UDC 681.3.06.

ANALYSIS OF THE FACTORIZATION METHOD BASED ON ELLIPTIC CURVES THEORY

I. Dermenzhy, G. Vostrov
Odessa national polytechnic university

Abstract. *The purpose of this paper is the analysis of elliptic curve method, in particularly the probabilistic approach of method research, and also the research of subexponential complexity definition. This paper describes the possibilities of the elliptic curves theory apparatus use. A comparative analysis of the existing factorization methods is given. The choice of the method based on the elliptic curves theory is substantiated. Special attention is paid to the method's probabilistic and subexponential character. The method's main directions of development and optimization are described. Possible approaches for described problems solving are given and analyzed. The possibility of creating a probabilistic model which applies the elliptic curves theory in factorization is considered. On the basis of the carried out analysis, further research is performed for all known subexponential factorization algorithms, taking into account the possibility of developing a probabilistic approach.*

Key words: *elliptic curve, elliptic curve method, factorization, smooth numbers, computational complexity, sub-exponential complexity.*

Introduction

Elliptic curves, which were originally a subdivision of the functions theory, managed to find their use in both "pure" mathematics and applied fields. The elliptic curves theory is used in classical analysis, in abstract and computational number theory and occupies a fundamental position in these areas [1]. From an applied point of view, the elliptic curves theory apparatus has many applications. Its significance is difficult to assess. In particular, it has proved to be extremely useful in cryptography, for solving the problem of discrete logarithm and factorization, for creating primality tests, and also for constructing cryptographic protocols [2, 3]. In addition, it has good prospects for use in the construction of complex dynamic systems. Elliptic curves played an important role in the proving of the last Fermat theorem. It was proved as a generalization of the Taniyama-Shimura theorem in the 1990s. Nevertheless, this direction is relatively "young" and has many gaps and unresolved problems.

1. Comparative analysis of factorization methods

The one of works aims, is reasonable choice of the factorization problem solution way. The factorization universal algorithm developing problem is that for different composite numbers classes the algorithms behave differently. So therefore it is expedient to create a method for their construction that is resistant to natural numbers properties possible variations. As shown by the results of experimental modeling, it is possible to significantly reduce the effect of such number classes on the computational factorization algorithms efficiency with the help of probabilistic processes. The use of such an approach may allow the creation of algorithms whose computational speed can be similar to the speed of the most advanced sub-exponential factorization methods. That is, it is significantly lower than the average time spent on the algorithms of the subexponential class. The development of such an approach is absolutely justified due to the fact that this number class has not been studied theoretically. It is also not known how to find these numbers, and it is also completely unknown how they are distributed among the whole set of natural numbers. In the literature, all assumptions in relation to them are built only on the basis of empirical results, and different authors have different views on this problem. In fact, this moment partially reflects the probabilistic nature of all subexponential methods principles. This question remains open and under-researched. This problem is complex and intertwined with number theory.

Among the subexponential algorithms the following algorithms should be highlighted: Dixon's factorization method, continued fraction factorization method (CFRAC), the quadratic sieve method (QS), elliptic curve factorization method (the Lenstra's method, or ECM) and the numerical field sieve method (NFS). The NFS is considered the most effective algorithm for factoring large numbers (more than 10^{110}), as was shown by Crandall and Pomerance [2]. There are two types of the numerical field sieve method: general (GNFS) and special

(SNFS). The special method is obviously more effective than the general, however it can be used only to factorize the numbers of a special type: $r^e \pm s$, where $r \in N$, $s \in Z$, r and s are small. Table 1 below shows the computational complexity of each method in L-notation.

Table 1.
The computational complexity of subexponential factorization methods

Name of method	Computational complexity
Dixon's factorization method	$L_n(\frac{1}{2};2\sqrt{2})$
Continued fraction factorization method	$L_n(\frac{1}{2};\sqrt{2})$
Quadratic sieve method	$L_n(\frac{1}{2};1)$
Elliptic curve method	$L_p(\frac{1}{2};\sqrt{2})$
General number field sieve method	$L_n(\frac{1}{3};(\frac{64}{9})^{\frac{1}{3}})$

Where n – is a composite number that factorized, and p – is smallest factor of this number. Computational complexity is indicated in the L-notation. L-notation, is an asymptotic notation, similar to O-notation, used for approximate estimation of the algorithm computational complexity and is determined by the formula: $L_p[\alpha, c] = e^{(c+o(1))(\ln p)^{\alpha}(\ln\ln p)^{1-\alpha}}$, when $p \to \infty$, and $c = const$, $c \in (0;\infty)$ and $\alpha = const$, $\alpha \in [0;1]$.

