La factorisation des grands nombres

JOHANNES BUCHMANN

Les systèmes modernes de cryptage de données seront sûrs tant que la décomposition des nombres de plus de 100 chiffres en facteurs premiers restera difficile. Toutefois les techniques de factorisation progressent rapidement.

Le nombre 114 381 625 757 888 867 669 235 779 976 146 612 010 218 296 721 242 362 562 561 842 935 706 935 245 733 897 830 597 123 563 958 705 058 989 075 147 599 290 026 879 543 541 est le produit de deux nombres premiers ; lesquels? Martin Gardner posa cette question aux lecteurs de *Pour la Science* en octobre 1977, dans sa rubrique de «Jeux mathématiques», mais une réponse ne fut donnée que 16 ans plus tard : en avril 1994, Paul Leyland, de l'Université d'Oxford, Michael Graff, de l'Université de l'Iowa, et Derek Atkins, de l'Institut de technologie du Massachusetts, identifièrent les deux facteurs, après avoir distribué des parties de la tâche, grâce au réseau *Internet*, à quelque 600 volontaires, qui laissèrent fonctionner sur leurs ordinateurs, pendant de nombreuses nuits, le programme écrit par Arjen Lenstra, du Centre de recherches de la Société *Bell Communications*.

La multiplication de deux nombres, même très grands, n'est pas compliquée : avec du papier et un crayon, on calcule le produit de deux nombres de 65 chiffres en une heure environ ; par ordinateur, le calcul est immédiat. En revanche, l'opération inverse, c'est-à-dire l'identification des facteurs d'un produit, est très difficile, même avec les calculateurs les plus rapides.

Les opérations mathématiques telles que la multiplication et la factorisation sont à la base des systèmes cryptographiques modernes : le cryptage est rapide, mais le décryptage est quasi impossible en pratique.

En 1978, Ronald Rivest, de l'Institut de technologie du Massachusetts, Adi Shamir, de l'Institut Weizmann à Rehovot (Israël), et Leonard Adleman, de l'Université de Californie du Sud, ont mis au point un protocole de cryptage, nommé RSA d'après leurs initiales, qui est fondé sur la factorisation : une personne voulant recevoir des messages cryptés choisit deux nombres premiers p et q, c'est-à-dire deux nombres entiers naturels qui ne sont divisibles que par 1 et par eux-mêmes ; il calcule leur produit $p \times q$ et le rend public, tout en conservant secrets les facteurs *(voir l'encadré de la page 90)*. Pour crypter un message, il suffit de connaître ce produit, nommé la clé publique, tandis que, pour décrypter un message, il faut connaître les facteurs premiers p et q : si ceux-ci ont plus de 150 chiffres, même les meilleures méthodes connues et les ordinateurs les plus puissants mettront 2 000 ans pour les trouver. Ainsi, on fabrique facilement des problèmes de factorisation, mais on ne sait pas, aujourd'hui, les résoudre en un temps raisonnable, si les facteurs premiers sont trop grands.

Trouvera-t-on un jour une méthode de factorisation rapide? Les Laboratoires RSA ont organisé un concours mondial de factorisation. Ils publient des produits de grands nombres premiers et récompensent leur factorisation *(on peut consulter leur site sur le réseau Internet : voir la bibliographie)*.

L'idée de mettre le public à contribution est judicieuse : on ignore si la factorisation est difficile par essence ou si les mathématiciens n'ont pas encore trouvé la méthode la plus habile. Aussi la seule garantie de la sécurité des procédés de cryptage est l'ignorance d'une méthode rapide de factorisation des nombres entiers.

