Research Journal of Applied Sciences 10 (8): 365-370, 2015
ISSN: 1815-932X

Lenstra Factorization Method Convergence Investigation on Elliptic Curves

Ismail Amer, Shamil T. Ishmukhametov and Ramilya G. Rubtsova
Kazan Federal University, Kremlevskaya Street 18, Kazan, Russia

Abstract: It is well known that the process of natural numbers decomposition in a product of primefactors (factorization) is a time-consuming computational procedure. This property is widely used in cryptography. In particular, the known RSA encryption method uses a composite number n of 1024 bits or more which is the product of two prime numbers as the secret key. One of the most effective methods of integer factorization is H. Lenstry Method based on the arithmetic of elliptic curves. This method has the following feature: its capacity does not depend on the size of the original number n but on the size of the smallest divisor n. Therefore, the Lenstra allows to factoring the numbers that are inaccessible for other methods. The peculiarity of Lenstra Method is its heavy dependence on the choice of an elliptic curve. More precisely, the algorithm selects an arbitrary curve over a prime field of the characteristic p where p is the unknown divisor n. Let t is the number of points on a selected curve. The rate of algorithm convergence depends on the greatest prime number dividing the number t. For example if $t = p_1^{t_1} \times p_2^{t_2} \times \ldots \times p_k^{t_k}$, the decomposition of t in the product of prime factors, then the method complexity depends on $B = \max_i \; p_i^{t_i}$. Due to the fact that the method success depends heavily on t value and its factors which which is the subject of luck. The worst case occurs if t is a prime number. To eliminate obviously bad cases, you must start the factorization procedure simultaneously on several different curves. Such parallelism allows you to find the curve on which the process will converge faster than the others. The problem is that the selection of too many curves will affect the overall performance of the method and an insufficient number of curves does not guarantee a result. In this study, we investigate the convergence of Lenstra factorization method, depending on the choice of an elliptic curve on which the factorization procedure is performed more effectively. We study the statistical distribution of "good" curves on which the factorization procedure is performed more efficiently.

Key words: Factorization of integers, H. Lenstra algorithm, elliptic curves, RSA encryption, smooth integers, distribution of smooths

INTRODUCTION

An integral number factorization is the procedure for this number expansion in the product of prime factors. Factorization is a complex computational task requiring significant computing resources. A known method of encryption and cryptography (RSA Method) is developed on the complexity of this task. To date, the length of composite numbers which can not be decomposed into a product of prime factors makes about 1024 bits or 300 decimal places. In 2009, the project of 768 bits number successful decomposition was completed which lasted about 10 years. Let's consider the basic methods of factorization used in the modern theory of numbers.

MATERIALS AND METHODS

The factorization problem has no polynomial solution algorithm, although, this theory was not proven. There are several subexponential factorization algorithms, including the fastest modern ones (in order of speed decrease) the Number Field Sieve (NFS), the Quadratic Sieve (QS) and Lenstra Factorization Method and the Elliptic Curves Method (ECM) (Ishmukhametov, 2014; Crandall, 2006). The complexity of the first two methods depends on the length of the factorisable number (we will denote it throughout the study by the letter n). ECM rate does not depend on the length of n but on the length of the smallest divisor of n. Therefore, although ECM is worse than the first two methods by speed, it is applicable for the decomposition of large dimension numbers with relatively small dividers for which the first two methods are useless. For example, a complete decomposition of the tenth Fermat number F10 with the length of 300 decimal digits and some dividers larger than Fermat numbers (Brent, 1999). The number F10 was decomposed into a product of 4 length factors: 8, 10, 40 and 252 of

Corresponding Author: Ismail Amer, Kazan Federal University, Kremlevskaya Street 18, Kazan, Russia

Table 1: The convergence of asymptotic estimate

NFS	QS	ECM
$\exp\left(C_1(\log n)\bar{3}(\log\log n)\bar{3}\right)$	$\exp\left(C_2(\log n)(\log\log n)^{1/2}\right)$	$\exp\left(C_3(\log p\log\ \log p)\right)\bar{2}$ p the least divider n

discharge. Let's put down asymptotic estimates of convergence in Table 1 of these methods (Crandall, 2006).

We see that the theoretical ECM assessment is even better than QS, however for RSA modules, when the smallest divisor is comparable to $\sqrt{n}$, the convergence QS is significantly higher than ECM.

