Factorisation de Lenstra par les courbes elliptiques

Projet Crypto

Natacha GILLAIZEAU-SIMONIAN

2020

Sommaire

1 Introduction : Le problème de la factorisation

Soit n un entier. On sait par le théorème fondamental de l'arithmétique que n se décompose de manière unique en produit de nombres premiers.

On ne connaît aucun algorithme de factorisation d'entier en temps polynomial, mais il n'est pas prouvé que le problème de factorisation ne soit pas dans la classe de complexité P. En revanche, ce problème est clairement dans NP.

Nous n'avons donc aucune certitude que le problème de factorisation d'entiers soit un problème "difficile" et pourtant beaucoup cryptosystèmes s'appuient sur la difficulté du problème de factorisation. On peut par exemple citer le cryptosystème RSA dont toute la sécurité repose sur le secret des deux facteurs premiers de la clé publique.

Les entiers de la forme $n = pq$ avec p et q premiers sont les plus difficiles à factoriser. Ce sont des nombres de cette forme qui sont utilisés dans RSA. En revanche, pour les nombres qui sont le produit de plusieurs "petits" facteurs, on a des algorithmes qui permettent de les factoriser rapidement.

2 Méthode $p-1$ Pollard

2.1 Principe

La méthode $p-1$ Pollard est une méthode de factorisation d'entiers possédant de "petits" facteurs.

L'idée est d'utiliser le petit théorème de Fermat :

$a^{p-1} \equiv 1 \bmod p$ si p|n, c'est-à-dire si p|pgcd($a^{p-1}-1$,p).

C'est aussi vrai si l'exposant est juste un multiple de $p-1$: si $p-1|M$ alors $p|pgcd(a^m-1,n)$.

Il est facile de trouver un tel multiple si $p-1$ n'a que des facteurs premiers "petits".

C'est un algorithme spécifique. Soit n un entier composé, $n = \prod_{i=1}^{k} p_i^{e_i}$. Il marche si les $p_i - 1$ sont friables.

Definition 2.1. Soient m et B deux entiers, on dit que m est B-FRIABLE si
$m = \prod_{q<B} q^{e_q}$, q premier
Et B-SUPERFRIABLE si $m = \prod_{q^{e_q}} q^{e_q}$, q premier.

Proposition 2.1. *Soit n un entier divisible par un nombre premier p tel que $p-1$ soit B-friable.*
Alors en prennant $m = \prod_{q \leq B} q^{\lfloor log_q n \rfloor}$
On a $\forall a \in \mathbb{Z}/n\mathbb{Z} \qquad p | pgcd(a^m - 1, n)$

Démonstration. $p-1 = \prod_{q \leq B} q^{e_q}$, or $p-1 < m$ donc $e_q \leq \lfloor log_q n \rfloor$ donc $p-1|n$. □

2.2 Pseudo-Algorithme

Entrées : n un entier composé, B la borne de friabilité, a dans $\mathbb{Z}/n\mathbb{Z}$
Sortie : un facteur non trivial de n ou échec.

```
Pollard(n,B,a) :
   aM = Mod(a,n)
   Pour tout premier q < B
      aM = aM^(p^logint(n,p));
      Si gcd(aM-1,n) > 1
        Alors rendre gcd(aM-1,n)
   Fin Pour
   Rendre échec
```

3 Courbes elliptiques sur un corps fini

3.1 Définition

Soit $K = \mathbb{F}_q$ un corps fini à q éléments.
Une courbe elliptique $E(K)$ sur le corps K est de la forme : $y^2 = x^3 + ax + b$ $a, b \in K$

Definition 3.1. Soit K un corps de caractéristique différente de 2 et 3, et $x^3 + ax + b \in K[x]$ un polynôme sans carré. Alors

$$E = \{(x, y) \in \mathbb{K}^2 : y^2 = x^3 + ax + b\} \cup \{\mathcal{O}\} \qquad (1)$$

est une courbe elliptique sur K. $\mathcal{O}$ représente le "point infini" de E.

Remarque. Le polynôme $x^3 + ax + b$ est sans carré si $4a^3 + 27b^2 \neq 0$.