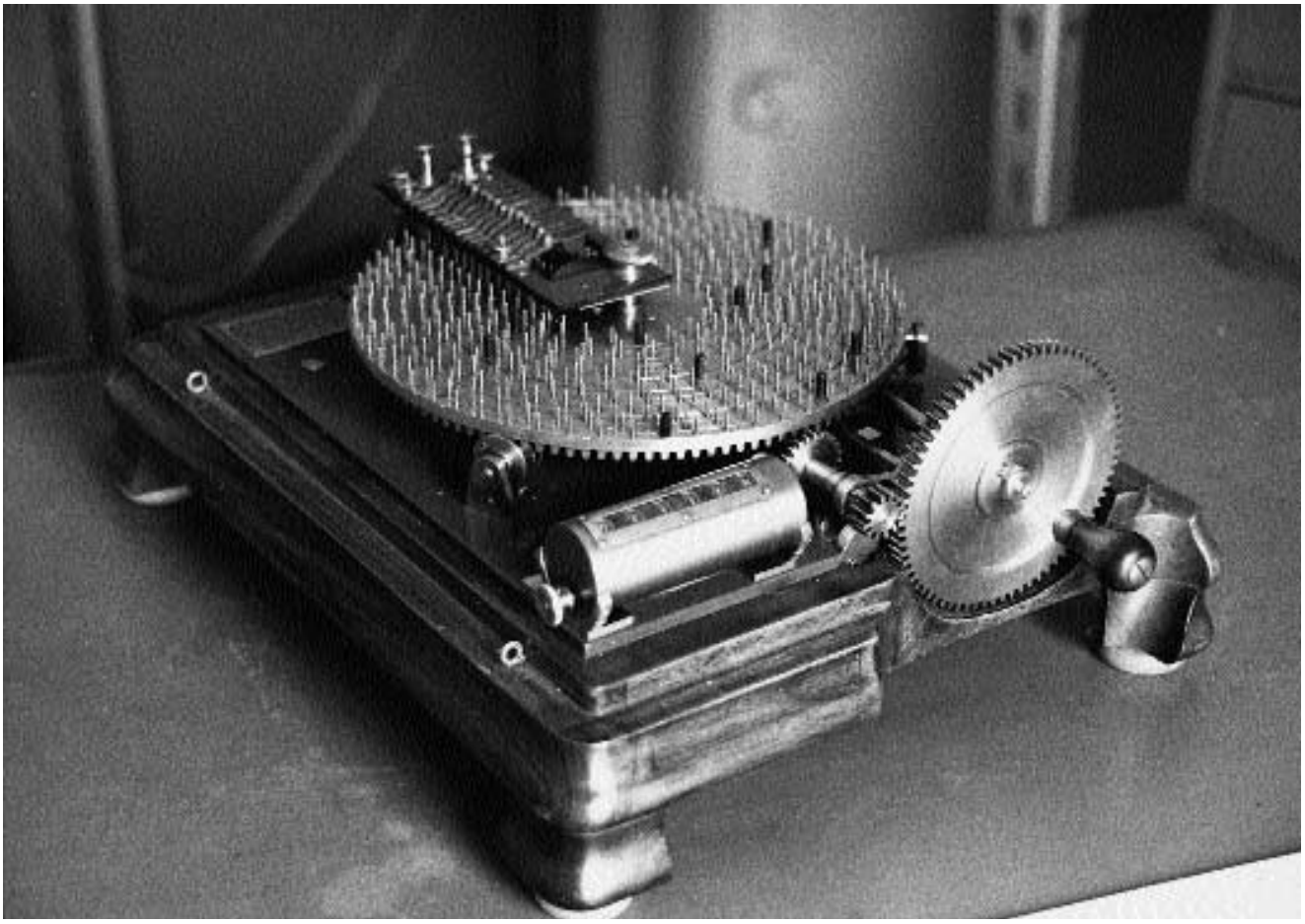
L'étude de la factorisation date de l'Antiquité : les mathématiciens d'alors savaient déjà que chaque nombre naturel est un produit de nombres premiers, et que la décomposition en facteurs premiers est unique, à l'ordre près. Par exemple, 12 se décompose seulement en $2 \times 2 \times 3$. L'étude des propriétés des nombres entiers naturels impose souvent la décomposition en facteurs premiers.

