Известия вузов. Математика
2014, №8, с. 53–59

http://old.kpfu.ru/journals/izv_vuz/
e-mail: izvuz.matem@kpfu.ru

Ш.Т. ИШМУХАМЕТОВ, Ф.Ф. ШАРИФУЛЛИНА

О РАСПРЕДЕЛЕНИИ ПОЛУПРОСТЫХ ЧИСЕЛ

Аннотация. Натуральное число называется полупростым, если оно является произведением двух простых чисел (возможно равных между собой). Обозначим через $g(y)$ вероятность того, что случайно выбранное натуральное числа y является полупростым. В этой статье выведем асимптотическую формулу для вычисления $g(y)$ и оценим ее точность при различных значениях y. Также вводим понятие сильно полупростого числа как произведения двух простых чисел большой размерности и изучим распределение сильно полупростых чисел.

Ключевые слова: полупростые числа, сильно полупростые числа, распределение полупростых чисел, факторизация целых чисел, метод шифрования RSA.

УДК: 519.711

Под *гладкостью* натурального числа n будем понимать возможность разложения этого числа в произведение большого числа простых сомножителей. *B-гладкое* число — это число, все простые делители которого ограничены сверху константой B. Понятие гладкости играет важную роль в теории чисел и криптографии.

Возможность использования этого понятия в криптографии основана на том, что процедура разложения целого числа на множители (факторизация) является трудоемкой вычислительной процедурой, требующей использования значительных вычислительных ресурсов [1], [2]. При разложении B-гладкого числа время разложения зависит не от величины самого числа, а от значения B. Чем больше значение этой константы, тем больше времени требуется для разложения числа. Известный метод двухключевой криптографии RSA основан на трудноразрешимости задачи факторизации, и ключом шифрования RSA является большое полупростое число. Поэтому знание распределения чисел различной степени гладкости является важной задачей теории чисел и ее приложений в криптографии.

Введем здесь понятие *сильной полупростоты*, определяя полупростое число n сильно полупростым, если оба сомножителя n являются большими.

Определение. Число n называется *сильно полупростым,* если n является произведением двух возможно равных простых чисел p и q, каждое из которых больше $n^{1/4}$.

В рамках определения гладкости простые и полупростые числа являются антиподами гладких чисел. Отметим, что для криптографии в первую очередь важны сильно полупростые числа, в окрестности которых нет гладких чисел, иначе проблема факторизации может быть решена значительно быстрее при использовании методов типа алгоритм факторизации Ленстры на эллиптических кривых и $(p-1)$-метода Полларда ([2], сс. 87, 133), поэтому необходимо знать распределение не только полупростых, но и гладких чисел [3].

Поступила 31.01.2013

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект №10-01-00399-а).

Основной проблемой, рассмотренной в этой статье, является изучение распределения полупростых и сильно полупростых чисел.

1. Точная формула для $\pi_2(y)$

Пусть $P = \{2 < 3 < \cdots < p_k < \cdots\}$ — упорядоченное множество всех простых чисел, а $\pi(x)$ — количество простых чисел, меньших или равных x. По теореме Чебышева ([4], с. 87)

$$\lim_{x\to\infty} \frac{\pi(x)\cdot \ln x}{x} = 1,$$

поэтому для больших x можно принять аппроксимацию $\pi(x) \approx \frac{x}{\ln x}$.

Уточнить формулу не представляется возможным, поскольку оценка погрешности этой формулы эквивалентна решению знаменитой проблемы Римана о распределении нулей ζ-функции [5].

Для действительного аргумента x определим значение $\lfloor x \rfloor$ равным ближайшему к x целому числу с меньшей стороны. Число полупростых чисел, меньших y, может быть точно вычислено по формуле

$$\pi_2(y) = \sum_{k=1}^{\sqrt{y}} \left\lfloor \pi\left(\frac{y}{p_k}\right) - k + 1 \right\rfloor,$$

полученной И. Ноэлом и Г. Паносом в 2005 г. и независимо Р. Вилсоном в 2006 г. [6]. Очевидно, эта формула не может быть использована для оценки числа полупростых чисел при больших y.

Определим две вероятностные функции $g(y)$ и $g^*(y)$, полагая их значения для произвольного натурального y равными вероятности того, что число y окажется полупростым и сильно полупростым. В этой статье выполним аналитический вывод асимптотических формул для функций $g(y)$ и $g^*(y)$ и проверим их экспериментально при $y \le 10^{10}$.

