Un logiciel open-source établit un nouveau record de factorisation

Fabrice Boudot, Pierrick Gaudry, Aurore Guillevic, Nadia Heninger,
Emmanuel Thomé, **Paul Zimmermann**

Journée Scientifique EXPLOR 2020
17 décembre 2020

Plan

Plan

Définition du problème

Multiplication : $17 \times 41 \to 697$

Factorisation : $697 \to 17 \times 41$

$p = 499311167684973822389059665671 \times q = 920830787846073284631890875649 \to$ $N = pq$ instantané

$N = 459781095919697252641526682462311030892270546862555375145479 \to p \times q$ prend 10 secondes avec SageMath (60 chiffres)

Multiplier est *facile*, factoriser est *difficile* $\Longrightarrow$ fonction à sens unique

Problème mathématique sous-jacent au cryptosystème RSA (CB, `https`, ...)

Algorithmes de factorisation d'entier

- *trial division* : diviser par tous les nombres premiers jusqu'une borne B. Pour $B = 10^6$ il y en a 78498 !
- ECM (méthode des courbes elliptiques) : trouve les facteurs de taille moyenne. Record actuel : facteur premier de 83 chiffres.
- crible quadratique (QS) : jusque 100 chiffres décimaux.
- crible algébrique (NFS) : au delà de 100 chiffres décimaux.

Nombres RSA : $N = pq$ avec p et q premiers de même taille.

Historique

- 1991 : lancement du *RSA Factoring Challenge*
- 1991 : factorisation de RSA-100 (QS)
- 1992 : factorisation de RSA-110 (QS)
- 1993 : factorisation de RSA-120 (QS)
- 1993 : factorisation de RSA-130 (NFS)
- 1999 : factorisation de RSA-512 (NFS, 155 chiffres)
- 2003 : factorisation de RSA-576 (NFS, 174 chiffres)
- 2005 : factorisation de RSA-640 (NFS, 193 chiffres)
- 2009 : factorisation de RSA-768 (NFS, 232 chiffres)
- **2020 : factorisation de RSA-250 (NFS, 250 chiffres)**

Plan

1. Historique de la factorisation d'entier
2. Le crible algébrique
3. Le logiciel CADO-NFS
4. Le record RSA-250

Le crible algébrique en un slide

1. *Sélection polynomiale* : trouver deux polynômes irréductibles $f(x)$ et $g(x)$ à coefficients entiers, ayant une racine commune modulo N ;
2. *Crible* : trouver plein de paires d'entiers (a, b) telles que $f(a/b)$ et $g(a/b)$ sont simultanément friables ;
3. *Algèbre linéaire* : résoudre une immense matrice creuse à coefficients 0 ou 1 ;
4. *Racine carrée* : en déduire deux entiers X et Y tels que $X^2 \equiv Y^2 \bmod N$, et calculer $\gcd(X - Y, N)$.

Plan

Le logiciel CADO-NFS

- Développé depuis 2007 ;
- 250 000 lignes de code C/C++, dont 60 000 pour la phase de *crible* ;
- Plusieurs améliorations importantes depuis 2016 :
  - amélioration du parallélisme (suppression des *bulles*) ;
  - plus de latitude dans le choix des paramètres ;
  - outils de simulation ;
- Logiciel libre (LGPL), modèle de développement ouvert (gitlab).
  Nos résultats peuvent être reproduits !

Utilisation de CADO-NFS

```
$ git clone https://gitlab.inria.fr/cado-nfs/cado-nfs.git
$ cd cado-nfs
$ make
...
$ ./cado-nfs.py 459781095919697252641526682462311030892270546862555375145479
...
920830787846073284631890875649 499311167684973822389059665671
```

Plan

Factorisation de RSA-250

$N = \text{RSA-250}$

Sélection polynomiale

$$
\begin{aligned}
f &= 86130508464000x^6 \\
&\quad -66689953322631501408x^5 \\
&\quad -52733221034966333966198x^4 \\
&\quad +46262124564021437136744523465879x^3 \\
&\quad -3113627253613202265126907420550648326x^2 \\
&\quad -1721614429538740120011760034829385792019395x \\
&\quad -81583513076429048837733781438376984122961112000 \\
g &= 185112968818638292881913x \\
&\quad -3256571715934047438664355774734330386901
\end{aligned}
$$

$\mathrm{Res}(f, g) = 48N$

À quoi ressemble une relation ?

