ISSN 1995-0802, Lobachevskii Journal of Mathematics, 2018, Vol. 39, No. 7, pp. 1010–1015. © Pleiades Publishing, Ltd., 2018.

The Error Probability of the Miller–Rabin Primality Test

S. T. Ishmukhametov[1]*, R. Rubtsova[1], and N. Savelyev[1]*****

(Submitted by F. M. Ablayev)

[1]Institute of Computer Mathematics and Informational Technologies,
Kazan (Volga region) Federal University, ul. Kremlevskaya 18, Kazan, Tatarstan, 420008 Russia

Received December 6, 2017

Abstract—In our paper we give theoretical and practical estimations of the error probability in the well-known Miller–Rabin probabilistic primality test. We show that a theoretical probability of error 0.25 for a single round of the test is very overestimated and, in fact, error is diminishing with the growth of length of numbers involved by a rate limited with $\ln n/\sqrt{n}$.

DOI: 10.1134/S1995080218070132

Keywords and phrases: *Primality testing, Miller–Rabin primality test, the Rabin Theorem, probability error of primality testing.*

1. INTRODUCTION

The Miller–Rabin primality test MRT has a wide application in Cryptography to distinguish composite numbers from primes ones. Garry Miller in [1] suggested a deterministic polynomial test which was based on the unproved Riemann Hypothesis while Michael Rabin [2] refused from the use of the RH and obtained a modern version of the test which became probabilistic. The error probability α in the MRT depends on the number of iterations (called rounds) each of which is diminishing α in 4 times. More exactly, let n be an odd integer which we need to check for primality, and $n-1=2^st$, where t is odd. At each iteration an individual integer a is chosen to check if some Boolean expression $R(a,n)$ holds:

$$R(a,n):\ n\text{mod}a\neq 0\ \&\ a^t\text{mod}n=1,\ \text{or},\ a^{t2^i}\equiv -1\text{mod}n,\quad 0\leq i<s.$$

This a is called *the base of iteration*. If after k rounds with different bases $a_1,a_2,\ldots,a_k$ all counted values $R(a_1,n),R(a_2,n),\ldots,R(a_k,n)$ are true, then n is called probable prime (that is, prime with a possibility of small error not exceeding $1/4^k$). But if a base a is found such that $R(a,n)$ is false, then the testing n is definitely composite. The MRT improved some previous known primality test of Fermat and Solovay–Strassen [3, 4] but contrary to the last it was able to check correctly Carmichael integers that were composite but defined by error as prime [5]. Our investigation concerns the possible probability of the MRT errors. Practical experiments show a diminishing number of false application of the MRT for larger and larger integers involved. The real probability of error is much lesser that 0,25 in a single iteration and this probability is diminishing with the growth of considered integers. In our paper we show that the real probability of composite n to be defined as prime is less that 10^{-6} when $n>10^9$. This allows us to reduce the number of rounds required to successfully separate composite numbers from primes at an essentially lesser number of rounds.

The latter plays an important role in connection with the grow length of primes used in Cryptographical Protocols like as the RSA Ciphering Algorithm [6] and Elliptic Curves Algorithm [7, 8].

*E-mail: Shamil.Ishmukhametov@kpfu.ru

**E-mail: Ramilya.Rubtsova@kpfu.ru

***E-mail: savelyevno@gmail.com

Our investigation closely interacts with the study of strong pseudoprime integers that are exactly those integers that are composite but pass successfully the MRT at a fixed set of bases containing first prime integers [10–12]. Let ψ_k be a least odd composite integer for which requirements $R(2,n), R(3,n), \ldots, R(p_k,n)$ hold. The study of this sequence allows the cryptographers to transform a probabilistic test into a deterministic one. Indeed, if a tested number n does not exceed ψ_k for some k, then it is sufficient to perform only k iterations with bases $2, 3, 5, \ldots, p_k$. Then if n passes all checks successfully, then it can be composite only it is equal to ψ_k. The values of already calculated ψ_k are growing very quickly so to find each next value is a hard computational task. The last two values ψ_k at $k = 12$ and $k = 13$ are predicted by Zang [12] and checked by Sorenson and Weber [13]. Some relevant information can be found in Ishmukhametov and Mubarakov [14].

