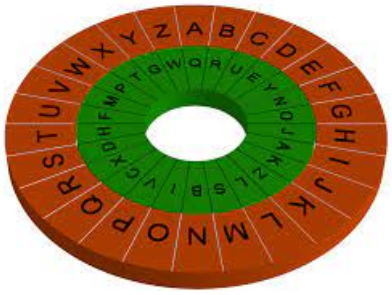
CY Cergy-Paris Université — **Enseignant : Kévin Carrier**
Introduction à la Cryptographie — kevin.carrier@cyu.fr

TD 1 : Chiffrements ancestraux

Exercice 1. (Chiffrement de César – 100-44 AV-JC)



1. Décryptez le message suivant :

```
bu nyhuk nblyyply ?  wlyzvuul why sh nblyyl ul klcplua nyhuk.
sh nblyyl klz lavpslz.
```

2. Décrivez le chiffrement qui a été utilisé. Expliquez comment vous avez pu le *casser*. Quel est le coût maximal de votre attaque?

Exercice 2. (Chiffrement par substitution – IV$^{\text{ème}}$ siècle AV-JC)

Le chiffrement de César consiste à substituer chaque lettre de l'alphabet par une autre. Cependant, ce chiffrement ne considère qu'un tout petit sous ensemble de toutes les substitutions possibles. Cette fois-ci, pour chiffrer un message constitué des lettres de `'a'` à `'z'`, on remplace chaque lettre par une autre que l'on aura lue dans une table de correspondance. La clé secrète sera cette table de correspondance. Par exemple, le mot `'vert'` aura pour chiffré `'bleu'` avec la clé de chiffrement suivante :

clair	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
chiffré	t	h	g	x	l	m	w	z	y	q	v	d	s	o	i	n	a	e	c	u	f	b	k	r	p	j

1. Combien de clés possibles existe-t-il? Pensez-vous que l'ordinateur le plus puissant au monde puisse tester toutes les clés en un temps raisonnable?

2. Implémentez le chiffrement (et le déchiffrement) par substitution. Pensez à implémenter également un générateur de clé. Le paramètre `key` sera un tableau à une dimension tel que `key[0]` contient le chiffré de `'a'`, etc.

Dans un texte en français, le nombre d'occurrences de la lettre `'e'` sera nettement supérieur au nombre d'occurrences de la lettre `'z'`. Pour une langue donnée, on dresse la table des fréquences des lettres en mesurant statistiquement sur un texte long clair, la fréquence d'apparition de chaque lettre.

3. Implémentez un script qui compte le nombre d'apparitions de chaque lettre de l'alphabet dans un texte passé en argument. Attention aux accents, majuscules, ponctuations... On exprimera les valeurs relativement au nombre total de lettres (utiles) lues.

   Remplissez les tables de fréquences suivantes (on pourra aussi représenter ces tables de fréquences à l'aide d'un histogramme) :

Français :

lettre	a	b	c	d	e	f	g	h	i	j	k	l	m
fréquence d'apparition													

n	o	p	q	r	s	t	u	v	w	x	y	z

Anglais :

lettre	a	b	c	d	e	f	g	h	i	j	k	l	m
fréquence d'apparition													

n	o	p	q	r	s	t	u	v	w	x	y	z

4. Proposez une méthode permettant de détecter la langue d'un texte chiffré avec un chiffrement par substitution.

5. Proposez une attaque par analyse fréquentielle sur le chiffrement par substitution pour décrypter le cryptogramme suivant (le clair est en français) :

```
wt dorxaplotxuci ina yvi ecndcxwcvi xopqtqwikiva tynnc
tvdcivvi syi wi qinpcv ei dpkkyvcsyio.  xpyoatva di v ina xtn
wt niywi ktvcioi ei aotvnkiaaoi nidoiaikiva ei w cvgpoktacpv.
eibt ty j i ncidwi tjbd, wi loid uiopepai eidoca, etvn nin
duopvcsyin ein lyiooin ivaoi win loidn ia win xionin, yvi
kiaupei apya tyatva cvlivciyni sy iapvvtvai xpyo dtduio
ein kinntlin :  w ceii ina ei otnio wt aiai e yv indwtji
ei dpvgctvdi, ei atapyio wi kinntli nidoia nyo npv dotvi xycn
e taaiveoi syi win duijiyz oixpynniva, ecnnckywtva tcvnc wi
kinntli.  w indwtji xpyjtca twpon eiwcjoio wi kinntli.  wi
einacvtatcoi v tjtca twpon sy t otnio ei vpyjity wi dotvi
ei w indwtji xpyo xpyjpco wcoi wi kinntli.  cw vi gtwwtca
xtn iaoi xoinni...  w toa ei ecnnckywio wi nyxxpoa kiki e yv
kinntli xpyo wi oiveoi nidoia ina txxiwi nailtvplotxuci.
```