It should be noted that the ECM convergence is strongly influenced by the choice of a curve on which the factorization is performed. Since, it is impossible to determine which curve will provide a greater convergence, it is necessary to conduct the statistical research of "good" and "bad" curves ratio, starting the procedure on several curves to provide a high probability of successful factorization.

Exponential methods: In order to estimate ECM successfully, it is also necessary to describe some of the slower but more simple methods of factorization.

Trial divisions TR Method: This method consists in the fact that a test division of the number n into prime numbers is performed $<\sqrt{n}$. If you appreciate the complexity of the division operation to 1, the complexity of the algorithm is estimated by the number of iterations, i.e., = $\sqrt{n}$. In other words:

$$C(TR) = 0(n^{1/2})$$

Pollard Rho Method: This method was developed by John Pollard in 1975 and is based on a statistical paradox of "birthdays". Its asymptotic complexity is estimated as:

$$C(\rho P) = 0(n^{1/4})$$

There are several methods that are similar to Rho-Pollard Method by difficulty and which have the same or a similar assessment of convergence. Pollard Method (p-1), Fermat Method, (p+1) method of Williams, the method of Shanks quadratic forms (Ishmukhametov, 2014) is among them. The effectiveness of these methods depends on the ratio of divider values or other special conditions (e.g., on the smoothness of p±1 numbers, where p is the divisor n).

ECM Lenstra algorithm: In this study, we will give a detailed description of Lenstra Method and perform the assessments of its complexity. Let's n is a factorisable number and p is its smallest divisor as before.

The elliptic curve in Weierstrass form over the field Fp is denotes the set of points that satisfy the equation:

$$y^2 = x^3 + Ax + B(\operatorname{mod} p) \tag{1}$$

together with a special point O called an infinitely remote point with uncertain coordinates.

The procedure of addition is determined on the points of an elliptic curve EC where Q = kP point may be calculated relatively easily for any P point and an integral number k (using a polynomial algorithm) but the solution of finding k factor according to given points P and Q is much more complicated. The latter problem is called the problem of discrete logarithm calculation on an elliptic curve and has no fast algorithms to solve it now a days. The complexity of the discrete logarithm taking pr on EC is widely used in cryptography and encryption algorithms and the development of digital signatures (Ishmukhametov and Rubtsova, 2014).

ECM algorhithm

Initialization stage: Let's choose the arbitrary positive integers A and B, the smaller n and consider a curve of Eq. 1, but choosing the n number as a module:

$$y^2 = x^3+Ax+B(\operatorname{mod} n) \tag{2}$$

The curve (2) is not an elliptic one in the usual sense as the main set Fn = {0,1,2, ... n-1} is not a field. However, the standard procedures of points addition and doubling may be performed on the set Fn. Moreover if P(x, y) is the point on the curve (2), P'(x mod p, y mod p) is the point on the conventional elliptic curve (Eq. 1). Therefore, working with a custom curve one may always keep in mind that the coordinates of all the points of the curve (1) may be found by the coordinates of the curve (1) if p module is known.

Let's choose any point on the curve P_0 (Eq. 2) and a number of B<p where p is the alleged divider. The details about the selection of B will be described. Let's turn from affine to projective coordinates of the curve EC (Ishmukhametov, 2014; Crandall, 2006):

$$P_0(x, y) \rightarrow P_0(x, y, 1)$$

The corresponding formulas for the sums of points are given. All points in the projective coordinates have three coordinates x, y, z.

The procedure of factorization consists of one or two performed stages

Stage 1: During an iterative procedure let's calculate a new point P_1 on the curve (2):

$$P_1 = B(B\text{-}1)\times\ldots\times 2P_0 = B!P_0$$

Let's the point P_1 has the following coordinates $P_1(x_1, y_1, z_1)$. Let's calculate the greatest common divisor:

$$d = GCD(n, z_1)$$

Let's check the condition 1<d<n. If it is performed, then the desired divisor of the number n is found. Otherwise, let's move on to the second stage.

Stage 2: Let's choose the number B_2, $B<B_2<p$. Let's $q_0<q_1<\ldots<q_k$ are prime numbers, located within the interval $[B; B_2]$. Let's calculate the points consistently $C_i = q_iP_1$, $C_i(u_i, v_i, w_i)$ and check the following term:

$$d = GCD(n, w_i) > 1 \qquad (3)$$

Once the condition (Eq. 3) is satisfied, we stop the procedure. The required divider is found. If all numbers $q_i<B_2$ are passed and the condition (Eq. 3) is not performed, then the procedure is failed.