The numerical field sieve method (both special and general) can be represented as the improvement of a simpler method - the rational sieve method, or the quadratic sieve method [6].

Algorithms of this class require to find smooth numbers. A mathematically substantiated definition of smooth numbers is absent in the literature, which causes a free interpretation of this concept. Many authors believe that these numbers value grows exponentially with increasing of n, such an assumption is not always justified. The numerical field sieve method, requires to find smooth subexponential numbers with the size near of $\sqrt{n}$ [6]. Due to the fact that these numbers are smaller, the probability that a number of such size turns out to be smooth is higher. This moment is the reason of the numerical field sieve method effectiveness [6]. To achieve acceleration, calculations within the method performance are carried out in numerical fields, which complicates the algorithm, compared to a simpler rational sieve method.

At the current stage of consideration, an important question arises about the concept of a number smoothness. The definition given in number theory can have a very free interpretation, since it states that a number is called smooth if all its divisors are “small”. Because of such definition of this term, many problems arise. In the case of the NTFS method, for simplicity, it is assumed that the term “smooth” means the number divisors of which are around $\sqrt{n}$ [6]. Nevertheless, such a definition does not introduce the necessary accuracy, but only limits the range of numbers that satisfy a given criterion. Thus the question of the number "smoothness" stays open in this case as well. One of the number smoothness determination is as follows: a number can be called a smooth, if the distance between this number ordered divisors is monotonic.

Crandall and Pomerance showed an important feature: none of the considered subexponential methods is strictly justified [2]. But precisely because of this methods class heuristic nature, the lower expected complexity of this type algorithms is achieved. Although the use of such methods is somewhat strange, nevertheless Crandall and Pomerance argue that in practice there is no need to constantly follow any of them, it is only necessary that the frequency of their implementation be sufficient for such an algorithm to be effective [2].

So, from the analysis of current most effective methods, due to perspective, complexity dependence mostly on smallest factor value, ease of the method parallelization, opportunity of using probabilistic approach of method research and wide specter of existing optimization, it seems that ECM is most promising among the factorization algorithms.

2. Survey of the factorization method based on the theory of elliptic curves.

First thing to do before starting complex analysis is to describe in details the ideas of method, it’s most important problems, and the possible ways of their solutions.

The elliptic curve method is an algorithm of stochastic nature, which first finds small factors.

The most important tasks when considering this point are the following:

- Questions about the probabilistic nature of the method and its determination.

- Features of the method, that allows to factor large composite numbers, and which distinguish this method from other existing ones.

- Questions about the concrete method computational complexity definition, in particular the problem of the precise subexponential complexity definition.

- How to determine the parameters of the method (the parameters of the curve itself and the coordinates of a point on it). Features of the curve random selection, and possible ways to reduce this approach randomization [7].

- In case of the method one full cycle unsuccessful execution the with given parameters, how these parameters should be changed for the subsequent run [7].

- How the execution time of the method and its specific implementations depends on the size of the already found factors of a given composite number [7].

- When the elliptic curve method will be most effective, and how should knowledge of this be used when combining this method with other factorization methods [7].

It’s seems that, the probabilistic nature of the method is difficult to interpret unambiguously. On the one hand, the stochastic character of the method leads to the fact that in most cases the method finds divisors faster than in the average search time by using other exponential algorithms [8]. On the other hand, it can even exceed this value. This also leads to the need of estimating the algorithm time by using probabilistic methods. Which show that, on average, it is still faster than exponential algorithms performance time [8].

Nevertheless, a more detailed analysis, make it obvious that getting rid of accidents by moving to a deterministic (or at least less randomized) choice of curves and points on them that, in a given class of numbers, will ensure that the divisor will be found much earlier. So it will greatly optimize this method. This approach has priority in the research of the ECM, but its development and implementation is a very time-consuming task, which still remains unsolved and moreover poorly studied. However, the importance of such research is quite obvious, from the method description.

The question about probabilistic nature of ECM is raised quite rarely. Nevertheless, in this paper, it occupies the fundamental position. The fact of this moment covering and approach of its consideration and research differs this article among other works that describes and research factorization problem, and its solutions based on elliptic curves theory.

The general form of the elliptical curve equation is as follows:

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \quad (1)$$

where $a_1, a_2, a_3, a_4, a_5, a_6$ - some constants belonging to the field over which this elliptic curve is being considered. On the basis of these coefficients specific values and the choice of the field with specific characteristic over which elliptic curve is constructed, a classification of elliptic curves is considered. Also a frequently-used criterion for the classification of elliptic curves is the curve "singularity". Those curves whose discriminant is different from 0 are not singular. Whereas curves with a discriminant equal to 0 are called singular. Thus, the classification of the entire set of curves is unusually rich, and at the same time it is far from concrete systematization due to the large number of criteria on the basis of which it can be carried out.