2. NOTATIONS AND DEFINITIONS

We begin with notations and definitions. Let for a natural n $odd(n)$ denote the maximal odd factor of n and $bin(n)$ denote the maximal power of 2 dividing n. Let n be an odd natural, $n > 9$. An integer a, $1 \leq a < n$, co-prime to n, is called a *primality witness* for n if the following condition $R(a,n)$ holds:

$$a^t \mathrm{mod} n = 1, \text{ or, } a^{t2^i} \equiv -1 \mathrm{mod} n, \quad 0 \leq i < s,$$

where $t = odd(n-1)$, $s = bin(n-1)$ (we replaced original Rabin's definition of the compositeness witnesses by the opposite relation).

For generality, we consider 1 and $n-1$ as primality witnesses and we call them trivial witnesses since they satisfy MR condition $R(a,n)$ for any n. Let $W(n)$ be a set of primality witnesses of n. Rabin [1] proved the following theorem estimating number of primality witnesses:

Theorem 1. *If odd number n is prime then $|W(n)| = n - 1$. If odd $n \geq 5$ is composite, then $|W(n)| \leq \varphi(n)/4$, where $\varphi(n)$ is Euler's totient function.*

As a corollary of Rabin's theorem one can deduce that for composite n the probability of random number $a < n$, be a primality witness is less than or equal to 0.25. So in order to check if a given odd number n is composite we may repeatedly choose numbers $a, 2 \leq a < n$, and check if they are primality witnesses for n. If we succeed to find an a that is not a primality witness for n, then n is definitely composite. But if we repeated this procedure k times with different a, and all them appeared to be primality witnesses, then n is probable prime with a probability error not exceeding 4^{-k}.

3. METHODS AND ALGORITHMS

We introduce a function Fr, defined at natural numbers as $Fr(n) = |W(n)|/\varphi(n)$, which characterizes the probability error of a random number a to appear in the set of primality witnesses of n. We call it *the frequency function*. By Rabin's theorem, $Fr(p) = 1$ for prime p, and $Fr(n) \leq 0.25$ for any composite n. So, $Fr(n)$ can be considered as a measure of primality of number n.

We note that there exist infinitely many composite integers with maximal $Fr(n) = 0.25$, so we can not find an upper border B for integers n with $Fr(n) = 0.25$. But we can show that an average value $Fr(n)$ is decreasing with the growth of numbers n involved.

Below we estimate frequency of numbers of form $n = p_1p_2, \ldots, p_k$, where all p_i are different prime numbers. Beginning with $k = 2$ such numbers can have a maximum measure of primality while for n divided by a prime power this measure takes sufficiently less values. So studying only such integers we can reduce the number of experiments but preserving the upper bound for the average frequency measure valid for all n exceeding investigated bounds. For example, we can deduce that for any $n > 10^9$ the average frequency measure does not exceed the value 10^{-6}, which is essentially less than 0.25, the upper bound given by Rabin's theorem.

In our practical experiments we group considered numbers in segments of form $\sigma = [A_\sigma, B_\sigma]$. For each segment σ, we take the average ratio among all products of k different primes located in such segments. Let $S^{(k)}(\sigma) = \{p_1p_2 \ldots p_k \mid p_i < p_{i+1},\ p_1p_2 \ldots p_k \in \sigma\}$ be the set of products of k different primes locating in the considered segment σ. We wish to compute an average frequency of numbers locating in intervals σ for different $k = 2, 3, \ldots$. The average frequency of class $S^{(k)}$ is defined as a sum of frequencies of all elements in $S^{(k)}$ divided by the number of such elements:

$$Fr(S^{(k)}) = \frac{1}{|S^{(k)}|} \sum_{n \in S^{(k)}} Fr(n).$$

Segments are defined as follows: $\sigma_0 = [1; 2^7)$, $\sigma_i = [2^i; 2^{i+1})$, $i \geq 7$. Due to such definition, length of segments grows by exponent, and, as we show further, it is convenience to present data of average frequency.

In order to find frequency of a product of k primes we use a formula proved by Ishmukhametov and Mubarakov in the forthcoming paper [15].

Let n be the product of k different odd primes $p_1 p_2 \ldots p_k$, and

$$d_i = GCD\left(p_i - 1; \prod_{j \neq i} p_j - 1\right), \quad u_i = odd(d_i), \quad s_i = bin(d_i), \quad i = 1, 2, \ldots, k.$$

Then, the number of all witnesses of n is calculated as $|W(n)| = u_1 u_2 \ldots u_k \left(1 + \sum_{i=0}^{s-1} 2^{ik}\right)$.

