Le chiffre de Merkle-Hellman

Le problème du sac à dos

En 1978, Ralph Merkle et Martin Hellman proposèrent un cryptosystème à clé publique basé sur un problème célèbre, le problème du sac à dos (*Knapsack problem*).

Imaginons une collection de cailloux de poids $a_1, a_2, \ldots, a_n$ connus.
Supposons que l'on place certains de ces cailloux dans un sac à dos et que l'on pèse le tout.

Est-il possible, connaissant ce poids total, de savoir quels cailloux sont dans le sac ?

Le problème peut s'exprimer ainsi :

Etant donné une suite d'entiers positifs $S=(a_1, a_2, \ldots, a_n)$ et un nombre entier s,

existe-t-il $(x_1, x_2, \ldots, x_n)$ tel que $\sum_{i=1}^{n} x_i a_i = s$, avec x_i égal à 0 ou à 1 ?

Si n est grand, le problème du sac à dos s'avère très difficile à résoudre et la fonction qui à un ensemble $(x_1, x_2, \ldots, x_n)$ associe s est une fonction à sens unique, s est facile à calculer et retrouver $(x_1, x_2, \ldots, x_n)$ à partir de s est difficile.

Martin Hellman propose l'exemple suivant :

S = (14, 28, 56, 82, 90, 132, 197, 284, 341, 455)

s = 516 ne peut s'obtenir comme somme d'élément de S.

s = 515 peut s'obtenir trois fois (dont
$515=14+28+132+341=1\cdot14+1\cdot28+0\cdot56+0\cdot82+0\cdot90+1\cdot132+0\cdot197+0\cdot284+1\cdot341+0\cdot455$)

Le cas d'une suite super-croissante

Dans un cas cependant, le problème du sac à dos ne présente aucune difficulté, celui où les éléments de S forment une suite super-croissante (chaque élément est supérieur à la somme des éléments précédents).

Prenons, par exemple, la suite super-croissante S = (2, 5, 9, 20, 42, 90, 250) et *s* = 56.

42 est le plus grand élément de S inférieur à s.

42 doit intervenir dans le calcul de *s*.

Sinon, ou la somme comporte un élément supérieur à *s* et elle est supérieur à 56
ou la somme ne comporte que des éléments inférieurs à 42 et elle est inférieure à 56.

Le solde vaut 14.

9 étant le plus grand élément de S inférieur à 14, il doit, selon le même raisonnement, intervenir dans le calcul de *s.*

Le nouveau solde étant 5, le problème est résolu : 56 = 5 + 9 + 42.

Le chiffre de Merkle-Hellman

L'idée de base du système consiste à construire une suite non super-croissante à partir d'une suite super-croissante, en conservant une clé secrète permettant de retrouver la suite initiale.

- Alice choisit une suite super-croissante S= (a_1, a_2, ... , a_n),

  un nombre m supérieur à $\sum_{i=1}^{n} a_i$ et un entier e, 1 < e < m, premier avec m.

- Pour chaque élément a_i de S, Alice calcule $b_i = a_i \cdot e \mod m$.
  Elle ordonne les éléments b_i dans l'ordre croissant pour obtenir une nouvelle suite
  S' = (b_1, b_2, ... , b_n) qui n'est plus super-croissante.

- Elle publie cette suite en conservant comme clés secrètes m, l'inverse d de e modulo m, la suite super-croissante S et la permutation p ayant permis d'ordonner S'.

- Pour chiffrer un message, Bernard le représente en code binaire et le décompose en blocs de longueur n au plus.

  Pour chaque bloc $m_1\ m_2\ \ldots\ m_n$, il calcule $M = \sum_{i=1}^{n} m_i b_i$.

- Pour déchiffrer le message M reçu, Alice calcule M' = M · d modulo m et détermine $x_1, x_2, \ldots, x_n$ tels que $\sum_{i=1}^{n} x_i a_i = M'$ (Il s'agit d'un problème simple de sac à dos, la suite $(a_1, a_2, \ldots, a_n)$ étant super-croissante).

  Elle retrouve finalement le message $m_1\ m_2\ \ldots\ m_n$ en appliquant à $x_1, x_2, \ldots, x_n$ la permutation p [1].

Exemple

- Alice choisit S= (1, 3, 5, 11, 25, 53, 101, 205, 512), m = 960 et e = 143.
  (l'inverse d de 143 mod 960 est 47)

- Pour chaque élément a_i de S, Alice calcule $b_i = a_i \cdot e$ mod m, ce qui donne (143, 429, 715, 613, 695, 859, 43, 515, 256).

  En ordonnant b_i, elle obtient la clé publique S' = (43, 143, 256, 429, 515, 613, 695, 715, 859) .

- Pour exprimer le message « RAS » en code binaire, Bernard peut, par exemple, utiliser le code ASCII à 8 bits.

  R correspond à 01010010, A à 01000001 et S à 01010011.

  Le message à coder est 0101001 0010000 0101010 011.

  Il le décompose en blocs de longueur convenue (7 par exemple) et chiffre chacun des blocs :

  0101001 se code 43 + 429 + 613 = 1085,

  0010000 se code 515,

0101010 se code 143 + 429 + 613 = 1185,

011 s'écrit 0110000 et se code 515 + 613 = 1128.

Il transmet à Alice le message 108 515 1185 1128

- Alice va déchiffrer ce message élément par élément, en calculant $M \cdot d \bmod m$ et en déterminant la solution du problème du sac à dos correspondant.

$1085 \cdot 47 \bmod 960 = 115$

$515 \cdot 47 \bmod 960 = 205$

$1185 \cdot 47 \bmod 960 = 15$

$1128 \cdot 47 \bmod 960 = 216$

115 = 101+11+3 correspond à 0000001 + 0100000 + 0001000 = 0101001

205 = 205 correspond à 0010000

15 = 11+3+1 correspond à 0100000 + 0001000 + 0000010 = 0101010

216 = 205+11 correspond à 001000 + 0100000 = 0110000

Alice retrouve le message   0101001001000001010100110000 : RAS.

Remarquons que si la longueur des blocs de chiffrement est égale à celle des caractères en code ASCII à 8 bits, chaque lettre sera codée par le même nombre. Le système est alors vulnérable à une attaque à l'aide d'une analyse de fréquence.
Il convient donc de choisir des blocs de chiffrement de longueur inférieure à celle de la clé.

Le système a été cassé en 1982 par Adi Shamir et n'est plus utilisé de nos jours.