Cryptanalyse

Monoalphabétique et Vigenère

Guénaël Renault

SALSA - LIP6/UPMC

22 février 2012

Part I

Cryptanalyse du Chiffrement Monoalphabétique

Cryptanalyse

Analyse de la sécurité d'un cryptosystème (théorique, complexité)

Attaquer une instance particulière d'un cryptosystème (retrouver la clé, un message clair)

Plusieurs niveau d'analyse

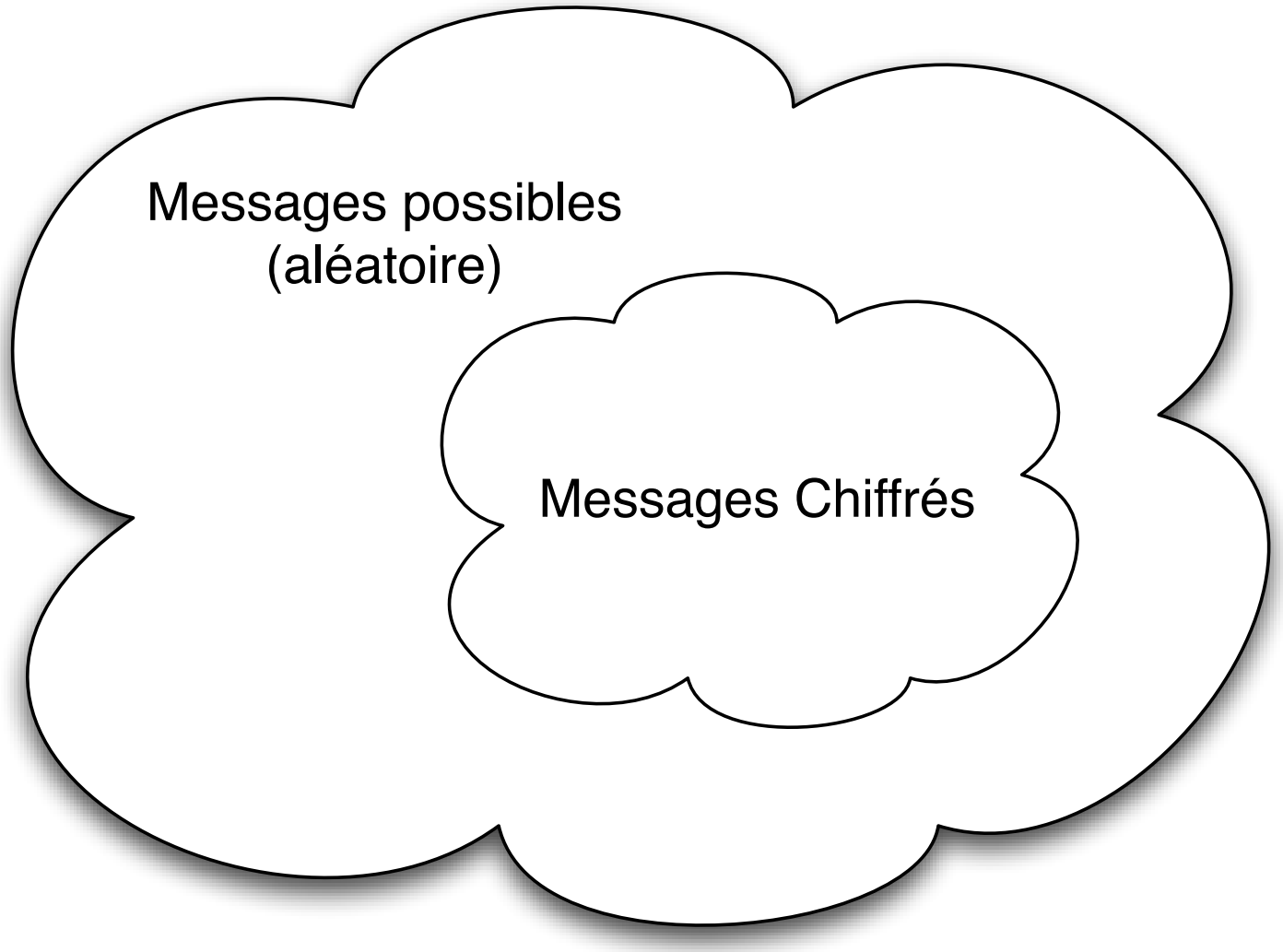
Clair/Chiffré inconnu

Couple Clair/Chiffré connu

Chiffré choisi (on a accès à un appareil de déchiffrement, boite noire)

Clair choisi (on a accès à un appareil de chiffrement, boite noire)

Cryptanalyse : Distingueur



⇒Comment distinguer les messages chiffrés des messages aléatoires ?

Distingueur : Rappels Probabilité 1

Définitions

Etant un ensemble fini Ω d'*évènements atomiques*. Une fonction $\mathbb{P}$ définie pour tout $\omega \in \Omega$ et à image dans $\mathbb{R}$ est appelée *probabiblité* dès que $0 \leqslant \mathbb{P}(\omega) \leqslant 1, \forall \omega \in \Omega$ et $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$.
Un *évènement E* est un sous-ensemble de Ω et sa probabilité d'apparition sera donné par

$$\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(E)$$

en particulier $\mathbb{P}(\emptyset) = 0$ et $\mathbb{P}(\Omega) = 1$

Exemples

- Pile ou face : $\Omega = \{P, F\}, \quad \mathbb{P}(F) = \frac{1}{2}, \mathbb{P}(P) = \frac{1}{2}$
- Deux dés de 6 :
  $\Omega = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}, \mathbb{P}((m,n)) = \frac{1}{36}, \mathbb{P}((1,m) = \frac{6}{36}$