4. EXPERIMENTAL CALCULATIONS

We performed two series of experimental calculation. In the first series we counted an exact average frequency for segments σ_i and different k.

In the second series (for larger segments) we calculated average frequency for a representative sampling of integers which was chosen so large that an additional adding of new composites into the sampling does not change first significant decimal digits. So we can ensure that our average ratings differ from exact no more than in 0.1 percent.

5. EXACT CALCULATIONS

We computed exact values of average frequencies up to $i = 28$. So the upper bound for exacts calculations was $2^{29} \approx 10^9$. Each calculation was performed with k taking values from 2 to 5.

Let us consider the case $k = 2$. At this case considered numbers have a form $n = pq$, where p and q are different primes. The common algorithm is transforming to the following ones.

Algorithm of counting $Fr(n)$ for $n = pq$

1. Compute $d = GCD(p - 1; q - 1)$ and find $u = odd(d)$ and $s = bin(d)$.
2. Compute the number of witnesses of n: $|W(n)| = u^2 \left(1 + \sum_{i=0}^{s-1} 4^i\right) = u^2 \left(4^s - 1\right)/3$.
3. Find $Fr(n)$: $Fr(n) = |W(n)|/\varphi(n) = |W(n)|/[(p - 1)(q - 1)]$.

Algorithm of counting average frequency at segment $\sigma = [A, B]$

1. Define variables s and *count* setting their values to 0.
2. Arrange a double cycle over set of odd prime integers $P = \{3, 5, 7, \ldots\}$ considering $n = pq \in \sigma$.
3. For each pair of primes $\langle p, q \rangle$ compute $Fr(n)$ and add it to s. Set $count = count + 1$.
4. After the cycle finished compute average frequency as $Fr^{\sigma}_{avg} = s/count$.

A similar procedure works for other k. We gathered all obtained results at Fig. 1. It can be noted that the average frequency is diminishing by a line low. For each k the frequency line corresponding to $(k + 1)$-tuples is located below the k-tuple line. That means that when integer n has many factors its primality measure $Fr(n)$ takes a smaller value.

The latter implies that if we count an average frequency only for semiprimes $n = pq$ located in a segment σ, this value can serve as an upper bound for the average frequency for all odd numbers in the considered segment.

6. APPROXIMATE CALCULATIONS

We continue our investigation by approximate estimations of the average frequency. At this section we continue our investigation up to 10^{15} and give a prognosis up to 10^{30} and further. Remind that prime integers used in Cryptography begins with length 160 bits for the Elliptic Curves which equivalent to about 10^{50}. So our prognosis is close to the below boundary of the Cryptography integers.

Nevertheless, due to fact that the average frequency of all numbers in the interval is much less that the average frequency of semiprimes in the same interval, our upper boundary remains valid to all prime numbers used in Cryptography.

In our approximate experiments we step by step refine previous estimations for a chosen segment till an extra addition of integers preserve first three significant digits, then we stop the calculation and print the result (see Fig. 2 and Fig. 3).

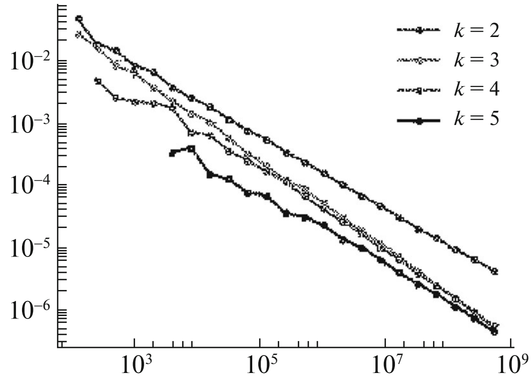

Fig. 1. Average frequency for $n = p_1p_2 \dots p_k$.

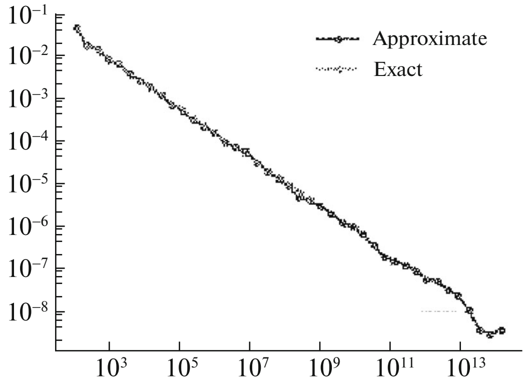

Fig. 2. Case $k = 2$.