Remarque. On supposera que E(K) est un groupe abélien.

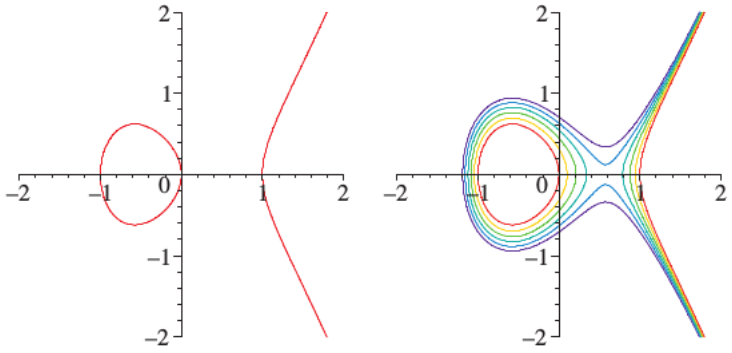

Figure 1: La courbe elliptique $y^2 = x^3 - x$ *(diagramme de gauche)* et courbes elliptiques $y^2 = x^3 - x + b$ pour b = 0, 0.1, 0.2, 0.3, 0.4 ,0.5 *(diagramme de droite)*

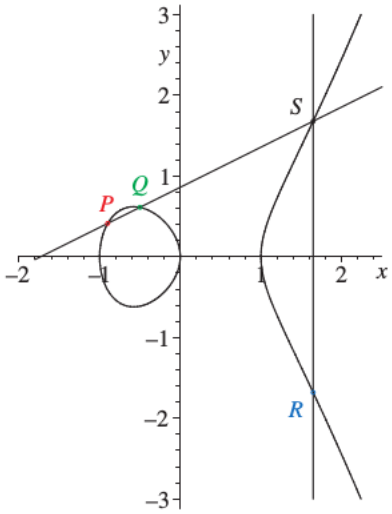

Figure 2: On additionne P et Q sur la courbe elliptique.

3.2 Addition sur E(K)

Quand on a 2 points P et Q sur E, on calcule les coordonnées de S le point d'intersection de la droite qui joint P et Q avec la courbe.
En posant R le symétrique de S, on a $P + Q = R$ donne une loi de groupe abélien.

- Si la droite (PQ) est verticale (*ie* $Q = -P$), alors $P + (-P) = -\mathcal{O} = \mathcal{O}$
- Si la droite (PQ) est tangente à un des 2 points (disons Q ici), alors on

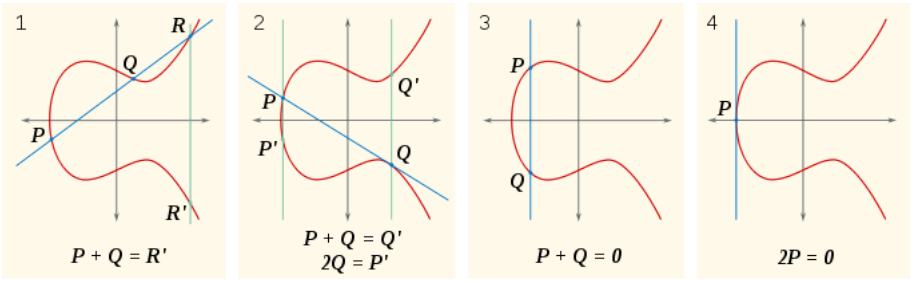

Figure 3: Différents cas pour l'addition

pose S = Q et donc R est le symétrique de Q.

- Si $P = Q$, on utilise la tangente au point. Elle coupe la courbe en un 3^{eme} point (S). On pose R son symétrique.
- $\mathcal{O}$ est l'élément neutre $(P + \mathcal{O} = -(-\mathcal{P}) = \mathcal{P})$

Supposons maintenant que $P = (x_1, y_1)$ et $Q = (x_2, y_2)$, $x_1 \neq x_2$ et $R = (x_3, y_3) = P + Q \in E\backslash\{\mathcal{O}\}$.