In the case of the Lenstra’s method, curves in the form $y^2 = x^3 + ax + b$, constructed over the field modulo n, where n is the factorized composite number. Curves of this type, built over a finite field, are called Weierstrass curves.

The method is based on the fact that, according to the arithmetic of elliptic curves, if n - is a prime number, the point at infinity means an unique additional projective point on an elliptic curve, which does not correspond to any affine point [2]. If number n - is composite, then there are other projective points to which no affine points correspond. However, only one additional point is allowed, it corresponds to the projective solution. $[0,1,0]$ [2]. Due to such a restriction in the definition of the elliptic curve group, the pseudo-elliptic curve no longer forms a group with a composite n [2]. There will be always such pair of points P and Q, for which sum $P+Q$ - is undefined. This is due to the angular coefficient structure:

$$m = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1}, if \rightarrow x_1 \neq x_2 \\ \dfrac{3x_1^2 + 2cx_1 + \lambda}{2y_1}, if \rightarrow x_1 = x_2 \end{cases}, \quad (2)$$

where $P = (x_1, y_1)$, $Q = (x_2, y_2)$.

These results are transferred to the elements of the set $E_{a,b}(Z_n)$, which differs from elliptic curves in the case when n - is composite number. In this case, the concept of an elliptic pseudo-curve is used, which is determined by the conditions:

1. $a,b \in Z_n$
2. $G.C.D.(a,b)=1$
3. $G.C.D.(4a^3+27b^2,n)=1$
4. $E_{a,b}(Z_n)=\{(x,y)\in Z_n \times Z_n : y^2=x^3+ax+b\}\vee\{O\}$,

where O - point at infinity, and $E_{a,b}(Z_n)$ - is an Weierstrass elliptic curve, constructed over the field modulo n which is determined by coefficients a and b, that belongs to this field.

In strict mathematical formulation, this curve is not considered as an elliptic curve (such a curve is also called a pseudo-curve), since F_p is not a field, and, based on this, it does not always able to perform the operations of finding the reverse element, which is necessary to find the sum curve points sum. It is due to the inability to calculate the sum of two points $P(x_1,y_1)$ and $Q(x_2,y_2)$, it turns out that the difference of the first coordinates x_2-x_1 should be equal 0 by modulo of n some divisor. So by computing the greatest common divisor of (n,x_2-x_1), the divisor of composite n is found. Lenstra's algorithm is based on the choice of elliptic pseudo-curve $E_{a,b}(F_p)$, random base point on it P_0, and its subsequent multiplication by various prime numbers and their degrees until we get:

$$kP_0=\infty(\mathrm{mod}\,p), \qquad (3)$$

where p – is one of n divisors.

Since none of n divisors it is not known beforehand, then there is no possibility of checking whether condition (3) is being fulfilled, on this basis, a sign of the algorithm successful completion is the fulfillment of the $G.C.D.(n,c)=d>1$ in the operation of doubling or adding points when calculating the next multiple c of the point P_0 when calculating the angular coefficient.

From the above methods, ECM is closest to a strict justification [2]. This is due to the Lenstra’s hypothesis [9] on the smooth numbers distribution in short intervals. By adopting this hypothesis, he showed that the expected number of arithmetic operations with integers of the order n, required to find the smallest divisor p of composite n by using elliptic curve method is equals: $\exp((2+O(1))\sqrt{\ln p \ln\ln p})$, where $O(1)\to 0$, when $p\to\infty$ [9].

At this point, the question of the number “smoothness” definition, considered at the beginning, arises again; in this case, the definition is similar to the definition for the NTFS method, with the following of the formulation inaccuracy.

Thus, there is only one heuristic gap in ECM, while QS and NFS have several similar gaps in their justification, as Crandall and Pomerance showed [2].

Also the concept b-smooth number is a fundamental in the elliptic curve theory. In the theory of elliptic curves, the term b-smooth number, is in some way a child in relation to the definition of a smooth number, but it has a concrete definition. According to Leonard Adleman’s definition, an integer number is called smooth if it consists of small simple factors. At this stage it is assumed that b-smooth number if none of the composite number n prime divisors does not exceed b [9]. Parameter b, in the case of the elliptic curve method, is the so-called method boundary and is determined directly by the researcher empirically.

When comparing these three most effective subexponential methods: ECM, QS and NFS, the size of the composite number smallest divisor is the main criterion.

