



Introduction à la sécurité

Chapitre 1

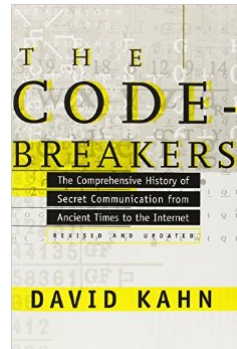
Le chiffrement par substitution



Pablo Rauzy <pr@up8.edu>
pablo.rauzy.name/teaching/is

Le chiffrement par substitution

- ▶ 16ème siècle AEC : premier document chiffré connu.
C'est une tablette d'argile, retrouvée en Irak, avec la recette d'un potier.
L'orthographe des mots y est changée (notamment, il manque les consonnes).



- ▶ *16ème siècle AEC* : premier document chiffré connu.
C'est une tablette d'argile, retrouvée en Irak, avec la recette d'un potier.
L'orthographe des mots y est changée (notamment, il manque les consonnes).
- ▶ *10ème à 8ème siècles AEC* : les scytales, chez les grecs.



- ▶ *16ème siècle AEC* : premier document chiffré connu.
C'est une tablette d'argile, retrouvée en Irak, avec la recette d'un potier.
L'orthographe des mots y est changée (notamment, il manque les consonnes).
- ▶ *10ème à 8ème siècles AEC* : les scytales, chez les grecs.
- ▶ *5ème siècle AEC* : plusieurs façons de chiffrer dont *atbash*, chez les Hébreux.
Il s'agit d'utiliser l'alphabet à l'envers (A devient Z, B devient Y, etc.)

- ▶ *16ème siècle AEC* : premier document chiffré connu.
C'est une tablette d'argile, retrouvée en Irak, avec la recette d'un potier.
L'orthographe des mots y est changée (notamment, il manque les consonnes).
- ▶ *10ème à 8ème siècles AEC* : les scytales, chez les grecs.
- ▶ *5ème siècle AEC* : plusieurs façons de chiffrer dont *atbash*, chez les Hébreux.
Il s'agit d'utiliser l'alphabet à l'envers (A devient Z, B devient Y, etc.)
- ▶ Après ça, arrivent des les premiers "vrais" systèmes de chiffrement.

Les premiers “vrais” cryptosystèmes

- ▶ Les premiers *cryptosystèmes* fonctionnent essentiellement par substitution.
- ▶ Il existe trois types de substitutions :
 - *monoalphabétique* : chaque lettre est remplacée par une autre,
 - *polyalphabétique* : une suite de substitutions monoalphabétiques est réutilisées en boucle,
 - *polygramme* : substitue des groupes de lettres par d'autres.

Chiffrement de César

- ▶ C'est le chiffrement par substitution le plus ancien connu (1er siècle AEC).
- ▶ Il était utilisé dans l'armée romaine (d'où son nom).
- ▶ Très faible, mais fonctionne bien en pratique grâce au faible taux d'alphabétisation dans la population.
- ▶ La clef est un nombre en 1 et 26 (A et Z).
- ▶ On opère un décalage circulaire de chaque lettre par la clef.
- ▶ Encore utilisé de nos jours avec ROT13 :).

Carré de Polybe

- ▶ Décrit pour la première fois vers 150 AEC.
- ▶ Utilisé par plusieurs civilisations de différentes manières.
- ▶ Dans sa forme la plus simple, il s'agit de substituer chaque lettre par ses coordonnées dans un carré de 5×5 .
- ▶ Une variante avec clef existe, où on se sert de la clef pour commencer à remplir le tableau.
- ▶ Une utilisation intéressante encore de nos jours de ce système est la communication simple à distance avec torches, drapeaux, ou du son.

	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I/J	K
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

"informatique" devient "243321344232114424414515"

Carré de Polybe

- ▶ Décrit pour la première fois vers 150 AEC.
- ▶ Utilisé par plusieurs civilisations de différentes manières.
- ▶ Dans sa forme la plus simple, il s'agit de substituer chaque lettre par ses coordonnées dans un carré de 5×5 .
- ▶ Une variante avec clef existe, où on se sert de la clef pour commencer à remplir le tableau (ci-dessous avec la clef "wikipedia").
- ▶ Une utilisation intéressante encore de nos jours de ce système est la communication simple à distance avec torches, drapeaux, ou du son.

	1	2	3	4	5
1	W	I/J	K	P	E
2	D	A	B	C	F
3	G	H	L	M	N
4	O	Q	R	S	T
5	U	V	X	Y	Z

"informatique" devient "123525414334224512424115"

- ▶ Au 9ème siècle, Al-Kindi écrit *Risalah fi Istikhraj al-Mu'amma* (Manuscrit sur le déchiffrement de messages cryptographiques).
- ▶ Il y fait la plus ancienne description de la méthode d'analyse de fréquences, et l'applique au chiffrement de César.