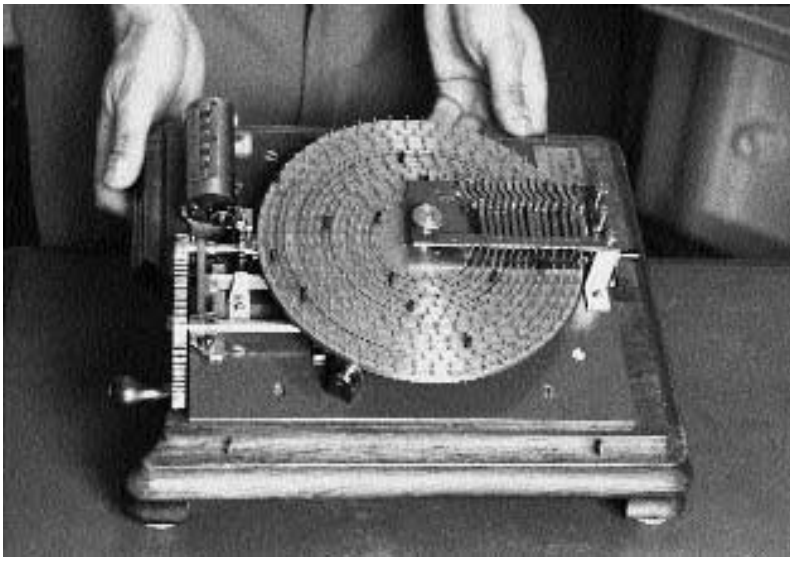
Les ordinateurs ont beaucoup apporté à la théorie des nombres. Dans cet article, nous verrons comment fonctionnent les algorithmes modernes de factorisation. Souvent les indications seront suffisamment précises pour qu'une programmation soit possible sur un ordinateur personnel ; en outre, une bibliothèque de programmes nommée LiDIA est proposée sur le réseau *Internet (voir la bibliographie)*.

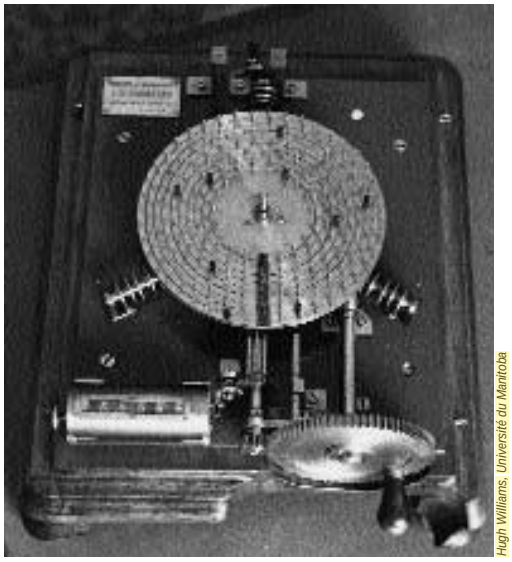
Les nombres de Fermat

Le juriste français Pierre de Fermat (1601-1665), notamment célèbre pour sa conjecture démontrée en 1997, pensait avoir trouvé une méthode pour fabriquer des nombres premiers aussi grands que l'on voulait : à partir des nombres entiers naturels i, il construisait les nombres $F_i = 2^{2^i} + 1$, nommés aujourd'hui nombres de Fermat. Les quatre premiers nombres de Fermat sont premiers : $F_0 = 2^{2^0} + 1 = 3$, $F_1 = 2^{2^1} + 1 = 5$, $F_2 = 2^{2^2} + 1 = 17$, $F_3 = 2^{2^3} + 1 = 257$ et $F_4 = 2^{2^4} + 1 = 65\,537$.

