Histoire de l'informatique

Chapitre 6

Histoire de la cryptologie

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

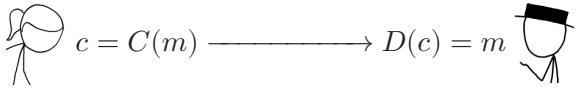
Histoire de la cryptologie

La cryptologie

- Étymologiquement, c'est "la science du secret".
- Un art ancien : les premiers documents chiffrés qu'on retrouve datent de l'Antiquité.
- Une science récente : sujet de recherche académique seulement depuis les années 60.
- La cryptologie étudie notamment
  - la *confidentialité*,
  - l'*authentification*,
  - la *non-répudiation*,
  - l'*intégrité*,
  - la *preuve à divulgation nulle de connaissance*, et
  - l'*anonymat*.
- Ses deux branches principales sont
  - la *cryptographie*, et
  - la *cryptanalyse*.

La cryptographie

- Étymologiquement, “l’écriture secrète”.
- Le but de cette discipline est de protéger les messages, en assurant leurs
  - confidentialité,
  - authenticité, et
  - intégrité.
- La cryptographie moderne utilise des *clefs*.
- Il y a deux grandes familles :
  - la cryptographie *symétrique* (à clef *secrète*), et
  - la cryptographie *asymétrique* (à clefs *publique* et *privée*).
- La cryptographie s’occupe principalement de la mise au point d’*algorithmes* et de *protocoles* permettant le chiffrement, de déchiffrement, et l’échange de messages.

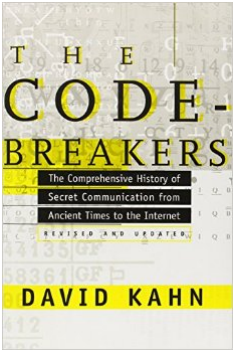


La cryptanalyse

- Étymologiquement, “défaire le secret”.
- Le but de cette discipline est de casser la cryptographie.
  - C’est à dire décrypter un message chiffré, sans connaître la clef de chiffrement.
  - On appelle ce processus une *attaque*.
- On catégorise souvent les attaques par ce à quoi l’attaquant a accès :
  - seulement des messages chiffrés,
  - certaines correspondances entre messages clairs et chiffrés,
  - certaines correspondances entre messages clairs choisis et leur chiffré,
  - certaines correspondances entre messages chiffrés choisis et leur clair,
- Il existe de nombreuses techniques d’attaques, mais on peut distinguer deux familles :
  - la cryptanalyse *classique*, et
  - les attaques par canaux auxiliaires.

Les prémices

- *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).



Les prémices

- *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.

Les prémices

- *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.)

Les prémices

- *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 jour avec ROT13 :).

- 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 “24332134423211442441451 5”

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	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”

Le premier manuscrit traitant de cryptanalyse

- 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.

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 et 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.
  - En pratique, 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.

- Message : “histoire de l’informatique”
- Clef : “Vigenère”

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

Cryptanalyse de Vigenère

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

Cryptanalyse de Vigenère

- 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 $i \equiv d \mod l$,
  - 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) :

```
KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYL
SKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJR
GPMURSKHFRSEIUEVGOYCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEE
DCLDHWTYYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYNP
WEBFNLFYNAJEBFR
```

- Distances entre les répétitions :

Trouver des séquences répétées

- Prenons un exemple (tiré de Wikipédia) :

```
KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYL
SKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJR
GPMURSKHFRSEIUEVGOYCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEE
DCLDHWTYYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYNP
WEBFNLFYNAJEBFR
```

- Distances entre les répétitions :
  - WUU : 95

Trouver des séquences répétées

- Prenons un exemple (tiré de Wikipédia) :

```
KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYL
SKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJR
GPMURSKHFRSEIUEVGOYCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEE
DCLDHWTYYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYNP
WEBFNLFYNAJEBFR
```

- Distances entre les répétitions :
  - WUU : 95
  - EEK : 200

Trouver des séquences répétées

- Prenons un exemple (tiré de Wikipédia) :

```
KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYL
SKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJR
GPMURSKHFRSEIUEVGOYCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEE
DCLDHWTYYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYNP
WEBFNLFYNAJEBFR
```