On a, si $P \neq Q$:

$$x_3 = (\tfrac{y_2 - y_1}{x_2 - x_1})^2 - x_1 - x_2 \qquad y_3 = -y_1 + \tfrac{y_2 - y_1}{x_2 - x_1} \qquad (2)$$

Et si $P = Q$:

$$x_3 = (\tfrac{3x_1^2 + a}{2y_1})^2 - 2x_1 \qquad y_3 = -y_1 + \tfrac{3x_1^2 + a}{2y_1} \cdot (x_1 - x_3) \qquad (3)$$

si $y_1 \neq 0$ et $2P = \mathcal{O}$ si $y_1 = 0$

3.3 Nombre de points sur $E(\mathbb{F}_q)$

Proposition 3.1. *Soit $E : y^2 = x^3 + ax + b$ une courbe elliptique sur $\mathbb{F}_p$*

- $p + 1 - 2\sqrt{p} \leq \#E(\mathbb{F}_p \leq p + 1 + 2\sqrt{p}$ *(Th. de Hasse)*
- *Pour tout nombre premier q tel que $|p + 1 - q| \leq 2\sqrt{p}$ il existe une courbe de cardinal q.*

Exemple. *Prenons $p = 7$. Par le théorème de Hasse, chaque courbe elliptique E sur $\mathbb{F}_7$ on a $3 \leq \#E \leq 13$*

3.4 E($\mathbb{Z}/n\mathbb{Z}$), N composé

Supposons n un entier composé, et un nombre premier p facteur de n.

On suppose $E : y^2 = x^3 + ax + b$ une courbe elliptique sur $\mathbb{Z}/n\mathbb{Z}$, c'est-à-dire que $a, b \in \mathbb{Z}/n\mathbb{Z}$.

On pose E_p la courbe elliptique avec les coefficients a et b de E réduits modulo p.

On a un morphisme de "groupes" entre E et E_p.

Ainsi, si on prend un point P = (x,y) sur E et qu'on considère le point P_p = (x mod p, y mod p), alors en prenant $\mathcal{O}_p$ correspondant à $\mathcal{O}$, on a que $P_p \neq \mathcal{O}_p$ pour tout $P \in E\backslash\{\mathcal{O}\}$ et ainsi,

$$P_p = \mathcal{O}_p \Longleftrightarrow P = \mathcal{O} \qquad (4)$$

4 La méthode de Lenstra

4.1 Principe général

La méthode de factorisation de Lenstra par les courbes elliptiques, ou en anglais *elliptic-curve factorization method* **(ECM)** est un algorithme de factorisation probabiliste qui doit son nom à Hendrik Lenstra.

La méthode ECM repose sur les mêmes principes que la méthode $p-1$ de Pollard, à l'exception qu'on travaille sur des courbes elliptiques au lieu de l'anneau $\mathbb{Z}/n\mathbb{Z}$.

En effet, l'idée est de choisir aléatoirement une courbe elliptique sur $\mathbb{Z}/n\mathbb{Z}$. On prend ensuite un point P sur cette courbe. Pour faire tout cela, on tire aléatoirement x,y et a et on résout sur b pour avoir une équation de courbe elliptique. C'est-à-dire qu'on veut : $y^2 = x^3 + ax + b$ et P = (x,y). On a aussi une borne B, comme dans l'algorithme $p-1$. On va tester tous les premiers inférieurs à B. On a aussi une borne C sur les puissances des premiers.

On tente ensuite de calculer les kP jusqu'à la borne B (comme p-1). Comme n est un entier composé, $\mathbb{Z}/n\mathbb{Z}$ n'est pas un corps. Ainsi, si lors du calcul de l'addition de points, on tombe sur un dénominateur non inversible dans $\mathbb{Z}/n\mathbb{Z}$, on a trouvé un facteur de n. On calcule donc les kP en espérant tomber sur un dénominateur non inversible.

Si on ne trouve pas de tel point, il suffit d'essayer avec une autre courbe elliptique.

La méthode ECM est très efficace pour trouver les "petits" facteurs. En effet, son efficacité dépend de la taille des facteurs plutôt que de celle du nombre à factoriser. Ainsi, pour factoriser les grands nombres, l'algorithme ECM permet de sortir rapidement les petits facteurs.