Crandall and Pomerance showed that in the case when the factoring number has a size greater than the record value for these methods, the only way to find a divisor is factorization by using elliptic curves [2]. This is justified by the dependence of its computational complexity in the first place on the size of the smallest divisor. However, it is obvious that such an approach to factorization cannot be considered as universal due to the fact that a certain part of the numbers (despite that this part is a statistically small) will not contain “small” divisors. In this case their factorization by using ECM is an extremely time-consuming task. Also the method itself cannot be considered as the most effective. Moreover, its effectiveness is much lower, for example, than the NTFS method effectiveness in this case.

ECM does an excellent job of finding divisors of 10 to 15 digits, divisors of 20 to 25 digits with complications, sometimes taking about ten hours to successfully complete the algorithm. And for divisors up to 35-digits with considerable difficulty [8].

ECM has properties that make it ideal for parallelization. Parallelization of the method is the concurrent attempt to decompose a single composite number with the help of many different curves. Such concept is very clear, and its implementation isn’t great problem too.

An important feature of the ECM, which has already been mentioned, is the dependence of the method primarily computational complexity on the value of the factorized number smallest divisor. And although this method is considered only the third in terms of efficiency among the factorization methods for large numbers in general (it takes a position after the general numerical field sieve and the quadratic sieve method) [2], in the future, due to a wide range of possibilities for optimization described more detailed in article [10] and effective parallelization which was mentioned above, this method looks the most perspective [10]. Consequently, this causes its further consideration and analysis.

3. The problem “sub-exponential” computational complexity definition

The main criterion in the problem of one concrete method choice from a set of methods, is the computational complexity of the method. That is, the number of operations performed as a function of the task size. At the current stage, it is this factor that determines the speed of performance and, accordingly, the effectiveness of a particular approach. So it is required to accurately assessment the computational complexity of the method. If it is impossible to determine an exact estimate, then usually proceeds to the estimates of the method upper and lower bounds computational complexity, i.e. to computational complexity in the best and worst possible cases for this method, as well as to averaged and heuristic estimates. Thus, the analysis and research of any method should be justified, including its computational complexity. So, the research and analysis of subexponential complexity of ECM are very important goals of this paper.

The author of the elliptic curve method, showed that the expected number of arithmetic operations with integers that are around the order of n, required to find smallest divisor p of composite n by using ECM is equal: $\exp((2+O(1))\sqrt{\ln p \ln\ln p})$, where $O(1)\to 0$, when $p\to\infty$ [9]. In his asymptotical analysis Lenstra gives estimation $B(p)=L(p)^{1/\sqrt{2}+O(1)}$, for optimal choice of boundary B_1, for one-step version of method [9]. Brent’s analysis suggests that this value can be reduced to $\log p$ for two-step method version [11].

The complexity of the EC method is defined as subexponential [2]. A large number of questions are connected with this moment, and before proceeding to their consideration it is necessary to find out from the beginning what the exponential computation complexity represents.

Exponential complexity - in the case of consideration from the algorithms complexity theory, is a complexity of the problem that is limited by a polynomial degree of the problem size, that is, it is limited by the function $\exp(P(n))$, where P – some polynomial, and n – is size of the task [12].

There are algorithms among the class of exponential, the execution time of which is still higher than the polynomial (“superpolynomial”), but is on average the smallest among this class (“subexponential”). Thus, algorithms with subexponential complexity in some way are the most effective among the exponential. The question about this type complexity the exact definition is still open [13]. There are currently two main definitions.

The first definition: the complexity of the problem is defined as subexponential, in the case when it is solved by an algorithm, which logarithm of the operation time grows less than any given polynomial. This definition is given at Complexity Zoo [14].

The second definition: the running time of the sub-exponential algorithm is determined by the value $2^{O(n)}$. This definition implies a larger computational complexity for this type of algorithm. It is given by Oded Regev [15]. An example of an algorithm with subexponential time that fits this definition is the generalized numeric field sieve algorithm for integer factorization.

The inaccuracy of the definition that the subexponential computational complexity is intermediate between polynomial and exponential, is not accessible. There is a fundamental need for a clear measure of the difference between this complexity and the exponential. Which will show how much faster the algorithms of this class are.

In the case of factoring algorithms, the subexponential nature is expressed in the L-notation of the computations complexity. In this case, the algorithm complexity is an exponent of a certain constant product by the natural logarithm of the task size for a degree less than one multiplied by the double natural logarithm of the size of the problem to a degree less than one. That is, it is determined by the formula: $L_n[\alpha,c]=e^{(c+O(1))(\ln n)^{\alpha}(\ln\ln n)^{1-\alpha}}$, where

$n \to \infty$, $\alpha = const$, $\alpha \in [0;1]$, according to statements of Oded Regev [15].

