TP 1
Cryptanalyse d'un schéma de chiffrement de Vigenère

Pierre Loidreau

Mardi 11 septembre 2012

Le but de ce TP est de réaliser la cryptanalyse d'un système de chiffrement de Vigenère. Cet algorithme sera ensuite utilisé pour décrypter des fichiers *Chiffre_ sw.txt* et *Vigenere_ Singh.txt*.

La cryptanalyse d'un *Vigenère* s'opère en deux étapes.

Étape 1 : Détermination de la longueur de la clé par la méthode de l'indice de coïncidence

On désigne par $\mathbf{y}$ le texte chiffré. C'est une suite de lettres majuscules de longueur n. Soit m un entier restant à définir. On decoupe $\mathbf{y}$ en m sous-suites $\mathbf{y}^{(i)},\ 0 \leq i < m$ de la manière suivante :

Si on écrit le texte chiffré sous la forme

$$\mathbf{y} = (y_1, y_2, \ldots, y_n),$$

alors la sous-suite d'indice $\mathbf{y}$ de position i, est définie par

$$y_k \in \mathbf{y}^{(i)} \iff k \equiv i \bmod m.$$

Ensuite, on détermine l'indice de coïncidence de chaque sous-suite $\mathbf{y}^{(i)},\ 0 < i \leq m$ défini par :

$$I_c(\mathbf{y}^{(i)}) = \frac{\sum_{k=0}^{25} f_k^{(i)}(f_k^{(i)} - 1)}{|\mathbf{y}^{(i)}|(|\mathbf{y}^{(i)}| - 1)},$$

où $|\mathbf{y}^{(i)}|$ désigne la longueur de la sous-suite, et $f_k^{(i)}$ correspond à la fréquence d'apparition de la $k$ième lettre de l'alphabet dans la sous-suite $\mathbf{y}^{(i)}$.

Si l'entier m est égal à la longueur de la clé secrète, alors, pour tout i, $I_c(\mathbf{y}^{(i)}) \approx 0.071$ qui est l'indice de coïncidence du français. Sinon, les $I_c(\mathbf{y}^{(i)})$

sont plus proches de la valeur $0.038 = 1/26$ correspondant à une répartition aléatoire des lettres.

Étape 2 : Indice de coïncidence mutuelle

Une fois que la longueur de la clé est connue égale à m, on s'applique à déterminer la clé elle-même. On considère les sous-suites

$$\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(m)}.$$

1. On calcule les valeurs $f_k^{(i)}$ des fréquences d'apparition de la $k$ième lettre de l'alphabet dans la sous-suite $\mathbf{y}^{(i)}$,
2. On détermine les indices de coïncidence mutuelle des sous-suites $\mathbf{y}^{(i)}$ par rapport aux sous-suites $\mathbf{y}^{(j)}$:

$$\mathbf{M}_g\left(\mathbf{y}^{(i)}, \mathbf{y}^{(j)}\right) = \frac{\sum_{k=0}^{25} f_k^{(i)} f_{k-g}^{(j)}}{|\mathbf{y}^{(i)}| \cdot |\mathbf{y}^{(j)}|}$$

   où $0 \leq i < j \leq m$ et $0 \leq g \leq 25$.

   L'indice de coïncidence mutuelle permet de déterminer les positions relatives des lettres de la clé dans l'alphabet. En effet, si g correspond au bon décalage de suites alors $\mathbf{M}_g\left(\mathbf{y}^{(i)}, \mathbf{y}^{(j)}\right) \approx 0.071$, sinon $\mathbf{M}_g\left(\mathbf{y}^{(i)}, \mathbf{y}^{(j)}\right) \approx 0.038$.
3. On détermine ensuite la clé utilisée, ainsi que le texte clair original.

Applications :

On programmera les algorithmes en langage C.

1. Retrouvez la clé et le texte clair correspondant au fichier *Chiffre_ sw.txt*. Accessoirement retrouvez l'auteur du texte.
2. Dans son livre *Histoire des codes secrets*, Simon SINGH propose un prix de 10.000 Livres Sterling à la personne capable de résoudre les 10 challenges proposés dans l'appendice. Le quatrième challenge est un texte chiffré par un cryptosystème de Vigenère, dont le texte se trouve dans le fichier *Vigenere_ Singh.txt*.

   Retrouvez la clé, le texte original, ainsi que l'auteur.