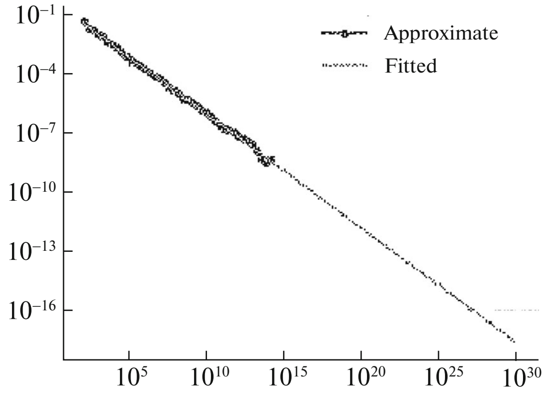

Fig. 3. Prognosis at $k = 2$.

In the next section we give some theoretical basics on these prognosis values.

7. THEORETICAL ESTIMATIONS FOR AN AVERAGE FREQUENCY

Let B be some integer. We give estimations for an upper boundary for the average frequency of semiprime integers in interval $[1; B]$. A distribution law of such integers was studied in [17].

Theorem 2. *The average frequency at interval* $[1; B]$ *of semiprimes* $n = pq$ *tends to 0 when* B *is tending to infinity with speed bounded by*

$$B^{-1/2} \ln B. \tag{1}$$

Proof. Let B be given. We divide set $[1; B]$ into the parts M_p consisting of numbers $n = pq$, $p < q < B/p$.

1. First we consider $p = 3$ and q, $p < q < B/3$. Then $d = GCD(p-1; q-1) = 2$. The number of witnesses of $n = 3q$ does not depend on q and equal to 2. The number of such pair is equal to $\pi(B/3) - 2$, where $\pi(x)$ is the prime-counting function. So, average frequency of numbers $3q < B$ can be estimated as

$$Fr(p=3) \approx \frac{1}{\pi(B/3)} \sum_{5 \le q < B/3} \frac{2}{\varphi(3q)} = \frac{1}{\pi(B/3)} \sum_{5 \le q < B/3} \frac{1}{q-1}$$
$$\approx \frac{3\ln(B/3)}{B} \left(\ln\ln(B/3) - 1/2 - 1/3\right) < \frac{3 \ln B \ln\ln B}{B}.$$

We used here approximate formulas ([18], p. 352): $\pi(x) \approx x/\ln x$, $\sum_{q<x} q^{-1} \approx \ln\ln x$. Clearly, $\lim_{B\to\infty} Fr(p=3) = 0$.

2. Let $p = 5$, and $7 \le q < B/5$. The GCD of $p-1$ and $q-1$ is equal either 2/ or 4, depending on if $q \equiv 3 \mathrm{mod} 4$ is valid. In the first case the number of witnesses is equal to two and in the second case to six. We can assume that the frequency of both cases are equal, so the average number of witnesses is four. The number of $n = 5q < B$ is $\pi(B/5) - 3$. The average frequency can be evaluated now as above with replacement everywhere 3 by 5:

$$Fr(p=5) \approx \frac{1}{\pi(B/5)} \sum_{7 \le q < B/5} \frac{4}{\varphi(5q)} = \frac{1}{\pi(B/5)} \sum_{7 \le q < B/5} \frac{1}{q-1}$$
$$\approx \frac{5\ln(B/5)}{B} \left(\ln\ln(B/5) - \frac{1}{2} - \frac{1}{3} - \frac{1}{5}\right).$$

3. For arbitrary $p < \sqrt{B}$ we have the formula

$$Fr(p) = \frac{p w_p}{\pi(B)} \left(\ln\ln(B/p) - \frac{1}{2} - \frac{1}{3} - \ldots - \frac{1}{p}\right) \approx \frac{p w_p}{\pi(B)} \left(\ln\ln(B/p) - \ln\ln p\right),$$

where w_p is average frequency in the set $\{pq_i \mid p < q_i < B/p\}$. Most part of q_i are co-prime to p, and $|W(pq_i)| = 2$. Function $|W(pq_i)|$ takes its maximum at $q_i = k(p-1)+1$, $k \ge 2$, but even at such integers the average frequency is diminishing when $B \to \infty$. Indeed, let H_p be set