The analysis of given problem leads to the following, as can be seen, complexity is no longer an exponent of a polynomial function (as with the classical representation of exponential complexity). But it is a product of a double logarithmic function of the task size and a logarithmic function. In this case, due to the restriction on the constant, both of these functions have a degree less than 1 (that is, a fractional degree). Thus, the growth rate of the function representing the degree in a given formula is significantly less than the polynomial function and even than the linear function. Because of this difference the subexponential complexity is a function whose complexity is significantly lower than a polynomial function, it stands out in a separate subclass that differs from the traditional representation of the exponential complexity class. So $(c + O(1))(\ln n)^{\alpha}(\ln\ln n)^{1-\alpha} << Poly(n)$ when $n \to \infty$, based on this statement, it follows that: $\exp((c + O(1))(\ln n)^{\alpha}(\ln\ln n)^{1-\alpha}) << \exp(Poly(n))$ when $n \to \infty$. Thus, the features of subexponential computational complexity becoming clear. The difference between more general exponential algorithm's class and more specific subexponential algorithm's class is shown.

4. Ways to solve probabilistic nature problem of the method

Due to probabilistic character of method it seems that using probabilistic number theory approach is very promising in this case. The main purpose of paper is finding the cases when this approach can be implemented, and the ways how to implement it.

The Dickman's function $\rho(\alpha)$ - it is the possibility that greatest prime divisor of integer number $x \to \infty$ has value less than $x^{1/\alpha}$. For each interval $[n-1, n]$ with integer n exists an analytical function ρ_n , such that $\rho_n(u) = \rho(u)$. When $0 \le u \le 1$, $\rho(u) = 1$. When $1 \le u \le 2$, $\rho(u) = 1 - \log u$. When $2 \le u \le 3$, $\rho(u) = 1 - (1 - \log(u-1))\log u + Li_2(1-u) + \frac{\pi^2}{12}$, where Li_2 - dilogarithm. Rest of ρ_n can be computed by using the infinite series [16]. Silverman and Wagstaff in their work designate through $\mu(\alpha, \beta)$ the probability, that x has the second largest prime divisor less than $x^{1/\alpha}$, and its greatest prime multiplier is less than $x^{\beta/\alpha}$, where $\alpha > \beta > 1$ [7]. Then the functional equations for ρ and μ are the follows $\rho(\alpha) = 1/\alpha \int\limits_{\alpha-1}^{\alpha} \rho(t)dt$ and $\mu(\alpha, \beta) = \frac{1}{\alpha - \beta} \int\limits_{\alpha-\beta}^{\alpha-1} \rho(t)dt$ [7].

For successful execution of the ECM algorithm, the order of the group formed by the points of a curve must be B_1-smooth, except for one additional prime divisor between B_1 and B_2, where B_1 and B_2 are the first and second boundaries of the ECM respectively. B_1-smoothness means b-smoothness of Adleman, definition of which was made earlier, but in this case the divisors should not exceed B_1 respectively. Designating $P(B_1, B_2)$ as the probability of successful method execution when choosing B_1 and B_2 as a boundaries , where $B_2 > B_1$. In this case: $P(B_1, B_2) \cong \mu(\alpha, \beta)$, when $\alpha = \frac{\log p}{\log B_1}$, $\beta = \frac{\log B_2}{\log B_1}$ [7].

The important point is to determine the belonging of a prime divisor p, of composite n, to the specific interval. The probability that at least one divisor p, belongs to interval $y^{\varepsilon} < p < y^{\varepsilon+1}$ approximately equals $1 - \prod\limits_{y^{\varepsilon} < p < y^{\varepsilon+1}} (1 - \frac{1}{p})$ [7].

The primary task is to estimate the order of the generated curve, for an effective curves iteration from a set of Weierstrass curves, those actually brute force pairs of parameters a and b that belongs to the field modulo n, of the elliptic curve until, we get a curve of B_1-smooth order. Such algorithm with the use of an effective curve order estimating method will lead to a significant increase in its effectiveness.

Since the generation of a curve is a computationally non-expensive operation, the entire complexity of the calculations is associated with finding primes in a given interval, and then multiplying the curve points by the given prime numbers and their degrees.

The problem is that there is no algorithm to pseudo-curves order calculation. The existing Schoof's curves order calculating algorithm, in addition to its complexity and complexity of its implementation, is intended for curves constructed over finite fields modulo prime number [2]. Knowing the divisor of a composite number n, it is possible to calculate the order of pseudo-curve by using Schoof's algorithm [2]. However, none of the divisors is known, moreover, the search of divisors is the Lenstra's method main goal.