Une fois les petits facteurs sortis, on peut factoriser le nombre restant avec un algorithme plus général comme par exemple la méthode du crible quadratique. En effet, pour cet algorithme, le temps d'exécution dépend uniquement de la taille de l'entier à factoriser. On peut donc estimer le temps d'exécution de ce genre d'algorithme et tenter de le réduire en sortant d'abord des petits facteurs avec la méthode ECM.

Il est important de noter que cet algorithme se parallèlise très bien, ce qui le rend d'autant plus intéressant.

4.2 Pseudo-Algorithme

Entrée : N un entier composé, B borne sur la taille des premiers, C borne sur les puissances des premiers testés.
Sortie : un facteur non trivial de n ou échec.

```
ECM(N,B,C) :
Choisir aléatoirement (a,u,v) ∈ {0,...,N-1}^3
b ← v^2 - u^3 - au,        g ← pgcd(4a^3 + 27b^2, N)
si 1 < g < n alors rendre g
si g = N rendre "échec"      \\ réessayer avec d'autres a,u et v

Soit E une "courbe elliptique" sur ℕ/nℤ d'équation y^2 = x^3 + ax + b
Soient p_1 = 2 < p_2 < ... < p_h, les nombres premiers < B

P ← (u,v),    Q ← P,    t ← 1

Pour i allant de 1 à h
   e_i = ⌊log_{p_i}(C + 2√C + 1)⌋
   Pour j allant de 0 à e_i - 1
     {Invariants de boucle :  t = p_i^j ∏_{1≤r≤i} p_r^{e_r}  et  Q = tP}
     Tenter de calculer p_i Q
     Si un dénominateur ω ∈ {1,...,N-1} n'est pas inversible mod N
     Alors rendre pgcd(ω, N)
     Sinon Q ← p_i Q,       t ← p_i t
Rendre "échec"
```

On met courbe elliptique entre guillemets car pour un entier composé, la définition d'une courbe elliptique a proprement parlé est plus compliquée. Mais cela n'a aucune importance pour l'alogrithme, on peut supposer que E est bien une courbe elliptique.

Pour calculer p_iQ, on utilise les formules (2) et (3) d'addition de points.

Dans la pratique, plutôt que de rendre échec, on tente avec une autre courbe. Si après un certain nombre de courbes testées on n'a toujours pas de résultat, on augmente les bornes. Et ainsi de suite jusqu'à trouver un facteur de N.

4.3 Arrêt de l'algorithme

Si N possède un facteur plus petit que C, alors l'algorithme a des chances de s'arrêter.

En effet, supposons p un facteur premier de N inférieur à C. On sait que p ne divise pas $4a^3 + 27b^2$, sinon l'algorithme s'arrête à la première étape (soit avec succès, dans ce cas on a notre facteur ; soit sans succès, dans ce cas il suffit de relancer l'algorithme et on aura des a et b différents).

Grâce à (4), on peut supposer qu'on travaille sur E_p, puisque chaque résultat sur E donne un résultat modulo p sur E_p ($Q = tP \to Q_p = tP_p = (tP)_p$).

L'évènement qu'on attend arrive lorsque, pour 2 facteurs premiers p et q de N, on arrive sur un multiple de l'ordre de P_p sur E_p qui n'est pas multiple de l'ordre de P_q sur E_q.

Lemme 4.1. *Supposons (E,P) choisis, p,q deux facteurs premiers de N, l le plus grand facteur premier de l'ordre de P_p dans E_p, $p \leq C$, $\#E_p$ B-friable et $l \nmid \#E_p$. Alors l'algorithme factorise N.*

Démonstration. Posons $k = \prod_{1 \leq r \leq h} p_r^{e_r}$ avec $e_r = \lfloor log_{p_r}(C + 2\sqrt{C} + 1) \rfloor$ pour $1 \leq r \leq h$ comme dans l'algorithme (h est le nombre de nombres premiers strictement inférieurs à B).

Il est facile de vérifier que les invariants de boucle sont toujours vrais avec la double boucle sur i et j.