Distingueur : Étude statistique des caractères

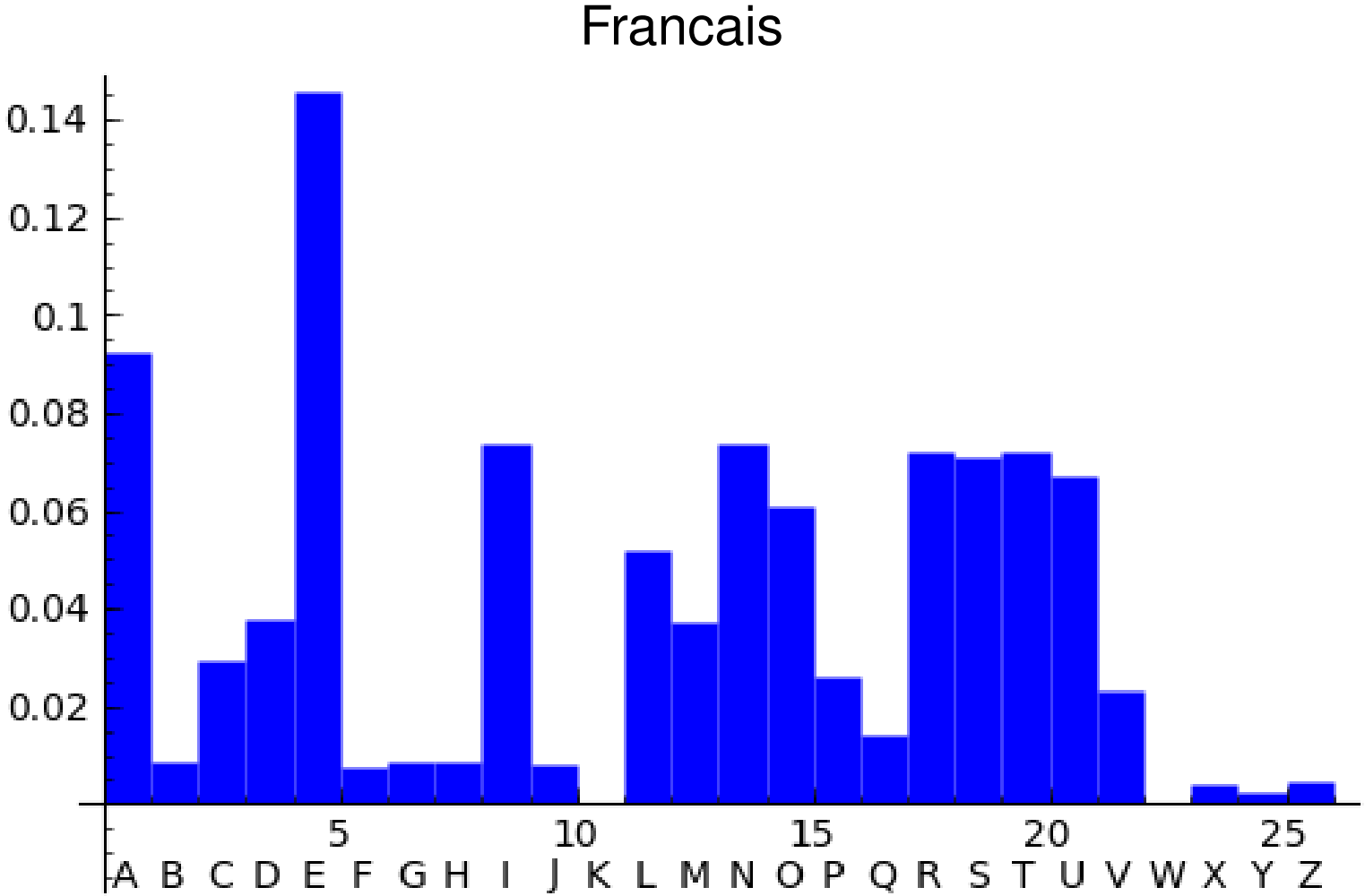
Soit c un caractère tiré aléatoirement. On cherche à analyser des évènements du genre $\{$*Le caractère c provient d'un texte français et est un A*$\}$.
$\Rightarrow$On ne peut pas faire une analyse sur tous les textes français possibles. Principe de la statistique, on effectue un *sondage* sur un ensemble fini et on extrapole les résultats.

Distingueur

- Dans un texte correspondant à un flux aléatoire de caractères on a $\mathbb{P}(\mathrm{A}) = 1/26$
- Dans un texte correspondant à un flux français de caractères on a $\mathbb{P}(\mathrm{A}) \simeq 9.2\%$

Distingueur : Étude statistique des caractères

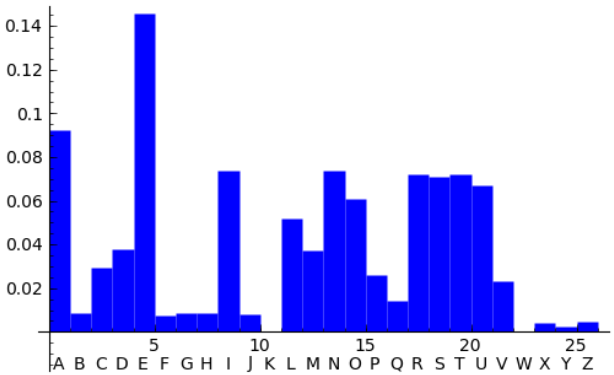
⇒La probabilité d'apparition des caractères est un distingueur entre les différentes langues et l'aléa.



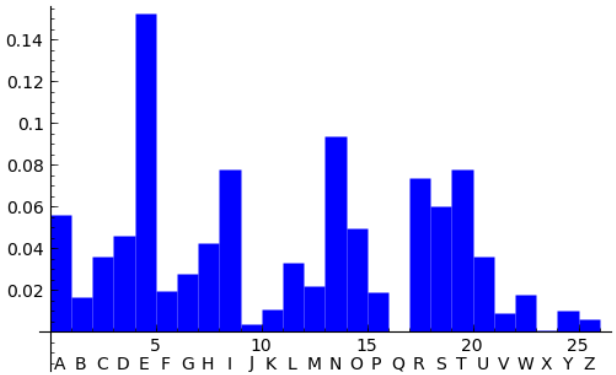
Distingueur : Étude statistique des caractères

⇒La probabilité d'apparition des caractères est un distingueur entre les différentes langues et l'aléa.