2. Оценки для $g(y)$ и $g^*(y)$

Пусть число y выбрано и зафиксировано, и $p \le \lfloor \sqrt{y} \rfloor$ — произвольное простое число. Рассмотрим действительное число $v = y/p$.

Это число является целым с вероятностью $1/p$ и простым с вероятностью

$$g_p(y) \approx 1/p \ln(y/p) = 1/p(\ln y - \ln p),$$

т. е. функция $g_p(y)$ равна вероятности того, что y окажется полупростым, один из сомножителей которого равен p.

Поэтому $g(y)$ можно аппроксимировать суммой ряда

$$g(y) \approx \sum_{p\le\sqrt{y}} \frac{1}{p(\ln y - \ln p)}, \tag{1}$$

где суммирование ведется по простым p.

Соответственно функцию $g^*(y)$ можно аппроксимировать рядом

$$g^*(y) \approx \sum_{y^{1/4}<p\le y^{1/2}} \frac{1}{p(\ln y - \ln p)}, \tag{2}$$

где суммирование также ведется по простым p.

3. Проверка формулы (1)

Сначала необходимо проверить точность формул (1) и (2). Для этого сравним значения функций $g(y)$ и $g^*(y)$, вычисленных по этим формулам, с относительной частотой распределения полупростых (сильно полупростых) чисел, полученной как отношение числа полупростых (сильно полупростых чисел) $\pi_2(y, L)$ и $\pi_2^*(y, L)$ на интервале $[y - L; y + L - 1]$ к длине этого интервала $2L$. Зададим значение $L = 100$ при $y = 1000$ и $L = 1000$ при $y = 10^k$, $k = 3, 4, \ldots, 10$. Поместим полученные значения и значения функции $g(y)$, вычисленной по формуле (1), в табл. 1.

Таблица 1

y	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}
$g(y)$	0.301	0.253	0.225	0.201	0.183	0.168	0.155	0.145
$\pi_2(y, L)/2L$	0.279	0.248	0.230	0.202	0.189	0.179	0.160	0.147
$g^*(y)$	0.125	0.107	0.086	0.073	0.066	0.059	0.052	0.0470
$\pi_2^*(y, L)/2L$	0.08	0.12	0.096	0.081	0.077	0.070	0.060	0.0473

Из табл. 1 видно, что $g(y)$ довольно точно приближает частоту случайного события "y — полупростое число". Расхождение функции $g^*(y)$ с точным значением было более значительным, поэтому для получения более точной аппроксимации увеличили длину интервала L до 2000. Теперь на рассматриваемом интервале при увеличении y отношение функции $g^*(y)$ к ее аппроксимации $\pi_2^*(y, L)/2L$ стремится к единице.

4. Проверка формул Мертенса

Для расчета функции $g(y)$ понадобится одно уточнение формулы Мертенса [4]. Пусть p_n — n-е простое число, $a_n = \ln p_n/p_n$ и $A(x)$ — частичная сумма ряда, задаваемого формулой

$$A(x) = \sum_{p<x} a_n = \sum_{p<x} \frac{\ln p}{p}, \tag{3}$$

где суммирование ведется по простым числам. По формуле Мертенса

$$A(x) = \ln x + B + O\left(\frac{1}{\ln x}\right),$$

где B – некоторая постоянная. В статье И. Шпарлинского [7] постоянная B принимается равной $1.33258\ldots$, но обнаружили, что в расчетах этой формулы перепутан знак — константа B должна быть отрицательной [8]: $\lim_{x\to\infty} A(x) - \ln x = -1.33258\ldots$.

Экспериментальным образом нами была предложена эмпирическая формула, уточняющая формулу Мертенса на интервале до 10^{10}:

$$A^*(x) = \ln x + \frac{1}{\ln x} - 1.34, \tag{4}$$

которая с большой точностью представляет сумму ряда (3) на интервале до 10^{10}. В табл.2 приведены значения функций $A(x)$ и $A^*(x)$, вычисленных по формулам (2) и (4).