Ecm algorithm analysis: Let's denote via t = #EC(p) the number of an elliptic curve points (Eq. 1). According to Hasse inequality, the following inequality is performed:

$$p+1-2\sqrt{p} < t < p+1+2\sqrt{p} \qquad (4)$$

and depending on the values of A and B coefficients t may take any value within the range (Eq. 4). Let's expand t in the product of prime factors:

$$t = p_0^{r_0}\times p_1^{r_1}\times\ldots\times p_k^{r_k}$$

Where:

$p_0 = 2$, $p_1 = 3$ = Successive primes, $r_i \geq 0$

p_k = The greatest divider t

Let's denote the greatest factor $p_i^{r_i}$ via M. For example if t is a prime number, the only non-zero value in this formula will be the last value $r_k = 1$ and M = t.

The initial point $P_0(x_0, y_0)$ of the curve (2) is represented by the point $P'_0(x_0 \bmod p, y_0 \bmod p)$, with the order t' which is the divider t of the point group order (Eq. 1), thus $tP'_0 = 0$, where O is infinitely remote point of the curve (2). In the projective coordinates of the point O the third coordinate is equal to 0. If you go back to the curve (2), the third coordinate z of the point tP is a multiple of p, i.e., z = hp for some integer h (due to the relation z^ '= z mod p = 0). Therefore:

$$GCD(n, z) = p \qquad (5)$$

Hence, the strategy of ECM Method first stage becomes clear: to provide the multiplier B! for the point P_0 large enough that the number of points t on the curve (1) turned to be the divisor B!, then:

$$tP'_o = 0 \rightarrow B!P'_0 = 0$$

and the third coordinate of the point B! P_0 will be multiple to p that will ensure the method success at the first stage.

If success is not achieved at the 1st stage of the algorithm, then the following conditions should be performed at the 2nd stage:

- $B \geq \max\{p_i^{r_i} \mid 0 \leq i < k\}$ where $p_i^{r_i}$ are t dividers, except the last one
- $r_k = 1$ is the greatest divisor t in the first degree
- $B_2 \geq p_k$

The first condition means that the order of the point P'_1, located at the first stage p_k. The terms 2 and 3 mean that p_k is within the range $[B; B_2]$. Then the calculations performed during the second stage of the method will inevitably lead to to the factor $q_i = p_k$ and the calculation will be completed successfully.

Definition: The number t is called B-smooth if every prime divisor t does not exceed the number of B. The number t is called gradually B-smooth if the degree of every prime divisor t, included in the expansion t does not exceed B.

Example: The number $t = 360 = 2^3\times 3^2\times 5$ is 5-smooth and 9-grade smooth. Now, we can formulate the general condition of ECM algorithm convergence.

ECM Method convergence term: The dimension t of an elliptic curve according to the module p is either B-gradually smooth number or is represented as the product of two factors $t = s\times p_k$. Where the number s is B-gradually smooth and p_k is a simple factor which does not exceed the limit B_2.

Smooth numbers are discussed in detail in a review study (Granville, 2004). Let's denote $\psi(x, y) \leq x$. At the values y, comparable by order with x the value of the function $\psi(x, y)$ may be calculated according to the following asymptotic formula:

Table 2: Dickmann de Bruin function value distribution for u≤10

$\rho(s)$	u
0.31	2
4.9×10^{-2}	3
4.9×10^{-3}	4
3.5×10^{-4}	5
2.0×10^{-5}	6
8.7×10^{-7}	7
3.2×10^{-8}	8
1.0×10^{-9}	9
2.8×10^{-11}	10

$$\psi(x, y) = x\times\rho(u)$$

where $u = \ln x/\ln y$, $\rho(u)$ is Dickmann de Bruin function (Granville, 2004). Let's demonstrate the table of this function values $u\leq10$ (Table 2).

RESULTS AND DISCUSSION

Analysis of ECM algorithm complexity: To estimate the total number of operations, let's note that one operation of addition or doubling of an elliptic curve point in projective coordinates consists of a constant number of operations (about 20) in the ring Zn. Thus, it is enough to count the number of addition and point doubling operations at the 1st and the 2nd stage of ECM algorithm.