- Distances entre les répétitions :
  - WUU : 95
  - EEK : 200
  - WXIZAYG : 190

Trouver des séquences répétées

- Prenons un exemple (tiré de Wikipédia) :

```
KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYL
SKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJR
GPMURSKHFRSEIUEVGOYCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEE
DCLDHWTYYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYNP
WEBFNLFYNAJEBFR
```

- Distances entre les répétitions :
  - WUU : 95
  - EEK : 200
  - WXIZAYG : 190
  - NUOCZGM : 80

Trouver des séquences répétées

- Prenons un exemple (tiré de Wikipédia) :

```
KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYL
SKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJR
GPMURSKHFRSEIUEVGOYCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEE
DCLDHWTYYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYNP
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) :

```
KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYL
SKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJR
GPMURSKHFRSEIUEVGOYCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEE
DCLDHWTYYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYNP
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) :

KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYL
SKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJR
GPMURSKHFRSEIUEVGOYCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEE
DCLDHWTYYIDGMVRDGMPLSWGJLAGOEEKJOFEKUYTAANYTDWIYBNLNYNP
WEBFNLFYNAJEBFR

- Distances entre les répétitions :
  - WUU : 95
  - EEK : 200
  - WXIZAYG : 190
  - NUOCZGM : 80
  - GMU : 90
  - DOEOY : 45
- PGCD(95, 200, 190, 80, 90, 45) = 5

- Soit c_i le texte des lettres à positions égale à i modulo 5 :
  - c_0 : KFJKKWMWFWSNKSKIAAWGFWYSYSJGWJAKDGMHZWGDEJWLSFWWMSSTZE KGSSVWYAEWMLWSWSDGWASKUMDMMATFXIDWWWTHWMMLWWJDWDDSAKKAWLWLJ
  - c_1 : QVUGJUKRGUVCQWOWXPPOQXGXNPNWUUPQUUTNGEKCCQVGKVWPTPKNGO CPKEGXGNOUEXNIUWGOGFIQCPGNVOQVKWCRGONTFGKGOGKCTGGWGJUNINEFE
  - c_2 : OJULIXJLHULMCRYCYXNJHIFCCNYZUUYHXYFUMYMTFCBOMJMNMXFUMY PMHIOIOYYNBYOOCKMCNFUHPVAYLYCFLCCIMVYCYYBFJNNLYMMGOOYYYYBYB
  - c_3 : WPNMNFBFUMPUTESTOPTBTZFSTTTGNQMUFTFORTEVBUPYTJFMHFFODE JUFUYZSDJHFVJFCVURMVDCFSVMFFUJNWUFUMBOTTBBFTEDYVPJEFTTBNFNF
  - c_4 : EUUEMQGNDBSEEESUTLCGDANESUGREEEIPSSCURERDSNLETMERSSCOE RRRECAAOLAELNREICUANEECUEAMNANEOLTSAUCNQNTTEEHIRLLEEADNPNAR

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.

- 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.

- 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
OSESSURLESPLANCHESQUECESROISDELAZURMALADROITSETHONTEUXL
AISSENTPITEUSEMENTLEURSGRANDESAILESBLANCHESCOMMEDESAVIR
ONSTRAINERACOTEDEUXCEVOYAGEURAILECOMMEILESTGAUCHEETVEUL
ELUINAGUERESIBEAUQUILESTCOMIQUEETLAIDLUNAGACESONBECAVEC
UNBRULEGUEULELAUTREMIMEENBOITANTLINFIRMEQUIVOLAITLEPOET
EESTSEMBLABLEAUPRINCEDESNUEESQUIHANTELATEMPETEETSERITDE
LARCHEREXILESURLESOLAUMILIEUDESHUEESSESAILESDEGEANTLEMP
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.

17ème siècle

- *1623 :* Dans son livre *De dignitate et augmentis scientiarum*, Francis Bacon expose une technique stéganographique qui consiste à représenter chaque lettre du texte en clair par un groupe de 5 lettres A ou B.
- *Grand chiffre du roi Louis XIV :* Les historiens disposent de quelques documents qui ont été chiffrés par ce qu'on nomme le Grand Chiffre du roi Louis XIV, et qui n'était en principe utilisé que pour des communications d'une importance extrême.
  - Un code utilisant 587 nombres différents et qui était si résistant qu'il déconcerta les cryptanalystes pendant des siècles.
  - Vers 1893, Étienne Bazeries réussi à casser ce code après 3 ans de travail.
  - En fait, chaque nombre représentait une syllabe de la langue française plutôt qu'une seule lettre comme les codes traditionnels.

