On the Complexity of Integer Factorization
N. A. Carella, February 2009.

Abstract: This note presents a deterministic integer factorization algorithm based on a system of polynomials equations. This technique exploits a new idea in the construction of irreducible polynomials with parametized roots, and recent advances in polynomial lattices reduction methods. The main result establishes a new deterministic time complexity bench mark.

1 Introduction

This note presents a deterministic integer factorization algorithm based on a system of polynomials equations. This technique combines a new irreducible polynomials construction technique and recent advances in lattice reduction methods to obtain a new result. The main result establishes a new deterministic time complexity bench mark. Background materials in the theory of integer factorization are given in [CE], [CP], [LA], [MZ], [RL], [S], [W], and similar sources.

The second section recalls the known results on the time complexity of integer factorization. It continues with the main contributions, Lemma 3 and Theorem 5, and concludes with an algorithm and a detailed analysis of the coefficients of the polynomial in Section 3.

2 Main Contributions

This work builds on the earlier successful applications of the theory of polynomials equations and lattice reduction methods to integer factorization.

Previous Results

The previous works claim the followings.

***Theorem* 1.** ([CR]) If the $\log_2(N)/4$ least significant bits of a factor p of N are known, then the factorization of the integer $N = pq$, $p < q < 2p$, has deterministic logarithmic time complexity $O((\log N)^c)$, $c > 0$ constant.

In the Summer of 2007 this result was improved to the following.

***Theorem* 2.** ([C]) Let $N = pq$, $p < q < 2p$. If the $(1/6)\log_2(N)$ most significant bits of a factor p are known, then the factorization of N has deterministic logarithmic time complexity $O((\log N)^c)$, $c > 0$ constant.

Note. The standard term *polynomial time* has been replaced with the more descriptive term *logarithmic time*. This is patterned after the closely related term *exponential time*.

Construction of Irreducible Polynomials

The height of a polynomial $f(x, y, z) = \sum_{0 \le i,j,l \le d} a_{i,j,l} x^i y^j z^l \in \mathbb{Z}[x,y,z]$ of maximum degree $\deg(f) = d$ in the variables x, y and z is given by the expression $\| f(x,y,z) \|_\infty = \max\{ \mid a_{i,j,k} \mid : 0 \le i, j, k \le d \}$.

***Lemma* 3.** Let $\alpha, \beta > 0$ be a pair of parameters, and let $N = pq$ be a composite integer such that $p = O(N^\alpha)$ and $q = O(N^{1-\alpha})$. Then there exists an irreducible polynomial $f(x, y, z) = c_4xy + c_3x + c_2y + c_1z + c_0 \in \mathbb{Z}[x, y, z]$ with the following properties.

(i) The polynomial $f(x, y, z)$ has a small integer root (x_0, y_0, z_0) where $0 \le \mid x_0 \mid \le X \le N^\alpha$, $0 \le \mid y_0 \mid \le Y \le N^{1-\alpha}$, and $0 \le \mid z_0 \mid \le Z \le O((\log N)^B)$, $B > 0$ constant.
(ii) The factors of N can be written as $p = mx_0 + c$ and $q = ny_0 + d$, where the moduli n and m (possibly relatively prime) are of size $O((\log N)^A)$, $A > 0$ constant.
(iii) The height of $f(xX, yY, zZ)$ satisfies the inequality $\| f(xX, yY, zZ) \|_\infty \ge N^{1+\beta}$.
(iv) The polynomial can be generated in deterministic logarithmic time $O((\log N)^c)$, $c > 0$ constant.

Proof: Let n and m be (possibly relatively prime) moduli of sizes $O((\log N)^A)$, and let $k = O(N^\beta)$ be an integer, with $A > 0$ and $\beta > 0$ constants. Next rewrite the equation $f(x, y) = xy - N$ as an equation of three variables

$$f(x, y, z) = (mx + c)(ny + d)(kz + e) - rN = c_4xy + c_3x + c_2y + c_1z + c_0, \qquad (1)$$

where $1 \le c, d < O((\log N)^A)$, and $r = kz_0 + e$ is prime (or nearly prime) with $\mid z_0 \mid \le Z \le O((\log N)^B)$. The coefficients c_i are obtained after a selective replacement of the known variable $z = z_0$, see Section 3 for more details. Clearly this is an irreducible polynomial over the integers and of height $\| f(xX, yY, zZ) \|_\infty \ge rN + O(N) = O(N^{1+\beta})$, and has a small integral root $0 \le \mid x_0 \mid < N^\alpha$, $0 \le \mid y_0 \mid < N^{1-\alpha}$, $0 \le \mid z_0 \mid < O((\log N)^B)$. ■

Apparently, it is a very difficult proof for those that have not seem it before. Nevertheless, it is an elementary transformation/deformation of the most important polynomial $f(x, y) = xy - N$ in integer factorization.

