Stage Février 2016 Du 08/02/16 au 12/02/16

Codage par substitution

Le but du TP est de voir certaines méthodes de cryptographie simples et déceler leur faiblesse en terme de sécurité. Cela montrera l'intéret d'avoir recours à d'autres méthodes de cryptographie comme la cryptographie RSA, la cryptographie par le chaos, la cryptographie à l'aide des courbes elliptiques ....

Vocabulaire

- La *Cryptologie* est la science des messages secrets. Elle se décompose en deux disciplines :
  - la *Cryptographie*, art de transformer un message clair en un message inintelligible par celui qui ne possède pas la clé de déchiffrement. Cependant, on utilise souvent le mot cryptographie comme synonyme de cryptologie.
  - la *Cryptanalyse*, art d'analyser un message chiffré afin de le décrypter quand on ne possède pas la clé de déchiffrement.
- *Chiffre* : anciennement *code secret*, par extension désigne aussi un algorithme utilisé pour le chiffrement ;
- *Chiffrer* : transformer à l'aide d'une clé de chiffrement un message en clair en un message chiffré, incompréhensible si on ne dispose pas de la clé de déchiffrement correspondante ;
- *Déchiffrer* : retrouver à l'aide de la clé de déchiffrement correspondante le message en clair d'origine à partir d'un message précédemment chiffré à l'aide d'une clé de chiffrement ;
- *Clé de chiffrement* : méthode permettant de chiffrer un message en clair ;
- *Clé de déchiffrement* : méthode associée à une clé de chiffrement et permettant de déchiffrer un message précédemment chiffré ;
- *Décrypter* : retrouver le message en clair correspondant à un message chiffré sans posséder la clé de déchiffrement ni la clé de chiffrement ;
- *Cryptogramme* : messge chiffré (incompréhensible si on ne dispose pas de la bonne clé de d'echiffrement).

Consignes générales

- On écrit les messages en lettres majuscules et sans accents. Ceci a pour but de simplifier le chiffrement et le déchiffrement des messages.
- Lorsque les messages sont constitués de plusieurs mots ou de plusieurs phrases, on conserve les espaces entre les mots et les signes de ponctuation. Le résultat est de faciliter encore le déchiffrement.
- Ce n'est pas conforme à la réalité, puisque dans la vraie vie, on veut au contraire compliquer au maximum le cryptogramme, pour empêcher l'adversiare de le décrypter, si possible...
- Plus un cryptogramme est long, plus il est facile à décrypter, car l'adversaire dispose de plus d'indices. Ceci explique pourquoi, dans les exercices, les messages à décrypter seront significativement plus longs que les messages à chiffrer ou à déchiffrer.
- On utilise la correspondance ci-dessous entre les lettres de l'alphabet {A, B, C,..., Z} et les nombres {0, 1, 2, ..., 25}

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

1 Chiffrement par décalage

- ☞ Un chiffrement par décalage consiste á décaler les lettres de l'alphabet d'une valeur fixe.
- ☞ Et bien sûr, si on dépasse Z, on reprend á partir de A.
- ☞ Autrement dit, on travaille "modulo 26"

Exercice 1. (Chiffre de César)
Les lettres de l'alphabet sont chiffrés à l'aide de la clé de chiffrement : $f : \{0, 1, \cdots, 25\} \mapsto \{0, 1, \cdots, 25\}$ qui à x associe $f(x) \equiv x + 4[26]$.
Cette méthode est réputé avoir été utilisé par Jules César pour communiquer secrétement avec ses généraux pendant la guerre des Gaules, d'où son nom.

1. Chiffrer le message : ALLEZ-Y
2. Quelle est la clé de déchiffrement du chiffre de César ?
3. Déchiffrer la réponse du général : SR C ZE

Exercice 2. (Une faiblesse dangereuse). Combien y-a-t-il de chiffrements par décalage différents ?

Exercice 3. (Une faille de sécurité). Le cryptogramme suivant a été obtenue à l'aide d'un chiffrement par décalage inconnu : TZTVIFE VJK LE REV. Un espion habile a réussi à découvrir que la lettre S est chiffrée par la lettre J.

1. Décrypter le message proposé.
2. Préciser quelles sont les clés de chiffrement et de déchiffrement correspondantes.

Exercice 4. (Décryptage). Le cryptogramme suivant a été obtenue à l'aide d'un chiffrement par décalage inconnu :

TOX VXLTK !
C'TB VTIMNKX NG XLIBHG.
T UBXGMHM,
LBZGX : FTKV-TGMHBGX

1. Proposé deux méthodes de décryptage, l'une tirant parti de l'exercice 2, et l'autre de l'exercice 3 .
2. Décrypter le message proposé.
3. Préciser quelles sont les clés de chiffrement et de déchiffrement correspondantes.