Thus, the only way to solve this problem is a theoretical research of the elliptical curves structures and of their various classes. Also, according to empirical results, there is a relationship between the curve parameters values, for which the total required boundaries of the first and second stages value reaches its minimum, in the case where optimization in the form of having the second stage is used. Ishmukhametov in his textbook displays this empirical result. [8].

Since the theoretical apparatus for choosing these parameters is not currently developed, the only way out is to randomly generate an elliptic curve by randomly selecting its parameters, and the most effective way of optimization is to use several curves concurrently. Since the value of the divisor p is unknown, then the boundary selection is carried out empirically, which reduces the method practical assessment convergence reliability.

In this work we suggest that in this case the use of effective pseudo-random number generators (PRNG) is highly promising approach of ECM optimization. Efficiency in this case is primarily determined by the compliance of these generators distribution laws with the smooth order curves distribution laws. Since the choice of b-smooth order curve is a sufficient condition for successful completion of the algorithm. Thus the perspective of proposed approach becoming justified.

By default, PRNG software implementations use a uniform distribution law, sometimes a normal distribution law. As for the case of the ECM method, none of the authors gave either clear instructions or specific recommendations on the choice of the PRNG distribution law. It is obvious that this law must correspond to the smooth elliptic curves distribution law among the set of all curves. This question has a very clear wording, but it is rarely raised, and its significance is fundamental. Thus, one of the main issues, the solution of which can lead to significant success and a breakthrough in the development of the ECM is the determination of the distribution law for smooth elliptic curves.

We offer that the way to this problem solution, is the analysis of empirical results set for the different number structures factorization. It may provide direction for further theoretical study of this issue. The possible approach in this case is number analysis of smooth order curves parameters distribution among curves parameters in general for all integer composite numbers from 4 to as high values, as it be required to the patterns of such distribution. However, details should have been clarified, since these ideas are the first steps in this direction.

Significant results were obtained by Kowalski are described in his article [17]. He advanced in solving the elliptic curves distribution law problem by using the analytic number theory apparatus. Author considered this problem as an analogy to the classical the primes distribution problem. Also he involves the equidistribution of Frobenius elements to uniform and large moduli, especially on totally split primes in such extensions [17]. In particular, the author gives the definition of the "elliptic twins" [17], which in their essence resemble prime twins. This leads to some local results on the elliptic curves group structures distribution defined over a prime finite field, demonstrating an interesting dichotomy for the possible groups appearing [17]. Author analyze this problem on generalized Riemann hypothesis and discuss the new difficulties which arise in comparison with the case of primes inarithmetic progressions [17]. Kowalski says that many of the questions raised in this article [17] seem to be very complex. On the other hand, they seem very interesting from the analytical number theory point of view. Since the problems are amenable to experimentation, Kowalski present some numerical data and further remarks. Despite the fact that the author himself claims that the results obtained in his article [17] are modest. Nevertheless, these results undoubtedly are the foundation for the whole set of elliptic curves research by means of analytic number theory.

In addition, the important point is the possibility of obtaining a divider as G.C.D. of curve discriminant and a factorized number [2]. Such cases for large n, are quite rare. However for $n<10^{11}$, their frequency is sufficient to occur by a simple curves iterating. Such an approach for numbers of small size is often a much more effective way of factorization, as evidenced by the empirical results described in the article [18].

However, also almost nothing is known about the elliptic curve discriminant properties. In particular, the authors do not give clear estimates of the divisor obtaining probability by this method, and the justification for this approach is not given. Nevertheless, the empirical results [18] suggests that for small n, this approach is more efficient in terms of computational complexity. This is partly due to a significant increase in the algorithm computational complexity with an increase of the boundary b. The main difficulties in this case are associated with an increase in the number of considered primes and, accordingly, the operations number during each cycle curve generation and the point on it generation with further point multiplication. This empirical results have great importance for that research, and probably are fundamental to understand the nature of elliptic curve discriminant. However, this problem does not seem such important and perspective as other problems described in this paper.

Also the research of the base ECM relationship between the number of curves generated and the required boundary problem is important. Since the computational costs for the algorithm performance depend significantly on the selected boundary, the more it is, the more calculations are performed (in the case of a complete cycle pass). But this way increases the probability of divisor obtaining in fewer passes (due to the requirement b- smoothness for the successful completion of the method is less strict with a larger value of the selected boundary). A more detailed statement of the problem and the first steps of its resolving with the corresponding experiment results, are described in the article [18].

