Cryptographie1

Historique

Auguste Kerckhoffs (1835-1903)

(1) Le système doit être matériellement, sinon mathématiquement, indéchiffrable ;

(2) Il faut qu'il n'exige pas le secret, et qu'il puisse sans inconvénient tomber entre les mains de l'ennemi ;

(3) La clef doit pouvoir en être communiquée et retenue sans le secours de notes écrites, et être changée ou modifiée au gré des correspondants ;

(4) Il faut qu'il soit applicable à la correspondance télégraphique ;

(5) Il faut qu'il soit portatif, et que son maniement ou son fonctionnement n'exige pas le concours de plusieurs personnes ;

(6) Enfin, il est nécessaire, vu les circonstances qui en commandent l'application, que le système soit d'un usage facile, ne demandant ni tension d'esprit, ni la connaissance d'une longue série de règles à observer.

Légèrement adaptés, ces points restent les axiomes fondamentaux de la cryptographie. En particulier

- *l'ennemi possède tous les détails de l'algorithme et il ne lui manque que la clef spécifique pour le chiffrement.*
- *Un chiffre basé uniquement sur le secret de l'algorithme n'a aucun intérêt et ne résiste pas à une fuite d'information.*

La sécurité est basée sur la clef. La méthode sera découverte tôt ou tard. De plus si l'algorithme est connu la communauté peut tester la solidité.

MÉTHODES HISTORIQUES

Les exemples suivants sont en français et utilisent un alphabet de 26 lettres.
Pour la facilité de lecture les espaces et ponctuations ont été conservées.
Le texte en clair est en minuscules, le texte chiffré en majuscules.
Récupérez les fichiers `mono.txt` (chiffré avec une permutation de lettres) et `poly.txt` (chiffré par la méthode de Vigenère).

(1) **Méthode de Jules César**
Il s'agit d'un décalage circulaire des lettres de l'alphabet.
Exemple
Clair abcdefghijklmnopqrstuvwxyz
Chiffre BCDEFGHIJKLMNOPQRSTUVWXYZA
(hal devient IBM).
Cassage :
• force brute : combien de possibilités ?
• début d'analyse de fréquences.
Particulièrement simple : en français la lettre la plus fréquente est le 'e'. Il suffit donc de compter les occurences des caractères.

Déchiffrez :
OQFFQ CGQEFUAZ QEF FDQE RMOUXQ, XM CGQEFUAZ EGUHMZFQ QEF BXGE PURRUOUXQ.
Ecrire un programme permettant le décryptage par force brute.
Ecrire un programme affichant l'histogramme (classé) des caractères présents dans un texte.
Utiliser un histogramme des caractères du texte chiffré pour choisir un décalage cohérent.

(2) **Cryptage mono-alphabétique**
Au lieu d'une simple transposition, on utilise une permutation de l'alphabet. L'analyse des fréquences et quelques tatonnements permettent de casser le code (c'est à dire retrouver la permutation utilisée).
Appliquer sur le fichier `mono`.
Lettres par fr équence décroissante dans la pluspart des textes en français :
`esaitnrulodc..`

(3) **Cryptage poly-alphabétique(Vigenère 1586)**
L'idée est de changer de transposition à chaque lettre du code. On utilise un décalage, défini par exemple par les lettres de la clef.
Par exemple, "Cryptage poly-alphabetique(vigenere)" chiffré avec la clef `secret` :

cryptage	poly	-alphabetique	(vigenere)
secretse	cret	secretsecret	secretse

devient :
UVAGXTYI RFPR-SPRYEUWXKHYX(NMIVRXJI)
c :2 s :18 2+18=20 soit U
q :16 r :17 16+17= 7 modulo 26 soit H
(*Informatiquement, on fera plutôt un XOR (codage et décodage équivalent)*).

Cette méthode a resisté aux cryptanalystes du XVI ème siècle jusqu'à la fin du XIX$^{\text{ème}}$. Charles Babbage, mathématicien anglais, et le major prussien Friedrich Kasiski ont utilisé le fait que si le mot de la clef est répété au long du texte, certains groupes de lettres du texte (digrammes ou trigrammes fréquents de la langue *le la ent ..*) se trouvent plusieurs fois associés aux mêmes lettres de la clef et produisent les mêmes groupes de lettres cryptées.
On effectue une recherche de tous les digrammes (ou trigrammes) du texte et on mémorise les distances entre positions successives des digrammes identiques. La longueur n de la clef est un diviseur du PGCD de (la plupart) ces distances. On traite ensuite séparément n remplacements mono-alphabétiques.
Une méthode beaucoup plus rapide est apparue au début du 20$^{\text{ème}}$ siècle : le calcul du coefficient de corrélation, appelé aussi indice de coïncidence, basé sur la probabilité que deux lettres d'un texte soient identiques.
Si une lettre α apparait n_1 fois dans un texte de n lettres, la probabilité pour que 2 lettres tirées au hasard soient des α est $\frac{n_1(n_1-1)}{n(n-1)}$.
Le coefficient de corrélation est la somme de ces probabilités pour toutes les lettres de l'alphabet.
Cette valeur est caractéristique d'une langue naturelle.
En français sa valeur est ≈ 0.074.
Déterminez le coefficient de corrélation pour le texte `mono` et pour le texte chiffré correspondant. Ceci caractérise les chiffrements monoalphabetiques.
Déterminer le coefficient de corrélation pour le texte `poly`.

Utilisation du coefficient de corrélation pour déterminer la longueur de la clef utilisée pour chiffrer `poly` :

- on calcule le coefficient pour des distances de lettres de $1, 2, \ldots,$.

- lorsque les valeurs s'approchent toutes du coefficient de corrélation pour la langue du texte(0.074), on a trouvé la longueur l de la clef.
- on traite ensuite le texte chifré comme l chiffrements monoalphabetiques.

Déchiffrez le texte `poly`.

Le chiffrement de Vigenère fournit la seule méthode (chiffrage à masque unique) algorithmiquement sûre si :

- la clef est de longueur égale à celle du texte ;
- la clef est choisie aléatoirement ;
- la clef est utilisée une seule fois.

L'exercice suivant est emprunté au cours de Ron Rivest au MIT. Voici deux mots anglais de 8 lettres chiffrés (en XOR octet par octet) avec la même clef :

```
u=['99', 'f0', '11', '31', '3f', 'f6', '9d', '52']
v=['89', 'fa', '1f', '34', '38', 'eb', '8c', '59']
```

Saurez-vous retouver ces deux mots ? (Indication : voir ce qu'il y a dans `/usr/share/dict/`).