Pour vous aider dans cette tache, vous pouvez utiliser une clé temporaire contenant le symbole $*$ pour chaque caractère que vous ne savez pas encore déchiffrer.

Exercice 3. (Chiffrement de Vigenère – XVI$^{\text{ème}}$ siècle)

Pour la suite, on définit les opérations + et − sur l'ensemble $\{\mathtt{a}, \mathtt{b}, \cdots, \mathtt{z}\}$. Pour cela, on commence par associer chaque lettre à un nombre avec le mapping suivant :

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

La somme (resp. la différence) de deux lettres est alors définie par la somme (resp. la différence) modulo 26 de leurs valeurs numériques données par le mapping précédent. Par exemple :

$$\mathtt{r} + \mathtt{w} = \mathtt{n}$$

car

$$\begin{gathered} \mathtt{r} \longleftrightarrow 17, \\ \mathtt{w} \longleftrightarrow 22, \\ 17 + 22 \equiv 13 \mod 26 \\ \text{et } 13 \longleftrightarrow \mathtt{n}. \end{gathered}$$

La somme (resp. différence) de deux mots contenant le même nombre de lettres est leur somme (resp. différence) lettre par lettre. Par exemple :

	tigre	⟷	19	8	6	17	4
+	zebre	⟷	25	4	1	17	4
=	smhii	⟷	18	12	7	8	8

Pour chiffrer un message **m** de longueur arbitraire avec une clé **k** de taille t, on concatène n fois la clé **k** pour obtenir une chaine de même longueur que le message; puis on somme **m** avec la chaine obtenue. Si la taille du message à chiffrer n'est pas un multiple de t, alors on pourra compléter le message avec un certain nombre de lettres pour que ce soit le cas. Cet ajout est appelé *bourrage* (ou *padding* en anglais).

1. Combien de clés de taille t existe-t-il? Pourquoi vaut-il mieux choisir une clé secrète en tirant aléatoirement 8 lettres plutôt que d'utiliser un mot dans le dictionnaire?
2. Un utilisateur du système propose d'utiliser un bourrage consistant à compléter le message avec autant de `'a'` que nécessaire. Pourquoi est-ce une mauvaise idée? Suffit-il de choisir une autre lettre que `'a'` pour régler le problème? Proposez une solution plus sûre.

Il est très courant d'avoir recours à des *padding* en cryptographie. Cet exercice montre l'importance de ne pas les choisir n'importe comment. Dans le cas présent, nous aurions pu éviter d'en utiliser ; en effet, pour chiffrer le dernier bloc du message de taille $\leq t$, nous aurions pu tronquer la clé.

3. Supposons que la taille t de la clé est exactement la taille du message. Montrez que pour tous textes $\mathbf{c}$ et $\mathbf{m}$ de taille t, il existe une clé $\mathbf{k}$ telle que $\mathbf{c}$ soit le chiffré de $\mathbf{m}$. En déduire qu'un attaquant qui a accès à un message chiffré ne peut pas le déchiffrer car il n'aura aucun moyen de valider sa réponse. Ce constat est-il toujours vrai lorsque l'on a une clé de taille strictement inférieure à la taille du message ou, de façon équivalente, lorsque l'on réutilise plusieurs fois la même clé pour chiffrer des messages différents?