Cependant, dès 1732, le mathématicien suisse Leonhard Euler trouva que

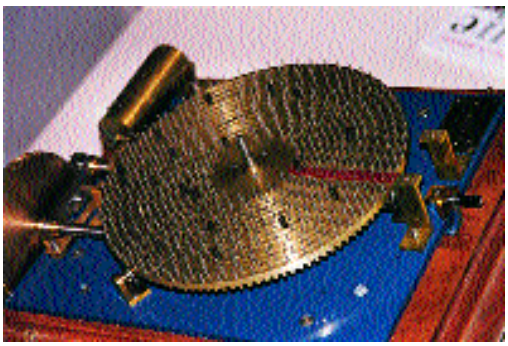


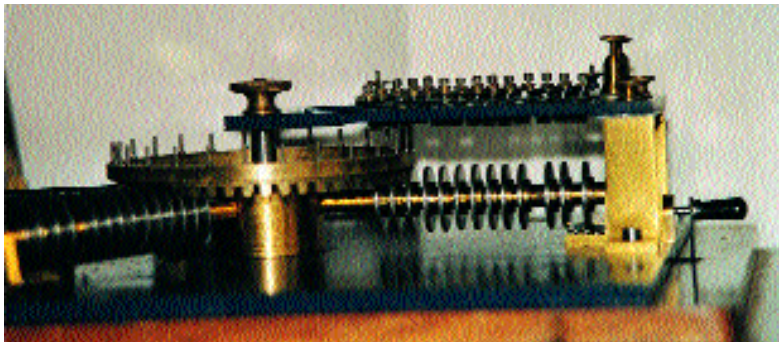




Hugh Williams, Université du Manitoba

1. LA «MACHINES À CONGRUENCES» de l'officier français Eugène Carissan (1880-1925). La manivelle fait tourner une roue dentée qui entraîne un compteur *(au premier plan sur la photographie supérieure)* et 14 couronnes indépendantes. Sur la photographie inférieure, à droite, les quatre couronnes externes ont été enlevées et le mécanisme apparaît. À chaque couronne est associé un nombre, de l'intérieur vers l'extérieur : 19, 21, 23, 26, 29, 31, 34, 37, 41, 43, 47, 55 et 59. Sur la face supérieure, chaque couronne porte un nombre de picots égal à ce nombre. Soit *n* un nombre à factoriser. On utilise la relation $n = x^2 - y^2$ et l'on cherche des conditions sur *x* pour que y^2 soit un carré parfait. Si les valeurs permises de *x* sont 1, 7, 8, 13, 14 et 20 modulo 21, c'est-à-dire si le reste de la division de *x* par 21 peut être 1, 7, 8, 13, 14, 20, alors on place un capuchon métallique sur les picots 1, 7, 8, 13, 14, 20 de la couronne 21. De même pour les autres couronnes utilisées. On tourne la manivelle et, quand les capuchons métalliques (les plots noirs) des différentes couronnes sont alignés sur la barrette, un contact électrique s'établit et un signal sonore retentit. Le nombre indiqué sur le cadran a alors des chances d'être un carré parfait. La machine est particulièrement adaptée pour la recherche de nombres convenant pour le crible quadratique, utilisé pour la factorisation des nombres. Carissan prouva avec elle, en tournant la manivelle pendant dix minutes, que 708 158 977 est un nombre premier, en montrant qu'il s'écrit d'une seule manière comme somme de deux carrés : $19\,224^2 + 18\,401^2$. La machine est aujourd'hui au Conservatoire national des arts et métiers, à Paris.





2. TOUTES LES COURONNES de la «machine à congruences» ont une position zéro (visible en rouge sur la photographie de gauche, où l'on a enlevé la barrette métallique). Les couronnes ne roulent pas seulement sur la roue dentée commune, mais aussi sur des disques, dont les axes, fixés à l'axe de la roue dentée, font des angles de 120 degrés les uns avec les autres *(à droite)*. Seules les pointes portant un capuchon produisent le contact dans la barrette métallique.

F_5, égal 4 294 967 297, était égal au produit de 641 par 6 700 417 : F_5 n'est donc pas un nombre premier (on dit que c'est un nombre composé). Puis, 150 ans plus tard, en 1880, F_6 fut factorisé et, encore 90 ans plus tard, en 1970, les deux facteurs premiers de F_7 ont été obtenus. Le nombre F_8 a été décomposé en 1980, et le nombre F_9 en 1990.

Cette histoire des nombres de Fermat montre combien la factorisation est difficile : les mathématiciens mirent 240 ans pour factoriser le nombre de Fermat F_7, à 39 chiffres, et, même équipés d'ordinateurs ils ont mis 20 ans à décomposer le nombre F_9, de 155 chiffres.

En revanche, il savaient que F_7, F_8 et F_9 n'étaient pas premiers bien avant de connaître leurs facteurs : on peut déterminer, grâce à un test spécial, si un nombre est composé, sans calculer ses facteurs. On s'assure ainsi que la factorisation est possible, avant de l'entreprendre.

Un test simple de primalité

Comment détermine-t-on qu'un nombre est premier? On peut utiliser le test de Fermat, c'est-à-dire examiner si n est un diviseur de $2^{n-1}-1$: Fermat a prouvé que tout nombre premier n (sauf 2) est diviseur de 2^{2n-1}. Par exemple, 3, qui est premier, est un diviseur de $2^{3-1}-1$, soit 3. Inversement, si n n'est pas un diviseur de $2^{n-1}-1$, n n'est pas premier. Par exemple, le nombre 6 n'est pas un diviseur de $2^{6-1}-1$, soit 31 ; il n'est pas premier.

