TP Sécurité – Cryptanalyse statistique

Julien Bernard

1 Introduction

1.1 Objectifs

L'objectif de ces travaux pratiques de sécurité est de se mettre dans la peau d'un cryptanalyste et d'expérimenter des méthodes de cryptanalyse statistique. Ce sujet de TP couvre l'ensemble des six séances de trois heures. L'évaluation prendra en compte la qualité de vos programmes (découpage en fonction, pertinence des structures de données, etc), ainsi qu'un challenge pendant le dernier TP.

1.2 Textes

Les textes en clair considérés dans ce TP sont écrits uniquement en majuscule et sans accent. Les espaces ne doivent pas être pris en compte, c'est-à-dire qu'ils sont conservés dans les textes chiffrés.

1.3 Programmation

Les programmes demandés doivent être réalisés en C. Les programme de chiffrement prennent le texte clair sur l'entrée standard, la clef en paramètre et produise le texte chiffré sur la sortie standard. Les programmes de déchiffrement prennent le texte chiffré sur l'entrée standard, la clef en paramètre et produise le texte en clair sur la sortie standard. Les programmes de décryptage prennent le texte chiffré sur l'entrée standard et produise le texte en clair sur la sortie standard.

Voici un morceau de code qui vous permettra de lire un texte sur l'entrée standard, caractère par caractère (voir `fgetc(3)`).

```
#include <stdio.h>
#include <stdlib.h>

int main(int argc, char *argv[]) {
  int i;
  while ((i = fgetc(stdin)) != EOF) {
    unsigned char c = (unsigned char) i;
    printf("%c\n", c);
  }
  return 0;
}
```

2 Chiffre de César

2.1 Chiffrement et déchiffrement

Écrire un programme `caesar_encrypt` qui prend en paramète l'indice de décalage et qui réalise un chiffrement par décalage.

Écrire un programme `caesar_decrypt` qui prend également en paramètre l'indice de décalage et qui réalise le déchiffrement.

```
$ caesar_encrypt 3 < plaintext > ciphertext
$ caesar_decrypt 3 < ciphertext > plaintext
```

2.2 Décryptage

Écrire un programme `caesar_break` qui décrypte un texte chiffré à l'aide d'un chiffrement par décalage. Pour cela, on cherchera le `E` grâce à une analyse fréquentielle, et on en déduira la clef.

3 Chiffre de Vigenère

3.1 Chiffrement et déchiffrement

Écrire un programme `vigenere_encrypt` qui prend en paramètre le mot-clef et réalise un chiffrement de Vigenère.

Écrire un programme `vigenere_decrypt` qui prend également en paramètre le mot-clef et qui réalise le déchiffrement.

3.2 Décryptage

Écrire un programme `vigenere_break` qui décrypte un texte chiffré à l'aide d'un chiffrement de Vigenère. Pour cela, on utilisera la méthode décrite en cours du calcul de l'indice de coïncidence pour déterminer la taille de la clef, puis on utilisera une analyse fréquentielle pour déterminer chaque lettre de la clef. Ensuite, on pourra aussi utiliser une autre méthode pour déterminer la taille de la clef, la méthode des diviseurs (voir la page Wikipédia sur le chiffrement de Vigenère).

4 Chiffrement par substitution

4.1 Chiffrement et déchiffrement

Pour créer la table de substitution, on utilisera la méthode de construction de la clef à partir d'un mot-clef vue en cours.

Écrire un programme `subst_encrypt` qui prend en paramètre un mot-clef et qui réalise un chiffrement par substitution.

Écrire un programme `subst_decrypt` qui prend également en paramètre un mot-clef et qui réalise le déchiffrement.

4.2 Décryptage

Écrire un programme `subst_break` qui décrypte un texte chiffré à l'aide d'un chiffrement par substitution. Pour cela, on pourra s'appuyer sur les heuristiques suivantes :

- Analyse fréquentielle des lettres
- Les lettres les plus souvent doubles en français sont `E`, `M`, `L`, `N`, `F`, `T` et `C`
- Les seuls mots de une lettre (sauf avant une apostrophe) sont le `A` et le `Y`
- Un mot d'une lettre devant une apostrophe est `T`, `S`, `D`, `J`, `L`, `M`, `C` ou `N`
- Analyse fréquentielle des séquences de taille 2
- Utilisation d'un mot probable apparaissant dans le texte
- Assistance humaine

Il est très difficile (long) de faire un programme qui décrypte très bien tout seul. Il est très facile de faire un programme qui décrypte à peu près. L'objectif des séances de TP est de faire un programme qui marche le mieux possible. Pour vous aider, vous avez à votre disposition un ensemble de paires de texte clair/chiffré.