The basic algorithm of Lemma 3 is sketched below, it is designed to work in tandem with Algorithm II, also note that the data m, n, c, d can be either inputted or internally generated.

Algorithm I
Input: $\alpha, \beta > 0$, and $N = pq$ such that $p = O(N^\alpha)$.
Output: $f(x, y, z) = c_4xy + c_3x + c_2y + c_1z + c_0$ and m, n, c, d.
1. Set $T = (\log N)^A$, $A > 1$, and generate a pair of primes or nearly primes moduli n and $m < T$.
2. Generate a prime $r = kz_0 + e$ (or nearly prime) with $k = O(N^\beta)$ and $\mid z_0 \mid \le Z \le O((\log N)^B)$.
3. Select a pair $c < m$, and $d < n$.
4. Compute the coefficients of the irreducible polynomial $f_{c,d}(x, y, z) = c_4xy + c_3x + c_2y + c_1z + c_0$.
5. Return $f(x, y, z) = f_{c,d}(x, y, z)$.

Algebraically Independent Polynomials

Although the technique of Lemma 3 can generate one or more irreducible polynomials, these polynomials are not algebraically independent. Accordingly, lattice reduction method is utilized to generate another algebraically independent polynomial. Further, since the third variable is known or its value is very small and can be determined by brute force search, just one additional algebraically independent polynomial is required.

***Theorem* 4.** ([ER]) Let $f(x, y, z) = c_4xy + c_3x + c_2 y + c_1z + c_0 \in \mathbb{Z}[x, y, z]$ be an irreducible polynomial of height $\| f(xX,yY,zZ) \|_\infty = W$ and with a small integer root (x_0, y_0, z_0) such that $| x_0 | < X$, $| y_0 | < Y$ and $| z_0 | < Z$. Suppose that the inequality

$$X^{3+3\tau}Y^{3+6\tau+3\tau^2}Z^{2+3\tau} < W^{2+3\tau-\varepsilon}, \tag{2}$$

where $\tau > 0$ is a lattice parameter, holds. Then there exists a pair of linearly independent polynomials $f_1(x, y, z)$ and $f_2(x, y, z)$ not multiple of $f(x, y, z)$, with a common root. Furthermore, the polynomials are generated in deterministic logarithm time.

The complete analysis of this and other special cases of polynomials in three and four variables and the corresponding polynomials bases of the polynomials lattices are given in [ER], and [JM].

The two polynomials $f_1(x, y, z)$ and $f_2(x, y, z)$ are linearly independent, but not guaranteed to be algebraically independent. However, the two pairs of polynomials $f(x, y, z), f_1(x, y, z)$ and $f(x, y, z), f_2(x, y, z)$ are guaranteed to be algebraically independent. Recent advances in the construction of three algebraically independent polynomials are discussed in [BA].

The Main Result

In the last decades the techniques of the theory of polynomials equations and lattice reduction methods have emerged as powerful tools in the theory of integer factorization.

***Theorem* 5.** The factorization of a composite integer $N \in \mathbb{N}$ has deterministic logarithmic time complexity $O((\log N)^c)$, $c > 0$ constant.

Proof: Without loss in generality, let $N = pq$ be a balanced integer, $p < q < 2p$. Put $\alpha = 1/2$, and let $\beta = 1/2 + \gamma$ for some $\gamma > 0$, and fix a pair of moduli n, $m = O((\log N)^A)$, where $A > 0$ is a constant. Then it is clear that the integer N has its factors in some residue classes

$$p = mx + c \quad \text{and} \quad q = ny + d, \tag{3}$$

where $0 \le | c |, | d | < O((\log N)^A)$. At most $O((\log N)^{2A})$ pairs (c, d) has to be tested to determine the correct residues classes (3) of the factors. Given the correct pair (c, d), there exists an irreducible polynomial

$$f(x, y, z) = c_4xy + c_3x + c_2y + c_1z + c_0 \tag{4}$$

over the integers $\mathbb{Z}$, which has a small integer solution (x_0, y_0, z_0) such that

$$p = mx_0 + c \quad \text{and} \quad q = ny_0 + d, \tag{5}$$

where $0 \le | x_0 |, | y_0 | < N^{1/2} = N^\alpha$ and $0 \le | z_0 | \le O((\log N)^B)$, $B > 0$ constant, see Lemma 3. Moreover, the height satisfies the inequality $W = \| f(xX, yY, zZ)) \|_\infty \ge N^{1+\beta} \ge N^{3/2+\gamma}$. Now by Theorem 4, there exists another algebraically independent polynomial $g(x, y, z)$ that shares the same root (x_0, y_0, z_0) and it can be determined using lattice reduction techniques whenever the inequality

