Better Polynomials for GNFS

(joint work with Shi Bai, Cyril Bouvier and Alexander Kruppa)

Paul Zimmermann

UNSW, September 21, 2016

RSA-768

$$
\begin{aligned}
N &= 123018668453011775513049495838496272077285356959533479219732245215172640050726365751874520219978646938995647494277406384592519255732630345373154826850791702612214291346167042921431160222124047927473779408066535141959745985690214341 3\\
f &= 265482057982680\,x^6\\
&+ 127650936076832188\,x^5\\
&- 5006815697800138351796828\,x^4\\
&- 464778544717278542717267745 0\,x^3\\
&+ 652543726193598939710966737189 4785\,x^2\\
&- 1818577935208859435672601886243 4803054\,x\\
&- 2775652667915438819952161997138 01103343120\\
g &= 346610035504925018514458 29\,x\\
&- 129118745658002122316354779157 4810881
\end{aligned}
$$

How does one find f and g from N?

Polynomial Selection for GNFS

Given N, we want to find two polynomials f and g such that:

- f and g have integer coefficients
- f and g are irreducible over $\mathbb{Z}$
- f and g have a common root m modulo N
- f and g have small coefficients
- f and g have many roots modulo small primes

In this talk we will only speak about *linear* polynomial selection, i.e., g has degree one.

Non-linear polynomial selection is not yet fully understood, except when $\deg(f) = \deg(g) = 2$ (Montgomery two quadratics), and in the DLP case.

Size Property

Usually we measure the size of the non-linear polynomial f only.

$$f(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_0$$

Let $F(x, y)$ be the homegeneous polynomial

$$F(x, y) = a_d x^d + a_{d-1} x^{d-1} y + \cdots + a_0 y^d$$

$$\mathrm{lognorm}(f) = \min_{s>1} \frac{1}{2} \log \left(s^{-d} \int_0^{2\pi} \int_0^1 F^2(s\cos\theta, \sin\theta) r^{2d+1} \mathrm{d}r\mathrm{d}\theta \right)$$

A lower bound for the norm comes from the middle coefficient of f

Polynomial Selection 1.0

Brian Murphy, *Polynomial Selection for the Number Field Sieve Integer Factorisation Algorithm*, Australian National University, 1999.

Base m method:

1. Input: N, a positive integer m
2. Decompose N in base m with $0 \leq a_j < m$:

$$N = a_d m^d + a_{d-1} m^{d-1} + \cdots + a_1 m + a_0$$

3. Take $f = a_d x^d + \cdots + a_0$ and $g = x - m$

In general, we have $a_j = O(m)$.

However, if one first chooses a degree d and a leading coefficient a_d, and takes $m = \lfloor (N/a_d)^{1/d} \rfloor$, then $a_{d-1} = O(da_d)$.

Root Property

Brian Murphy defined $\alpha(f)$ to measure the root property of f.

The smaller $\alpha(f)$, the better f is.

With a given α value, the average norm of f is reduced by $\exp^{\alpha}$.

RSA-768: we had $\alpha(f) \approx -7.30$, which reduced the norm by a factor about 1500.

One thus often considers the *combined score*:

$$E(f) = \mathrm{lognorm}(f) + \alpha(f)$$

Rootsieve

Polynomial selection consists in two stages:

1. size optimization: find a few polynomials f with small norm
2. rootsieve: optimize the root properties of those f and keep the best one

$$
\begin{aligned}
f &= 265482057982680\,x^6 \\
&+ 1276509360768321888\,x^5 \\
&- 5006815697800138351796828\,x^4 \\
&- 464778544717278542717726774 50\,x^3 \\
&+ 652543726193598939710966737189 4785\,x^2 \\
&- 1818577935208859435672601886243 4803054\,x \\
&- 2775652667915438819952161997138 01103343120 \\
g &= 346610035504925018514458 29\,x \\
&- 1291187456580021223163547791574 810881
\end{aligned}
$$

If f is skewed, one can add multiples of g without increasing much the norm of f:

$$f' = f + (ux^2 + vx + w)g$$

Rootsieve: find integers u, v, w such that f' has many roots modulo small primes, i.e., a good α value.

