TD1 — Chiffrements historiques
Corrections

Ex3. L'œil qui voit tout L'étrange alignement de graffitis reproduit en fin d'exercice a été découvert sous la peinture sur le mur de la salle E11. Le message semble ancien, il est composé à partir de 22 symboles distincts.

(a) L'utilisation de symboles exotiques rend-elle le chiffrement d'un message plus sûr ?

- Non, pas en tant que tel.

  Les outils de base de cryptanalyse (tels que l'indice de coïncidence ou l'analyse par fréquence) considèrent un alphabet de symboles quelconques. Si le chiffrement est mono- ou poly-alphabétique, ces outils permettront d'analyser et « déchiffrer » le message, quel que soit l'alphabet utilisé.
  L'utilisation d'un alphabet exotique nécessite tout de même une étape supplémentaire lors du déchiffrement : retrouver la correspondance entre entre les symboles exotiques et les lettres de l'alphabet d'origine.

(b) Calculer l'indice de coïncidence du message. Que peut-on en déduire ?

- Pour calculer l'indice de coïncidence, on applique la formule suivante :

$$IC = \sum_c \frac{n_c(n_c-1)}{n(n-1)}$$

avec :

c : symbole présent dans le message
n : nombre total de symboles dans le message
n_c : nombre d'occurrences du symbole c dans le message

Remarquez que l'alphabet complet peut contenir plus de symboles que ceux utilisés dans le message chiffré. L'indice de coïncidence ne tient compte que des symboles présents dans le message chiffré.

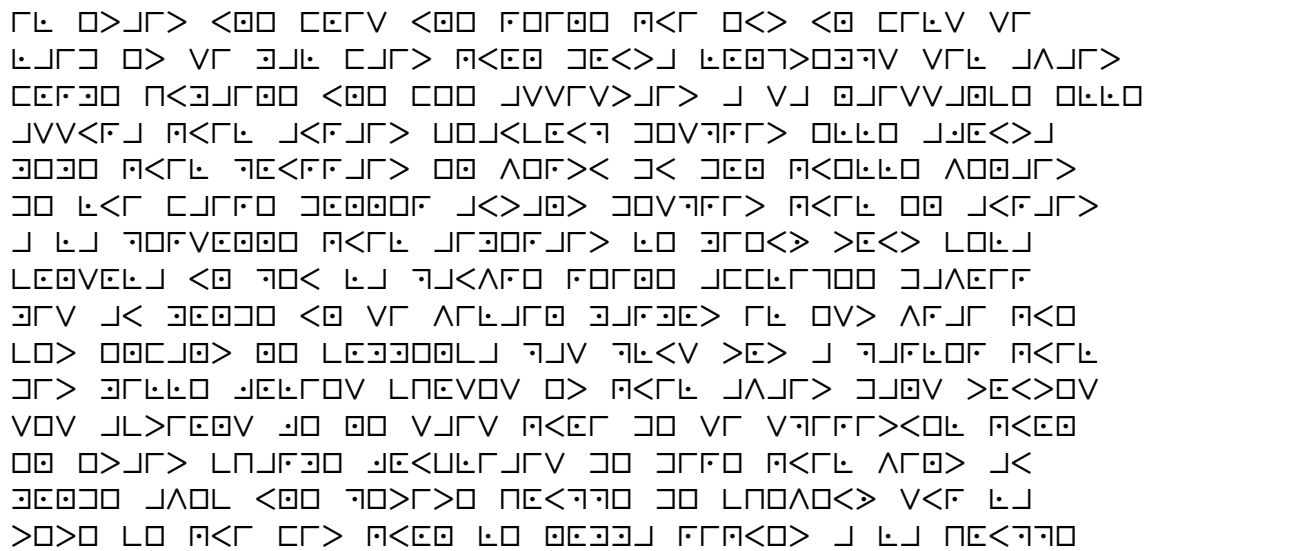
On obtient l'indice de coïncidence suivant :

$$IC = \frac{62\times61+38\times37+82\times81+45\times44+\cdots+2\times1+4\times3+2\times1}{591\times590} = 0{,}0744$$

L'indice de coïncidence nous permet de différencier un texte en langue naturelle (quand sa valeur est supérieure à $0{,}05$) d'un texte aléatoire (quand sa valeur est inférieure à $0{,}05$).

- Puisque l'indice de coïncidence est supérieur à $0{,}05$ en considérant l'ensemble des symboles du message, on en déduit qu'il s'agit d'un chiffrement mono-alphabétique.