$$X^{3+3\tau}Y^{3+6\tau+3\tau^2}Z^{2+3\tau} < W^{2+3\tau-\varepsilon}, \tag{6}$$

holds. Replacing $X < N^{1/2}$, $Y < N^{1/2}$, $Z < O(N^\delta)$ and $N^{3/2+\gamma} \le W$ in (6) returns

$$X^{3+3\tau}Y^{3+6\tau+3\tau^2}Z^{2+3\tau} < N^{3+9\tau/2+3\tau^2/2+(2+3\tau)\delta} \le N^{(3/2+\gamma)(2+3\tau-\varepsilon)} \le W^{2+3\tau-\varepsilon}, \tag{7}$$

where $\delta > 0$ is an arbitrarily small number, and all the relevant constants has been omitted. These data in turn imply that (6) and (7) holds if and only if

$$\frac{3\tau^2/2+(2+3\tau)\delta+3\varepsilon/2}{2+3\tau-\varepsilon} < \gamma \tag{8}$$

holds. Further, since there is almost no restriction on the parameter $\gamma > 0$, for example, $\gamma = 1/4$ or $1/3$ or $1/2$, etc is feasible, the previous inequalities (6), (7) and (8) hold for any appropriate choice of γ.

Ergo the solution $(x, y, z) = (x_0, y_0, z_0)$ of the system of equations

$$f(x, y, z) = 0,\ g(x, y, z) = 0,$$

can be recovered by means of resultants or Grobner bases calculations. Specifically, computing the roots of the polynomials

$$r_1(x) = \mathrm{Res}_y(f(x, y, z_0), g(x, y, z_0)) \quad \text{and} \quad r_2(y) = \mathrm{Res}_x(f(x, y, z_0), g(x, y, z_0)), \tag{9}$$

where z_0 is known. Next observe that and each of these algorithms above has logarithmic time complexity. In particular, the running time of the entire algorithm is dominated by at most $O((\log N)^{2A})$ lattice reduction steps, one for each pairs (c, d). Thus, the overall time complexity of the integer factorization algorithm is deterministic logarithmic time $O((\log N)^c)$, $c > 0$ constant. Quod erat demonstrandum ■

The choice of parameter $\alpha > 0$ assumes a priori knowledge on the sizes of the factors $p = N^{\alpha}$ and $q = N^{1-\alpha}$ of $N = pq$. Furthermore, since the subset of balanced integers $\mathcal{B} = \{ N = pq : p < q < ap$, with p, q primes $\}$ is the most important case in integer factorization, it was set to $\alpha = 1/2$. Balanced integers are the most difficult to factor. However, the probability of an arbitrary integer of being balanced is negligible. Indeed, the subset of balanced integers has zero density in the set of integers. More precisely, it has cardinality $\mathcal{B}(x) = \#\{ N = pq \le x : p < q < ap \} = c_0 x/\log(x)^2$, $c_0 = c_0(a)$ constant, see [DM].

Numerical experiments will have to be performed to determine the best choices of the parameters $\beta = 1/2 + \gamma$ and $\tau > 0$. The first controls the height of the polynomial $f(x, y, z)$ and the second is part of the lattice basis, see Theorem 4.

The last algorithm below encodes the basic procedure of Theorem 5. In step 2.1, the data m, n, c, d is passed on to Algorithm I.

Algorithm II
Input: $N = pq$, and $\alpha, \beta > 0$ such that $p = O(N^{\alpha})$.
Output: p, q.
1. Set $T = (\log N)^A$, $A > 0$, and select a pair of primes or nearly prime moduli n and $m < T$.
2. For $c, d < T$ do
2.1 Construct an irreducible polynomial $f_{c,d}(x, y, z) = c_4xy + c_3x + c_2y + c_1z + c_0$ such that $0 \le |x_0| \le N^{\alpha}$, $0 \le |y_0| \le N^{1-\alpha}$, and $0 \le |z_0| \le O((\log N)^B)$, $B > 0$ constant, see Lemma 3 and Algorithm I.
2.2 Construct an algebraically independent polynomial $g_{c,d}(x, y, z)$, see Theorem 4.
2.3 Compute the root (x_0, y_0, z_0) of the system of equations $f(x, y, z) = 0$, $g(x, y, z) = 0$, using resultants or Groebner bases methods, here z_0 is known.

2.4 Compute the potential factors $p_{c,d} = mx_0 + c$ and $q_{c,d} = ny_0 + d$.
2.5. If $1 < \gcd(p_{c,d}, N) < N$ or $1 < \gcd(q_{c,d}, N) < N$, then halt.
3. Return $p = p_{c,d}$, and $q = N/p_{c,d}$ or $q = q_{c,d}$, and $p = N/q_{c,d}$.

Ultimately an algorithm that accepts an arbitrary integer and internally generates all its parameters seems to be feasible in the near future.