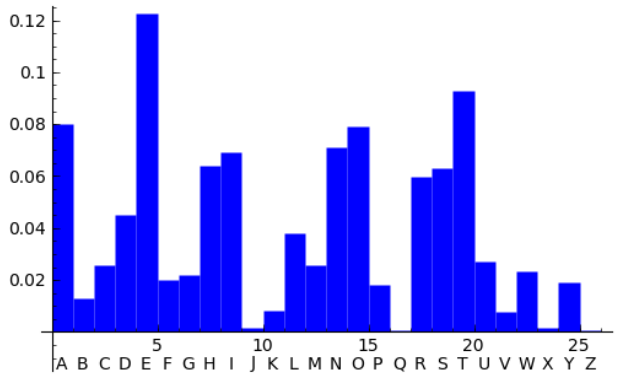
Francais



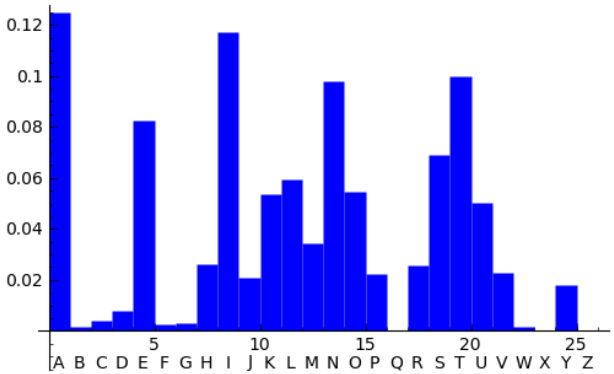
Allemand



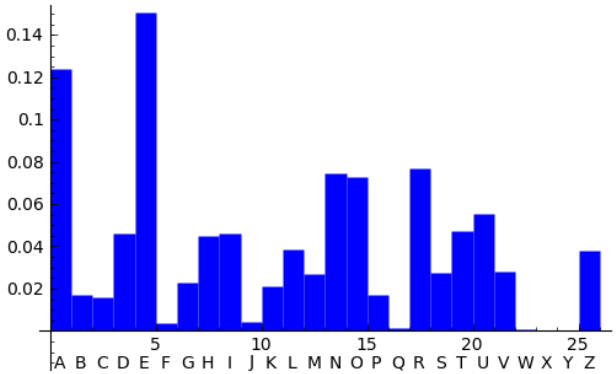
Anglais



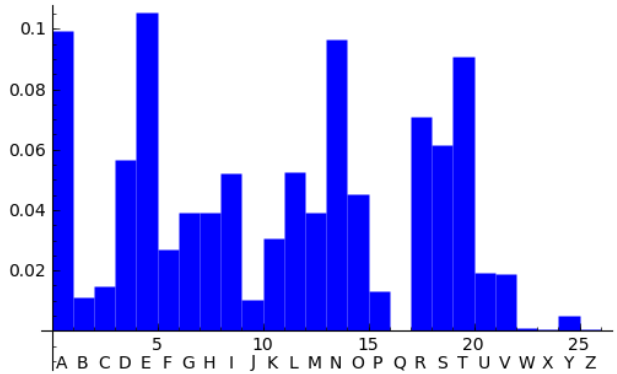
Finnois



Breton



Suedois



Cryptanalyse du mono-alphabétique

$\Rightarrow$Al Kindi ($\simeq$ 800) explique dans son ouvrage de cryptanalyse la méthode de l'étude des fréquences.
Il explique exactement ce que nous venons de voir : la distribution des caractères dans un texte permet de le caractériser.