Polynomial Selection 1.0 for c59

$N = 7164152076175143545513361647566709043406333222824787179542 9$

Take $d = 4$, $a_4 = 17$.

$m = \lfloor (N/a_4)^{1/4} \rfloor = 254787999387056$

$$\begin{aligned} f &= 17x^4 + 14x^3 + 9216283939705 0x^2 \\ &+ 112990054008451x + 67766784591317 \\ g &= x - 254787999387056 \end{aligned}$$

lognorm 30.05, $\alpha(f) = 0.56$, combined score 30.61

Polynomial Selection 1.0 for RSA-1024

A. Lenstra, E. Tromer, A. Shamir, W. Kortsmit, B. Dodson, J. Hughes, P. Leyland, *Factoring Estimates for a 1024-Bit RSA Modulus*, Asiacrypt 2003.
Appendix A gives polynomials of degree 5 to 9.

$$
\begin{aligned}
f &= 2180047385355840\,x^6 \\
&- 3142872579455569636\,x^5 \\
&- 1254155662796860036208992514969847001569768\,x^4 \\
&- 1234618459668212931188535497431179367033899 9\,x^3 \\
&+ 3268536304983015875268773778111527849449995205 22\,x^2 \\
&+ 4609395911122979440239635705733809071478223546768\,x \\
&- 1107469276875825996795501758167470636492551999659 0997 \\
g &= x - 629042860635589902725572332002739171597034508 8070
\end{aligned}
$$

lognorm 100.02, $\alpha = -5.56$, combined score 94.46

Polynomial Selection 2.0

Thorsten Kleinjung, *On polynomial selection for the general number field sieve*, Mathematics of Computation, 2006

Thorsten Kleinjung, *Polynomial selection*, CADO workshop on integer factorization, Nancy, France, 2008

Choose a degree d, a leading coefficient a_d, a parameter P

Find m, ℓ such that $d^d a_d^{d-1} N - m^d$ is divisible by ℓ^2

Polynomial Selection 2.0 for c59

$$\begin{aligned} f &= 3000x^4 + 1026311x^3 - 3186148008x^2 \\ &- 5531789032354x + 412291028949240 \\ g &= 211092420527x - 13846210151047 \end{aligned}$$

lognorm 20.93, $\alpha = -3.37$, combined score 17.56

Polynomial Selection 2.0 for RSA-1024

T. Kleinjung, *Cofactorisation strategies for the number field sieve and an estimate for the sieving step for factoring 1024 bit integers*, SHARCS 2005.

$$
\begin{aligned}
f &= 1000000001002023904806000\,x^6 \\
&+ 269697895236768163056606416340\,x^5 \\
&- 6212838818608524196100227896844747498\,x^4 \\
&- 847105251394275537650757048185246266813 6\,x^3 \\
&+ 738608916851310255504408252889378679701231117 95\,x^2 \\
&+ 10323950425845926908896158377241426163762406505 3206\,x \\
&- 1139431985616391987769376205036438729670911719012775559 12 \\
g &= 514662055961724717752552412597334861\,x \\
&- 22651198301463826278447637231994318097020553454 5
\end{aligned}
$$

lognorm 93.15, $\alpha = -8.20$, combined score 84.95

Polynomial Selection 3.0

Shi Bai, Cyril Bouvier, Alexander Kruppa, P. Z., *Better Polynomials for GNFS*, Mathematics of Computation, 2016.