small primes, special-q, *large primes*

✓	$5^2 \cdot 11 \cdot 23 \cdot 287093 \cdot 870953 \cdot 20179693 \cdot 28306698811 \cdot 47988583469$	$2^3 \cdot 5 \cdot 7 \cdot 13 \cdot 31 \cdot 61 \cdot 14407 \cdot 26563253 \cdot 86800081 \cdot 269845309 \cdot 802234039 \cdot 1041872869 \cdot 5552238917 \cdot 12144939971 \cdot 15856830239$
✓	$3 \cdot 1609 \cdot 77699 \cdot 235586599 \cdot 347727169 \cdot 369575231 \cdot 9087872491$	$2^3 \cdot 3 \cdot 5 \cdot 13 \cdot 19 \cdot 23 \cdot 31 \cdot 59 \cdot 239 \cdot 3989 \cdot 7951 \cdot 2829403 \cdot 31455623 \cdot 225623753 \cdot 811073867 \cdot 1304127157 \cdot 78955382651 \cdot 129320018741$
✓	$5 \cdot 1381 \cdot 877027 \cdot 15060047 \cdot 19042511 \cdot 11542780393 \cdot 13192388543$	$2^4 \cdot 5 \cdot 13 \cdot 31 \cdot 59 \cdot 823 \cdot 2801 \cdot 26539 \cdot 2944817 \cdot 3066253 \cdot 87271397 \cdot 108272617 \cdot 386616343 \cdot 815320151 \cdot 1361785079 \cdot 12322934353$
✓	$2^3 \cdot 5^2 \cdot 173 \cdot 971 \cdot 613909489 \cdot 929507779 \cdot 1319454803 \cdot 2101983503$	$2^7 \cdot 3^2 \cdot 5 \cdot 29 \cdot 1021 \cdot 42589 \cdot 190507 \cdot 473287 \cdot 31555663 \cdot 654820381 \cdot 802234039 \cdot 19147596953 \cdot 23912934131 \cdot 52023180217$
✗	$2^2 \cdot 15193 \cdot 232891 \cdot 19514983 \cdot 139295419 \cdot 540260173 \cdot 606335449$	$2^2 \cdot 3^4 \cdot 13 \cdot 19 \cdot 74897 \cdot 1377667 \cdot 55828453 \cdot 282012013 \cdot 802234039 \cdot 3350122463 \cdot 35787642311 \cdot 37023373909 \cdot 128377293101$
✗	$2^2 \cdot 5^4 \cdot 439 \cdot 1483 \cdot 13121 \cdot 21383 \cdot 67751 \cdot 452059523 \cdot 33099515051$	$2^2 \cdot 3^3 \cdot 11 \cdot 13 \cdot 19 \cdot 5023 \cdot 3683209 \cdot 98660459 \cdot 802234039 \cdot 1506372871 \cdot 4564625921 \cdot 27735876911 \cdot 32612130959 \cdot 45729461779$

small primes : nombreux $\rightarrow$ colonnes denses de la matrice

large primes : rares $\rightarrow$ colonnes creuses (au plus 3 de chaque côté ici).

Quelques chiffres

	RSA-250
sélection polynomiale	114 core-years
$\deg f, \deg g$	6, 1
crible	2450 core-years
relations brutes	8 745 268 073
relations uniques	6 132 671 469
filtrage	qq jours
après *singleton removal*	2 739 226 048
après *clique removal*	1 816 698 332
après *merge*	405M lignes, densité 252
algèbre linéaire	250 core-years
racine carrée	qq jours

Contribution d'EXPLOR

EXPLOR a été utilisé pour la phase de *crible*, du 25 novembre 2019 au 21 janvier 2020.

Sur 602551 tâches, 60381 ont été faites sur EXPLOR, soit environ 107 core-years.

Règles de *fair sharing* : nous pouvions utiliser au maximum 25% de la plate-forme EXPLOR à un instant donné.

Phase de crible *embarrassingly parallel* : un nœud = une tâche (le logiciel se débrouille tout seul pour utiliser de manière optimale les k cœurs du nœud).

Autres clusters utilisés : Juwels/PRACE (1380 core-years), Grid-5000 Nancy (590 core-years), autres sites Grid-5000 (235 core-years), UC San Diego (140 core-years).

Et finalement...

$$
\begin{aligned}
\text{RSA-250} &= 2140324650240744961264423072839333563008614715144755017797754920\\
&\quad 8814180234471401366433455190958046796109928518724709145876873 9\\
&\quad 6261921557363047454770520805119056493106687691590019759405693 45\\
&\quad 7452230589325976697471681738069364894699871578494975937497937,\\
p &= 6413528947707158027879019017057738908482501474294344720811685 96\\
&\quad 32024532344630238623598752668347708737661925585694639798853367,\\
q &= 3337202759497815655622601060535511422794076034476755466678452 09\\
&\quad 87023841729210037080257448673296881877565718986258036932062711
\end{aligned}
$$

Et l'ordinateur quantique ?

En 1994, Peter Shor a inventé un algorithme pour factoriser un entier sur un ordinateur quantique

Cet algorithme est plus rapide que NFS sur un ordinateur classique

La factorisation d'un entier de n bits nécessite un ordinateur quantique *parfait* avec $2n$ qubits (bits quantiques)

Un ordinateur quantique est très difficile à construire, et encore plus un ordinateur quantique *parfait*

Record actuel avec un ordinateur quantique (2018) : $4\,088\,459 = 2017 \times 2027$

RSA-1024 (1024 bits) sera sans doute factorisé avant que l'ordinateur quantique devienne compétitif (pour ce problème).