Comme $\#E$ est B-friable et que $p \leq C$, on a par le théorème de Hasse que $\#E | k$.

Soit d l'ordre de P_p dans E_p. Par le théorème de Lagrange, on a que $d | E_p$. Ainsi $l \leq B$ et $d | k$.

Soit $1 \leq i \leq h$ tel que $l = p_i$, et soit e l'exposant de l dans E_p.

On a donc, quand $j = e - 1$:

$$t = l^{e-1} \prod_{1 \leq r \leq i} p_r^{e_r} \qquad \text{et} \qquad Q = tP.$$

au début de la boucle sur j. $t \not\equiv 0 \bmod d$ et $lt \equiv 0 \bmod d$.

On a donc que $Q_p = tP_p \neq \mathcal{O}_p$ et $lQ_p = ltP_p = \mathcal{O}_p$.

Par conséquent, même si l'algorithme fini par réussir à calculer lQ, alors le résultat ne peut-être que $\mathcal{O}$. On voit qu'en fait, il termine en trouvant un facteur avant d'atteindre cette situation.

Supposons par l'absurde que $lQ = \mathcal{O}$ est calculé par l'algorithme. Puisque que les résultats de ce calculs sont aussi valables dans E_q, on a calculé $lQ_q = (ltP)_q = \mathcal{O}_q$. Mais alors, puisque l ne divise pas $\#E_q$ ou l'ordre de P_q, le point $Q_q = tP_q$ est déjà égal à $\mathcal{O}_q$.

On a alors, par (4) que $Q = \mathcal{O}$ et $Q_q = \mathcal{O}_q$, ce qui est absurde.

□

4.4 Probabilités de réussite et complexité de l'algorithme

Nous allons à présent essayer évaluer la probabilité que les hypothèses du **Lemme 4.1** soient satisfaites.

Supposons que $\mathbb{P}(l \nmid \#E_q)$ est très proche de 1. Nous allons donc considérer uniquement la probabilité de choisir des paramètres donnant une courbe elliptique E_p telle que $\#E_p$ est B-friable.

Theorème 4.2. *Il existe $c \in \mathbb{R}_+^*$ tel que que la propriété suivante est vérifiée. Soit p un premier, $S \subseteq]p+1-\sqrt{p}, p+1+\sqrt{p}[$ et $\#S \geq 3$, et $a, b \in \mathbb{F}_p$ choisis aléatoirement. Soit*

$$E_p = \{(u, v) : v^2 = u^3 + au + b\}$$

une courbe elliptique sur $\mathbb{F}_p$. Alors :

$$\mathbb{P}(\#E_p \in S) \geq \frac{c \cdot \#S}{\sqrt{p} log(p)}$$

Notons que S est la moitié du milieu de l'intervalle donné par le théorème de Hasse.

Exemple. *Prennons $p = 7$. On a vu dans l'exemple de la partie 2.3 que pour chaque courbe elliptique dans $\mathbb{F}_7$, on a $3 \leq \#E \leq 13$.*
Ainsi, on a $S = \{6, 7, 8, 9, 10\}$.

Corollaire 4.2.1. *Il existe $c \in \mathbb{R}_+^*$ tel que que la propriété suivante est vérifiée. Soit $p \leq C$ un facteur premier de N, et*

$$\sigma = \#\{\text{nombres B-friables dans }]p+1-\sqrt{p}, p+1-\sqrt{p}[\}$$

Si $\sigma \geq 3$, alors le nombre M de triplets $(a, u, v) \in \{0, ..., N-1\}^3\}$ pour lesquels l'algorithme factorise N satisfait

$$\frac{M}{N^3} \geq \frac{c \cdot \sigma}{\sqrt{p} log(p)}$$

On note $s = \frac{\sigma}{2\sqrt{p}}$, la probabilité qu'un nombre aléatoire dans $]p+1-\sqrt{p}, p+1-\sqrt{p}[$ soit B-friable.
Si on lance m fois l'algorithme, alors la probabilité d'échec est au plus :

$$(1 - \frac{M}{N^3})^m \leq (1 - \frac{sc}{ln(p)})^m \leq (1 - \frac{sc}{ln(C)})^m \leq e^{-msc/ln(c)} \leq \epsilon$$

quand on choisit $m \geq ln(\frac{1}{\epsilon})\frac{ln(C)}{sc}$ avec c comme dans le **Corollaire 4.2.1**.