Conclusions

In this paper, the place of elliptic curves in the applied mathematics fields was described, in particular, when solving the factorization problem. A comparative characterization of the factorization methods is given, the importance of the analysis and research of the method based on the elliptic curves theory is substantiated. It is necessary to direct as much effort as possible to the development of the Lenstra method for the reasons that the factorization problem is one of the fundamental in modern mathematics and number theory. It is important both in their theoretical aspects and in the applied sense. A brief description of the method is given. The concept of computational complexity of the method is analyzed. Special attention is paid to the definition of subexponential complexity, since it is directly related to the evaluation of the computational complexity of the ECM. Fundamental questions are posed in the study of the Lenstra’s method, and approaches for their solution are described. Analyzed the probabilistic nature of the method, given the prospects and ways to determine it, in particular by using the "effective" in this case, pseudo-random number generators to generate a curve and a point on it. Probabilistic algorithm’s class based on random elliptic curve choice is built. Proved that computational complexity of such algorithm’s class is highest among existing exponential algorithms. Also the problem of smooth numbers is considered and researched.

It seems that ECM has a great field for research, and a great approaches number for this researches. This paper gives some of such approaches. Perspective and expediency of them is justified. In some sense this paper has a survey character. It shows new approaches of factorization problem solving by using probabilistic number theory, in particularly. Part of suggestions are obvious enough, and don’t need detailed and strict justification. However, great part of ideas described in this paper are brave enough, and should be considered more in future. Their consideration should include the results of empirical researches and deep theoretical analysis.

References

1. Koblitz, N. (2001), A Course in Number Theory and Cryptography. — M.: Scientific Publishing House "TVP", P. 188–200, — ISBN 5-85484-014-6.

2. Crandall, R. E., Pomerance, C. B. (2001), Prime numbers: A Computational Perspective. — New York: Springer-Verlag, 545 p., — ISBN 0-387-94777.

3. Wagon, S. (1986), Primality Testing. — Math. Intel, Vol.8, Num. 3, pp. 58–61.

4. Taniyama, Yutaka (1956), Problem 12. — The Bulletin of the London Mathematical Society, vol. 21 (2), pp. 186–196.

5. Soloviev, Yu. (1998), Taniyama’s hypothesis and Fermat last theorem [Gipoteza Taniyami i poslednyaya teoriya Ferma], — Sorov education journal, 138 p.

6. Couveignes, J., (1993), A general number field sieve implementation, — Lecture Notes in Mathematics, pp. 103–126.

7. Silverman, R., Wagstaff, S. (1993), A practical analysis of the elliptic curve factoring algorithm — Mathematics of computation volume 61, number 203, pp. 445–462.

8. Ishmukhametov, Sh. T., (2011), Methods for the factorization of natural numbers [Metody faktorizacii naturalnyh chisel], Kazan. UN., Kazan, 190 p.

9. Lenstra, H. W., (1987), Factoting integers with elliptic curves, Annual of Mathematics Volume 126, New-Jersey, pp. 649–673.

10. Brent, R. (1998), Some integer factorization algorithms using elliptic curves, Australian Computer Science Communications 8, 149–163 p.

11. Vostrov, G., Dermenji, I., (2018), Algorithmic processes of the large numbers factorization based on the theory of elliptic curves, ELTECS ONPU, Астропринт, vol. 28, pp. 223-233. – ISSN 2221–3805.

12. John von Neumann (1958), A certain zero-sum two-person game equivalent to the optimal assignment problem, Contributions to the Theory of Games, Princeton Univ. Press, pp. 100–122.

13. Aaronson, S., (2009), A not-quite-exponential dilemma, Shtetl optimized, available at: https://www.scottaaronson.com/blog/?p=304.

14. Complexity Zoo, (2008), Wayback Machine Class SUBEXP: Deterministic Subexponential-Time, available at: https://complexityzoo.uwaterloo.ca/Complexity_Zoo:S#subexp

15. Regev, O., (2004), A Subexponential Time Algorithm for the Dihedral Hidden Subgroup Problem with Polynomial Space, available at: https://arxiv.org/abs/quant-ph/0406151.

16. Bach, E., (1996), Asymptotic Semismoothness Probabilities, Peralta, René, Mathematics of Computation, vol. 65 (216), pp. 1701-1715, DOI:10.1090/S0025-5718-96-00775-2

17. Kowalski, E., (2005), Analytic problems for elliptic curves, available at: https://arXiv.org:math/0510197.

18. Vostrov, G., Dermenzhy, I., (2018), Factorization Algorithm Based on the Elliptic Curve Theory, Proceedings of the Xth International Scientific and Practical Conference “Electronics and Information Technologies” ELIT, Lviv-Karpaty village, August 30 – September 2 2018. – Lviv: Ivan Franko National University of Lviv, Section A, pp. 117–121.