Exercice 5. (Programmation). On rappelle qu'une fonction et que la création d'une chaine de caractères peuvent en Python s'écrire sous la forme suivante :

```
def puissance(a,p):
  resultat = 1
  for i in range(p):
    resultat = a * resultat
  return(resultat)
```

```
Alphabet = ''
for k in range(65, 91):
  Alphabet = Alphabet + chr(k)
```

La fonction précédente calculant a^p lorsque l'on donne les arguments a et p à la fonction puissance. La variable Alphabet est à la fin de la boucle affectée de la valeur 'ABCDEFGHIJKLMNOPQRSTUVWXYZ'

1. En vous inspirant de la syntaxe de la fonction précédente, écrire une fonction **cesar(carac, decalage)** prenant en argument un caractère **carac** et un entier **decalage** et qui retourne la lettre obtenue par le décalage proposé dans le chiffre de César. Pour pouvoir utiliser cette fonction dans plus de cas, on pourra ajouter que si le caractère **carac** n'est pas une lettre majuscules (donc n'est pas dans la chaine Alphabet), la fonction retournera **carac** sans modification. On rappelle que si a et b sont deux entiers $a\%b$ envoie le reste de la division euclidienne de a par b (donc renvoie $a \equiv [b]$
2. Ecrire une fonction **chiffre(message, decalage)** prenant en argument une chaine de caractères **message** et un entier **decalage** et qui retourne le message codé obtenu par le décalage proposé dans le chiffre de César
3. Tester votre fonction sur les exemples des exercices préc{edents.

2 Chiffrement par substitution

- ☞ Un chiffrement par substitution simple consiste à remplacer chaque lettre par une autre selon une mt́hode décidée à l'avance.
- ☞ Le chiffre de César est un cas particulièrement simple de chiffrement par substitution

Exercice 6. (Chiffrer/Déchiffrer)
Dans le cas général, il n'y a pas de fonction mathématique simple permettant de chiffrer chaque lettre, c'est pourquoi on donne la table de chiffrement en entier. Par exemple dans cet exercice on utilise la substitution suivante (remarque : pour bien distinguer les messages chiffrés et les messages en clair, on emploiera les minuscules pour les messages codés et les majuscules pour les messages en clair) :

A	B	C	D	E	F	G	H	I	J	K	L	M
s	c	w	u	d	x	b	f	y	t	g	z	i

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
h	o	j	n	k	a	m	l	p	e	q	r	v

1. Chiffrer le message : facile.
2. Ecrire la table de déchiffrement correspondante.
3. Déchiffrer le message DAASY

Exercice 7. (Cette fois ce n'est pas faible). Combien y-a-t-il de chiffrements par substitution différents ?

Exercice 8. (une faille de sécurité). Les longs messages chiffrés par substitution peuvent être assez facilement lorsque l'on sait en quelle langue a été écrit le message d'origine. Pour cela on utilise la méthode d'analyse de fréquence.
En effet en français, la lettre E est beaucoup plus utilisée que les autres : le tableau suivant indique le pourcentage d'apparition des lettres dans un texte français :

— A : 9,4	— E : 15,9	— I : 8,4	— M : 3,2	— Q : 1,1	— U : 6,2	— Y : 0,2
— B : 1,0	— F : 0,9	— J : 0,9	— N : 7,2	— R : 6,5	— V : 2,2	— Z : 0,3
— C : 2,6	— G : 1,0	— K : 0,0	— O : 5,1	— S : 7,9	— W : 0,0	
— D : 3,4	— H : 0,8	— L : 5,3	— P : 2,9	— T : 7,3	X : 0,3	

On va utiliser cela pour effectuer le décryptage d'un texte (tiré de la nouvelle (traduite en Français) 'Le scarabée d'or' d'Edgar Allan Poe.
Le message codé d{ecrit le lieu où a été enterré un trésor par des pirates :

.

sp ckp qriir lfpm a ekora lr a rqrzsr

lfpm af uefgmr ls lgfcar

zsfifpor ro sp lrdirm ro oigyr

jgpsorm pkil rmo zsfio lr pkil

nigpugnfar ogdr mrnogrjr cifpuer

ukor rmo afuery lr a krga dfsuer

lr af oror lr jkio spr agdpr

l fcrgeer lr a ficir f aifqrim

af cfaar ugpzsfpor ngrlm fs afidr.

1. Effectuer l'analyse fréquentielle de ce texte crypté de 228 lettres.
2. Décryptage du texte.
   Répondre aux questions suivantes qui vous guideront dans le décryptage du texte :
   - ☞ Quelles sont les deux lettres (en minuscules) qui apparaissent le plus fréquemment dans le texte ?
   - ☞ Comparer avec les deux lettres (en majuscules) les plus fréquentes en français et en déduire leur décodage :
   - ☞ Ecrire en dessous des trois premières lignes du texte les lettres trouvées.
   - ☞ Apparait le mot 'fs'. Quelle lettre la lettre 's' peut-elle remplacer ?
   - ☞ Apparait le mot 'oror', quel mot peut-il remplacer (les accents ne comptent pas) ?
   - ☞ En déduire la traduction de 'oror lr jkio' :
   - ☞ Quel mot peut représenter le premier mot 'sp' du texte ?
   - ☞ Faire le bilan des nouvelles lettres trouvées puis écrire la deuxième et les trois dernières lignes du texte en remplaçant par les nouvelles lettres trouvées.
   - ☞ Décrypter aussi les expressions 'zUAiANTE ET UN' puis 'ugNzUANTE'
   - ☞ Ecrire le texte en remplaçant par les nouvelles lettres trouvées et continuer le décryptage.
3. Ecrire une fonction python qui donne les fréquences d'apparition des lettres dans un texte donné