19ème siècle

- *1854 :* Un pionnier du télégraphe, Charles Wheatstone, apporte sa contribution à la cryptologie en inventant le chiffrement de Playfair, du nom de celui qui l'a fait connaître.

19ème siècle

- *1854 :* Un pionnier du télégraphe, Charles Wheatstone, apporte sa contribution à la cryptologie en inventant le chiffrement de Playfair, du nom de celui qui l'a fait connaître.
  - L'idée est de chiffrer les lettres par groupe de deux en utilisant une table 5 par 5 (sans la lettre W).
  - Si les 2 lettres sont identiques (ou s'il n'en reste qu'une) mettre un 'X' après la première lettre. Chiffrer la nouvelle paire ainsi constituée et continuer avec la suivante.
  - Si les lettres se trouvent sur la même ligne de la table, il faut les remplacer par celles se trouvant immédiatement à leur droite (en bouclant).
  - Si les lettres apparaissent sur la même colonne, les remplacer par celles qui sont juste en dessous (en bouclant).
  - Sinon, remplacer les lettres par celles se trouvant sur la même ligne, mais dans le coin opposé du rectangle défini par la paire originale.

19ème siècle

- *1854 :* Un pionnier du télégraphe, Charles Wheatstone, apporte sa contribution à la cryptologie en inventant le chiffrement de Playfair, du nom de celui qui l'a fait connaître.
- *1854 :* l'anglais Charles Babbage décrypte le chiffrement par substitution polyalphabétique de Vigenère, exposé en 1586.

19ème siècle

- *1854 :* Un pionnier du télégraphe, Charles Wheatstone, apporte sa contribution à la cryptologie en inventant le chiffrement de Playfair, du nom de celui qui l'a fait connaître.
- *1854 :* l'anglais Charles Babbage décrypte le chiffrement par substitution polyalphabétique de Vigenère, exposé en 1586.
- *1883 :* Le hollandais Auguste Kerckhoffs publie un ouvrage sur la cryptologie :
  La cryptographie militaire.
  Il y expose notamment quelques règles à respecter pour concevoir un bon système cryptographique, toujours valables actuellement, dont la principale est la suivante :
  la sécurité d'un système ne doit pas reposer sur le secret de la méthode de chiffrement.

Première Guerre Mondiale

- Pendant la guerre de 14-18, la maîtrise cryptographique des Français les aident considérablement à décrypter les messages ennemis, leur procurant un avantage très important.
- Le télégramme Zimmermann, intercepté en 1917 par le Royaume-Uni qui cryptanalysa son contenu, a accéléré l'entrée en guerre des États-Unis.
- La rapidité des transmissions a bénéficié des progrès du 19ème siècle, et est désormais instantanée, mais le déchiffrage des messages chiffrés, réalisé à la main, reste très lent, souvent plusieurs heures.

Seconde Guerre Mondiale

- La cryptologie a joué un rôle décisif pendant la Seconde Guerre mondiale.
- Les exploits des alliés en matière de cryptanalyse auraient permis d'écourter la guerre (de un à deux ans, selon certains spécialistes).

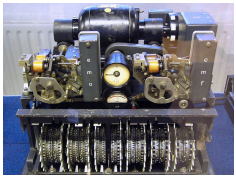
Seconde Guerre Mondiale

- La cryptologie a joué un rôle décisif pendant la Seconde Guerre mondiale.
- Les exploits des alliés en matière de cryptanalyse auraient permis d'écourter la guerre (de un à deux ans, selon certains spécialistes).
- La machine Enigma :
  - Originellement inventée pour les civils, elle retient l'attention des militaires.
  - Le chiffrement effectué par la machine Enigma est à la fois simple et astucieux ; il permet plus de 10^{16} combinaisons (clefs) possibles.