3 The Polynomial and its Coefficients

Let n and m be (possibly relatively prime) moduli of sizes $O((\log N)^A)$, $A > 0$ constant. Let $r = kz_0 + e \geq N^{\beta} > N^{1/2}$ be a prime (or nearly prime), where $k > N^{1/2}$ is a prime, $| z_0 | \leq O((\log N)^B)$, and $\beta > 1/2$ is a constant. Multiplying the three linear factors returns the polynomial equation

$$\begin{aligned}(mx + c)(ny + d)(kz + e) - rN &= kmnxyz + emnxy + dkmxz + demx + cknyz + ceny + dcdkz + cde - rN \\ &= (kz_0 + e)mnxy + (kz_0 + e)dmx + (kz_0 + e)kny + cdkz + cde + - rN \\ &= c_4xy + c_3x + c_2y + c_1z + c_0, \\ &= f(x, y, z),\end{aligned} \tag{10}$$

where a selective replacement of the known variable $z = z_0$ was used. The coefficients are

$$c_0 = cde - rN, \qquad c_1 = cdk, \qquad c_2 = knr, \qquad c_3 = dmr, \qquad c_4 = mnr,$$

where $0 \leq c < m$ and $0 \leq d < n$, and $0 < e < k < r$. By construction, the largest possible prime factors of the coefficients $c_0, \ldots, c_4$, are k and r. But since

$$c_0 \equiv cde - rN \bmod k, \qquad c_1 \equiv cdk \equiv 0 \bmod k, \qquad c_2 \equiv knr \equiv 0 \bmod k,$$
$$c_3 \equiv dmr \not\equiv 0 \bmod k, \qquad c_4 \equiv mnr \not\equiv 0 \bmod k,$$

and

$$c_0 \equiv cde \not\equiv 0 \bmod r, \qquad c_1 \equiv cdk \not\equiv 0 \bmod r, \qquad c_2 \equiv knr \equiv 0 \bmod r,$$
$$c_3 \equiv dmr \equiv 0 \bmod r, \qquad c_4 \equiv mnr \equiv 0 \bmod r,$$

the coefficients are relative prime or nearly relative prime. More precisely, $\gcd(c_0, \ldots, c_4) \leq mn = O((\log N)^{A+B})$ holds. This ensures that the height of the polynomial satisfies

$$\| f(xX, yY, zZ) \|_{\infty} \geq | c_0/mn | \geq (rN + O(N^{\beta+\varepsilon}))/mn = O(N^{1+\beta-\varepsilon}),$$

where $\varepsilon > 0$ is an arbitrarily small number.

References

[LA] A Lenstra, Integers Factoring, Designs, Codes, and Cryptography, 19, 101-128 (2000).

[S] Shoup, Victor. A computational introduction to number theory and algebra. Cambridge University Press, Cambridge, 2005.

[RL] Riesel, Hans. Prime numbers and computer methods for factorization. Second edition. Progress in Mathematics, 126. Birkhäuser Boston, Inc., Boston, MA, 1994.

[MZ] Menezes, Alfred J.; van Oorschot, Paul C.; Vanstone, Scott A. Handbook of applied cryptography. CRC Press, Boca Raton, FL, 1997.

[W] Williams, Hugh C. Édouard Lucas and primality testing. Canadian Mathematical Society Series of Monographs and Advanced Texts, 22. A Wiley-Interscience Publication. New York, 1998.

[CP] C Pomerance RE Crandall, Primes Numbers: A Computational Perspective, Springer-Verlag, 2006.

[CE] Cohen, Henri. A course in computational algebraic number theory. Graduate Texts in Mathematics, 138. Springer-Verlag, Berlin, 1993.

[DM] Andreas Decker, Pieter Moree, Counting RSA-integers, arXiv:0801.1451

[C] N. A. Carella, Note on Integer Factoring Methods III, arXiv:0707.4468

[CR] Coppersmith, Don Finding small solutions to small degree polynomials. Cryptography and lattices (Providence, RI, 2001), 20-31, Lecture Notes in Comput. Sci., 2146, Springer, Berlin, 2001.

[CR] Coppersmith, Don Small solutions to polynomial equations, and low exponent RSA vulnerabilities. J. Cryptology 10 (1997), no. 4, 233-260.

[ER] M. Ernst, E. Jochemsz, A. May and B. de Weger. Partial key exposure attacks on RSA up to full size exponents. Lecture Notes in Computer Science 3494, 371–387, EUROCRYPT 2005.

[BA] Aurelie Bauer, Antoine Joux, Toward a Rigorous Variation of Coppersmith's
Algorithm on Three Variables, EUROCRYPT 2007, LNCS 4515, pp. 361–378, 2007.

[JM] Ellen Jochemsz, Alexander May: A Polynomial Time Attack on RSA with Private CRT-Exponents Smaller Than N 0.073. CRYPTO 2007: 395-411.