АНАЛІЗ МЕТОДУ ФАКТОРИЗАЦІЇ НА ОСНОВІ ТЕОРІЇ ЕЛІПТИЧНИХ КРИВИХ

Дерменжи І. Д., Востров Г. М.
Одеський національний політехнічний університет

Анотація. *У даній статті описані можливості використання апарату теорії еліптичних кривих, зокрема, при вирішенні задачі факторизації складеного числа. Дана задача займає фундаментальне положення серед як чисто математичних, так і прикладних наук. Розглянуто проблему імовірнісного принципу роботи субекспоненційних методів факторизації. Дан порівняльний аналіз існуючих субекспоненційних методів факторизації. Описані основні ідеї, переваги та недоліки метода решета числового поля, окрема увага уділяється проблемі визначення гладкого числа для даного класу алгоритмів. Обґрунтовано вибір методу, заснованого на теорії еліптичних кривих. . Ідея методу ґрунтується на побудові псевдокривої над кільцем лишків складеного числа. Завдяки цьому вдається отримувати ситуації, коли неможливо знайти зворотний елемент в заданому кільці при складанні двох точок кривої, що сигналізує про знаходження дільника. Головною особливістю методу є залежність його обчислювальної складності від найменшого дільника числа, що факторизується, а не від безпосередньо нього самого. Обґрунтована перспективність дослідження та розвитку цього метода. Описані головні задачі, що виникають при докладному аналізі методу еліптичних кривих. Особлива увага приділяється імовірнісному і субекспоненціальному характеру методу. Описана та проаналізована проблема визначення терміну субекспоненційна складність алгоритму, зокрема аналізується субекспоненційна складність безпосередньо методу еліптичних кривих. Описано основні напрямки розвитку та оптимізації метода. Дано та проаналізовано можливі підходи до вирішення описаних проблем. Розглядана можливість створення імовірнісної моделі, що використовує теорію еліптичних кривих при факторизації. В основі заданого рішення полягає використання ефективних генераторів псевдо випадкових чисел. На основі проведеного аналізу проводяться подальші дослідження для всіх відомих алгоритмів факторизації субекспоненціального класу з урахуванням можливості розробки імовірнісного підходу. Також увага приділяється питанням щодо дискримінанту кривої, та його властивостям що дозволяють на його основі отримувати дільники складеного числа.*

Ключові слова*: еліптична крива, метод еліптичної кривої, факторизація, гладкі числа, обчислювальна складність, субекспоненціальна складність.*

АНАЛИЗ МЕТОДА ФАКТОРИЗАЦИИ НА ОСНОВЕ ТЕОРИИ ЭЛЛИПТИЧЕСКИХ КРИВЫХ

Дерменжи И. Д., Востров Г. Н.
Одесский национальный политехнический университет,

Аннотация. *В данной статье описаны возможности использования аппарата теории эллиптических кривых, в частности, при решении задачи факторизации составного числа. Дан сравнительный анализ существующих методов факторизации. Обоснован выбор метода, основанного на теории эллиптических кривых. Особое внимание уделяется вероятностному и субэкспоненциальному характеру метода. Описаны основные направления развития и оптимизации метода. Даны и проанализированы возможные подходы к решению описанных проблем. Рассмотрена возможность создания вероятностной модели, использующей теорию эллиптических кривых при факторизации. На основе проведенного анализа проводятся дальнейшие исследования для всех известных алгоритмов факторизации субэкспоненциального класса с учетом возможности разработки вероятностного подхода.*

Ключевые слова: *эллиптическая кривая, метод эллиптической кривой, факторизация, гладкие числа, вычислительная сложность, субэкспоненциальная сложность.*

Received: 00.00.0000.

George Vostrov, Ph. D. of Technical Sciences, Associate Professor of the Department of Applied Mathematics and Information Technologies, Odessa National Polytechnic University. Shevchenko ave., 1, Odessa, Ukraine.
E-mail: vostrov@gmail.com, mob. +380503168776

Востров Георгій Миколайович, кандидат технічних наук, доцент кафедри прикладної математики та інформаційних технологій Одеського національного політехнічного університету. Проспект Шевченко, 1, Одеса, Україна.
E-mail: vostrov@gmail.com, тел. +380503168776

ORCID ID: 0000-0003-3856-5392

Ivan Dermenzhy, Student of the Department of Applied Mathematics and Information Technologies, Odessa National Polytechnic University. Shevchenko ave., 1, Odessa, Ukraine.
E-mail: ivandermenji97@gmail.com, mob. +380965824211

Дерменжи Іван Дмитрович, студент кафедри прикладної математики та інформаційних технологій, Одеського національного політехнічного університету. Проспект Шевченко, 1, Одеса, Україна.
E-mail: ivandermenji97@gmail.com, тел. +380965824211

ORCID ID: 0000-0003-0421-3372