Une addition de point sur la courbe elliptique avec les formules (2) et (3) prend

un nombre constant d'opérations arithmétiques modulo N. De plus, le temps d'exécution de l'intérieur de la double boucle est $\mathcal{O}(log(p_i))$ opérations modulo N.

Ainsi, le coût total de l'algorithme est de $\mathcal{O}(\sum_{i \leq h} e_i log(p_i))$ ou $\mathcal{O}(mBlog(C))$ opérations modulo N. Et le nombre d'opérations arithmétiques modulo N pour m exécutions est

$$\mathcal{O}(mBlog(C)) \qquad \text{ou} \qquad \mathcal{O}(mBlog(N))$$

Augmenter B rend σ et s plus grands, et donc m plus petit. Il faut donc trouver le bon équilibre.

4.5 Choix des bornes

Globalement, choisir de bons paramètres est difficile, il n'y a pas vraiment de règle. Il faut essayer et essayer de trouver un bon équilibre.

digits	optimal B1	expected curves (default parameters for GMP-ECM 6)
20	11,000	86
25	50,000	214
30	250,000	430
35	1,000,000	910
40	3,000,000	2,351
45	11,000,000	4,482
50	43,000,000	7,557
55	110,000,000	17,884
60	260,000,000	42,057
65	850,000,000	69,471
70	2,900,000,000	102,212
75	7,600,000,000	188,056
80	25,000,000,000	265,557

Figure 4: Paramètres optimaux selon *Optimal parameters for ECM* (ref **[6]**)

On a ici un choix du paramètre C et du nombre de courbes à tester, tout ça en fonction du nombre de chiffres du nombre à factoriser.

De plus, selon Joachim von zur Gathen and Gerhard, *Modern Computer Algebra* (ref **[3]**), on peut prendre $B = e^{\sqrt{\frac{ln(C) \cdot ln(ln(C))}{2}}}$.

References

[1] Richard P. Brent. "Some integer factorization algorithms using elliptic curves". In: *arXiv:1004.3366 [cs, math]* (Apr. 20, 2010). arXiv: 1004.3366. URL: http://arxiv.org/abs/1004.3366 (visited on 06/03/2020).

[2] *Integer factorization*. In: *Wikipedia*. Page Version ID: 951308860. Apr. 16, 2020. URL: https://en.wikipedia.org/w/index.php?title=Integer_factorization&oldid=951308860 (visited on 06/04/2020).

[3] Joachim von zur Gathen and Jürgen Gerhard. *Modern Computer Algebra*. 3e édition. Cambridge University Press, 2013. ISBN: 978-1-107-03903-2. URL: www.cambridge.org/9781107039032.

[4] *Lenstra elliptic-curve factorization*. In: *Wikipedia*. Page Version ID: 957951778. May 21, 2020. URL: https://en.wikipedia.org/w/index.php?title=Lenstra_elliptic-curve_factorization&oldid=957951778 (visited on 06/03/2020).

[5] R Lercier. "Factoriser des entiers par la méthode des courbes elliptiques". In: (June 1993), p. 63.

[6] *Optimal parameters for ECM*. URL: https://members.loria.fr/pzimmermann/records/ecm/params.html (visited on 06/03/2020).

[7] Pascal Molin. *Factorisation d'entiers — Documentation Crypto M1 MIC 2020*. URL: https://webusers.imj-prg.fr/~pascal.molin/cours/crypto/cours_factorisation.html (visited on 06/03/2020).

[8] *Pollard's p-1 algorithm*. In: *Wikipedia*. Page Version ID: 955643709. May 8, 2020. URL: https://en.wikipedia.org/w/index.php?title=Pollard%27s_p_%E2%88%92_1_algorithm&oldid=955643709 (visited on 06/03/2020).