Pseudoprime Numbers: Basic Concepts And The Problem Of Security

Vladimir Pevnev

National Aerospace University „KhAI“, Kharkiv, Ukraine
V.Pevnev@csn.khai.edu

Abstract. The analysis of existing methods of construction of prime numbers is carried out in the presented paper. New concepts related to pseudoprime numbers are introduced. Theorems are formulated and proved on the basis of which such numbers are constructed. Using pseudoprime numbers can significantly reduce the number of candidates viewed as primes. The proposed corollaries of the theorems allow to significantly increase the ranges of unambiguous determination of such candidates. The results of the conducted experiments are presented, their analysis is carried out. The most important result of applying pseudoprime numbers is the estimation of the number of numbers to be checked. For large numbers with a size of more than 200 decimal signs, they make up less than 8 percent of the numbers on a given interval, and as the size of the numbers increases, this percentage decreases. Based on experimental data, basic properties of pseudoprime numbers are formulated.

Keywords. Prime, pseudoprime, theorems of the theory of pseudoprime numbers, properties of pseudoprime numbers, cardinality of a pseudoprime number.

Key-Terms: FormalMethod, MathematicalModel, Research.

1 Introduction

The development of infocommunication technologies has led to a decrease in the protection of all high-tech systems. Particularly acute is the issue of ensuring functional safety in such systems as the Internet of Things, Cloud storages, big data. Any computer that performs the functions of managing, processing or storing data, becomes available from anywhere in the world. Ensuring information security, as an integral part of the functional safety, is extremely reluctant, because it costs a lot of money, but there is no profit. Numerous materials in the press confirm the necessity of such works, the most famous became an example of the virus Stuxnet, which paralyzed the Iranian nuclear program [1].

One of the areas of information security is ensuring the confidentiality of information circulating in the system. As the most general way to solve this problem is cryptographic protection. As is a well-known resistance of cryptosystems is determined by the secrecy of the key. In asymmetric encryption systems, this secrecy depends on the size of the key. One of the most general public key systems is the

RSA system, which is based on the factorization problem, which belongs to the class of NP-complete tasks [2]. At the time of writing of this paper, the size of the used keys is 2048 or more bits. And here another problem of finding a prime number of large dimension arises.

In number theory, the problem of finding and proving the primality of a number is one of the main. The complexity of finding prime numbers is due to the fact that there is no function of their distribution on the number axis.

To build large prime numbers the following method is used in many sources [3-5]. A sequence of primes $p_1 < p_2 < p_3 < \ldots$ is constructed until the prime number of the required quantity will be found. The prime odd number p_1 is selected randomly. After the prime number p_{i-1} is constructed, a random number r, $1 \le r \le p_{i-1} - 1$ need to be selected. Let $r = 2s \cdot t$, t is odd. Then, as a candidate for the next prime number p_i is taken $n = 2rp_{i-1} + 1 = 2^{s+1} * p_{i-1} \cdot t + 1$. Further, n is checked for primality by known methods. The disadvantage of this approach is obvious - the probability of guessing at large prime numbers (> 200D) is too small.

Another group of methods [4,5] is based on the choice of the arithmetic sequence or the sum of products of primes with unity. The main disadvantage of all considered methods is obtaining predictable prime numbers, which are easy enough to repeat. If the thus obtained prime number is used as keys in encryption systems, then it becomes possible to build a pool of keys most usually used by users [6]. This leads to a fairly quick compromise. This problem is already openly stated by specialists in the field of information security [7].

In papers [6, 8, 9] the practical, theoretically grounded possibility of determining the minimum distance between two adjacent prime numbers is shown. This is achieved through pseudo-prime numbers that can be used to determine the possible location of prime numbers. With the proposed approach, it is possible to sharply reduce the number of checked numbers by the primality with no single prime number being skipped.

The next step in the search for a prime number is the definition of the primality of the number being checked. As mentioned above, none of the existing exact algorithms can solve this problem within an acceptable time. It should be noted that in [10] the deterministic polynomial algorithm AKS of the definition of the primality of numbers is given, but its complexity is $O(\log^{18} n)$ [11]. This fact makes this algorithm inconvenient for practical application.