4. Implémentez le chiffrement et le déchiffrement de Vigenère. Pensez à implémenter également un générateur de clé.

5. Déchiffrez le message suivant qui a été chiffré avec la clé `odyssee` :

```
rdlkdijwokvwybaljdwyrzrbqkwisgcdwwtofcvwwxoqjwqoypugucet
ddpsaxybvshwvsfggfsxiiuyhhipskydiymsvrdwglwidjwhiweksnig
zhagvihsfcksvc
```

Exercice 4. (Indice de coïncidence)

Définition 1 : L'indice de coïncidence $I_C(\mathbf{m})$ d'un texte $\mathbf{m}$ est la probabilité que deux caractères choisis uniformément dans $\mathbf{m}$ soient égaux.

$$I_C(\mathbf{m}) = \sum_{x=\mathtt{a}}^{\mathtt{z}} \frac{n_x(n_x-1)}{n(n-1)}$$

où n est la longueur de $\mathbf{m}$ et n_x est le nombre d'occurrences de la lettre x dans $\mathbf{m}$.

La valeur de $I_C(\mathbf{m})$ est en quelque sorte une signature de la langue utilisée. Si $\mathbf{m}$ est un texte en français, $I_C(\mathbf{m})$ aura une certaine valeur typique, si c'est un texte en anglais, alors il aura une autre valeur ou encore, si $\mathbf{m}$ est une chaine de lettre aléatoire, alors $I_C(\mathbf{m})$ sera très différent.

1. Calculez $I_C(\mathbf{m})$ lorsque :
   - $\mathbf{m}$ est un texte en français ;
   - $\mathbf{m}$ est un texte en anglais ;
   - $\mathbf{m}$ est un texte aléatoire.

Une propriété importante de l'indice de coïncidence est qu'il est invariant par un chiffrement de César. On va utiliser cette propriété pour retrouver la taille de la clé dans un chiffrement de Vigenère.

Soit $\mathbf{c}$ le message chiffré avec une clé de taille t suivant :

```
hrixysthtweczxfkwegskaizdzilhrixysthtweczxfxzyfjxeyvk
ybxnyxyfferwiwpbxexwedsmqevcwslgivomfxfpmcwwxevotlox
```

On note $\mathbf{c}_i$ le texte composé de toutes les lettres de $\mathbf{c}$ dont la position est de la forme $i + tn$ avec $n \in \mathbb{N}$.

2. Montrez que $\mathbf{c}_i$ correspond à un chiffré de César.

3. Remplissez le tableau suivant avec les indices de coïncidence correspondants :

t	$\mathbf{c}_0$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$	$\mathbf{c}_4$	$\mathbf{c}_5$	$\mathbf{c}_6$	$\mathbf{c}_7$	$\mathbf{c}_8$	$\mathbf{c}_9$	$\mathbf{c}_{10}$	$\mathbf{c}_{11}$	$\mathbf{c}_{12}$
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													

4. En déduire une valeur probable de t.

5. Une source d'information parallèle nous informe être quasi certain que le mot `anneau` est contenu dans le texte chiffré $\mathbf{c}$. Avec cette information supplémentaire, décryptez le cryptogramme.

Exercice 5. (Le chiffrement de la Scytale – 404 AV-JC)

Une alternative au chiffrement par substitution est le chiffrement par permutation. On chiffre un clair en mélangeant les lettres selon une permutation définie par une clé secrète.

Soit un tableau bidimensionnel de p colonnes et q lignes. Pour chiffrer un message $\mathbf{m}$ de taille $n \in [\![p(q-1)+1, pq]\!]$, il suffit de remplir le tableau avec les lettres du texte de gauche à droite et de haut en bas. Le message chiffré est construit en lisant les caractères du tableau de haut en bas et de gauche à droite. Pour le déchiffrement on remplit le tableau de haut en bas puis de gauche à droite avec les caractères du texte chiffré et on le lit de gauche à droite et de bas en haut.

La clé secrète est donc le couple (p, q).

1. Avec un chiffrement par permutation, la table des fréquences des lettres est-elle modifiée? Proposez un test permettant de deviner que l'on a affaire à un chiffrement par permutation.

2. Implémentez le chiffrement et le déchiffrement par permutation ligne/colonne.