Outils Mathématiques

TP Crypto n° 1 : Chiffrements élémentaires par substitution monoalphabétique

Objectifs :

1. réaliser des opérations de chiffrement/déchiffrement par substitution
2. cryptanalyser des cryptogrammes chiffrés par de tels systèmes
3. réaliser des opérations de chiffrement/déchiffrement par substitution polyalphabétique
4. cryptanalyser des cryptogrammes chiffrés par de tels systèmes

Outils :

— les commandes UNIX `sed`, `od`, `hexedit`
— un langage de programmation (R, C, Java, ...)
— éventuellement un tableur

Webographie

— Ars Cryptographica : un site riche en présentation et démonstration de systèmes cryptographiques de l'antiquité à nos jours.
— Handbook of applied cryptography : un ouvrage technique de référence. Fichiers PDF disponibles en ligne.

1 Outil d'analyse de fichiers

Le fichier `compteLettres.R` contient quelques commandes du langage `R` pour compter le nombre d'occurrences des 26 lettres de l'alphabet latin contenues dans un fichier (codé en ASCII par exemple).
Question 1 Chargez ce fichier dans `R` avec la commande `source(<nom fichier>)`, et en utilsant la fonction `compteLettres`, analysez le contenu de différents fichiers.
Question 2 Analysez le fichier `cigale.txt`. Tracez un histogramme de répartition des 26 lettres. En quoi cet histogramme est-il caractéristique de la langue utilisée dans ce fichier.

2 Substitutions sur l'alphabet latin

2.1 César

Question 3 Comment utiliser `sed` pour chiffrer un fichier texte avec un code de César ? Quel est le script pour le chiffrement ? pour le déchiffrement ?
Question 4 Chiffrez le texte de votre choix à l'aide d'un code de César et communiquez–le à votre voisin sans lui donner la longueur du décalage afin qu'il tente de le cryptanalyser.

2.2 Substitutions quelconques

Il est évidemment possible d'utiliser la commande `sed` pour chiffrer et/ou déchiffrer des fichiers à l'aide de substitutions monoalphabétiques quelconques.
Question 5 Traitez quelques exemples de votre choix sur des fichiers de texte dont vous aurez préalablement converti toutes les lettres en leur équivalent majuscule non accentué.

La cryptanalyse des substitutions monoalphabétiques quelconques est plus difficile que celle du code de César car le nombre de clés est incomparablement plus grand ($26! \approx 4.10^{26}$ au lieu de 26) et une recherche exhaustive est absolument hors de portée même du plus puissant des ordinateurs. Mais heureusement [1], toutes les langues vivantes présentent des particularités statistiques facilement identifiables. Ainsi en français, la lettre `E` est de loin la plus fréquente ($\approx 16\%$) suivie des lettres `SANTIRULO`.
Question 6 Décryptez le fichier `cryptogram1`.
Question 7 Tentez ensuite l'exercice un peu plus difficile de décrypter un texte chiffré duquel on a ôté tous les espaces et signes de ponctuation (le fichier `cryptogram2` en est un exemple).

1. du point de vue du cryptanalyste !

3 Substitutions sur l'alphabet des 256 octets

De nos jours, l'informatique aidant, on peut chiffrer n'importe quel fichier en le considérant comme un texte écrit avec un alphabet comprenant 256 lettres : les 256 octets.

Question 8 Combien y a-t-il de substitutions monoalphabétiques sur l'alphabet des octets ? Utilisez un logiciel comme `bc` pour calculer des grands nombres (`man bc` pour obtenir de l'aide).

Est-il possible d'envisager une recherche exhaustive de la substitution monoalphabétique ?

On peut concevoir un système de substitution monoalphabétique "moderne" à l'aide de l'opération du ou-exclusif bit à bit entre deux octets. On transforme chaque octet d'un fichier en faisant un ou-exclusif bit à bit de cet octet avec l'octet qui sert de clé. Par exemple si l'octet à coder est `0x41` (c'est-à-dire la lettre `A`), `01000001` en binaire, et l'octet clé est `0x61` (c'est-à-dire la lettre `a`), `01100001` en binaire, on obtient l'octet `00100000` qui correspond au caractère ESPACE.

Question 9 Un grand avantage de cette façon de chiffrer les messages est que l'algorithme qui sert à chiffrer est rigoureusement identique à celui qui sert à déchiffrer. Pourquoi ?

Question 10 Le programme `crypxor` écrit en C effectue le chiffrement et le déchiffrement de fichiers. Pour l'utiliser, il suffit de taper la ligne de commande

```
> crypxor <cle> <fichier a (de)chiffrer> <fichier (de)chiffre>
```

Par exemple pour chiffrer le fichier `cigale.txt` avec l'octet `0x61` correspondant au caractère `a`, on utilise la commande

```
> crypxor a cigale.txt cigale.chiffre
```

Utilisez ce programme pour chiffrer et déchiffrer les fichiers de votre choix.

Attention les fichiers chiffrés obtenus ne sont en général plus des fichiers textes : à cause du ou exclusif, certains caractères ne sont pas des caractères affichables (codes inférieurs à 32 par exemple). Pour visualiser le contenu d'un fichier, on peut utiliser la commande UNIX `xxd`.

```
> xxd  <fichier>
```

Question 11 Tentez de décrypter le fichier `cryptogram3` qui est une version chiffrée du source du programme `crypxor` écrit en C.