To assess the complexity of the algorithm 1st stage, we use Stirling's formula to estimate the value B!:

$$B!\sim\sqrt{2\pi B}\left(\frac{B}{e}\right)^{B}$$

Where:

$$\log_2 B!\sim B\log_2 B$$

Hence, the number of operations with EC points is equal to the value $2B\log_2 B$. At the second stage of the algorithm, the following operations are performed sequentially $C_i = q_iP_1$. Before the proceeding with this step, it is necessary to make a table of point values in order to speed up these operations $2P_1, 4P_1,...,2gP_1$.

For a small value g and then the calculation of the next point $C_i = q_i P_1$ may be performed according to the following Eq .6:

$$C_i = q_{i-1}P_1 + (q_i - q_{i-1})P_1 = C_{i-1} + 2jP_1 \qquad (6)$$

using only one addition operation as the value $2j P_1$ will be taken from the table. Thus, the number of operations at the 2nd stage is the sum of prime numbers within the interval $[B; B_2]$ and some constant. This number is asymptotically $= B_2/\ln B_2$.

Calculation of operations number due to the curve size smoothness: Let the number of points on the considered curve (1) is still equal to t. Let's consider the different options for t number smoothness.

Case A; t is a prime number: In this case, the first step of the method may be omitted and only the second stage is performed. The number of operations makes the value of $O(t/\ln t)$ and in the worst case, when the divisor is approximately $= \sqrt{n}$, it is estimated by the following value $O(\sqrt{n}/\ln n)$.

In this case, ECM Method is comparable to the trial division method and is worse than most other exponential methods listed in the study.

The probability of such an event performance is equal to the frequency of prime numbers occurrence and makes, $p(A)\sim1/\ln n^{1/2} = 2/\ln n$. In fact, this probability will be slightly lower as the given number is equal to the ratio of prime numbers in the range from 1 to t to the interval length and prime numbers appear more frequently at the beginning of the interval and rarely at the end of it.

Case B; t is smooth $\sqrt{t}$ number: In this case, all the divisors t are $<\sqrt{t}$. If you take the constant $B = n^{1/4}$, then the procedure will end at the first stage and will be assessed by the following value $O(n^{1/4}\times\ln n)$.

In this case, ECM Method does not win by comparison with the most exponential methods listed in the study. Even if we accelerate the algorithm performance by the second stage, the total score will be no better than $O(n^{1/4})$.

Let's note that the probability of event B performance is equal to the allocation frequency $\sqrt{t}$ of smooth numbers within the interval close to t. This probability is estimated as the following value:

$$\psi\left(t,\sqrt{t}\right)/t\sim\rho(2) = 0{,}3069$$

Case C; t is $t^{1/s}$ a smooth number for some natural number s: This case summarizes the case B. It is easy to understand that the constant B in this case may be less or equal to $t^{1/s}$ and the total number of transactions is estimated by the following value $O(n^{1/2s})$.

This estimate is an exponential one and exceeds the asymptotic estimate for the majority of section 3 algorithms at $s\geq3$. The frequency of C cases occurrence depends on the number $t^{1/s}$ the smooth numbers and are estimated by the following value:

$$\frac{\psi\left(t,t^{\frac{1}{s}}\right)}{t}\sim\rho(s)\sim\frac{1}{s^s}$$

The analysis of the function values ρ(u) indicates that the achievement of method performance high values with the increasing s is available for a small percentage of curves of a determined interval. For example if $s = 4$ the difficulty of factorization is estimated by a sufficiently effective assessment of $O(n^{1/8})$. At that the share of curves on which this estimate is achieved is equal approximately to 4.9×10^{-3}, i.e. such a curve occurs once for 200 cases. If two hundred curves are started simultaneously, it will increase the overall assessment of the complexity in 200 times. The question of such a decision effectiveness is determined by the divisor p of the number n. The greater the value p, the better to take a greater number of curves.

In general case, at a fixed value of u the frequency of a desired curve appearance is equal to ρ(u), so the average number of curves required for the emergence of such a curve is equal to 1/ρ(u) and the overall assessment of performance makes:

$$f(u) = \frac{p^{1/u}}{\rho(u)} \tag{7}$$

The minimum of f(s) function depends on the dimension p and determines the number of curves to be selected, so that the evaluation is optimal. In the next study, we present some numerical calculations for fixed lengths of the argument p.