[illegible]

L'analyse de fréquences

- ▶ Dans chaque langue, il y a des lettres qui reviennent plus souvent que d'autres.
- ▶ Par exemple en français le 'E' est la lettre la plus courante.

L'analyse de fréquences

- ▶ Dans chaque langue, il y a des lettres qui reviennent plus souvent que d'autres.
- ▶ Par exemple en français le 'E' est la lettre la plus courante.
- ▶ On peut donc supposer que la lettre qui revient le plus dans le chiffré est un 'E', si on sait que le clair est en français.
- ▶ Si on ne connaît pas la langue d'origine du message, on peut calculer la fréquence de chaque lettre, et comparer cela avec les fréquences connues pour différentes langues.
- ▶ Il est aussi possible de le faire pour des tuples de lettre afin d'être encore plus précis.

Le chiffrement de Vigenère

- ▶ Cryptosystème polyalphabétique décrit en 1586 dans le *Traité sur les chiffres* de Blaise de Vigenère.
 - On trouve aussi une méthode similaire dans un texte de Giovan Battista Bellaso datant de 1533.
- ▶ Premier chiffrement à réellement introduire la notion de *clef*.
- ▶ La clef pouvant être au moins aussi longue que le message (en prenant comme référence un livre par exemple), le chiffrement de Vigenère résiste complètement à l'analyse de fréquences.
- ▶ En pratique, ce chiffrement a tenu près de 300 ans, jusqu'en 1863 quand Friedrich Kasiski publie une cryptanalyse du système.

Exemple

► Message : “introduction à la sécurité”

► Clef : “Vigenère”

m	I	N	T	R	O	D	U	C	T	I	O	N	A	L	A	S	E	C	U	R	I	T	E
k	V	I	G	E	N	E	R	E	V	I	G	E	N	E	R	E	V	I	G	E	N	E	R
c	D	V	Z	V	B	H	L	G	O	Q	U	R	N	P	R	W	Z	K	A	V	V	X	V

	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
A	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
B	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	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	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

- On suppose qu'on a uniquement les 26 lettres majuscules de l'alphabet.
1. Écrire une fonction C qui permet de chiffrer avec la méthode de Vigenère.
 2. Écrire une fonction C qui permet de déchiffrer avec cette même méthode.
 3. À quelle catégorie d'attaques ce cryptosystème peut résister ? Pourquoi ?

- ▶ On suppose qu'on a uniquement les 26 lettres majuscules de l'alphabet.
- 1. Écrire une fonction C qui permet de chiffrer avec la méthode de Vigenère.
- 2. Écrire une fonction C qui permet de déchiffrer avec cette même méthode.
- 3. À quelle catégorie d'attaques ce cryptosystème peut résister ? Pourquoi ?
- 4. Écrire une fonction C qui affiche la clef si on lui donne une paire clair/chiffré.

- ▶ Il est possible de casser le chiffrement de Vigenère, mais il faut réduire le problème à l'analyse de fréquences.

- ▶ Il est possible de casser le chiffrement de Vigenère, mais il faut réduire le problème à l'analyse de fréquences.
- ▶ C'est à dire se ramener dans le cas du chiffrement de César :
 - si la clef est de longueur l , $\forall d < l$, soit c_d la suite des lettres à un indice $i \equiv d \pmod l$,
 - alors c_d est un chiffré de César utilisant comme décalage celui correspondant à la d ème lettre de la clef de Vigenère.
- ▶ Il faut donc retrouver la longueur de la clef.

Retrouver la longueur de la clef

- ▶ Pour retrouver la longueur de la clef, il y a besoin d'avoir un chiffré bien plus long que la clef.
- ▶ Ensuite il faut trouver des séquences de lettres qui se répètent.
- ▶ Si une séquence de lettre se répète plusieurs fois cela veut dire :
 - soit que la même séquence du clair a été chiffré avec la même partie de la clef,
 - soit que par hasard des séquences différentes du clair se retrouve chiffré de manière identique.
- ▶ En pratique sur les textes en langue naturelle, il suffit de trouver des séquences d'au moins 3 lettres pour que la probabilité de la seconde option soit très faible.

