Cours Turing
Semaine 9

1 Cryptographie à clé secrète : un peu d'histoire

Dans cette première section, nous allons passer en revue quelques méthodes de chiffrement historiques, qui ne sont plus utilisées aujourd'hui, essentiellement à cause de la puissance des outils informatiques actuels, qui en permettent un décryptage quasi-instantané... Néanmoins, il est intéressant de se plonger dans les principes des premiers systèmes de cryptographie, car les systèmes modernes s'en inspirent fortement.

Tout au long de ce nouveau chapitre, trois personnages-clés vont nous accompagner : Alice, Bob et Eve, dans le scénario suivant : Alice cherche une méthode pour communiquer un message à Bob, tout en évitant qu'Eve, qui écoute leur conversation, soit capable de décrypter le message.

1.1 Préliminaire : l'addition modulo 26

Dans ce qui suit, nous allons manipuler l'alphabet latin, composé de 26 lettres, en nous restreignant pour simplifier aux majuscules. Sur cet alphabet, il est possible de définir une règle d'addition en identifiant chaque lettre à sa position dans l'alphabet (en commençant avec le A en position 0) :

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Ainsi, on a par exemple :

`A + A = A, A + B = B, B + B = C, B + C = D, ..., E + F = J, ...`

car

`0 + 0 = 0, 0 + 1 = 1, 1 + 1 = 2, 1 + 2 = 3, ..., 4 + 5 = 9, ...`

Mais alors, qu'advient-il par exemple de Y + D ou encore P + S ? En effet, $24 + 3 = 27$ et $15 + 18 = 33$, or il n'y a que 26 lettres dans l'alphabet latin. C'est pour cela qu'on définit une *addition modulo 26* sur l'ensemble des lettres, ce qui revient à dire que si le résultat de l'addition est plus grand que 25, on lui soustrait 26.

Exemples :

Y + D = B, car 24 + 3 = 27 et 27 - 26 = 1

P + S = H, car 15 + 18 = 33 et 33 - 26 = 7

L'avantage de cette règle est que l'addition de n'importe quelle paire de lettres redonne toujours une lettre.

1.2 Le chiffre de César (I^{er} siècle av. J.-C.)

La première idée pour chiffrer un message remonte à loin... Pour Alice, il s'agit simplement de choisir une des 26 lettres de l'alphabet, disons K pour l'exemple, et d'additionner celle-ci à chaque lettre du message qu'elle désire envoyer. Par exemple, pour chiffrer le message BONJOUR, elle effectue les additions :

```
   B  O  N  J  O  U  R
+  K  K  K  K  K  K  K
----------------------
=  L  Y  X  T  Y  E  B
```

et envoie donc le message chiffré LYXTYEB à Bob. Deux questions se posent alors :

1. Bob est-il capable de déchiffrer son message, et si oui, comment ?
2. Et si par hasard Eve intercepte le message envoyé par Alice, que peut-elle en faire ?

1. Pour que Bob puisse déchiffrer facilement le message transmis par Alice, il importe que celle-ci lui ait transmis *au préalable* la lettre K qu'elle a utilisé pour effectuer les additions (on appelle cette lettre la *clé secrète*). Bob peut alors effectuer les soustractions correspondantes (toujours modulo 26, ce qui veut dire qu'il rajoute 26 lorsque le résultat de la soustraction est plus petit que 0) pour retrouver le message d'origine :

```
   L  Y  X  T  Y  E  B
-  K  K  K  K  K  K  K
----------------------
=  B  O  N  J  O  U  R
```

Notez que Bob pourrait de manière équivalente *additionner* une lettre à chaque lettre du message reçu pour retrouver le message d'origine : voyez-vous laquelle ?

2. Que peut faire Eve de son côté si elle intercepte le message LYXTYEB ? Elle ne connaît pas la lettre secrète choisie par Alice, mais d'un autre côté, l'alphabet latin ne comporte que 26 lettres... Rien n'empêche donc Eve d'effectuer les mêmes soustractions que Bob, chaque fois avec une lettre différente, jusqu'à ce qu'elle tombe sur un message cohérent. Eve effectue ici ce qu'on appelle une *attaque par force brute*, ce qui veut dire qu'elle teste toutes les possibilités

de chiffrement. Notez qu'avec beaucoup de malchance, il se pourrait qu'Eve tombe sur deux messages cohérents avec deux choix différents de lettre (et ne puisse donc pas décrypter le message d'origine), mais cette probabilité tombe rapidement à zéro pour de longs messages.

Alice et Bob doivent donc trouver autre chose pour assurer le secret de leurs communications...

1.3 Le chiffre par substitution (monoalphabétique)

Une variante du chiffre de César est la suivante : plutôt que d'additionner toujours la même lettre à chacune des lettres du message, Alice pourrait décider de substituer chaque lettre de l'alphabet avec une autre, par exemple :

```
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
| | | | | | | | | | | | | | | | | | | | | | | | |
B K Z Y M F E G A R S W C Q J O N V T H L I D U X P
```

Remarques :

- Certaines lettres peuvent rester inchangées avec ce système, comme ici le F.

- Le chiffre de César est un cas particulier du chiffre de substitution.