Enigma

Seconde Guerre Mondiale

- La cryptologie a joué un rôle décisif pendant la Seconde Guerre mondiale.
- Les exploits des alliés en matière de cryptanalyse auraient permis d'écourter la guerre (de un à deux ans, selon certains spécialistes).
- La machine Enigma :
  - Originellement inventée pour les civils, elle retient l'attention des militaires.
  - Le chiffrement effectué par la machine Enigma est à la fois simple et astucieux ; il permet plus de 10^{16} combinaisons (clefs) possibles.
- Le chiffre de Lorenz :
  - La machine Enigma est la plus connue, mais l'attaque de la machine de Lorenz, est à l'origine d'immenses progrès des sciences et des techniques.
  - Enigma n'a jamais été cassée, contrairement à Lorenz.



Lorenz

La cryptologie moderne

- Claude Shannon est considéré par plusieurs comme le père de la cryptographie mathématique, depuis la publication de son premier article en 1949.
- En formalisant certains concept fondamentaux, il fait la transition de la cryptologie de l'art à la science.
- Il donne notamment deux objectifs à la cryptologie : le secret et l'authentification.
- Il définit aussi deux types de secrets en fonction des adversaires :
  - adversaires disposant de ressources infinies : secret *théorique*,
  - adversaires ayant des ressources limitées : secret *pratique*.
- C'est principalement le secret théorique qu'il étudie en développant sa *théorie de l'information*.
- Il démontre notamment que le secret parfait ne peut être obtenu qu'avec une clef secrète dont la longueur est égale à la longueur de l'information à chiffrer.

Masque jetable

- Avant de continuer sur les chiffrements par bloc et par flot, voyons ce qu'on appelle le *masque jetable* (ou *chiffrement de Vernam*).
- Le principe est celui d'un chiffrement de Vigénère, mais avec une clef aussi longue que le message (en pratique on utilise des bits et un ou-exclusif plutôt qu'une addition modulaire sur l'alphabet).
- Ce type de chiffrement est *incassable*, à condition que :
  - la clef soit bien aussi longue que le message à chiffrer,
  - les caractères (ou bits) composant la clef soient choisis aléatoirement,
  - chaque clef ne soit utilisée qu'une seule fois (d'où le "jetable").

Masque jetable

- Avant de continuer sur les chiffrements par bloc et par flot, voyons ce qu'on appelle le *masque jetable* (ou *chiffrement de Vernam*).
- Le principe est celui d'un chiffrement de Vigénère, mais avec une clef aussi longue que le message (en pratique on utilise des bits et un ou-exclusif plutôt qu'une addition modulaire sur l'alphabet).
- Ce type de chiffrement est *incassable*, à condition que :
  - la clef soit bien aussi longue que le message à chiffrer,
  - les caractères (ou bits) composant la clef soient choisis aléatoirement,
  - chaque clef ne soit utilisée qu'une seule fois (d'où le "jetable").
- En effet, si on ne connaît que le texte chiffré, et que toutes les clefs sont équiprobables, alors tous les textes clairs de cette longueur sont possibles avec la même probabilité.
- Cette sécurité est inconditionnelle (elle ne repose pas sur une difficulté de calcul).

En pratique

- En pratique, il est presque impossible de mettre ce type de chiffrement en application.
- Il est très difficile de générer des clefs parfaitement aléatoires.
- La distribution des clefs est fortement problématique (sauf avec la cryptographie quantique).

Les années 1970

- Dans les années 1970, l'utilisation des ordinateurs a permis trois avancées majeures publiques (c'est-à-dire non secrètes ni contrôlées par les services de renseignements) :
  - le développement d'un standard public de chiffrement ;
  - le développement de l'échange de clefs Diffie-Hellman ;
  - le développement du chiffrement asymétrique.

Standards de chiffrements

- *1975 :* un groupe de recherche de IBM publie le projet *Data Encryption Standard* (DES) à l'invitation du NIST.
- Après des conseils et des modifications par la NSA, le chiffrement a été adopté et publié en tant que Federal Information Processing Standard (FIPS).
- Après des remplacement successifs par des dérivés de DES (DESX, 3DES), un concours ouvert a été lancé pour trouver le prochain standard.
- Le concours *Advanced Encryption Standard* (AES) est lancé en 1997.
- Il est remporté en 2000 par *Rijndael*, choisi parmi 15 propositions.

- Le concept de cryptographie asymétrique a été présenté publiquement pour la première fois par Whitfield Diffie et Martin Hellman à la *National Computer Conference* en 1976.
- En 1974, Ralph Merkle a travaillé sur des puzzles qui constituent la première construction à clef asymétrique, mais ses travaux ne sont publiés qu'en 1978.