Déchiffrons le message suivant :

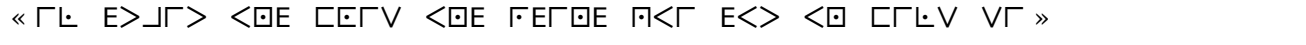


Puisque le chiffrement est mono-alphabétique, chaque symbole du message chiffré correspond à une lettre de notre alphabet. Nous supposons que le message est en français.
Le déchiffrement du message complet peut s'effectuer en utilisant utilisant des tables de fréquences : on fait correspondre des suites de symboles les plus fréquents dans le message chiffré, avec des suites de lettres (ou des mots complets) les plus fréquentes dans la langue du message en clair.

Nous pourrions tenter d'utiliser une table de fréquences des lettres simples : les lettres les plus fréquentes en langue française sont E, A, S, I, N. Toutefois, le résultat ne sera pas probant pour ce message particulier. En effet, en substituant les 5 symboles les plus fréquents dans le message chiffré par les 5 lettres les plus fréquentes en langue française, la première ligne du message devient : « AL ENSAN IOE CEAV NOE FEAOE PIA EIN IO CALV VA »
Mais « ENSAN » et « EIN » ne sont pas des mots de la langue française.

□ → E
Conservons tout de même la substitution « □ → E », car en français la lettre « E » apparaît sensiblement plus souvent que toutes les autres lettres, et le symbole « □ » apparaît également sensiblement plus grand que les autres symboles dans le message chiffré.

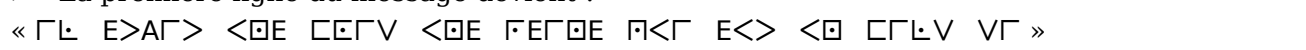
- La première ligne du message devient :

«  »

Complétons le déchiffrement en utilisant des tables de fréquences de suites de lettres, voire de mots entiers.

⌟ → A
Le symbole « ⌟ » a un nombre d'occurrences important dans le message chiffré, et il apparaît plusieurs fois seul dans le message. Il est fort probable qu'il s'agisse de la lettre « A », qui est la seule à avoir une signification en tant que mot en français (avec la lettre « Y » mais dont la fréquence est beaucoup plus faible en français).

- La première ligne du message devient :

«  »

> → T, ⌜ → I

Le deuxième mot « E>A⌜> » pourrait correspondre à « ETAIT » ou « ETANT ».
Essayons la substitution « > → T » et « ⌜ → I ». Il s'agit d'un choix « empirique » qui pourrait s'avérer incorrect ; si la suite du déchiffrement devient impossible, nous pourrons revenir sur notre choix.

- La première ligne devient :

« Il⌞ ETAIT <⊡E ⊏⊡IV <⊡E ⌜EI⊡E ⊓<I E<T <⊡ ⊏Il⌞V VI »

⌞ → L, < → U, ⊡ → N

Les mots « <⊡E » et « Il⌞ » pourraient respectivement correspondre à « UNE » et « IL ».
Essayons la substitution « ⌞ → L », « < → U », et « ⊡ → N ».

- La première ligne devient :

« IL ETAIT UNE ⊏⊡IV UNE ⌜EINE ⊓UI EUT UN ⊏ILV VI »

Le reste du déchiffrement du message complet s'effectue en poursuivant la méthode appliquée ci-dessus : identifier une suite de lettres quasiment déchiffrées, déterminer la substitution la plus probable pour les symboles restants, et poursuivre le déchiffrement.
Une fois entièrement déchiffré, le message devient :

« IL ETAIT UNE FOIS UNE REINE QUI EUT UN FILS SI
LAID ET SI MAL FAIT QUON DOUTA LONGTEMPS SIL AVAIT
FORME HUMAINE UNE FEE ASSISTAIT A SA NAISSANCE ELLE
ASSURA QUIL AURAIT BEAUCOUP DESPRIT ELLE AJOUTA
MEME QUIL POURRAIT EN VERTU DU DON QUELLE VENAIT
DE LUI FAIRE DONNER AUTANT DESPRIT QUIL EN AURAIT
A LA PERSONNE QUIL AIMERAIT LE MIEUX TOUT CELA
CONSOLA UN PEU LA PAUVRE REINE AFFLIGEE DAVOIR
MIS AU MONDE UN SI VILAIN MARMOT IL EST VRAI QUE
CET ENFANT NE COMMENCA PAS PLUS TOT A PARLER QUIL
DIT MILLE JOLIES CHOSES ET QUIL AVAIT DANS TOUTES
SES ACTIONS JE NE SAIS QUOI DE SI SPIRITUEL QUON
EN ETAIT CHARME JOUBLIAIS DE DIRE QUIL VINT AU
MONDE AVEC UNE PETITE HOUPPE DE CHEVEUX SUR LA
TETE CE QUI FIT QUON LE NOMMA RIQUET A LA HOUPPE »