2 Terminology

Let us consider the concept of a pseudoprime number. This concept was introduced for numbers successfully passing the Fermat test. Depending on the primality tests that check numbers for primality, there are different types of pseudo-prime numbers. These types include pseudoprime numbers of Fermat, Fibonacci, Lucas, Euler-Jacobi. What is common to all these numbers? All these numbers, being composite, successfully pass the corresponding tests for primality.

Definition of the pseudoprime number. The pseudoprime is a number N whose primality is not proved.

Definition of a cardinality of the pseudoprime number. Let us introduce the concept of the cardinality of the pseudoprime number. The cardinality *P* of a pseudoprime number *N* is the guaranteed number of prime numbers that is not divisible by a given number. Obviously, the greater cardinality of such number, the greater the probability that this number will be prime.

Definition of a prime number. Based on the cardinality of the pseudoprime number, we can formulate the concept of the prime number. The pseudoprime number *N* will be proved as prime if and only if its cardinality is equal to $P = \pi(N^{1/2})$ on the interval of natural numbers from 1 to $N^{1/2}$. In this definition, a $\pi(N)$ refers, as in the theory of numbers, the number of primes on the interval of natural numbers from 1 to *N*.

Definition of a factorial of the prime number. Let us introduce the concept of the «factorial of the prime numbers» and denote it as follows $\pi(n)!$. This entry denotes the product of all primes not greater than *n*. For example, $\pi(7)!= \pi(8)!= \pi(9)!= \pi(10)!=2*3*5*7=210$. In fact, the $\pi(n)!$ means the number of prime numbers in the product. This term in its meaning coincides with a term primorial, introduced by H Harvey Dubner in 1987 and received $p_n\#$ designation [11]. In the author's opinion, the proposed designation $\pi(n)!$ is more obvious, therefore in this paper, such designation form is used.

3 Basic Theorems Of The Theory Of Pseudoprime Numbers

Theorem 1. The sum (difference) of products of two disjoint sets of prime numbers is a prime number with each of the elements of these sets.

$$gcd\left(\left(\prod_{i=1}^{k} a_i \pm \prod_{j=1}^{l} b_j\right),\ \forall x \in X = A \bigcup B\right) = 1,$$

where $a \in A; |A| = k;$
$b \in B; |B| = l;$
$A \subset P;$
$B \subset P;$
$A \bigcap B = 0;$
P – a set of prime numbers.

Proof of the theorem 1.

Consider any element *a* belonging to the set A ($a \in A$). It is obvious that for any *a*, the following expressions are true:

$$gcd\left(\prod_{i=1}^{k} a_i\ ,\ \forall\ a \in A\right) = a,$$

$$gcd\left(\prod_{j=1}^{l} b_j\ ,\ \forall\ a \in A\right) = 1.$$

Therefore, based on the above arguments,

$$gcd\left(\left(\prod_{i=1}^{k} a_i \pm \prod_{j=1}^{l} b_j\right),\ \forall a \in A\right) = 1.$$

A similar proof can be given for any element *b* belonging to the set B ($b \in B$) and, as a result, we get that

$$gcd\left(\left(\prod_{i=1}^{k} a_i \pm \prod_{j=1}^{l} b_j\right),\ \forall b \in B\right) = 1.$$

Since the element *x* belongs to the union of two disjoint sets of primes *A* and *B*, then it can be stated that

$$gcd\left(\left(\prod_{i=1}^{k} a_i \pm \prod_{j=1}^{l} b_j\right),\ \forall x \in X = A \bigcup B\right) = 1.$$

The theorem is proved.

A corollary of Theorem 1. If we take $\pi(n)!$ and add 1 to it, the resulting number will be relatively prime with all prime numbers not greater than *n.* As a matter of fact, this formulation almost completely repeats Euclid's theorem on the infinity of primes. On the basis of the Theorem 1, the next relatively prime is the number obtained as the sum of $\pi(n)!$ and the first prime number that is greater *n*. If to $\pi(n)!$ Add the next prime number, and then we get a number that is relatively prime to numbers, taking part in this action.