Mise en œuvre

- En fait, dans l'article de 1976, Diffie et Hellman n'ont pas donné d'exemple de cryptosystème asymétrique (ils n'en avaient pas trouvé).
- C'est en 1978, que Ronald Rivest, Adi Shamir, et Leonard Adleman présentent *RSA*.
- Le système Merkle-Hellman est généralement considéré comme la première réalisation pratique d'un système de chiffrement à clef publique, mais il a été cassé par Shamir en 1982.

- En parallèle des recherches publiques, la GCHQ auraient mené des recherches secrètes ayant abouties dès le début des années 1970 aux concepts et outils de la cryptographie asymétrique.
  - En 1970, James Ellis invente le concept.
  - En 1973, Clifford Cocks invente l'algorithme de RSA.
  - En 1974, Malcolm Williamson invente un protocole d'échange de clef très proche de celui de Diffie et Hellman.
- Ces découvertes n'ont été rendues publiques par le GCHQ qu'en 1997.

Petite introduction à la cryptographie asymétrique

- La cryptographie *asymétrique* (aussi appelée *à clef publique et privée*) est une technique de chiffrement relativement récente.
- L'idée est de pouvoir communiquer de manière sûre sans dépendre de la communication d'une clef secrète.

La paire de clef

- Le principe est d'avoir deux clefs, de telle sorte que
  - un message chiffré avec la première clef ne peut être déchiffré qu'avec la seconde,
  - un message chiffré avec la seconde clef ne peut être déchiffré qu'avec la première.
- Par convention, l'une de ces clefs est appelée la *clef privée* et l'autre la *clef publique*.

Double fonctionnalité

- Un tel système permet de faire deux choses majeures.

1. Assurer la confidentialité du message :

2. Assurer l'authenticité de l'émetteur :

Double fonctionnalité

- Un tel système permet de faire deux choses majeures.

1. Assurer la confidentialité du message :
   - L'émetteur chiffre un message avec la clef publique du destinataire.
   - Seul le destinataire peut le déchiffrer avec sa clef privée.
2. Assurer l'authenticité de l'émetteur :
   - L'émetteur chiffre un message avec sa clef privée.
   - Le destinataire peut déchiffrer ce message avec la clef publique de l'émetteur.

- La première idée de mise en œuvre d'un tel système sont les *puzzles de Merkle*.
- Le principe est le suivant :
  1. L'émetteur crée plein de messages de la forme (id, key) avec id un identifiant unique et key une clef aléatoire pour un cryptosystème symétrique.

Exemple : puzzles de Merkle

- La première idée de mise en œuvre d'un tel système sont les *puzzles de Merkle*.
- Le principe est le suivant :
  1. L'émetteur crée plein de messages de la forme (id, key) avec id un identifiant unique et key une clef aléatoire pour un cryptosystème symétrique.
  2. L'émetteur chiffre tous ces messages avec des petites clefs de telle sorte qu'il soit possible de les déchiffrer en attaquant par force brute.

- La première idée de mise en œuvre d'un tel système sont les *puzzles de Merkle*.
- Le principe est le suivant :
  1. L'émetteur crée plein de messages de la forme (id, key) avec id un identifiant unique et key une clef aléatoire pour un cryptosystème symétrique.
  2. L'émetteur chiffre tous ces messages avec des petites clefs de telle sorte qu'il soit possible de les déchiffrer en attaquant par force brute.
  3. L'émetteur envoie tous les chiffrés au destinataire.

- La première idée de mise en œuvre d'un tel système sont les *puzzles de Merkle*.
- Le principe est le suivant :
  1. L'émetteur crée plein de messages de la forme (id, key) avec id un identifiant unique et key une clef aléatoire pour un cryptosystème symétrique.
  2. L'émetteur chiffre tous ces messages avec des petites clefs de telle sorte qu'il soit possible de les déchiffrer en attaquant par force brute.
  3. L'émetteur envoie tous les chiffrés au destinataire.
  4. Le destinataire en choisi un au hasard, et le casse par force brute.

