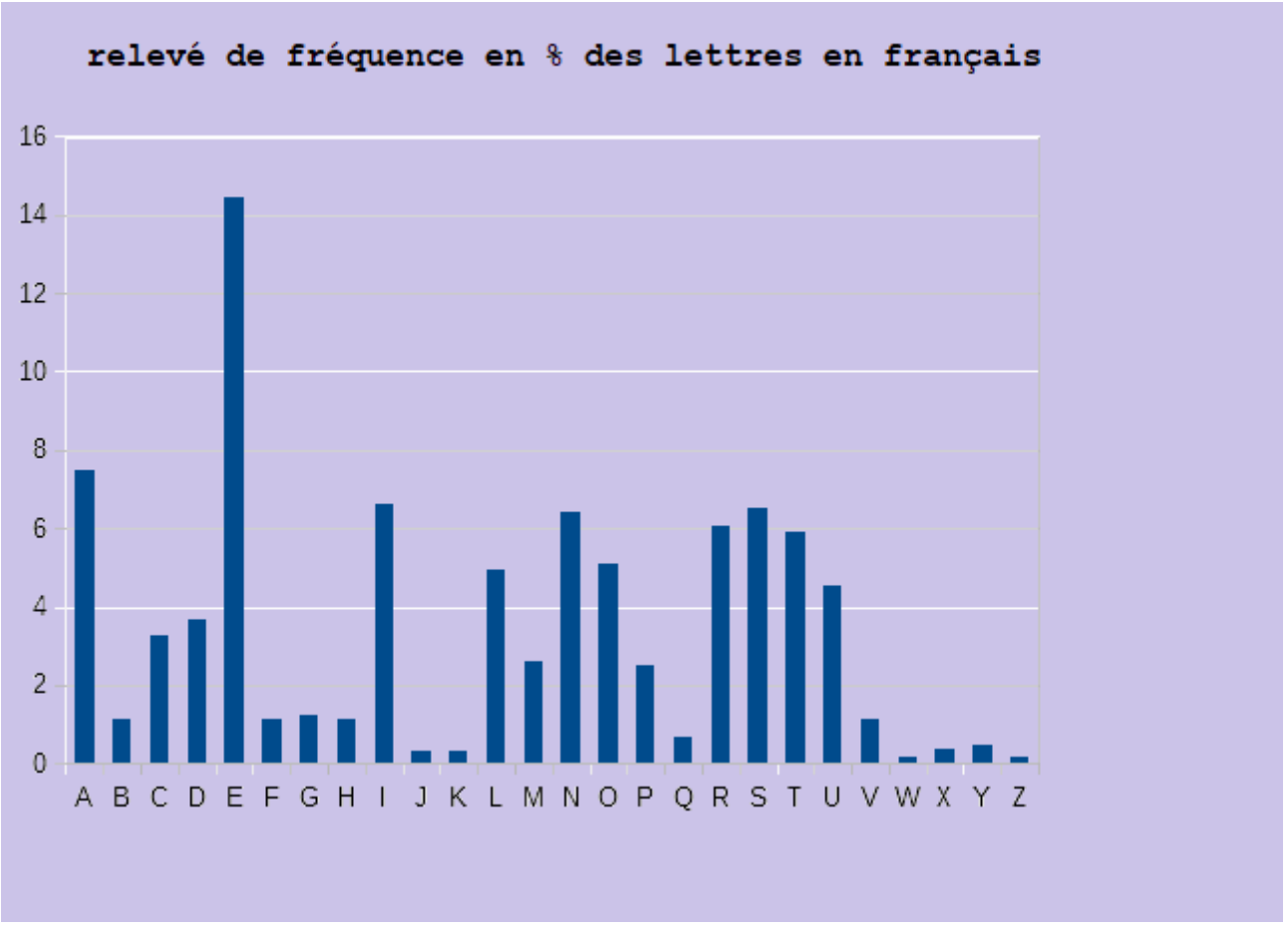
Les répétitions dans le texte peuvent indiquer qu'ils ont été chiffrés avec le même bout de clé.

KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUD**WUU**MBSVLPS
NCMUEKQCTESWR**EEK**OYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTD**WXIZA**
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGR**WUU**NEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSH**NUOCZGM**RUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNL**GOY**L
SKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFST**NUOCZGMEEK**CPJR
GPMURSKHFRSEIUEV**GOY**C**WXIZAYG**OSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVID**GMU**CGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
W**GMU**SWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEE
DCLDHWTYYIDGMVRDGMPLSWGJLAGO**EEK**JOFEKUYTAANYTDWIYBNLNYNP
WEBFNLFYNAJEBFR

Distance entre les répétitions :
GMU : 90
NUOCZGM : 80
WUU : 95
WXIZAYG : 190
EEK : 200
GOY: 70
Le **pgcd** de ces distances est 5.
La clé est donc de 5

Texte trouvé sur la page Wikipédia
cryptanalyse du chiffre de Vigenère
Poème: l'albatros de Baudelaire

II) cryptanalyse du chiffrage de Vigenère



Détermination de la clé

-La lettre **E** est la plus présente
dans un texte en français.
-la lettre **la plus fréquente** dans chaque subdivision correspond à la lettre E
-le **décalage relatif** entre ces deux lettres permet d'avoir la clé

II) cryptanalyse du chiffrage de Vigenère

Déchiffrer :

-Après avoir obtenu le décalage relatif, on applique le **chiffrage de César** à chaque subdivision, puis on réassemble toute les subdivisions

b	j	r	/	w	w	/	h	o			
1	9	17	/	22	22	/	7	14			
a	a	a	/	s	s	/	g	g			
0	0	0	/	18	18	/	6	6			
b	j	r	/	o	o	/	n	u			
1	9	17	/	14	14	/	13	20			

→ bonjour clé: a i u (0 8 20) avec 8 = 26 - 18, 20 = 26 - 6

III) interprétation des résultats :

la lettre majoritaire correspond-elle à E dans une subdivision ?

Intervalle de fluctuation :

Soit Xn et Yn 2 variables aléatoires
qui suivent ß(n,p) et ß(n,q)
Avec p=0,164 la probabilité d'avoir un E,
q=0,085 la probabilité d'avoir un A,
n le nombre de lettres dans la subdivision

On veut éviter que les 2 intervalles se croisent. On trouve n au moins :
-**n=181 (seuil de 90%)**
-**n=260 (seuil de 95%)**
-**n=441 (seuil de 99,7%)**

$$I_n = \left[p - \frac{1.96 \times \sqrt{p(1-p)}}{\sqrt{n}} ; p + \frac{1.96 \times \sqrt{p(1-p)}}{\sqrt{n}} \right]$$

De même l'indice de coïncidence a besoin d'un n assez grand (n>50)

$$\sqrt{n} > 1.96 \frac{\sqrt{q(1-q)} + \sqrt{p(1-p)}}{p-q}$$:Un minorant de n

III) interprétation des résultats :

Une longue clé permet
au message d'être protégé.

chiffre de Vernam, ou masque jetable

La clé est
- de longueur égale à celle du message
- utilisée une seule fois

Ainsi, pour un texte de 10 lettres, on a $26^{10} = 10^{14}$ possibilités.