1. generate a pair (f, g) such that $\mathrm{Res}(f, g) = N$ with whatever method you like (Murphy, Kleinjung, ...)
2. if $f = f_d x^d + \cdots + f_0$ and $g = g_1 x + g_0$, reduce by LLL the following lattice (here for degree $d = 6$):

$$L = \begin{pmatrix} s^6 f_6 & 0 & 0 & 0 & 0 & 0 \\ s^5 f_5 & s^5 g_1 & 0 & 0 & 0 & 0 \\ s^4 f_4 & s^4 g_0 & s^4 g_1 & 0 & 0 & 0 \\ s^3 f_3 & 0 & s^3 g_0 & s^3 g_1 & 0 & 0 \\ s^2 f_2 & 0 & 0 & s^2 g_0 & s^2 g_1 & 0 \\ s f_1 & 0 & 0 & 0 & s g_0 & s g_1 \\ f_0 & 0 & 0 & 0 & 0 & g_0 \end{pmatrix}$$

The integer s represents the wanted polynomial "skewness"

A short vector obtained has the form

$$\begin{pmatrix} s^6(\lambda f_6) \\ s^5(\lambda f_5 + \mu g_1) \\ s^4(\lambda f_4 + \mu g_0 + \nu g_1) \\ s^3(\lambda f_3 + \nu g_0 + \delta g_1) \\ s^2(\lambda f_2 + \delta g_0 + \eta g_1) \\ s(\lambda f_1 + \eta g_0 + \rho g_1) \\ \lambda f_0 + \rho g_0 \end{pmatrix}$$

and corresponds to

$$f' = \lambda f + (\mu x^4 + \nu x^3 + \delta x^2 + \eta x + \rho)g$$

which also shares with g the root m modulo N

Analysis

By LLL's theory, short vectors have norm about $\det(L^t L)^{1/12}$.

Assuming $g_1 \ll g_0$, and $f_5 \ll s f_6$, $\det(L^t L) \approx f_6^2 g_0^{10} s^{32}$.

Thus short vectors have norm about $f_6^{1/6} g_0^{5/6} s^{8/3}$, and the middle coefficient of f' is about $f_6^{1/6} g_0^{5/6} s^{-1/3}$.

We want to avoid that g is a short vector, thus we want:

$$f_6^{1/6} g_0^{5/6} s^{8/3} \ll g_0$$

$$s^{8/3} \ll \left(\frac{g_0}{f_6}\right)^{1/6}$$

$$s \ll \left(\frac{g_0}{f_6}\right)^{1/16}$$

This gives the middle coefficient of f' about $f_6^{3/16} g_0^{13/16}$.

Now since $N = \mathrm{Res}(f, g) = f_6 g_0^6 + \cdots$, we have $g_0 \approx (N/f_6)^{1/6}$.

This gives the middle coefficient of f' about $f_6^{5/96} N^{13/96}$.

RSA-1024: $N^{13/96} \approx 5.3 \cdot 10^{41}$, lognorm ≈ 96.08

- Kleinjung's method controls the coefficients a_{d-2}
- our method controls all coefficients globally

Polynomial Selection 3.0 for c59

$$\begin{aligned} f_0 &= 540x^4 + 333x^3 - 23604460805x^2 \\ &+ 54548211699456x - 9766543585685 \\ g_0 &= 47249588263x - 106207542799191 \end{aligned}$$

$$\mathrm{Res}(f_0, g_0) = c59$$

$$\begin{aligned} f &= 1080x^4 + 540666x^3 + 142166403x^2 \\ &- 1839184384064x - 139953175243282 \\ g &= 47249588263x - 105000795043838 \end{aligned}$$

$$\mathrm{Res}(f, g) = 2 \cdot c59$$

lognorm 19.90, $\alpha = -1.77$, combined score 18.13

Polynomial Selection 3.0 for RSA-1024

After LLL reduction of polynomial found by Lenstra, Tromer, Shamir, Kortsmit, Dodson, Hughes and Leyland:

