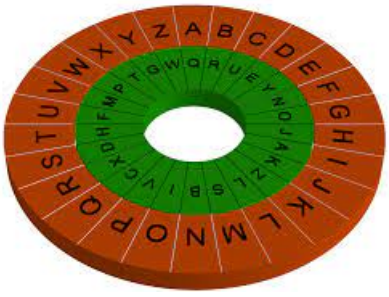
Network Security – Introduction to cryptography

Ancestral Encryption

Exercice 1. (Caesar cipher – 100-44 BC)



1. Decrypt the following ciphertext:

   `Twt rxewtg ztn xh uxuittc.`

2. Describe the encryption which has been used. Explain how you've been able to *break* it. What is the typical (maximal) cost of your attack?

Exercice 2. (Substitution encryption – IV$^{\text{th}}$ century BC)

The Caesar cipher consists of substituting each letter of the alphabet with another. However, this encryption method only considers a very small subset of all possible substitutions. This time, to encrypt a message made of letters from `'a'` to `'z'`, each letter is replaced by another that has been read in a mapping table. The secret key will be this mapping table. For example, the word `'white'` will be encrypted `'black'` with the following key:

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	k	m	w	l	a	q	v	d	s	o	i	n	y	e	u	c	f	z	b	r	p	j

1. How many possible keys are there? Do you think the most powerful computer in the world can test all the keys in a reasonable time?

2. Propose an algorithm to randomly generate a key.

In English, the number of occurrences of the letter `'e'` is clearly greater than the number of occurrences of the letter `'z'`. For a given language, the letter frequency table is drawn up by measuring statistically the frequency of appearance of each letter on a long text.

3. Propose a method to distinguish if a ciphertext comes from a French or English plaintext.

Here are the frequency tables for the 20 most frequent letters, pair of letters and triplet of letters (computed from English novels):

letter	E	T	A	O	N	I	S	R	H	L	D
frequency (%)	12.56	9.15	8.08	7.47	7.38	7.24	6.59	6.42	5.27	4.04	3.99

	U	M	F	P	W	G	B	Y	C
	2.79	2.6	2.17	1.91	1.89	1.8	1.67	1.65	3.18

pair	TH	HE	IN	ER	AN	RE	ES	ON	ST	NT
frequency (%)	3.02	2.496	2.078	1.821	1.676	1.467	1.345	1.318	1.29	1.267

	EN	ED	ND	AT	TI	TE	OR	AR	HA	OF
	1.243	1.187	1.16	1.131	1.115	1.062	1.023	0.948	0.948	0.945

triplet	THE	AND	ING	ENT	ION	NTH	TER	INT	OFT	THA
frequency (%)	2.069	0.819	0.607	0.487	0.428	0.381	0.367	0.357	0.355	0.355

	ERE	TIO	HER	FTH	ETH	ATI	HAT	ATE	STH	EST
	0.352	0.335	0.327	0.321	0.315	0.307	0.295	0.286	0.281	0.277

4. Use the previous tables to decrypt the following message:

```
gwn hjhngv uhc hjhn ukn tjgw cknusl, pahgnhg dqg gwn warn
ao gwn takec sucn hnt, jl gwn wqhckncgw suh twa jl mkjsev
dnhg ah subjhm gwaln cknusl pasn gkqn.  ncmuk ueeuh ran
```

Exercice 3. (Vigenère – XVI$^{\text{th}}$ century)

In what follows, we define the $+$ and $-$ operations on the set $\{\mathtt{a}, \mathtt{b}, \cdots, \mathtt{z}\}$. To this end, we start by associating each letter to an integer with the following mapping:

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

The sum (resp. the difference) of two letters is then defined as the sum (resp. the difference) modulo 26 of their numerical values given by the previous mapping. For instance:

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

because

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

Naturally, the sum (resp. the difference) of two words containing the same number of letters is the sum (resp. the difference) letter by letter. For instance:

	`tiger`	$\longleftrightarrow$	19	8	6	4	17
+	`zebra`	$\longleftrightarrow$	25	4	1	17	0
=	`smhvr`	$\longleftrightarrow$	18	12	7	21	17

To encrypt a message $\mathbf{m}$ of arbitrary length with a key $\mathbf{k}$ of size t, we concatenate n times the key $\mathbf{k}$ with itself to obtain a string of the same length as the message; then we add $\mathbf{m}$ to the obtained string. If the length of the message to encrypt is not a multiple of t, then we complete the message with the appropriate number of letters. This operation is called a *padding*.

1. Decipher the following ciphertext with the key `mandela` :

   `qdhfeeianvvxseyofwtzwqrsxphempbqasiohlrynazufhxzctaajiehqwbupon`

2. How many keys of size t are there? Why is it better to choose a key uniformly at random in a set of all the string of 8 rather than choosing it in the dictionary?

3. A user propose to use a padding by adding as `'a'` as necessary. Why is it a bad idea? What if he choose an other letter than `'a'`? Propose a padding protocol which is more secure.

It is very common to use *padding* in cryptography. This exercise shows the importance of not choosing them haphazardly. In this case, we could have avoided using it; indeed, to encrypt the last block of the message of size $\leq t$, we could have truncated the key.

4. Let assume the size t of the key is exactly the size of the message to encrypt. Show that for all ciphertext $\mathbf{c}$ and message $\mathbf{m}$ both of size t, there is a unique key $\mathbf{k}$ such that $\mathbf{c}$ is the ciphertext of $\mathbf{m}$. Deduce that an attacker which only knows the ciphertext $\mathbf{c}$ cannot recover the corresponding plaintext $\mathbf{m}$, even if he has an infinite computing power. Is it still true when the key is strictly lower than the message, or equivalently, if the same key is reused several times to encrypt various messages?

Exercice 4. (The Scytale encryption – 404 BC)

An alternative way to the substitution encryption is the permutation encryption. A plaintext is encrypted by mixing the letters according to a permutation defined by a secret key.

Given a bi-dimensional table of p columns and q rows, for encrypt a message $\mathbf{m}$ of size $n \in [\![p(q-1)+1, pq]\!]$, one has to fill the table left to right and top to bottom with the letters from $\mathbf{m}$. The ciphertext is simply the message that it is read when reading the table top to bottom and left to right. To decipher the message, one only has to fill the table top to bottom and left to right then read the message left to right and top to bottom.

The secret key is the pair (p, q).

1. Decipher the following message with the key $(6, 5)$:

   `Aati lttsgl e o grnltlsodhi t.`

2. With a permutation encryption, is the letter frequency table changed? Propose a test to guess that we are dealing with a permutation encryption.

Exercice 5. (Indice de coïncidence – Bonus)

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

– **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 **c** le message chiffré avec une clé de taille t suivant :

```
hrixysthtweczxfkwegskaizdzilhrixysthtweczxfxzyfjxeyvk
ybxnyxyfferwiwpbxexwedsmqevcwslgivomfxfpmcwwxeipxlqu
```

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

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

3. Remplir 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 **c**. Avec cet information supplémentaire, décrypter le cryptogramme.