UN ALGORITHME DE FACTORISATION DES ENTIERS

On sait que tout nombre entier naturel s'écrit de façon unique comme produit de nombres premiers. Ce théorème est effectif. On connaît de nombreux algorithmes, plus ou moins efficaces, pour factoriser un entier naturel en produit de facteurs premiers. Aucun n'est polynomial. La difficulté de factoriser des entiers garantit la sécurité de plusieurs protocoles cryptographiques. Il est donc utile de connaître les algorithmes les plus performants pour factoriser des entiers.

1. Le paradoxe des anniversaires

On suppose qu'il y a 40 étudiants dans un groupe de TD. La probabilité que deux d'entre eux fêtent leur anniversaire le même jour de l'année est (on suppose que personne n'est né un 29 février)

$$1-(1-\frac{1}{365})(1-\frac{2}{365})\cdots(1-\frac{39}{365})\geq 0.89$$

Il y a donc une forte probabilité que deux étudiants fêtent leur anniversaire le même jour. Et pourtant 40 est très éloigné de $365\ldots$

On va donner une explication.

Supposons que l'on dispose d'une urne contenant p boules numérotées de 1 à p. On procède à $n\geq 2$ tirages *avec remise*. On estime la probabilité $P(p,n)$ d'avoir tiré n boules deux à deux distinctes :

$$P(p,n)=\prod_{1\leq i\leq n-1}\left(1-\frac{i}{p}\right)\leq\prod_{1\leq i\leq n-1}\exp(-\frac{i}{p})=\exp(-\frac{n(n-1)}{2p})\leq\exp(-\frac{(n-1)^2}{2p})$$

Donc si n est plus grand que $1+\sqrt{p}$ la probabilité de tirer deux fois la même boule est minorée par la constante $1-\exp(-1/2)>0.39$.

Il suffit donc que le nombre de tirages soit proportionnel à la racine carrée du nombre de boules pour avoir une probabilité significative de tirer deux fois la même boule.

2. Application aléatoire d'un ensemble fini dans lui même

Soit F un ensemble fini à p éléments. Soit $\mathcal{A}(F)$ l'ensemble des applications de F dans F. On munit $\mathcal{A}(F)$ de la mesure uniforme. On fixe un élément O de F.

À toute application $f:F\to F$ on associe la suite $x_0=O$, $x_{i+1}=f(x_i)$ obtenue par itération de f. Cette suite est ultimement périodique. Cela signifie qu'elle est périodique à partir d'un certain rang. On note π_f la période et μ_f la prépériode. Donc π_f est le plus petit entier positif tel que pour tout i assez grand on ait $x_{i+\pi_f}=x_i$. Et μ_f est le plus petit entier tel que pour tout $i\geq\mu_f$ on ait $x_{i+\pi_f}=x_i$.

La somme $\rho_f=\mu_f+\pi_f$ est une variable aléatoire sur $\mathcal{A}(F)$. La probabilité de l'événement $\rho_f\geq n$ est $P(p,n)$. L'espérance (c'est-à-dire la moyenne) $E(\rho_f)$ vérifie

$$\begin{aligned} E(\rho_f) &= 1 + \sum_{n\geq 2} P(p,n) \leq \sum_{n\geq 0} \exp(-\frac{n^2}{2p}) \leq 1 + \int_0^\infty e^{-\frac{x^2}{2p}} dx \\ &= 1 + \sqrt{2p} \int_0^\infty e^{-x^2} dx = 1 + \sqrt{\frac{p\pi}{2}} \end{aligned}$$

car $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$.

3. Deux algorithmes simples pour factoriser

Observons que la vérification de la primalité d'un nombre entier est chose facile. Il suffit donc, pour factoriser, de savoir trouver, pour tout N, un facteur non trivial M de N (et son cofacteur $R = N/M$). On dit qu'on a *cassé* N. Si les facteurs trouvés M et R ne sont pas premiers, on peut leur appliquer à nouveau l'algorithme de *cassage*.

Une première méthode consiste à calculer la division euclidienne de l'entier N à factoriser par les entiers $r = 2$, 3, 5, 7, 9, 11, 13, 15 etc.

Si on parvient à $r \geq \sqrt{N}$ sans trouver de facteur, alors le nombre N est premier et l'algorithme s'arrête.

La complexité de cette méthode est en $O(N^{\frac{1}{2}+o(1)})$ ce qui n'est pas très bon.

Une méthode élégante et beaucoup plus efficace (mais heuristique) est due à Pollard.

On suppose pour simplifier que $N = pq$ avec p et q premiers distincts. On fixe un polynôme f à coefficients entiers (souvent $f(X) = X^2 + 1$) et on considère la suite itérée à valeur dans $\mathbb{Z}/N\mathbb{Z}$ définie par $x = x_0$ quelconque et $x_{k+1} = f(x_k) \bmod N$.