Exemple : Cryptanalyse du décalage

Soit à retrouver le texte clair francais correspondant au chiffré suivant

QIUWIXNHISYIXYRENSYISESYISAEGESGIXEZ

REWTGNQIXYEGGTRUTLSIHIXEKIRRIIYHIXIX

YWTNXISKESYXYTZYIXQIXGTRRZSNGEYNTSXH

IAWTSYIYWIGMNKKWIIXUEWQIGTHICYWZSELI

SYQIXZNAWEYTZYEZQTSLHZXIOTZWIYXIWEIV

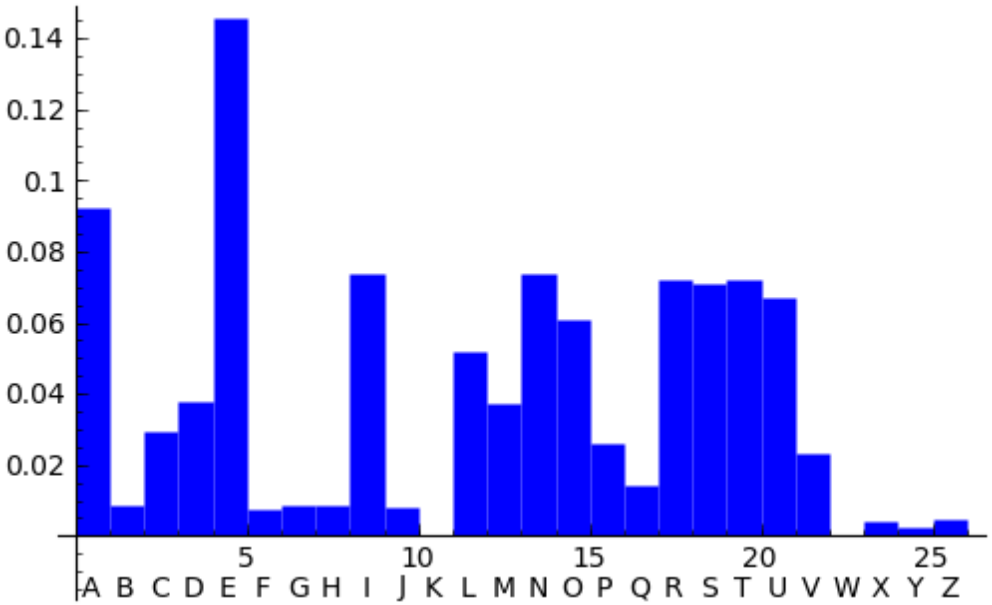
ZNUIHZSIAEQNXIHIGTRRZSNGEYNTSGMNKKWE

SYINQTGGZUIWEQIUEQEGIKQTYYESYKEGIEQE

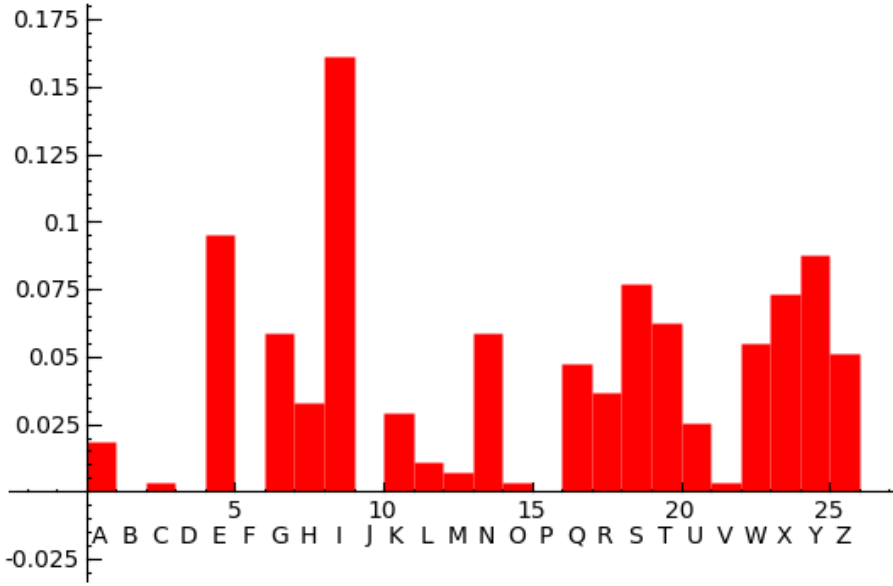
AISZIHIXYWTNXUEQRNIWX

Exemple : Cryptanalyse du décalage

Francais

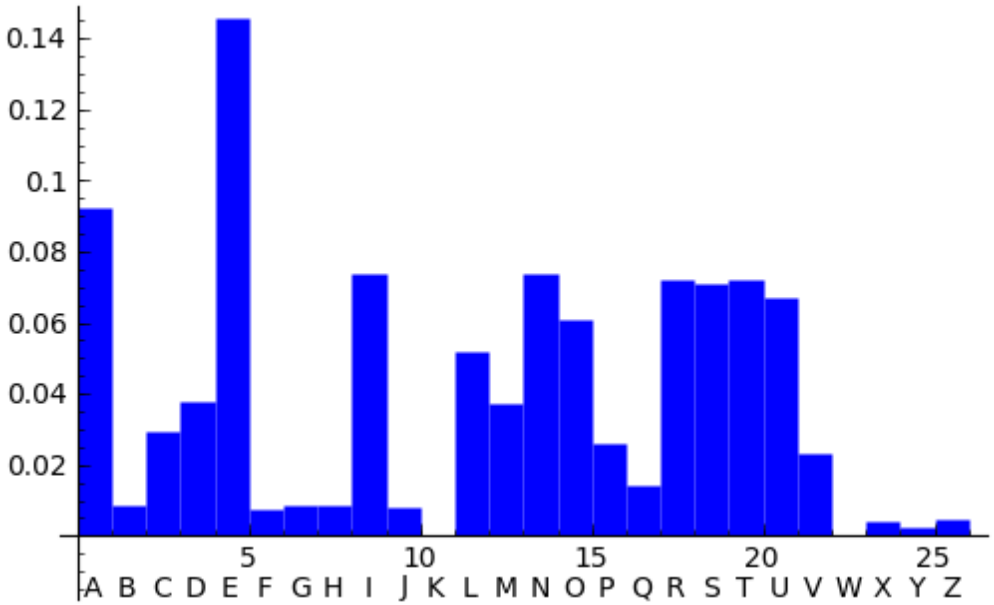


Chiffré

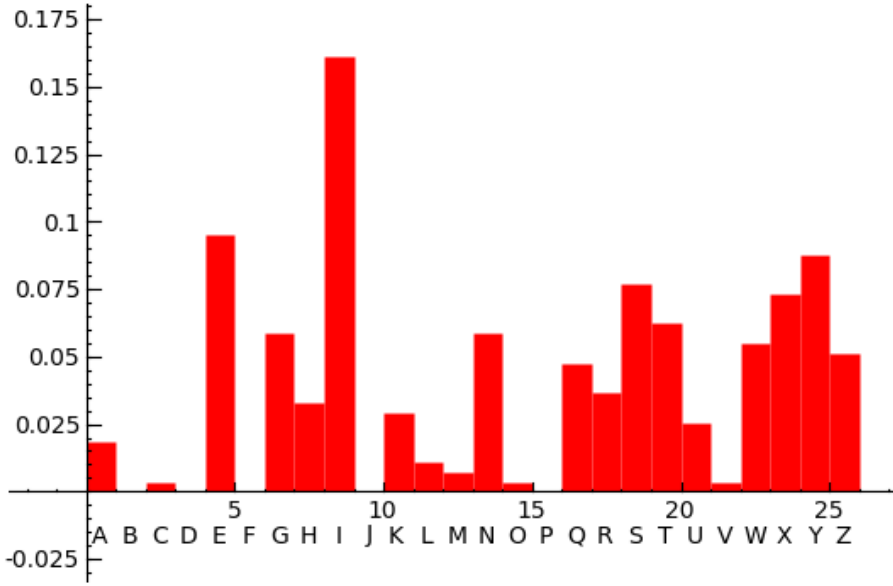


Exemple : Cryptanalyse du décalage

Francais



Chiffré



On lit la clé 4

Exemple : Cryptanalyse du mono-alphabétique

Soit à retrouver le texte clair francais correspondant au chiffré suivant

RDMJDQXIDBVDQVHUXBVDBUBVDBNUOUBODQUW

HUJPOXRDQVUOOPHMPGBDIDQUTDHHDDVIDQDQ

VJPXQDBTUBVQVPWVDQRDQOPHHWBXOUVXPBQI

DNJPBVDVJDOKXTTJDDQMUJRDOPIDSVJWBUGD

BVRDQWXNJUVPWVUWRPBGIWQDYPWJDVQDJUDZ

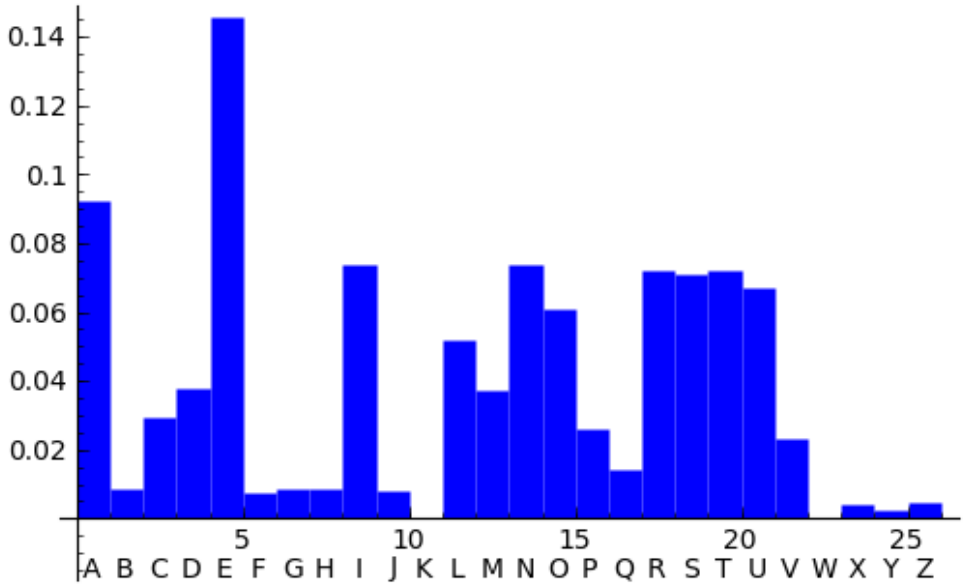
WXMDIWBDNURXQDIDOPHHWBXOUVXPBOKXTTJU

BVDXRPOOWMDJURDMURUODTRPVVUBVTUODURU

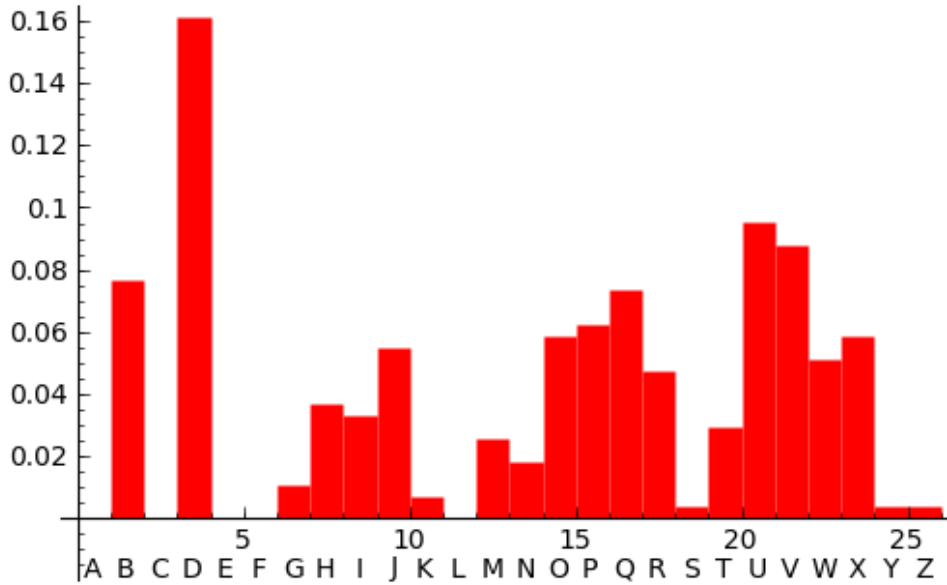
NDBWDIDQVJPXQMURHXDJQ

Exemple : Cryptanalyse du mono-alphabétique

Francais



Chiffré



La clé est la permutation :

[ULOIDTGKXYCRHBPMZJQVWNFSAE]

Exemple : Cryptanalyse du mono-alphabétique