Examinons un exemple plus compliqué, afin d'apprécier l'utilité du test de Fermat : le nombre 58 483 est-il premier? On pourrait commencer par calculer $2^{58\,482}-1$, puis chercher si le résultat est un multiple de 58 483, mais un calcul direct serait extrêmement long, car le nombre $2^{58\,482}-1$ s'écrit avec 17 604 chiffres. On utilise plutôt l'arithmétique des congruences *(voir l'encadré de la page 91)*, pour obtenir un algorithme efficace, fondé sur l'idée suivante : vérifier que 58 483 est un diviseur de $2^{58\,482}-1$ revient à chercher si le reste de la division de $2^{58\,482}$ par 58 483 est 1.

Pour calculer ce reste, on effectue une «exponentiation binaire». On commence par écrire 58 483 comme une somme de puissances de 2 : $58\,482 = 2^{15}+2^{14}+2^{13}+2^{10}+2^6+2^5+2^4+2$. Puis on utilise cette décomposition pour écrire le nombre $2^{58\,482}$ comme un produit de puissances de 2 : $2^{2^1}\times 2^{2^4}\times 2^{2^5}\times 2^{2^6}\times 2^{2^{10}}\times 2^{2^{13}}\times 2^{2^{14}}\times 2^{2^{15}}$. Ensuite on calcule successivement les restes des divisions par 58 483 de 2^{2^0}, 2^{2^1}, 2^{2^2}, 2^{2^3}... jusqu'à $2^{2^{15}}$. Le reste de la division de $2^{2^{i+1}}$ par 58 483, pour n'importe quel nombre entier i, s'obtient simplement à partir du reste précédent (celui de la division de 2^{2^i} par 58 483) : on l'élève au carré, on divise ce carré par 58 483, et on prend le reste de cette division *(voir la figure 3)*. Ainsi, de proche en proche, on détermine le reste par 58 483 de la division pour chacune des puissances qui interviennent dans le développement binaire de 58 482. On multiplie enfin tous ces restes, mais en ne gardant, après chaque multiplication, que le reste de la division par 58 483. De cette façon, les résultats intermédiaires restent de taille raisonnable. On détermine ainsi que le reste de la division de $2^{58\,482}$ par 58 483 est égal à celui du produit $4\times 7\,053\times 34\,259\times 42\,237\times 34\,763\times 53\,664\times 5010\times 10\,893$: il est égal à 11 669. Pour simplifier l'expression des résultats, les mathématiciens écrivent : $2^{58\,482}\equiv 11\,669 \bmod 58\,483$. Le signe ≡ se lit «congru à» ; il indique que les expressions à sa gauche et à sa droite ont le même reste dans la division par le nombre indiqué après l'abréviation mod, laquelle se lit «modulo».

Le système RSA

Le protocole de cryptographie RSA se fonde sur le résultat suivant :

Soit p et q deux nombres premiers. On pose $n = pq$. Si e est un nombre entier premier avec $(p-1)(q-1)$, alors il existe un nombre entier d positif, tel que, pour tout A, $A^{ed} - A$ soit divisible par n.

(X et Y sont dits premiers entre eux si leur seul diviseur commun est 1).

Le protocole est le suivant :

Alice choisit p, q, e (p et q premiers, et e premier avec $(p-1)(q-1)$.

Alice calcule $n = pq$ et d, ce qui est facile.

Alice rend publics n et e (par exemple en les publiant dans un annuaire).

Robert, qui veut communiquer une information secrète à Alice, transforme son information en un nombre entier A inférieur à n (ou en plusieurs nombres si nécessaire) avec un codage connu de tous.

Robert calcule le reste B de la division de A^e par n, envoie B à Alice par un canal qui n'a pas besoin d'être protégé (par exemple en publiant B dans un journal).

Alice, pour décoder B, va calculer B^d ; comme les restes de la division de B^d et de A^{ed} par n sont égaux à A, elle connaîtra A et donc le message de Robert.

Sécurité du système RSA

La sécurité de ce codage à clé publique repose sur la difficulté du calcul des facteurs de n : quiconque les connaît peut facilement calculer d et déchiffrer les messages envoyés par Robert à Alice.