Таблица 2

x	10	10^2	10^3	10^4	10^5	10^6	10^7	10^8	10^9
$A(x)$	1.3127	3.3695	5.6095	7.8909	10.1843	12.4836	14.7857	17.0882	19.3907
$A^*(x)$	1.3969	3.4823	5.7125	7.9789	10.2597	12.5479	14.8401	17.1350	19.4315
$A(x)/A^*(x)$	0.9397	0.9676	0.9820	0.9890	0.9926	0.9949	0.9963	0.9973	0.9979

Мы не смогли вычислить значение $A(x)$ для $x = 10^{10}$, так как это требует вычислений с точностью, превышающей точность стандартного формата данных double, однако для последнего найденного значения $A(4.3 * 10^9)$ отношение $A(x)/A^*(x)$ составило 0.9981, что подтверждает вывод о высокой точности аппроксимации функции Мертенса по крайней мере на интервале до 10^{10}.

5. Вывод выражения для $g(y)$

В этом разделе получим аналитическое выражение для $g(y)$. Понадобится

Теорема Абеля ([4], с. 106). *Пусть $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n \le \cdots$ — последовательность вещественных чисел такая, что $\lim \lambda_n = \infty$ при $n \to \infty$, и $\{a_n\}$, $n = 1, 2, \ldots,$ — последовательность комплексных чисел. Пусть $A(x) = \sum\limits_{\lambda_n < x} a_n$, а $\varphi(x)$ — комплекснозначная функция, определенная для $x \ge 0$, с непрерывной производной при $x > 0$. Тогда*

$$\sum_{n=1}^{k} a_n \varphi(\lambda_n) = A(x)\varphi(x) - \int_{\lambda_1}^{x} A(t)\varphi'(t)dt,$$

где $\lambda_k \le x < \lambda_{k+1}$.

Для фиксированного y определим функцию $\varphi(y,t)$ аргумента t, удовлетворяющую требованиям теоремы Абеля на интервале $2 \le t \le x$, где $x = \sqrt{y}$,

$$\varphi(y,t) = \frac{1}{\ln t(\ln y - \ln t)} = \frac{1}{\ln y}\left(\frac{1}{\ln t} + \frac{1}{\ln y - \ln t}\right).$$

Пусть $a_n = \ln p_n / p_n$, тогда

$$A(x) = \sum_{p_n < x} a_n = \sum_{p < x} \ln p / p, \quad a_n \varphi(y, p_n) = \frac{1}{p_n(\ln y - \ln p_n)}, \quad g(y) = \sum_{p_n < x} a_n \varphi(y, p_n),$$

и по теореме Абеля

$$g(y) = A(x)\varphi(y,x) - \int_2^x A(t)\varphi'_t(y,t)dt. \tag{5}$$

Производная $\varphi(y,t)$ по t равна

$$\varphi'_t(y,t) = \frac{1}{\ln y}\left(\frac{1}{\ln t} + \frac{1}{\ln y - \ln t}\right)' = \frac{1}{\ln y}\left(-\frac{1}{t\ln^2 t} + \frac{1}{t(\ln y - \ln t)^2}\right).$$

Для того чтобы вычислить (5), заменим функцию $A(x)$ на $A^*(x)$, используя приближение Мертенса (4):

$$g(y) \approx A^*(x)\varphi(y,x) - \int_2^x A^*(t)\varphi'_t(y,t)dt. \tag{6}$$

Первое слагаемое имеет вид

$$A^*(x)\varphi(y,x) = \frac{\ln x + C + 1/\ln x}{\ln y(\ln y - \ln x)},$$

где $C = -1.34$ есть константа Мертенса.

При $x = \sqrt{y}$ имеем $\ln x = \ln y/2$, и

$$A^*(\sqrt{y})\varphi(y,\sqrt{y}) = \frac{1}{\ln y} + \frac{2C}{\ln^2 y} + \frac{4}{\ln^3 y} = \frac{1}{\ln y} + O\left(\frac{1}{\ln^2 y}\right).$$

Рассмотрим интеграл в правой части (6)

$$I = \int_a^x A^*(t)\varphi'(y,t)dt = \frac{1}{\ln y}\int_2^x \left(-\frac{1}{t\ln t} + \frac{\ln t}{t(\ln y - \ln t)^2}\right)dt + C\int_2^x \varphi'(y,t)dt + \int_2^x \frac{\varphi'(y,t)}{\ln t}dt.$$