$$
\begin{aligned}
f &= 1173597989242921482240\,x^6 \\
&- 436081570202935700372728737578556 16\,x^5 \\
&+ 691958140341173987625035104743657545537\,x^4 \\
&+ 450511202161208734370957748132330118597 3519\,x^3 \\
&- 1730445251943964340375558511050751276493525 7500\,x^2 \\
&- 283131007738513042381019627129255517197411658 67633\,x \\
&+ 249963295649448077896029171367943737823089597 99485325 \\
g &= x \\
&- 629042860635589902725572332002739172216328869 9413
\end{aligned}
$$

$$\mathrm{Res}(f, g) = 538336 \cdot \text{RSA-1024}$$

lognorm 94.91, against 100.02 for the original polynomial

Polynomial pair found by Qingshu Meng using CADO-NFS revision 08f093c, with $P = 10^7$.

$$
\begin{aligned}
f &= 151448400\,x^6 \\
&- 203489670366601592122200\,x^5 \\
&- 960953651899092573770502050054669\,x^4 \\
&- 1407532974044695171973176982146771751477\,x^3 \\
&+ 3559156543814006546215736268738742949699710911724\,x^2 \\
&+ 29805585971485768554689000707583997601628133810281178012\,x \\
&- 211816194480849710702918231963075997264000726292020901360889688 0 \\
g &= 132450931230608801237\,x \\
&- 211371832060288225537316421302771767619243394668743
\end{aligned}
$$

$$\mathrm{Res}(f, g) = 100 \cdot \text{RSA-1024}$$

lognorm 90.82, $\alpha = -10.25$, combined score 80.57

An unpublished method

Input: an integer N to factor, a degree d, a leading coefficient f_d

1. choose a prime ℓ
2. for each root r of $f_d x^d = N \bmod \ell$
3. find m near $(N/f_d)^{1/d}$ such that $m = r \bmod \ell$
4. decompose N in base (ℓ, m)

$$N = f_d m^d + f_{d-1}\ell m^{d-1} + \cdots + f_1 \ell^{d-1} m + f_0 \ell^d$$

5. let $f = f_d x^d + \cdots + f_0$ and $g = \ell x - m$
6. reduce f, g by LLL

For RSA-1024 and $d = 6$, $f_6 = 3603600$ and $s = 5 \cdot 10^5$ are near from optimal.

LLL-reduced polynomials:

$$
\begin{aligned}
f &= 3744091168611775581600\,x^6 \\
&+ 3028827051075286908767242070 6\,x^5 \\
&- 251169282909731209144413611236809316 1\,x^4 \\
&- 185569605416477685517445559955827004956807 0\,x^3 \\
&+ 388206106439624853674170722836142293063019070452 00\,x^2 \\
&+ 755974433098785157927161619895559662452476923417 3008994\,x \\
&- 357716546146353270863663072098241863771250156240754792406487 47 \\
g &= 33083\,x \\
&- 182937102819972834389259308677783365493601039944917
\end{aligned}
$$

$$\mathrm{Res}(f, g) = 103898633827610 6 \cdot \text{RSA-1024}$$

Optimized polynomials:

$$
\begin{aligned}
f &= 374409116861177558160 0\, x^6 \\
&+ 474028631870471447780706911 06\, x^5 \\
&- 236372054780597398453512980239 0942191\, x^4 \\
&- 930090957816162504504234847555033442032 3766\, x^3 \\
&+ 259852522086903461039183054245735579564121 40395221\, x^2 \\
&+ 145779081089121830955601564853582667013031741639 55491788\, x \\
&- 128779423339812445819681686895154961834721676202854696437 41438 \\
g &= 33083\, x \\
&- 182937102819972834389259308677783365493575835694450
\end{aligned}
$$

lognorm 95.84, $\alpha = -8.72$, combined score 87.12

Questions

Is it better to use Kleinjung's method before our LLL-based method?

If so, what is the best P value to use in Kleinjung's method?

Can we find a better lattice that would lead to smaller norms?