Repeating this procedure, we can get a set of relatively prime numbers. Given that the obtained numbers are not guaranteed to be divisible by the first $\pi(n)$ prime numbers, it can be argued that these numbers are pseudoprime with the cardinality $\pi(n) + 1$.

Theorem 2. On the interval between $\pi(n)! + 1$ and added to $\pi(n)!$ prime numbers All numbers will be composite, except for those that have turned out as a result of addition, And the maximum added prime number must be less than the square of the first prime number to be added.

Proof of the theorem 2.

We begin the proof by considering the restrictive part of the theorem. Let us consider a prime number a_i, that belongs to the set of primes A.

$$a_i \in A; \quad |A| = B; \ B >> \pi(n),$$

where B – the cardinality of the set A.

If we consider a number equal to the sum of $\pi(n)!$ and the product of two primes greater than *n*, then, according to Theorem 1, it will be a pseudoprime number. The resulting number can be a prime number. Suppose that the minimal prime number, greater than *n*, is equal to *a*. Then the minimum product of two prime numbers, greater than the n, the number will be greater a^2. Thus, the maximum prime number that can be used to uniquely define a pseudo-prime number must be less than a^2. The restrictive part of the theorem is proved.

Consider the interval between numbers $\pi(n)!+1$ and $\pi(n)!+a$, where a is the nearest to n prime number. The smallest number on this interval is $\pi(n)!+2$. It is quite obvious that this number is divided by 2. Similarly with the following number $\pi(n)!+3$ that is divided by 3. In the general case, all numbers that are not greater than $\pi(n)!+ n$ have at least one common divisor with $\pi(n)!$. If we consider the following number $\pi(n)!+n+1$, then it will always be even, like all the others, which result from the addition of two odd numbers n and k. If we consider numbers of the form $\pi(n)!+n+2k$, then they can be either prime or composite.

If the number is composite, then it can be represented, in accordance with the main theorem of arithmetic, in the form of a product of primes, and at least one of the factors is not greater than the number n. If the number $\pi(n)!+n+2k$ is prime, then it will correspond to the number a. This means that on the interval between the numbers $\pi(n)!+1$ are $\pi(n)!+a$ all numbers are composite.

Let us consider the whole interval indicated in the Theorem 2. Since we consider we consider numbers less than a^2, then similar arguments can be applied to all the numbers of the interval under consideration. Thus we can conclude That on the interval between the numbers $\pi(n)!+1$ and $\pi(n)!+a^2$ All numbers will be guaranteed composite, except for those obtained as the sum between $\pi(n)!$ and prime numbers in the interval from a to a^2.

The theorem is proved.

It should be noted that the Theorem 2 can be formulated as follows: «On the interval between $\pi(n)!-1$ and subtracted from $\pi(n)!$ prime numbers, all numbers will be composite except those obtained as a result of subtraction, and the maximum deductible prime number must be less than the square of the first subtrahend prime number». The proof is similar to the above.

A corollary of Theorem 2. The above range concerns the unique definition of pseudoprime numbers. If the added prime number a_j is greater than a^2, part of numbers which will be within the range between $\pi(n)! + a_j$ and $\pi(n)! + a_{j+1}$ can be prime numbers. In order not to miss such locations of prime numbers, It is necessary to construct all possible combinations of primes greater than n, including the value of their degrees. At the same time, these combinations should not be included among which are the factors of $\pi(n)!$.

As it is proved in the Theorem 2, to uniquely define a pseudoprime number, the value of the prime number to be added must not exceed the square of the first prime number greater than n. Is this large or small? When considering large numbers $n >$ 200 D this will be a relatively small number. For example, $\pi(1000)!$ will correspond to the number 416 D. The interval of numbers on which it will be necessary to search for a prime number will be 1018081. There are 79682 pseudoprime numbers on it, which is less than 8 percent of all numbers in this interval. It should be noted the possibility of increasing the checked interval. According to the Theorem 1, by subtracting a prime number from $\pi(n)!$ it is possible to increase its size to 2036162, and the number of checked numbers to 159364. It is obvious that with the increase of n, the interval will increase rapidly.

Any number can be represented as a sum of two numbers. Proceeding from the fact that all pseudoprime numbers are odd, any of them can be represented as a sum of

even and odd numbers. Obviously, there are a lot of variants of representation and the larger the number are, the more such options are.