- La première idée de mise en œuvre d'un tel système sont les *puzzles de Merkle*.
- Le principe est le suivant :
  1. L'émetteur crée plein de messages de la forme (id, key) avec id un identifiant unique et key une clef aléatoire pour un cryptosystème symétrique.
  2. L'émetteur chiffre tous ces messages avec des petites clefs de telle sorte qu'il soit possible de les déchiffrer en attaquant par force brute.
  3. L'émetteur envoie tous les chiffrés au destinataire.
  4. Le destinataire en choisi un au hasard, et le casse par force brute.
  5. Le destinataire envoie l'identifiant id qu'il a trouvé à l'émetteur, et garde la clef key.

Exemple : puzzles de Merkle

- La première idée de mise en œuvre d'un tel système sont les *puzzles de Merkle*.
- Le principe est le suivant :
  1. L'émetteur crée plein de messages de la forme (id, key) avec id un identifiant unique et key une clef aléatoire pour un cryptosystème symétrique.
  2. L'émetteur chiffre tous ces messages avec des petites clefs de telle sorte qu'il soit possible de les déchiffrer en attaquant par force brute.
  3. L'émetteur envoie tous les chiffrés au destinataire.
  4. Le destinataire en choisi un au hasard, et le casse par force brute.
  5. Le destinataire envoie l'identifiant id qu'il a trouvé à l'émetteur, et garde la clef key.
  6. L'émetteur retrouve la clef key qui va avec l'identifiant id qu'il reçoit.

Exemple : puzzles de Merkle

- La première idée de mise en œuvre d'un tel système sont les *puzzles de Merkle*.
- Le principe est le suivant :
  1. L'émetteur crée plein de messages de la forme (id, key) avec id un identifiant unique et key une clef aléatoire pour un cryptosystème symétrique.
  2. L'émetteur chiffre tous ces messages avec des petites clefs de telle sorte qu'il soit possible de les déchiffrer en attaquant par force brute.
  3. L'émetteur envoie tous les chiffrés au destinataire.
  4. Le destinataire en choisi un au hasard, et le casse par force brute.
  5. Le destinataire envoie l'identifiant id qu'il a trouvé à l'émetteur, et garde la clef key.
  6. L'émetteur retrouve la clef key qui va avec l'identifiant id qu'il reçoit.
  7. Les deux peuvent maintenant communiquer en chiffrant leur message de manière symétrique.

De nos jours

- La cryptologie est un domaine de recherche très actif.
- Parmi les nombreux thèmes de recherche qui animent la communauté scientifiques, on va brièvement parler :
  - de *chiffrement homomorphe*,
  - d'*attaques par canaux auxiliaires*.

Chiffrement homomorphe

- Comment faire si on a besoin de déléguer un calcul (par exemple dans le cloud), mais qu'on veut garder confidentielles les données sur lequel il porte ?
  - Un cryptosystème est dit homomorphe si il possède certaines caractéristiques algébriques qui permettent de réaliser des opérations sur les chiffrés.
  - Soit $F : A \to B$ une fonction de chiffrement, et soient $\odot_A$ et $\odot_B$ des opérations sur A et B.
  - On dit que le cryptosystème (F, F^{-1}) est homomorphe pour $\odot_A$ si on a $\odot_B$ telle que $F^{-1}(F(x) \odot_B F(y)) = x \odot_A y$.
  - On parle de cryptosystème *partiellement homomorphe* quand il commute avec un ensemble restreint d'opérations.

Attaques par canaux auxiliaires

- Un algorithme cryptographique peut être vu de deux façons :
  - d'un côté, c'est un objet mathématique abstrait,
  - de l'autre, c'est un code qui va finir par être exécuté sur du matériel.
- Le premier point de vue correspond à celui de la cryptanalyse classique.
- Le second correspond à celui de la sécurité physique.
- Les attaques physiques tirent partie des caractéristiques spécifiques des implémentations pour retrouver les paramètres secrets utilisés pendant le calcul.
- Ces attaques sont donc moins générales que celles de la cryptanalyse classique, mais elles sont aussi beaucoup plus puissantes.

- Il existe de nombreux types d'attaques physiques.
- À haut niveau, on peut déjà les classer selon deux axes :
  - *invasives* ou *non-invasives* : faut-il ouvrir ou casser en partie l'implémentation, ou au contraire n'exploiter que des informations naturellement (bien que non-intentionnellement) émises ?
  - *active* ou *passive* : l'attaque agit-elle sur l'implémentation ou se contente-t-elle de l'observer ?