Trouver des séquences répétées

- Prenons un exemple (tiré de Wikipédia) :

```
KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLP  
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWIXIZ  
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUEAPYMEKQHUIDUXFP  
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBNLGOYL  
SKMTEFVJJTWMMFMPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEYOEEKCPJR  
GPMURSKHFRSEIUEVGOCWIXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL  
WNOJNSIOFRWUCCEWVKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC  
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT  
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTJKNNE  
DCLDHWTTYIDGMVRDGMPLSWGJLAGOEKJOFEKUYTAANYTDWIYBNLNYP  
WEBFNLFYNAJEBFR
```

- Distances entre les répétitions :

Trouver des séquences répétées

► Prenons un exemple (tiré de Wikipédia) :

KQWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWIXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBNLGOYL
SKMTEFVJJTWMMFMPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEYOYEKCPJR
GPMURSKHFRSEIUEVGOCWIXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCEWVKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTJEKNEE
DCLDHWTTYIDGMVRDGMPLSWGJLAGOEKJOFEKUYTAANYTDWIYBNLNYNP
WEBFNLFYNAJEBFR

► Distances entre les répétitions :

- WUU : 95

Trouver des séquences répétées

► Prenons un exemple (tiré de Wikipédia) :

KQWQEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKYOSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWIXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBNLGOYL
SKMTEFVJJTWMMFMPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEYEEKCPJR
GPMURSKHFRSEIUEVGOCWIXIZAYGOSAANYDOEYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCEWVKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTJKNNE
DCLDHWTTYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYNP
WEBFNLFYNAJEBFR

► Distances entre les répétitions :

- WUU : 95
- EEK : 200

Trouver des séquences répétées

► Prenons un exemple (tiré de Wikipédia) :

KQWQEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLP
NCMUEKQCTESWREEKYOSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBNLGOYL
SKMTEFVJJTWMMFMPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEYEEKCPJR
GPMURSKHFRSEIUEVGOCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCEWVKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTJKNNE
DCLDHWTTYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYP
WEBFNLFYNAJEBFR

► Distances entre les répétitions :

- WUU : 95
- EEK : 200
- WXIZAYG : 190

Trouver des séquences répétées

 Prenons un exemple (tiré de Wikipédia) :

KQWFEVJJPJJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLP
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBNLGOYL
SKMTEFVJJTWMMFMPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEYEEKCPJR
GPMURSKHFRSEIUEVGOCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCEWVKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTJKNNE
DCLDHWTTYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYP
WEBFNLFYNAJEBFR

 Distances entre les répétitions :

- WUU : 95
- EEK : 200
- WXIZAYG : 190
- NUOCZGM : 80

Trouver des séquences répétées

 Prenons un exemple (tiré de Wikipédia) :

KQWFEVJJPJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLP
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUEAPYMEKQHUIDUXFP
GUYTSMFTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBNLGOYL
SKMTEFVJJTWMMFMPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEYEEKCPJR
GPMURSKHFRSEIUEVGOCYWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTJKNNE
DCLDHWTTYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYP
WEBFNLFYNAJEBFR

 Distances entre les répétitions :

- WUU : 95
- EEK : 200
- WXIZAYG : 190
- NUOCZGM : 80
- GMU : 90

Trouver des séquences répétées

► Prenons un exemple (tiré de Wikipédia) :

KQWFEVJJPJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKYOSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBNLGOYL
SKMTEFVJJTWMMFWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEYOYEKCPJR
GPMURSKHFRSEIUEVGOCYWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTCJKNEE
DCLDHWTTYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYP
WEBFNLFYNAJEBFR

► Distances entre les répétitions :

- WUU : 95
- EEK : 200
- WXIZAYG : 190
- NUOCZGM : 80
- GMU : 90
- DOEOY : 45

Trouver des séquences répétées

► Prenons un exemple (tiré de Wikipédia) :

KQWFEVJJPJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKYOSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYL
SKMTEFVJJTWMMFMPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEYOYEKCPJR
GPMURSKHFRSEIUEVGOCYWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTJEKNEE
DCLDHWTTYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYP
WEBFNLFYNAJEBFR

► Distances entre les répétitions :

- WUU : 95
- EEK : 200
- WXIZAYG : 190
- NUOCZGM : 80
- GMU : 90
- DOEOY : 45

► $\text{PGCD}(95, 200, 190, 80, 90, 45) = 5$

Produire les chiffrés de César

► Soit c_i le texte des lettres à positions égale à i modulo 5 :