$\Rightarrow$La cryptanalyse du chiffrement mono-alphabétique peut être vue comme une méthode *force-brute-assistée* où l'on va faire des hypothèses tout au long de l'attaque.
$\Rightarrow$Dans le cas général, il nous faut plus que la fréquence des lettres pour y arriver efficacement.

Bigrammes	Pourcentages	Bigrammes	Pourcentages
ES	3,15	LE	2,46
EN	2,42	DE	2,15
RE	2,09	NT	1,97
ON	1,64	TE	1,63
ER	1,63	SE	1,55

Conclusion 1

⇒Le chiffrement mono-alphabétique est très facile à attaquer !

Comment le sécuriser ?

- Substitution homophonique (1500-1750)

Steno (Venise) 1411

Conclusion 1

$\Rightarrow$Le chiffrement mono-alphabétique est très facile à attaquer !

Comment le sécuriser ?

- Substitution homophonique (1500-1750)
- Passer à la transposition (voir TD, Polybe, ADGVX)
- Passer au poly-alphabétique (Vigenère, la suite du cours)

$\Rightarrow$Comment mesurer formellement la sécurité d'un cryptosystème ?

Part II

Chiffrement de Vigenère et sa Cryptanalyse

Chiffrement de Vigenère (1523-1596)

$\Rightarrow$Chiffrement multi-mono-alphabétique

Chiffrement vu sur l'alphabet :

	A	T	T	A	Q	U	E	E	N	E	M	I	D	E	M	A	I	N
+	C	R	Y	P	T	O	C	R	Y	P	T	O	C	R	Y	P	T	O
=	C	K	R	P	J	I	G	V	L	T	F	W	F	V	K	P	B	B

Chiffrement vue sur les entiers modulo 26 :

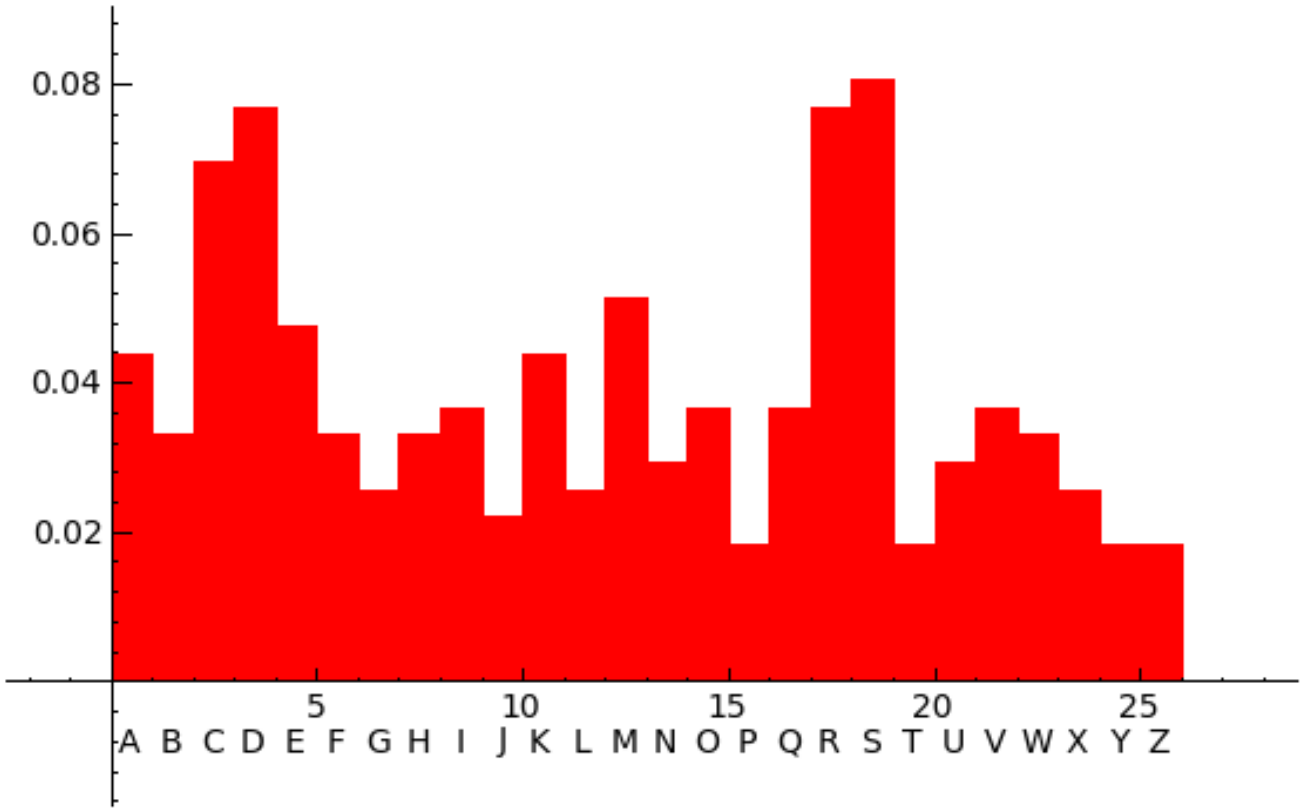
	0	19	19	0	16	20	4	4	13	4	12	8	3	4	12	0	8	13
+	2	17	24	15	19	14	2	17	24	15	19	14	2	17	24	15	19	14
=	2	10	17	15	9	8	6	21	11	19	5	22	5	21	10	15	1	1

$\Rightarrow$On note c_i le i-ème caractère dans l'alphabet avec $i \in \{0, \ldots, 25\}$.