Подстановка $z = \ln t$ в первый из трех интегралов даст

$$I_1 = \frac{1}{\ln y}\int_{\ln 2}^{\ln x}\left(-\frac{1}{z} + \frac{z}{(\ln y - z)^2}\right)dz = \frac{1}{\ln y}\left((-\ln z)\Big|_{\ln 2}^{\ln x} + \int_{\ln 2}^{\ln x}\frac{z\,dz}{(\ln y - z)^2}\right) =$$
$$= \frac{\ln\ln 2 - \ln\ln x}{\ln y} + \frac{1}{\ln y - \ln x} - \frac{1}{\ln y - \ln 2} + \frac{\ln(\ln y - \ln x) - \ln(\ln y - \ln 2)}{\ln y}.$$

Так как слагаемые $(-\ln\ln x)/\ln y$ и $\ln(\ln y - \ln x)/\ln y$ взаимно уничтожаются, то

$$I_1(x) = \frac{\ln\ln 2 - \ln(\ln y - \ln 2)}{\ln y} + \frac{1}{\ln y - \ln x} - \frac{1}{\ln y - \ln 2}. \tag{7}$$

Преобразуем слагаемое $\ln(\ln y - \ln 2)$, используя известное соотношение $\ln(1+\alpha) = \alpha + O(\alpha^2)$,

$$\ln(\ln y - \ln 2) = \ln\left(\ln y\left(1 - \frac{\ln 2}{\ln y}\right)\right) = \ln\ln y - \frac{\ln 2}{\ln y} + O\left(\frac{1}{\ln^2 y}\right).$$

Заменяя в (7) x на $y^{1/2}$, получим окончательное выражение

$$I_1(\sqrt{y}) = \frac{\ln\ln 2 - \ln\ln y + 2\ln 2/\ln y + O(1/\ln^2 y)}{\ln y} - \frac{1}{\ln y - \ln 2} = -\frac{\ln\ln y}{\ln y} + O\left(\frac{1}{\ln y}\right).$$

Соответственно второй интеграл, полученный из (6), равен

$$I_2(x) = C\int_2^x \varphi'(y,t)dt = C(\varphi(y,x) - \varphi(y,2)) = \frac{C}{\ln y(\ln y - \ln x)} - \frac{C}{\ln y(\ln y - \ln 2)} =$$
$$= \frac{2C}{\ln^2 y} - \frac{C}{\ln y(\ln y - \ln 2)} = O\left(\frac{1}{\ln^2 y}\right).$$

Наконец,

$$I_3(x) = \frac{1}{\ln y}\int_{\ln 2}^{\ln x}\left(-\frac{1}{z^2} + \frac{1}{(\ln y - z)^2}\right)dz = \frac{1}{\ln y}\left(\frac{1}{\ln x} - \frac{1}{\ln 2} + \frac{1}{\ln x} - \frac{1}{\ln y - \ln 2}\right).$$

При $x = \sqrt{y}$ получим $I_3(\sqrt{y}) = \frac{1}{\ln y}\left(\frac{4}{\ln y} - \frac{1}{\ln 2} - \frac{1}{\ln y - \ln 2}\right) = O\left(\frac{1}{\ln y}\right)$.

Сумма трех интегралов равна

$$I_1 + I_2 + I_3 = -\frac{\ln\ln y}{\ln y} + O\left(\frac{1}{\ln y}\right).$$

Окончательно имеем

$$g(y) = A(x)\varphi(y,x) - I_1 - I_2 - I_3 = \frac{\ln\ln y}{\ln y} + \frac{C_1}{\ln y} + \frac{C_2}{\ln^2 y} \; + \; O\left(\frac{1}{\ln^3 y}\right). \tag{8}$$

Нахождение точных значений постоянных C_1 и C_2 довольно сложно, так как они зависят от точности формулы Мертенса, которая имеет большое расхождение при малых значениях y, поэтому для лучшего приближения константы C_1 и C_2 можно подобрать перебором.

Серия экспериментов дала формулу

$$\widetilde{g}(y) = \frac{\ln\ln y}{\ln y} + \frac{0.265}{\ln y} - \frac{1.540}{\ln^2 y}. \tag{9}$$

Заметим, что соответствующие преобразования для функции $g^*(y)$ дают ту же оценку (9), однако константы C_1 и C_2 будут другими:

$$\widetilde{g}(y) = \frac{\ln\ln y}{\ln y} - \frac{2.65}{\ln y} + \frac{13.7}{\ln^2 y}. \tag{10}$$

В табл. 3 приведем значения $g(y)$, $\widetilde{g}(y)$ и $g^*(y)$, найденные по формулам (1), (9) и (10), и их соотношение.