(c) *(bonus)* Déchiffrer le graffiti suivant découvert sur le même pan de mur :

∀⊏⊐□∀⊓□⌜□ ⊏∧□⌜ >⊓□ ⌜⌟⌐⊡⊔⊏∀

- Le message bonus utilise la même substitution. Avec les symboles identifiés, nous obtenons le message suivant :

« SOME∀HERE OVER THE RAINBO∀ »

Il nous manque la substitution pour l'un des symboles ! En effet, le symbole « ∀ » n'apparaissait pas dans le long message complet, et nous n'avons pas encore déterminé la lettre correspondante.
Pour autant, avec le message bonus partiellement déchiffré, nous pouvons aisément déterminer la substitution : « ∀ → W ».

- Le message bonus entièrement déchiffré est : « SOMEWHERE OVER THE RAINBOW »

Ex4. Chiffrement par xor (noté $\oplus$)

(a) M , K , C sont des blocs de bits. Supposons que $\text{long}(M)=8$ et $\text{long}(K)=4$. On rappelle que $E(K,M)=M\oplus k$. Chiffrer le message 01011100 avec la clé 0101 .

- La longueur de la clé K est plus petite que la longueur du message M . Pour pouvoir chiffrer M avec la clé K avec un chiffrement par xor, il faut répéter la clé K autant de fois que nécessaire jusqu'à atteindre la longueur de M (on tronque éventuellement la dernière répétition de K , pour que la longueur totale soit exactement celle de M).

  Puis, on effectue l'opération xor bit à bit entre M et K répété, ce qui donne :

$$\begin{array}{ll} M & 0101\ 1100 \\ K & \underline{0101\ 0101} \\ \oplus & 0000\ 1001 \end{array}$$

(b) Malheureusement votre adversaire sait que le message clair M commence par 0101 , et il sait que $C=01101100$. Que peut-il faire pour trouver la clé ? Comment s'appelle ce type d'attaque ?

- Il s'agit d'une attaque à texte clair connu.

  L'adversaire peut retrouver la clé en effectuant l'opération $M\oplus C$ sur les 4 premiers bits du message chiffré. En effet, la longueur de la partie connue du message (4 bits) est au moins aussi grande que la longueur de la clé (également 4 bits dans notre cas).

$$\begin{array}{lll} M & 0101 & \\ C & \underline{0110} & \\ \oplus & 0011 & = K \end{array}$$

(c) L'attaque précédente a-t-elle encore un intérêt si $\text{long}(K)=\text{long}(M)=8$, et la clé K n'est plus utilisée par la suite (masque jetable) ?

- Si la clé K n'est pas réutilisée, alors l'adversaire ne pourra pas déchiffrer de futurs messages même s'il trouve cette clé K utilisée pour chiffrer le message M . À noter que dans l'attaque décrite précédemment, l'adversaire a besoin de connaître l'ensemble du message clair M et du message chiffré C pour trouver la clé K .

(d) Maintenant $\text{long}(K)=\text{long}(M)=4$. On veut comparer $E(K,M)=M\oplus K$ (xor bit à bit), et $E'(K,M)=(M+K)(mod\,16)$ (addition modulo 16 des entiers M et K codés en base 2). Est-ce la même chose ?

- Montrons que ces deux méthodes de chiffrement ne sont pas strictement identiques. Pour cela, montrons qu'il existe au moins un message M et une clé K tels que : $E(K,M)\neq E'(K,M)$

Considérons le message $M=1100_2=12_{10}$ et la clé $K=1011_2=11_{10}$.
Alors :

$$\begin{array}{lllllll} E(K,M) & = & M\oplus K & = & 1100_2\oplus 1011_2 & = & 0110_2\quad 6_{10} \\ E'(K,M) & = & (M+K)mod\,16 & = & 23_{10}\,mod\,16 & = & 0111_2\quad 7_{10} \end{array}$$