Chiffrement de Vigenère (1523-1596)

⇒Chiffrement multi-mono-alphabétique

⇒Ne conserve pas les proprétés statistiques du chiffrement monoalphabétique.



Cryptanalyse Etape 1 : longueur de la clé

⇒Charles Babbage (1792-1871)
⇒Friedrich Wilhelm Kasiski (1805-1881)

Longueur de la clé : Test de Kasiski

KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPS
NCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZA
YGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFP
GUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYL
SKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJR
GPMURSKHFRSEIUEVGOYCWXIZAYGOSAANYDOEOYJLWUNHAMEBFELXYVL
WNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFC
MPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEIWCWODCCULWRIFT
WGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEE
DCLDHWTVBUVGFBIJG

⇒Le pgcd (problable) entre les différentes distance

Cryptanalyse Etape 2 : Finalisation Attaque

Principe

Soit T un texte dans un langage usuel représenté sous la forme d'une suite finie $T = (k_i)_{i \in I}$ de caractères. Pour toute suite extraite de T définie par $T' = (k_j)_{j \in J}$ avec $J = \{i = d \mod \ell \| i \in I\}$ (d et ℓ fixés a priori). Si T' est suffisamment longue la probabilité d'apparition p_i du caractère c_i dans T' est la même que dans T.

$\Rightarrow$On découpe le texte chiffré en bloc de ℓ caractères et on applique une cryptanalyse par décalage sur les colonnes !
$\Rightarrow$Les colonnes deviennent les suites extraites T'

Cryptanalyse Etape 2 : Finalisation Attaque

Attaque du chiffrement par décalage : Basée sur les fréquences

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

$\Rightarrow$Sous-textes plus courts $\Rightarrow$ analyse fréquences plus fastidieuse

$\Rightarrow$Plus difficile de deviner la clé, 26^{ℓ} tests où ℓ est la longueur de la clé

Cryptanalyse automatique : Indice de Coincidence

Définition

L'indice de coïncidence d'un texte est la probabilité de tirer un couple de lettres identiques au hasard.

$$IC = \sum_{i=0}^{25} \frac{C_2^{n_i}}{C_2^n} = \sum_{i=0}^{25} \frac{n_i(n_i-1)}{n(n-1)}$$

où n_i est le nombre de caractère c_i dans le texte et n est la longueur total de ce dernier.

William F. Friedman (1891 - 1969)

Cryptanalyse automatique : Indice de Coincidence

$\Rightarrow IC = \sum_{i=0}^{25} \frac{n_i(n_i-1)}{n(n-1)}$ distingue l'aléatoire $\Rightarrow$ attaque longueur de la clé

- Lorsque le texte est suffisamment long ($n \to \infty$) l'indice IC est donné par

$$IC \simeq \sum_{i=0}^{25} p_i^2 = 0.074 \text{ pour l'alphabet français}$$

où p_i est la probabilité d'apparition de la lettre numérotée i dans un texte en français.

- Lorsque les lettres sont distribuées aléatoirement, l'indice de coïncidence est faible

$$IC \simeq \sum_{i=0}^{25} (\frac{1}{26})^2 \simeq 0.038$$

Cryptanalyse automatique : Indice de Coincidence

$\Rightarrow IC = \sum_{i=0}^{25} \frac{n_i(n_i-1)}{n(n-1)}$ distingue l'aléatoire $\Rightarrow$ attaque longueur de la clé

Key Length	Average Index	Individual Indices of Coincidence
4	0.038	0.034, 0.042, 0.039, 0.035
5	0.037	0.038, 0.039, 0.043, 0.027, 0.036
6	0.036	0.038, 0.038, 0.039, 0.038, 0.032, 0.033
7	0.062	0.062, 0.057, 0.065, 0.059, 0.060, 0.064, 0.064
8	0.038	0.037, 0.029, 0.038, 0.030, 0.034, 0.057, 0.040, 0.039
9	0.037	0.032, 0.036, 0.028, 0.030, 0.026, 0.032, 0.045, 0.047, 0.056

Cryptanalyse automatique : IC Mutuelle

$\Rightarrow$IC mutuelle distingue l'aléatoire sur deux textes $\Rightarrow$ attaque sur la clé.

Définition

L'indice de coïncidence mutuelle entre deux textes t_1 et t_2 est la probabilité de tirer au hasard la même lettre dans t_1 et t_2.

$$ICM = \sum_{i=0}^{25} \frac{m_i n_i}{mn}$$

où m_i (resp. n_i) est le nombre de caractères c_i dans le texte t_1 (resp. t_2) et m (resp. n) la taille de ce dernier.

Cryptanalyse automatique : IC Mutuelle

$\Rightarrow ICM = \sum_{i=0}^{25} \frac{m_i n_i}{mn}$ pour distinguer l'aléatoire, attaque du décalage

- Propriété idem à l'indice de coïncidence
- Attaque des chiffrements par décalage par analyse successive de décalés

Cryptanalyse automatique : IC Mutuelle

Blocks		Shift Amount												
i	j	0	1	2	3	4	5	6	7	8	9	10	11	12
1	2	.025	.034	.045	.049	.025	.032	.037	.042	.049	.031	.032	.037	.043
1	3	.023	**.067**	.055	.022	.034	.049	.036	.040	.040	.046	.025	.031	.046
1	4	.032	.041	.027	.040	.045	.037	.045	.028	.049	.042	.042	.030	.039
1	5	.043	.021	.031	.052	.027	.049	.037	.050	.033	.033	.035	.044	.030
1	6	.037	.036	.030	.037	.037	.055	.046	.038	.035	.031	.032	.037	.032
1	7	.054	.063	.034	.030	.034	.040	.035	.032	.042	.025	.019	.061	.054
2	3	.041	.029	.036	.041	.045	.038	.060	.031	.020	.045	.056	.029	.030
2	4	.028	.043	.042	.032	.032	.047	.035	.048	.037	.040	.028	.051	.037
2	5	.047	.037	.032	.044	.059	.029	.017	.044	.060	.034	.037	.046	.039
2	6	.033	.035	.052	.040	.032	.031	.031	.029	.055	.052	.043	.028	.023
2	7	.038	.037	.035	.046	.046	.054	.037	.018	.029	.052	.041	.026	.037
3	4	.029	.039	.033	.048	.044	.043	.030	.051	.033	.034	.034	.040	.038
3	5	.021	.041	.041	.037	.051	.035	.036	.038	.025	.043	.034	.039	.036
3	6	.037	.034	.042	.034	.051	.029	.027	.041	.034	.040	.037	.046	.036
3	7	.046	.023	.028	.040	.031	.040	.045	.039	.020	.030	**.069**	.042	.037
4	5	.041	.033	.041	.038	.036	.031	.056	.032	.026	.034	.049	.029	.054
4	6	.035	.037	.032	.039	.041	.033	.032	.039	.042	.031	.049	.039	.058
4	7	.031	.032	.046	.038	.039	.042	.033	.056	.046	.027	.027	.036	.036
5	6	.048	.036	.026	.031	.033	.039	.037	.027	.037	.045	.032	.040	.041
5	7	.030	.051	.043	.031	.034	.041	.048	.032	.053	.037	.024	.029	.045
6	7	.032	.033	.030	.038	.032	.035	.047	.050	.049	.033	.057	.050	.021