It should be noted that Eq. 7 is valid only for small values of u. For large u values one should use the estimate of the number ψ(x, y) given by Lenstra (1987) in his main study and based on the assessment of the number $x^{1/u}$ given in an earlier study by Canfield *et al.* (1983). Let's consider the following function:

$$L(x) = e^{\sqrt{\ln x \ln\ln x}}$$

Theorem (Canfield, Erdos, Pomerance Theorem consequence 3.1): Let $\alpha>0$ is a real number. The probability of the fact that a random integer $t\leq x$ has all simple dividers $\leq(L(x))^{\alpha} = L(x)^{-1/2\alpha+0(1)}$ at $x\rightarrow\infty$.

The necessary probability $1\text{-}e^{-g}$, g is a small positive number, the event $p\leq(L(x))^{\alpha}$ is achieved, when the number of curves is equal to $gL(x)^{1/2\alpha}$. The total time is estimated by the following value:

$$L(x)^{\alpha}\times gL(x)^{1/2\alpha} = gL(x)^{\alpha+1/2\alpha}$$

This value achieves its maximum if it is $(\alpha+1/2\alpha) = 0$, where $\alpha = 1/\sqrt{2}$. At that the value $gL(x)^{\alpha+1/2\alpha} = gL(x)^{\sqrt{2}}$.

Statement 1: ECM algorithm finds its smallest divisor p of the number n with the probability $1\text{-}e^{-g}$ during the period:

$$gL(x)^{\sqrt{2}}\times M(n) \tag{8}$$

where M(n) is the time of one addition operation performance for an elliptic curve points of n dimension (Lenstra, 1987).

Lenstra called the statement 1 a hypothesis, since it is based on the assumption that the smooth numbers are distributed within the following interval:

$$\left[p+1-2\sqrt{p};p+1+2\sqrt{p}\right]$$

with the same frequency as the entire interval [1; p] which is generally not true (smooth numbers are closer to the beginning of the interval [1; p] they are obviously located more frequently).

Calculation of curves effective number at fixed p: Let $p = 10^{50}$ which corresponds to the length of the argument RSA $n = 10^{100}$ approximately, 330 bits. According to the Eq. 8, we calculate the optimal value of ECM algorithm running time:

$$gL(x)^{\sqrt{2}}\times M(n) = ge^{\sqrt{2\ln10^{50}\ln\ln10^{50}}}\times M(n)\approx gM(n)\times10^{14}$$

At that the number of curves should be equal to:

$$gL(x)^{1/2\alpha} = gL(x)^{1/2\sqrt{2}}\approx10^{3.5}\times g = 3162\,g$$

With the argument increase the number of curves required for maximum performance is also increased, so in the boundary cases of ECM application it is difficult to achieve such a number of parallel calculations.

CONCLUSION

Lenstra algorithm analysis showed above, leaves a number of open issues related to the application of the algorithm.

The first one relates to the issue of uneven distribution of smooth numbers at the beginning and the end of the considered intervals and clarifying the formula. The second question relates to the selection of the optimum number of curves: which values of the p divisor may help to achieve a realistically optimal value at which the most efficient implementation of the algorithm is achieved?

The third issue concerns the experimental calculations of the method performance at different p divisors.

Another problem is related to the choice of the boundary B at the first stage of Lenstra algorithm. How an option B is associated with the probability of successful completion condition performance concerning the second stage of the algorithm?

ACKNOWLEDGEMENT

The research is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

REFERENCES

Brent, R.P., 1999. Factorization of the Tenth Fermat Number. Mathematics of Computation, 68 : 429-451.

Canfield, E.R., P. Erdos and C. Pomerance, 1983. On a Problem of Oppenheim Concerning "factorisatio numerorum". J. Numb. Theory, 17 (1): 1-28.

Crandall, R., 2006. Prime Numbers. A Computational Perspective [Text]. R. Crandall and C. Pomerance (Eds.). Springer, Berlin, pp: 597.

Granville, A., 2004. Smooth Numbers: Computational Number Theory and Beyond, Proc. of MSRI Workshop, pp: 268-363.

Ishmukhametov, Sh.T., 2014. Methods of integers factorization [Text]. Sh.T. Ishmuhametov (Eds.). LAP Lambert Academic Pub., pp: 256. ISBN: 978-3-659-17639-5.

Ishmukhametov, Sh.T. and R.G. Rubtsova, 2014. Mathematical Methods of Information Security: Manual, KFU.

Lenstra, Jr., 1987. H.W. Factoring Integers with Elliptic Curves. Annals of Mathematics, 126 (3): 649-673.