Таблица 3

y	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}
$g(y)$	0.387	0.2551	0.2252	0.2015	0.1829	0.1681	0.1555	0.1448
$\widetilde{g}(y)$	0.286	0.2517	0.2236	0.2012	0.1830	0.1680	0.1555	0.1448
Ratio	1.080	1.0135	1.0068	1.0015	0.9998	1.0003	1.0000	1.0001
$g^*(y)$	0.125	0.107	0.086	0.073	0.066	0.059	0.052	0.0470
$\widetilde{g}^*(y)$	0.183	0.1148	0.0854	0.0700	0.0608	0.0547	0.0503	0.0470
Ratio	0.682	0.9317	1.0068	1.0425	1.0856	1.0790	1.0338	1.0006

Видим, что формулы (9) и (10) достаточно точно отражают частотные функции $g(y)$ и $g^*(y)$ для $y \le 10^{10}$. Отметим также, что поскольку вероятность случайно выбранного натурального числа y быть простым равна приблизительно $1/\ln y$, значит, на интервале длины B полупростых чисел больше, чем простых, примерно в $\ln\ln B$ раз.

Литература

[1] Ишмухаметов Ш.Т. *Методы факторизации натуральных чисел* (КФУ, Казань, 2011).

[2] Бойко А.А., Зиятдинов Д.Б., Ишмухаметов Ш.Т. *Об одном подходе к проблеме факторизации натуральных чисел*, Изв. вузов. Матем., № 4, 15–22 (2011).

[3] Ишмухаметов Ш.Т., Шарифуллина Ф.Ф. *О гладкостепенных элементах*, Информационные технологии в системе социально-экономической безопасности России и ее регионов: Сб. трудов IV Всеросс. науч. конф. (Казань, 2012), с. 128–132.

[4] Чандрасегхаран К. *Введение в аналитическую теорию чисел* (Мир, М., 1974).

[5] Derbyshire J. *Prime obsession*: *Bernhard Riemann and the greatest unsolved problem in mathematics* (Joseph Henry Press, 2003).

[6] Weisstein Eric W. *Semiprime*, `http://mathworld.wolfram.com/semiprime.html`.

[7] Shparlinski I. *Anatomy of integers and cryptography*, `http://cosec.bit.uni-bonn.de/fileadmin/user_upload/teaching/08us/08us-cryptabit/shparlinksi_anatomy-of-integers-and-cryptography.pdf`.

[8] Sloane N.J.A. *Sequences* A083343, in: The on-line encyclopedia of integer sequences, `http://oeis.org//A083343`.

Ш.Т. Ишмухаметов

профессор, кафедра системного анализа и информационных технологий,
Казанский (Приволжский) федеральный университет,
ул. Кремлевская, д. 18, г. Казань, 420008, Россия,

`e-mail:` Shamil.Ishmukhametov@kpfu.ru

Ф.Ф. Шарифуллина

аспирант, кафедра системного анализа и информационных технологий,
Казанский (Приволжский) федеральный университет,
ул. Кремлевская, д. 18, г. Казань, 420008, Россия,

`e-mail:` sharifullinaff@gmail.com

Sh.T. Ishmukhametov and F.F. Sharifullina

On a distrubution of semiprime numbers

Abstract. A semiprime is a natural number which is the product of two (possibly equal) prime numbers. Let y be a natural number and $g(y)$ be the probability for a number y to be semiprime. In this paper we derive an asymptotic formula for counting $g(y)$ for large values of y and evaluate its correctness for different y. We also introduce a notion of strong semiprime as a product of two primes of large dimension and investigate a distribution of strong semiprimes.

Keywords: semiprime integer, strong semiprime, distribution of semiprimes, factorization of integers, the RSA ciphering method.

St.T. Ishmukhametov

Professor, Chair of System Analysis and Information Technologies,
Kazan (Volga Region) Federal University,
18 Kremlyovskaya str., Kazan, 420008 Russia,

`e-mail:` Shamil.Ishmukhametov@kpfu.ru

F.F. Sharifullina

Postgraduate, Chair of System Analysis and Information Technologies,
Kazan (Volga Region) Federal University,
18 Kremlyovskaya str., Kazan, 420008 Russia,

`e-mail:` sharifullinaff@gmail.com