Pour essayer toutes les possibilités,
l'ordinateur a besoin de beaucoup de temps.
Complexité: $O(n^n)$

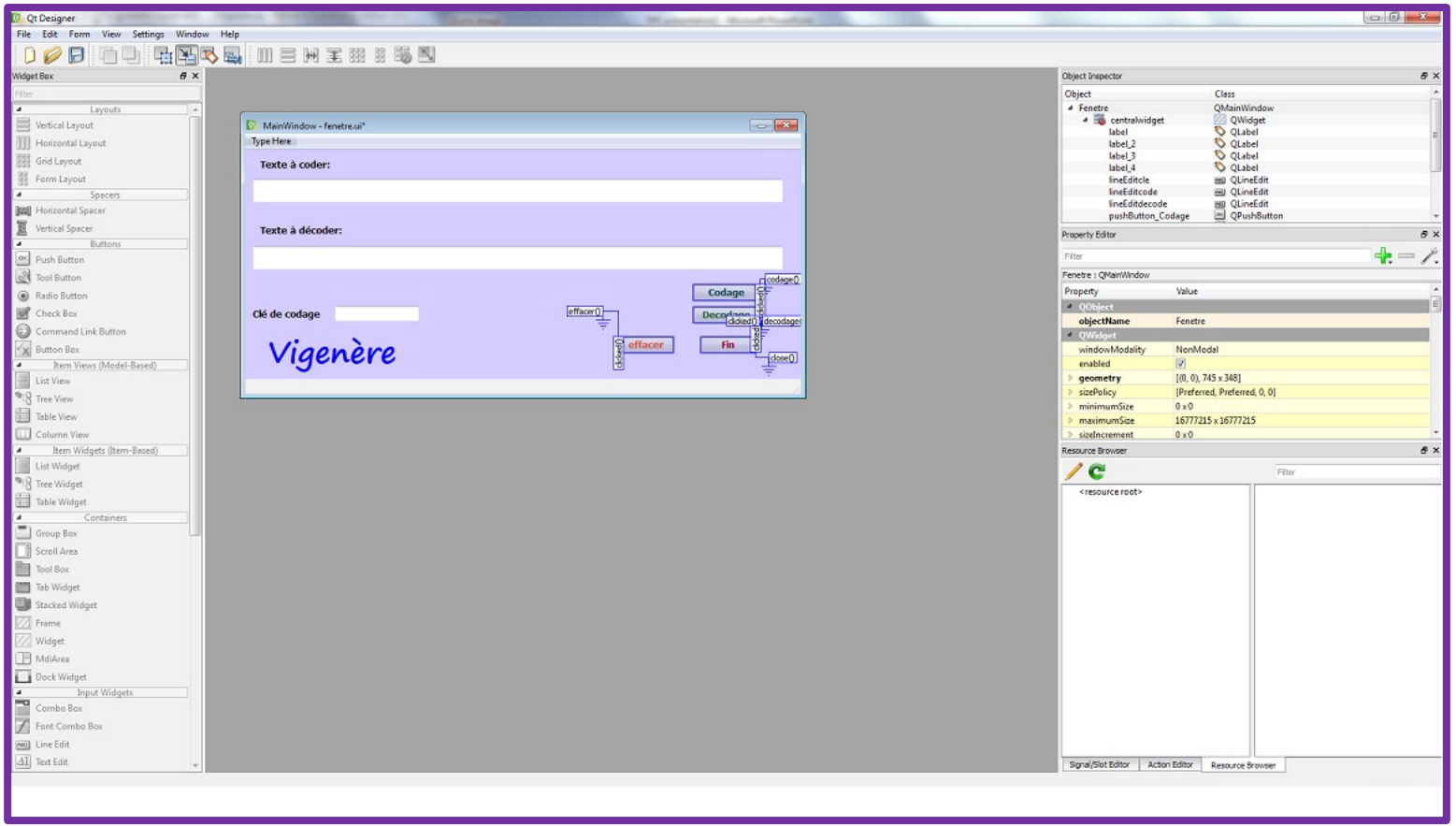
Conclusion:

Qu'est ce que le TIPE m'a apporté?

→ connaissance: logiciel QtDesigner et le module PyQt, programmation Python, cryptanalyse
→ m'impliquer sur un projet de plus longue duré
→ nécessité d'avoir une longue clé ou un cryptage plus robuste comme le RSA, DES….

Merci, pour votre attention

Annexe



Interface du programme :

-on a utilisé Qt Designer, avec le module Python PyQt4