$$H_p = \{p(k(p-1)+1) \mid 2 \le k < \sqrt{B/k}\}.$$

Then,

$$Fr_{avg}(H_p) = \frac{1}{\sqrt{B/p}} \sum_{q=k(p-1)+1} \frac{|W(pq)|}{\varphi(pq)} = \frac{1}{\sqrt{B/p}} \sum_{q=k(p-1)+1} \frac{p-1}{k(p-1)^2}$$
$$< \frac{1}{\sqrt{B/p}(p-1)} \sum_{q=k(p-1)+1} \frac{1}{k} < \frac{\ln\sqrt{B/p} + \gamma}{\sqrt{B/p}(p-1)} \to 0 \text{ at } B \to \infty$$

(we used a formula for the partial sum of Harmonic Series $\sum_{k<=x} 1/k = \ln x + \gamma + \varepsilon$).

Table

k	3	4	5	6
$Fr(k)$	0.0247	0.0064	0.0017	0.0004

We see that the average frequency $Fr(p)$ has a limit$=0$ at $B\to\infty$. Since, frequency of other numbers of form pq (fixed p) have lesser number of witnesses, then function $Fr(p)$ is bounded by the function $Fr_{avg}(H_p)$. This completes the proof of the theorem.

In the next table we presented experimental values of the average values of $Fr(n)$ counted at intervals $[1, 10^k]$ for $k = 3, \ldots, 6$:

By our theoretical estimate (1), we wait for a reduction rate bounded approximately by $\sqrt{10} \approx 3.16$ at each next k. Real values show even larger reduction an the rate equal approximately to 4.

8. FINAL SECTION

In our paper we gave theoretical and practical reasons of the need of improving the Miller–Rabin primality test. Estimations of diminishing of the probability errors of MRT lead us to more effective algorithms of primality testing. This is an important for Cryptography and its applications to the Information Security.

ACKNOWLEDGMENTS

The research was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, project no. 1.12878.2018/12.1.

REFERENCES

1. G. L. Miller, "Riemann's hypothesis and tests for primality," J. Comput. Syst. Sci. **13**, 300–317 (1976).
2. M. O. Rabin, "Probabilistic algorithm for testing primality," J. Numb. Theory **12**, 128–138 (1980).
3. R. M. Solovay and V. Strassen, "A fast Monte-Carlo test for primality," SIAM J. Comput. **6**, 84–85 (1977).
4. R. M. Solovay and V. Strassen, "V. Erratum: A fast Monte-Carlo test for primality," SIAM J. Comput. **7**, 118 (1978).
5. R. Baillie and S. S. Wagstaff, Jr., "Lucas pseudoprimes,"Math. Comp. **35** (152), 1391–1417 (1980).
6. R. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," Commun. ACM **21**, 120–126 (1978).
7. N. Ferguson and B. Schneier, *Practical Cryptography* (Wiley, Chichester, 2003).
8. S. Ishmukhametov, *Factorization Methods for Natural Numbers* (Kazan Federal Univ., Kazan, Russia, 2011) [in Russian].
9. C. Pomerance, J. L. Selfridg, and S. Wagstaff, Jr., "The pseudoprimes to $25 \cdot 10^9$," Math. Comp. **35** (151), 1003–1026 (1980).
10. G. Jaeschke, "On strong pseudoprimes to several bases," Math. Comp. **61** (204), 915–926 (1993).
11. J. Yupeng and D. Yingpu, "Strong pseudoprimes to the first eight prime bases," Math. Comp. **83** (290), 2915–2924 (2014).
12. Z. Zhang, "Two kinds of strong pseudoprimes up to 10^{36}," Math. Comp. **76** (260), 2095–2107 (2007).
13. J. Sorenson and J. Webster, "Strong pseudoprimes to twelve prime bases," arXiv:1509.00864 [math.NT] (2015).
14. S. Ishmukhametov and B. Mubarakov, "On practical aspects of the Miller–Rabin primality test," Lobachevskii J. Math. **34**, 304–312 (2013).
15. S. Ishmukhametov and B. Mubarakov, "On the number of witnesses in the Miller–Rabin primality test," (to be published).
16. `https://en.wikipedia.org/wiki/Coprime_integers`.
17. S. Ishmukhametov and F. F. Sharifullina, *On distribution of semiprime numbers*, Russ. Math. (Iz. VUZ) **58** (8), 43–48 (2014).
18. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers,* 4th ed. (Clarendon, Oxford, 1959).