Cryptanalyse automatique : IC Mutuelle

Blocks		Shift Amount												
i	j	13	14	15	16	17	18	19	20	21	22	23	24	25
1	2	.034	.052	.037	.030	.037	.054	.021	.018	.052	.052	.043	.042	.046
1	3	.031	.037	.038	.050	.039	.040	.026	.037	.044	.043	.023	.045	.032
1	4	.039	.040	.032	.041	.028	.019	**.071**	.038	.040	.034	.045	.026	.052
1	5	.042	.032	.038	.037	.032	.045	.045	.033	.041	.043	.035	.028	.063
1	6	.040	.030	.028	**.071**	.051	.033	.036	.047	.029	.037	.046	.041	.027
1	7	.040	.032	.049	.037	.035	.035	.039	.023	.043	.035	.041	.042	.027
2	3	.054	.040	.028	.031	.039	.033	.052	.046	.037	.026	.028	.036	.048
2	4	.047	.034	.027	.038	.047	.042	.026	.038	.029	.046	.040	.061	.025
2	5	.034	.026	.035	.038	.048	.035	.033	.032	.040	.041	.045	.033	.036
2	6	.033	.034	.036	.036	.048	.040	.041	.049	.058	.028	.021	.043	.049
2	7	.042	.037	.041	.059	.031	.027	.043	.046	.028	.021	.044	.048	.040
3	4	.037	.045	.033	.028	.029	**.073**	.026	.040	.040	.026	.043	.042	.043
3	5	.035	.029	.036	.044	.055	.034	.033	.046	.041	.024	.041	**.067**	.037
3	6	.023	.043	**.074**	.047	.033	.043	.030	.026	.042	.045	.032	.035	.040
3	7	.035	.035	.035	.028	.048	.033	.035	.041	.038	.052	.038	.029	.062
4	5	.032	.041	.036	.032	.046	.035	.039	.042	.038	.034	.043	.036	.048
4	6	.034	.034	.036	.029	.043	.037	.039	.036	.039	.033	**.066**	.037	.028
4	7	.043	.032	.039	.034	.029	**.071**	.037	.039	.030	.044	.037	.030	.041
5	6	.052	.035	.019	.036	.063	.045	.030	.039	.049	.029	.036	.052	.041
5	7	.040	.031	.034	.052	.026	.034	.051	.044	.041	.039	.034	.046	.029
6	7	.029	.035	.039	.032	.028	.039	.026	.036	**.069**	.052	.035	.034	.038

Cryptanalyse automatique : IC Mutuelle

Blocks		Shift Amount												
i	j	13	14	15	16	17	18	19	20	21	22	23	24	25
1	2	.034	.052	.037	.030	.037	.054	.021	.018	.052	.052	.043	.042	.046
1	3	.031	.037	.038	.050	.039	.040	.026	.037	.044	.043	.023	.045	.032
1	4	.039	.040	.032	.041	.028	.019	**.071**	.038	.040	.034	.045	.026	.052
1	5	.042	.032	.038	.037	.032	.045	.045	.033	.041	.043	.035	.028	.063
1	6	.040	.030	.028	**.071**	.051	.033	.036	.047	.029	.037	.046	.041	.027
1	7	.040	.032	.049	.037	.035	.035	.039	.023	.043	.035	.041	.042	.027
2	3	.054	.040	.028	.031	.039	.033	.052	.046	.037	.026	.028	.036	.048
2	4	.047	.034	.027	.038	.047	.042	.026	.038	.029	.046	.040	.061	.025
2	5	.034	.026	.035	.038	.048	.035	.033	.032	.040	.041	.045	.033	.036
2	6	.033	.034	.036	.036	.048	.040	.041	.049	.058	.028	.021	.043	.049
2	7	.042	.037	.041	.059	.031	.027	.043	.046	.028	.021	.044	.048	.040
3	4	.037	.045	.033	.028	.029	**.073**	.026	.040	.040	.026	.043	.042	.043
3	5	.035	.029	.036	.044	.055	.034	.033	.046	.041	.024	.041	**.067**	.037
3	6	.023	.043	**.074**	.047	.033	.043	.030	.026	.042	.045	.032	.035	.040
3	7	.035	.035	.035	.028	.048	.033	.035	.041	.038	.052	.038	.029	.062
4	5	.032	.041	.036	.032	.046	.035	.039	.042	.038	.034	.043	.036	.048
4	6	.034	.034	.036	.029	.043	.037	.039	.036	.039	.033	**.066**	.037	.028
4	7	.043	.032	.039	.034	.029	**.071**	.037	.039	.030	.044	.037	.030	.041
5	6	.052	.035	.019	.036	.063	.045	.030	.039	.049	.029	.036	.052	.041
5	7	.040	.031	.034	.052	.026	.034	.051	.044	.041	.039	.034	.046	.029
6	7	.029	.035	.039	.032	.028	.039	.026	.036	**.069**	.052	.035	.034	.038

$\Rightarrow$On termine en résolvant un système linéaire.

$$\begin{cases} \delta_3 = \delta_4 + 18 \\ \delta_3 = \delta_6 + 15 \\ \delta_4 = \delta_7 + 18 \\ \vdots \end{cases}$$

Conclusion 2

$\Rightarrow$Le chiffrement mono-alphabétique est très facile à attaquer !

Comment le sécuriser ?

- Substitution homophonique (1500-1750)
- Passer à la transposition (voir TD, Polybe, ADGVX)
- Passer au poly-alphabétique (Vigenère, la suite du cours)

Conclusion 2

$\Rightarrow$Le chiffrement mono-alphabétique est très facile à attaquer !
Le chiffrement de Vigenère ne semble pas beaucoup plus sûr !

Existe-t-il un cryptosystème inattaquable ?

Shannon

Claude Shannon (1916 - 2001) a publié deux articles de recherche en 1948 et 1949 donnant les fondations de la théorie de l'information et, plus généralement, de la cryptologie moderne. Il donne les première preuve de sécurité d'un cryptosystème en se basant sur des principes de probabilité et de statistique.

Définitions importantes

- Théorie de l'information
- Entropie d'un langage
- Chiffrement parfait

Chiffrement parfait

Intuition

Un cryptosystème sera dit chiffrement parfait lorsque la donnée d'un message chiffré ne révèle aucune fuite d'information sur la clé ou le message clair correspondant et aucune information non plus sur les textes chiffrés futurs.