En utilisant ce tableau (de haut en bas) pour chiffrer le message BONJOUR, Alice obtient le message chiffré KJQRJLV, qu'elle envoie à Bob.

Comment Bob fait-il dans ce cas pour déchiffrer le message d'Alice ? A nouveau, il est nécessaire qu'Alice lui ait communiqué au préalable une clé secrète, qui n'est ici rien d'autre que le tableau ci-dessus (dans son intégralité). A l'aide du tableau, Bob peut retrouver le message d'Alice en utilisant celui-ci de bas en haut.

Se pose maintenant à nouveau la question de savoir ce qu'Eve peut faire si elle intercepte le message KJQRJLV ? La première chose à remarquer est qu'ici le nombre de clés possibles est *beaucoup* plus élevé qu'avec le chiffre de César ! Ce nombre est égal au nombre de permutations possibles des 26 lettres de l'alphabet, qui vaut $26 \cdot 25 \cdot 24 \cdots 3 \cdot 2 \cdot 1 = 26! \simeq 4 \cdot 10^{26}$. Il est donc complètement impossible pour Alice de tester toutes les possibilités en un temps raisonnable, même avec le plus performant des ordinateurs actuels... Une attaque par force brute est infaisable ici (et serait du reste bien inutile : voyez-vous pourquoi ?).

Alice et Bob auraient-ils donc trouvé le système de chiffrement idéal ? Pas si sûr... En effet, si le message envoyé par Alice est relativement long, une attaque possible d'Eve, dite aussi *attaque par analyse fréquentielle*, est la suivante : il s'agit de compter le nombre d'occurences de chaque lettre dans le message chiffré et de diviser ce nombre par le nombre total de lettres du message, pour finalement obtenir la fréquence d'apparition de chaque lettre dans le message.

Or il se trouve que dans un long texte en français, la fréquence d'apparition de chaque lettre est relativement stable et surtout connue : par exemple, le E apparaît beaucoup plus souvent que toutes les autres lettres, et les lettres A, I, S apparaissent beaucoup plus fréquemment que X, Y ou Z, par exemple. Ainsi, en calculant les fréquences d'apparition de chaque lettre

dans le message chiffré, et en comparant celles-ci aux fréquences connues des lettres en français, Eve peut essayer d'établir une correspondance pour retrouver quelle lettre a été remplacée par quellle autre. Bien sûr, ce système n'est pas parfait, mais une fois que certaines lettres ont été identifiées (surtout les voyelles), il devient relativement facile pour Eve de deviner les autres lettres par déduction, en identifiant certains mots, et ainsi de retrouver le message d'origine.

Caramba, encore raté ! Essayons autre chose...

1.4 Le chiffre de Vigenère (XVI^{e} siècle)

L'idée suivante est de choisir non pas une lettre, comme dans le chiffre de César, mais une suite de lettres, comme par exemple la suite "MDR", et d'utiliser celle-ci de manière répétée pour chiffrer un message. Voici ce que cela donne sur un exemple :

```
   B  O  N  J  O  U  R  A  L  A  N  T  U  R  I  N  G
+  M  D  R  M  D  R  M  D  R  M  D  R  M  D  R  M  D
-------------------------------------------------------
=  N  R  F  V  R  L  D  D  C  M  Q  K  G  U  Z  Z  J
```

Remarques :

- On a évité d'utiliser des espaces ici pour se restreindre aux 26 lettres latines, mais il est tout à fait possible d'ajouter l'espace à l'ensemble des lettres utilisées.

- Cette méthode est appelée un chiffrement par substitution *polyalphabétique*, car contrairement au cas précédent, une lettre donnée n'est pas systématiquement remplacée par une même autre lettre ; la substitution s'applique à des motifs de 3 lettres consécutives.

Muni de la clé secrète MDR, Bob est capable de déchiffrer le message reçu en effectuant les soustractions correspondantes (il faut certes qu'Alice et Bob s'accordent bien sur la position du début du message). Ce système a clairement l'avantage de la simplicité.

Et si Eve intercepte le message chiffré, que peut-elle faire ?

- Une attaque par force brute ?

Si la clé secrète comporte d lettres, le nombre de clés possibles vaut $\underbrace{26 \cdot 26 \cdot 26 \cdots 26}_{d \text{ fois}} = 26^d$,
ce qui grandit (très) vite avec d (p. ex., si $d = 10$, alors $26^d \simeq 10^{14}$). Il faut essayer autre chose...

- Une attaque par analyse fréquentielle ?

Comme mentionné plus haut, ce système de chiffrement a pour propriété qu'une lettre donnée n'est pas systématiquement remplacée par un même autre lettre. Une analyse naïve des fréquences des lettres dans le message chiffré ne donne donc rien ici. Mais une analyse des fréquences des motifs de d lettres consécutives pourrait quant à elle donner quelque chose. Le problème pour Eve est qu'elle ne connaît même pas la longueur d de la clé...

Ceci dit, au XIX^{e} siècle (donc trois siècles plus tard...), Friederich Kasiski publie un test qui permet de deviner la valeur de d ; ce test met fin à l'utilisation du chiffre de Vigenère.