Comme $f(X)$ est un polynôme, l'application induite

$$\begin{aligned} f_N : \quad \mathbb{Z}/N\mathbb{Z} &\longrightarrow \mathbb{Z}/N\mathbb{Z} \\ x \bmod N &\longmapsto f(x) \bmod N \end{aligned}$$

est application *produit* des deux applications

$$\begin{aligned} f_p : \quad \mathbb{Z}/p\mathbb{Z} &\longrightarrow \mathbb{Z}/p\mathbb{Z} \\ x \bmod p &\longmapsto f(x) \bmod p \end{aligned}$$

et

$$\begin{aligned} f_q : \quad \mathbb{Z}/q\mathbb{Z} &\longrightarrow \mathbb{Z}/q\mathbb{Z} \\ x \bmod q &\longmapsto f(x) \bmod q \end{aligned}$$

Plus précisément, on note γ l'isomorphisme Chinois

$$\gamma : \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}$$

et on vérifie que $\gamma \circ f_N = (f_p \times f_q) \circ \gamma$.

On suppose que les applications f_p et f_q se comportent comme des applications aléatoires indépendantes. Plus précisément, on suppose que f_p suit la loi uniforme dans l'ensemble $\mathcal{A}(\mathbb{Z}/p\mathbb{Z})$ des applications de $\mathbb{Z}/p\mathbb{Z}$ dans lui même. On suppose aussi que f_q suit la loi uniforme dans $\mathcal{A}(\mathbb{Z}/q\mathbb{Z})$, et que f_p et f_q sont indépendantes.

C'est une hypothèse un peu folle puisque $f = x^2 + 1$ est déterminé donc choisi aléatoirement dans un ensemble à un seul élément...

On note $y_k = x_k \bmod p$ la classe de x_k modulo p. On note $z_k = x_k \bmod q$ la classe de x_k modulo q.

On vérifie que $y_{k+1} = f_p(y_k)$ et $z_{k+1} = f_q(z_k)$. L'isomorphisme Chinois γ envoie x_k sur (y_k, z_k).

On note π_p et μ_p la période et la prépériode de f_p. On note π_q et μ_q la période et la prépériode de f_q.

On a de bonnes raisons de penser que π_p et μ_p (qui dépendent de f et de p) sont $O(\sqrt{p})$. De même on a de bonnes raisons de penser que π_q et μ_q sont $O(\sqrt{q})$.

Ainsi on dispose d'une suite itérée dans $\mathbb{Z}/N\mathbb{Z}$ dont la composante en p (resp. q) a période et prépériode $O(\sqrt{p})$ (resp. $O(\sqrt{q})$).

Si k est assez grand on aura très vraisemblablement

$$pgcd(x_k - x_{k+\pi_p}, N) = p$$

ce qui donne un facteur non trivial de N. Bien sûr, cette formule n'est d'aucun secours puisqu'on ne connaît pas π_p. Mais elle implique que pour k assez grand et multiple de π_p mais non de π_q on a

$$pgcd(x_k - x_{2k}, N) = p.$$

L'algorithme de Pollard consiste à calculer itérativement $x_k = f(x_{k-1})$ et $X_k = x_{2k} = f(f(X_{k-1}))$ et le pgcd ci-dessus pour $k = 0, 1, 2, ...$, jusqu'à trouver un facteur de N.

Heuristiquement cette méthode trouve un facteur p en temps $O(p^{\frac{1}{2}+o(1)})$ soit $O(N^{\frac{1}{4}+o(1)})$.

4. QUESTIONS

(1) On pourra citer un protocole cryptographique qui repose sur la difficulté de factoriser et voir dans quelle mesure la sécurité de ce protocole est affectée par l'algorithme présenté dans le sujet.

(2) On pourra justifier le calcul de l'espérance de ρ_f.

(3) On pourra étudier l'intégrale $\int_{-\infty}^{\infty} e^{-x^2} dx$, soit en calculant une valeur approchée, soit en démontrant qu'elle est égale à $\sqrt{\pi}$.

Dans ce dernier cas, on pourra montrer que $\left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 = \pi$ en écrivant

$$\left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 = \int_{-\infty}^{\infty} e^{-x^2} dx \int_{-\infty}^{\infty} e^{-y^2} dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2-y^2} dxdy$$

et en introduisant les coordonnées polaires $r = \sqrt{x^2+y^2}$ et l'angle θ.

(4) On pourra voir que certains choix pour le polynôme $f(X)$ sont plus pertinents que d'autres. Par exemple on pourra se demander si $f(X) = X^2$ est meilleur ou pire que $f(X) = X^2 + 1$. Sachant que $(x + 1/x)^2 = x^2 + 1/x^2 + 2$ que peut-on dire du choix $f(X) = X^2 - 2$?

(5) On pourra essayer d'estimer la période π_N et la prépériode μ_N de $f_N : \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$ en fonction de π_p, μ_p, π_q, μ_q.

(6) On pourra expliquer pourquoi l'algorithme de Pollard est particulièrement efficace pour trouver les petits facteurs premiers de grands entiers et illustrer cette méthode par une implémentation.