Caractérisation

Supposons qu'un cryptosystème vérifie

$$\#\mathcal{K} = \#\mathcal{P} = \#\mathcal{C}$$

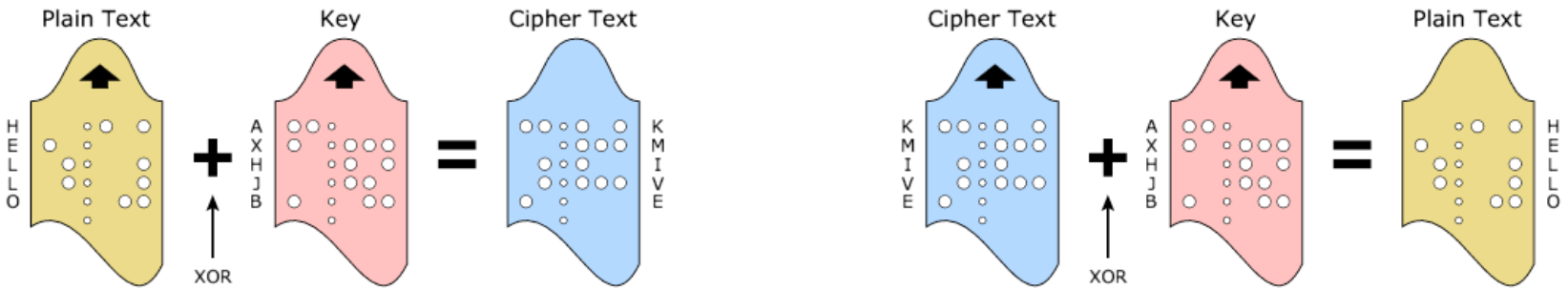
alors il sera chiffrement parfait ssi les deux conditions suivantes sont vérifiées:

- Toutes les clés sont utilisées avec même probabilité
- Pour tout couple $(m, c) \in \mathcal{P} \times \mathcal{C}$ il existe une unique clé k telle que $e_k(m) = c$.

Exemple : Vernam's One Time Pad

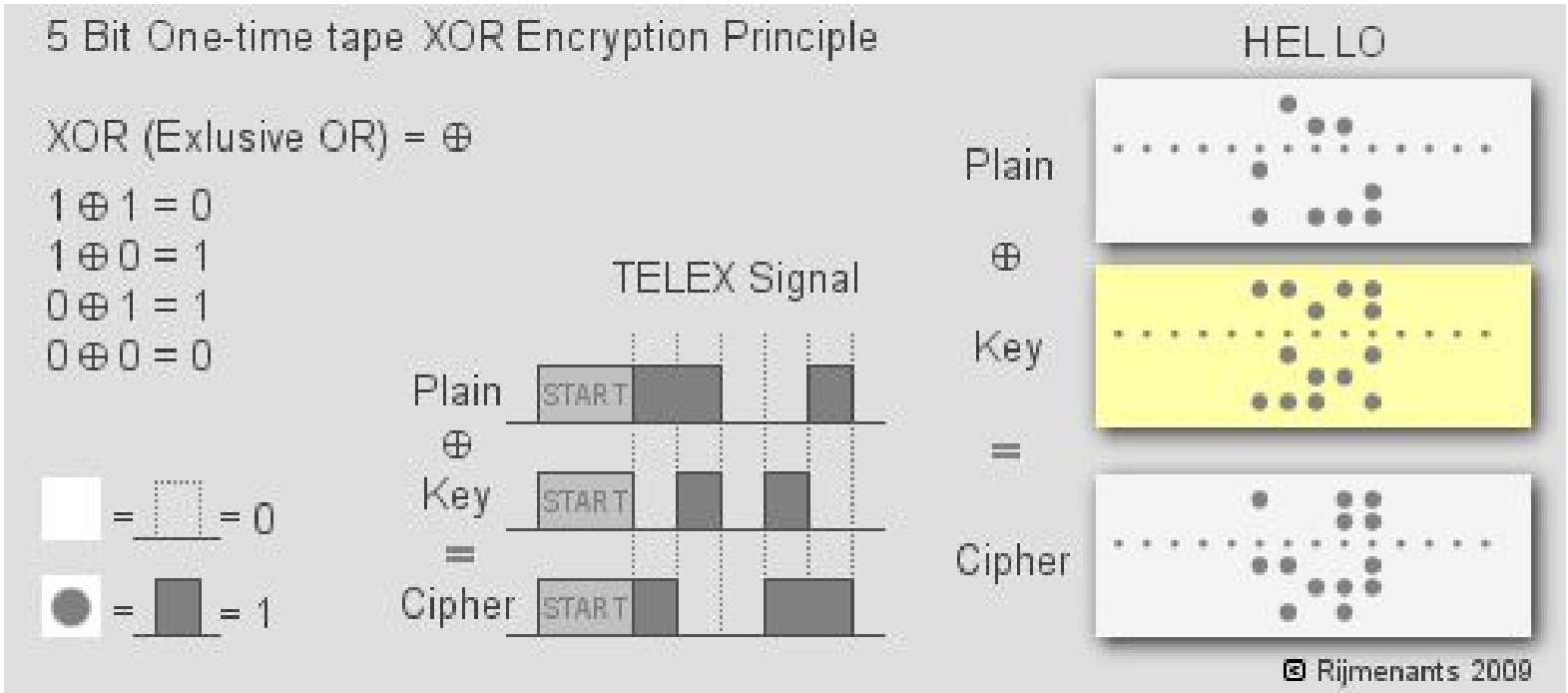
Gilbert S. Vernam (1890-1960), proposa le cryptosystème qui porte son nom en 1917 et fût déposé un brevet le concernant jusqu'en 1919 (US PATENT 1310719).
Le principe est simple : l'utilisation du XOR !
Les messages clairs et chiffrés, les clés seront des suites de bits de même longueur.

$$C[i] = M[i] \oplus K[i] \text{ et } M[i] = C[i] \oplus K[i]$$



Exemple : Vernam's One Time Pad

- C'est le seul cryptosystème à chiffrement parfait !
- Très peu pratique !
- Utilisé dans la cryptographie Top Secrète (téléphone rouge, valise diplomatique, militaire (Atomique)).
- Le principe est utilisé pour faire des chiffrements symétrique dépendant de générateur aléatoire.



Exemple : Vernam's One Time Pad

- La clé doit être aussi longue que le message
- Elle doit être aléatoire
- Elle doit être utilisée une unique fois
- Projet VENONA des USA pour écouter les discussions Russes utilisant un Two-Time Pad $\Rightarrow$ faiblesse !

Conclusion finale !

- La cryptographie parfaite est possible...
- mais impraticable !
- Les études sur le chiffrement parfait a permis de définir des standarts de chiffrement comme DES ou AES (voir le cours de Crypto/Secu en M1).