Each of the terms, according to the main theorem of arithmetic, can be represented as the product of primes.

Theorem 3. If at least in one of the variants of representing a pseudoprime number in the form of two terms, these terms turn out to be not mutually prime numbers, then the considered number is a composite.

Proof of the theorem 2. Let us take two disjoint sets of prime numbers A and B.

$$A = \{a_i\},\ B = \{b_j\}.$$

$$a_i,\ b_j \in N.$$

$$A \cap B = \varnothing.$$

Let us consider the products of the set A and B. Obviously, these are two relatively prime numbers.

$$gcd\ (\prod a_i,\ \prod b_j) = 1.$$

To ensure that the numbers are not relatively prime, they must have a common divisor. Введем в каждое из произведений по множителю C. In this case

$$gcd\ (C*\prod a_i,\ C*\prod b_j) = C.$$

The sum of these two numbers can be represented in the form

$$C*\prod a_i + C*\prod b_j = C\ (\prod a_i + \prod b_j).$$

In this case, the number under investigation will be the composite. As it should be proved.

4 Using Pseudoprime Numbers

When using primes, two problems usually occur: finding a prime number and checking the selected number for primality. The main disadvantage of the known algorithms for finding prime numbers is a simple search of candidates. In the best case, the numbers multiples of two are discarded. Using pseudoprime numbers, only the numbers that are obtained as a result of applying theorems 1 and 2 are checked. Organizing the checking of a number for primality, both deterministic and probabilistic methods are used. Both groups are well known and described in many sources [3, 13, 14].

The range of single-valued definition of pseudo-simple numbers will be denoted as the step of the definition (SD). An SD size can be easily calculated using the theorem 2. Table 1 presents a data that allows evaluating the performance of the proposed method for finding pseudoprime numbers. The table contains the following data:

- the $\pi(n)!$ size – Number of decimals in the product;
- the highest factor (HF) – the greatest prime number in $\pi(n)!$;

- the number of primes (NP) – number of primes on SD;
- the percentage of the checked numbers (% checked) – ratio of the number of prime numbers in SD to the total number of numbers on the SD.

As can be seen from Table 1, the growth rate of a quantity of the size $\pi(n)!$ Exceeds the growth rate of the factorial dependence that once again emphasizes the complexity of the problem of finding prime numbers of large dimensions. The most interesting data, which are given in the table, in the opinion of the author, is the percentage of the numbers being checked. For large numbers (D> 200), this index is less than 9% of the total number of numbers on the interval being checked. If you take any algorithm that will only check for numbers ending in 1,3,7,9, then the percentage of the checked numbers will be 40. When the cardinality of a pseudoprime number increases, the percentage of the numbers being checked decreases.

Table 1.

$\pi(n)$	Size	HF	SD	NP	% checked
10	9	29	961	152	22
20	25	71	5329	685	13
30	46	113	16129	1847	11,5
40	68	173	32041	3396	10,6
50	91	229	54289	5472	10
60	115	281	80089	7782	9,7
70	140	349	124609	11631	9,3
80	166	409	175561	15873	9
90	191	463	218089	19361	8,9
100	218	541	299209	25836	8,6
110	246	601	368449	31310	8,5
120	274	659	436921	36594	8,4

To increase the SD size, it is necessary to multiply the obtained product of prime numbers before their summation with the corresponding prime numbers by all the numbers less than the next prime number for $\pi(n)$. The obtained results for $\pi(11)$ are presented in Table 2.

Table 2.

Calculating the beginning of the interval	Begin of the interval	Number of checked PN	Number of PN
$\pi(11)!$	2310	34	23
2* $\pi(11)!$	4620	34	21
3* $\pi(11)!$	6930	34	19
4* $\pi(11)!$	9240	34	18
5* $\pi(11)!$	11550	34	16
6* $\pi(11)!$	13860	34	18
7* $\pi(11)!$	16170	34	16
8* $\pi(11)!$	18480	34	14
9* $\pi(11)!$	20790	34	15
10* $\pi(11)!$	23100	34	14
11* $\pi(11)!$	25410	34	15
12* $\pi(11)$	27720	34	21

Based on the theorem 3, a method of checking numbers for simplicity is proposed.