- c_0 : KFJKKWMWFWSNKSIAAWGFWYSYSJGWJAKDGMHZWGDEJWLSFWWMSSTZE KGSSVWYAEWMLWSWSDGWASKUMDMATFXIDWWWTHWMMLWWJDWDDSAKKAWLWLJ
- c_1 : QVUGJUKRGUVCQWOWXPPQXGXNPNWUUPQUUTNGEKCCQVGKVWPTPKNGO CPKEGXGNOUEXNIUWGOGFIQCPGNVOQVKWCRGONTFGKGOGKCTGGWGJUNINEFE
- c_2 : OJULIXJLHULMCRYCYXNJHIFCCNYZUUYHXYFUMYMTFCBOMJMNMXFUMY PMHIOIOYYNBYOOCKMCNFUHPVAYLYCFLCCIMVYCYYBFJNNLYMMGOOYYYYBYB
- c_3 : WPNMNFBFUMPUTESTOPTBTZFSTTTGNQMUFTEFORTEVBUPYTJFMHFFODE JUFUYZSDJHFVJFCVURMVDCFSVMFFUJNWUFUMBOTTBBFTEDYVPJEFTTBNBNF
- c_4 : EUUEMQGNDBSEESUTLCGDANESUGREEEIPSSCURERDSNLETMERSSCOE RRRECAAOLAE LNREICUANEECUEAMNANEOLTSaucnQNTTEHIRLLEEADNPnAR

Analyse de fréquences

- ▶ Pour simplifier, on va le faire très simplement en cherchant juste les 'E' (qui est la lettre la plus courante en français).
- ▶ Avec la commande **fold** je peux couper le chiffré par ligne de 5 lettres.
- ▶ Avec la commande **cut** je peux isoler les caractères de la colonne i pour obtenir c_i .
- ▶ Ensuite avec les commandes **sort** et **uniq**, je peux trouver la lettre la plus utilisée.
- ▶ Si cette lettre est 'E', je retrouve le décalage du chiffrement de César.

Analyse de fréquences

- ▶ Pour simplifier, on va le faire très simplement en cherchant juste les 'E' (qui est la lettre la plus courante en français).
- ▶ Avec la commande **fold** je peux couper le chiffré par ligne de 5 lettres.
- ▶ Avec la commande **cut** je peux isoler les caractères de la colonne i pour obtenir c_i .
- ▶ Ensuite avec les commandes **sort** et **uniq**, je peux trouver la lettre la plus utilisée.
- ▶ Si cette lettre est 'E', je retrouve le décalage du chiffrement de César.
 - c_0 : W qui donnerait S
 - c_1 : G qui donnerait C
 - c_2 : Y qui donnerait U
 - c_3 : F qui donnerait B
 - c_4 : E qui donnerait A
- ▶ La clef serait donc SCUBA.

Déchiffrement

- ▶ On essaye de déchiffrer, si ça ne donne rien de cohérent, on peut regarder si par malchance le 'E' est arrivé en seconde position lors de l'analyse de fréquences, sinon on a cassé le code.
- ▶ Essayons :

SOUVENTPOURSAMUSERLESHOMMESDEQUIPAGEPRENNENTDESALBATROS
VASTESOISEAUXDESMERSQUISUIVENTINDOLENTSCOMPAGNONSDEVOYA
GELENAVIREGLISSANTSURLESGOUFFRESAMERSAPEINELESONTILSDEP
OESSURLESPLANCHESQUECESROISDELAZURMALADROITSETHONTEUXL
AISSENTPITEUSEMENTLEURSGRANDESAILESBLANCHESCOMMEDESAVIR
ONSTRAINERACOTEDEUXCEVOYAGEURAILECOMMEILESTGAUCHEETVEUL
ELUINAGUERESIBEAUQUILESTCOMIQUEETLAIDLUNAGACESONBECAVEC
UNBRULEGUEULELAUTREMIMEENBOITANTLINFIRMEQUIVOLAITLEPOET
EESTSEMBLABLEAUPRINCEDESNUESQUIHANTELETEMPETEETSERITDE
LARCHEREXILESURLAUMILIEUDES HUEESSESAILESDEGEANTLEMP
ECHENTDEMARCHER

Remise en forme

Souvent, pour s'amuser, les hommes d'équipage
Prennent des albatros, vastes oiseaux des mers,
Qui suivent, indolents compagnons de voyage,
Le navire glissant sur les gouffres amers.

À peine les ont-ils déposés sur les planches,
Que ces rois de l'azur, maladroits et honteux,
Laissent piteusement leurs grandes ailes blanches
Comme des avirons traîner à côté d'eux.

Ce voyageur ailé, comme il est gauche et veule !
Lui, naguère si beau, qu'il est comique et laid !
L'un agace son bec avec un brûle-gueule,
L'autre mime, en boitant, l'infirme qui volait !

Le Poète est semblable au prince des nuées
Qui hante la tempête et se rit de l'archer ;
Exilé sur le sol au milieu des huées,
Ses ailes de géant l'empêchent de marcher.

L'Albatros, de Charles Baudelaire.