1. From the given number, the square root is calculated and the result is rounded down.
2. Calculates the difference between the target number and the number obtained.
3. These numbers are decomposed into factors.
4. If the factors have at least one the same number, then go to step 8.
5. From a smaller number, one is subtracted.
6. If the result is bigger than 1, then go to step 2.
7. The number is prime.
8. The number is composite.

By its idea, the presented method is similar to a quadratic sieve. The difference is that the number obtained as a result of the extraction of the square root in the proposed method decreases. This is due to the fact that more than one square root can have a maximum of one efficient, while the remaining ones will be less than the resulting root.

As an example, consider the number 996533. The square root of this number rounded down will be equal to 998. In Table 3t there are numbers equal and less than 998 and their decomposition, according to the main theorem of arithmetic, in columns one and two are. There are the numbers obtained as the difference between the checked number for primality and the numbers in the first column, in the third column. There is the decomposition of numbers from the third column in the fourth column. There is the greatest common divisor in the fifth column.

Table 3.

I number	Decomposition of the I-st number	II number	Decomposition of the II -nd number	gcd
998	2*499	995535	3*3*5*22123	1
997	PN	995536	2*2*2*2*43*1447	1
996	2*2*3*83	995537	17*157*373	1
995	5*199	995538	2*3*277*599	1
994	2*7*71	995539	PN	1
993	3*331	995540	2*2*5*7*13*547	1
992	2*2*2*2*2*31	995541	3*29*11443	1
991	PN	995542	2*497771	1
990	2*3*3*5*11	995543	19*151*347	1
989	23*43	995544	2*2*2*3*3*3*11*419	1
988	2*2*13*19	995545	5*199109	1
987	3*7*47	995546	2*497773	1
986	2*17*29	995547	3*7*47407	1
985	5*197	995548	2*2*248887	1
984	2*2*2*3*41	995549	PN	1
983	PN	995550	2*3*5*5*6637	1
982	2*491	995551	PN	1
981	3*3*109	995552	2*2*2*2*2*53*587	1
980	2*2*5*7*7	995553	3*3*13*67*127	1
979	11*89	995554	2*7*17*47*89	89

How effective is this way of confirming the number primality? Obviously, a sufficiently long time will be spent on decomposing a large number into multipliers. Proceeding from the stated problem of finding a common factor, it makes sense to use the well-known Euclidean algorithm for finding the greatest common divisor.

Before proceeding to the use of the Euclidean algorithm, it is necessary to note one more regularity. Most of the pairs under consideration have in their composition the first ten primes from 2 to 29. If we multiply these numbers, the result is 6469693230.

Define gcd of numbers 6469693230 and 996533. gcd (6469693230, 996533) is 1. This means that our investigated number is not divisible without a remainder by any prime number less than or equal to 29, that cardinality of a number 996533 is proved (P= 10). This means that as soon as the remainder of the division in any iteration becomes equal to or less than the number 29, it is possible to assert unequivocally, that gcd of this numbers will be equal 1. Using the product of the following 10 prime numbers from 31 to 71, it is possible to calculate the gcd of this pair of numbers.

If we consider table 3, then gcd should be searched between the I and II numbers represented in the first and third columns respectively. The result is presented in the fifth column.

The possibility of applying gcd detection methods is quite productive. Although this is nothing more than a method of trial division, but the speed of the proposed method is much higher. This can be achieved by first calculating the products of prime numbers, and their number in the product can be quite large. For example, if we multiply all the prime numbers that are less than 1000, and there are 169 such numbers, and gcd = 1, then the probability that the test number is composite, will be equal 2^{-169}[15,16]. It should be noted the fact that the results of the preliminary calculation can be stored in a database that will be constantly updated.

Quite interesting results, from the point of view of parallelization of the calculation process, is presented in Table 4. If you take the same number 996533 and repeat the process, the results of which are presented in Table 3, having previously increased the number several times. Table. 4 shows the number steps NS on which the desired result is obtained, depending on the multiplication factor MF.

Table 4.

MF	1	2	3	4	5	6	7	8	9	10
NS	20	79	40	40	8	43	62	66	59	43

Obviously, the obtained result will depend on the ratio of the two terms and the greatest common divisor.

5 Properties Of Pseudoprime Numbers

At this moment, the complexity of constructing prime numbers is due to the absence of the distribution laws of this numbers. Unlike prime numbers for pseudoprime numbers, based on the considered examples, it is possible to reveal some regularities conditioned by the properties of these numbers and the method for their construction.

The first property is repeatability of pseudoprime numbers. This is because it is possible to reuse prime numbers in the product of prime numbers. The cardinality of the pseudoprime number remains unchanged, and its size increases. This property is clearly seen from Table. 2.

The second property is specularity or symmetry. This property follows from the theorem 1. Pseudoprime numbers are symmetric with respect to $k\pi(n)!$.

6 Conclusions

The presented paper is the result of three years works on creating the theory of pseudoprime numbers. In this paper for the first time a definition of some terms is given, some theorems on pseudoprime numbers are formulated or slightly modified, a

correct proof is given. All the above theoretical calculations are confirmed by the carried out experiments, which confirm their correctness.

The use of pseudoprime numbers significantly reduces labor costs searching for prime numbers by specifying a location on the numeric axis where they can be located. As the size of the generated number increases, the relative number of the considered options decreases.

Based on the above theorems, it is possible to state quite reliably the possible numbers of twins, although this requires additional studies.

The use of a variety of combinatorial methods for the use of pseudoprime numbers makes it possible to assert a high potential for their use in finding prime numbers.

References

1. The Real Story of Stuxnet - IEEE Spectrum, www.spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet
2. Gary, M., Johnson, D., Computers intractability: a guide to the theory of NP-completeness. San Francisco: W.H.Freeman, (1979)
3. Emily Riemer. Pseudoprimes and Carmichael Numbers.MATH0420 (2016)
4. Couvreur, C., Quisquater J. J. An introduction to fast generation of large primes. Philips J. Res 37, pp. 231–264 (1982)
5. Mihailescu, P. Fast generation of provable primes using search in arithmetic progressions. Advances in cryptology—CRYPTO '94 -Santa Barbara, CA, pp. 282–293, (1994)
6. Pevnev, V. Ja. Metodika postroenija psevdoprostyh chisel (In Russian). In: Sistemi obrobki informacii. Kharkov, HUPS im. I. Kozheduba Publ., 3(140), pp. 30-32. (2016)
7. Mimoso, M. Prime Diffie-Hellman Weakness May Be Key to Breaking Crypto, threatpost.com/prime-diffie-hellman-weakness-may-be-key-to-breaking-crypt/115069/#sthash.wnLEv2 zR.dpuf.
8. Pevnev, V. Ja. Generator prostyh chisel (In Russian). Kafedra sistem informacii. Zbirnik naukovih prac', Kharkov, TOV «Shhedra sadiba pljus» Publ., 140-146 (2014)
9. Pevnev, V. Ja. The theoretical justification for methodology of constructing pseudo-primes (In Russian). In.J. Radioelektronni i komp'uterni sistemi, 6(80), 210-213 (2016)
10. Agrawal, M., Kayal N., Saxena N. PRIMES is in P J.Annals of Mathematics.,160, pp. 781–793 (2004).
11. Venturi D. Lecture Notes on Algorithmic Number Theory. Springer-Verlag, New-York, Berlin, (2009).
12. Harvey Dubner. Factorial and primorial primes In: Journal of Recreational Mathematics.19, pp. 197–203. (1987).
13. Prahar, K. Primzahlverteilung. Springer. Berlin, (1957).
14. Crandall R., Pomerance C. Prime nambers. A computational perspective Second edition Springer (2005).
15. Solovay R., V. Strassen. A fast Monte-carlo test for primality. In: SIAM J. Comput., V. 6, 84-85. (1977)
16. Jean Gallier. Notes on Public Key Cryptography And Primality.Testing Part 1: Randomized Algorithms Miller–Rabin and Solovay–Strassen Tests. Department of Computer and Information Science University of Pennsylvania. Philadelphia (2016)