Математика и ее приложения : ЖИМО. 2012. Вып. 1 (9). С. 107–110

УДК 512.543

С. И. Хашин[1]

Проверка эффективности методов факторизации Полларда

Ключевые слова: факторизация целых чисел, метод Полларда.

Сравнивается эффективность различных элементарных методов факторизации натуральных чисел, не превосходящих 2^{64}. Основным методом факторизации для чисел в этих пределах признается метод Полларда. $(p \pm 1)$-методы факторизации также показывают высокую эффективность в большом числе случаев и заслуживают реализации.

Key words: integer factorization, Pollard methods.

Efficiency of various elementary methods of factorization of the integers that do not exceed 2^{64} is compared. The main method for numbers admits these limits is the Pollard's method. $(p \pm 1)$-methods also show high efficiency in a large number of cases and realization deserves.

1. Введение

Проблема разложения на простые множители больших целых чисел до сих пор не имеет устраивающего всех решения и в общем виде весьма сложна [1, 5, 7]. В настоящей работе мы сравниваем эффективность различных методов факторизации применительно к числам, меньшим 2^{64}. Это потребовалось при разработке небольшой библиотеки на языке $C++$, реализующей теоретико-числовые функции для чисел типа $int64$. Библиотека используется как для учебных целей [3], так и для проверки некоторых алгоритмов при булевом сжатии файлов [2].

2. Метод последовательных делений и метод Ферма

Метод последовательных делений – наиболее простой и неэффективный метод факторизации. Однако следует отметить, что для нечетного числа n, взяв остаток от деления на $3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 = 15015$, мы можем сразу проверить делимость на числа $3, 5, 7, 11, 13$. Таким образом, за одно деление мы можем найти простой делитель у примерно 81% всех натуральных чисел. Этот шаг можно рассматривать как эффективную начальную проверку, после которой остаются только числа, не имеющие делителей, меньших 17.

[1]Ивановский государственный университет; E-mail: khash2@mail.ru. Работа выполнена при поддержке гранта РФФИ 10-07-00350а

Метод Ферма [1] также весьма прост и также неэффективен. Для чисел в рассматриваемых пределах (до 2^{64}) может потребоваться несколько миллиардов проверок, что совершенно недопустимо. Конечно, в отдельных случаях метод может очень быстро наткнуться на простой делитель, но вероятность этого крайне мала. Фактически нет ни одного аргумента в пользу реализации этого метода, кроме разве что учебных целей.

3. ρ-метод факторизации Полларда

Идея метода [1, 6, 5] заключается в следующем. Пусть n – натуральное число, которое мы хотим разложить на множители. Сначала выбираем некоторый многочлен степени не ниже 2, например, $F(x) = x^2 + 1$ и рассматриваем последовательность вычетов $\mod n$

$$x_0 = 0, x_1 = 1, \ldots, x_{n+1} = F(x_n)\,.$$

Если p – простой делитель n, то после приведения последовательности $\{x_n\}$ по модулю p она начнет повторяться, причем математическое ожидание периода повторений имеет порядок $\sqrt{p}$.

В наихудшем случае, когда число n является произведением двух простых сомножителей, оно имеет простой делитель, не превышающий 2^{32}, поскольку мы рассматриваем лишь числа, меньшие 2^{64}. Таким образом, ожидаемый период повторений последовательности $\{x_n \mod p\}$ будет иметь порядок не больше 2^{16}, то есть величина, вполне доступная для перебора. Следует отметить, что рассматриваемый алгоритм в значительной степени случаен, его эффективность сильно и непредсказуемо зависит от выбора многочлена и начального элемента в последовательности.

Рассмотрим теперь алгоритм подробнее. При выбранном многочлене $F(x) = x^2 + 1 \mod n$ рассматриваем две последовательности чисел:

$$x_0 = 0, x_1 = 1, \ldots, x_{n+1} = F(x_n)\,,$$

$$y_0 = 1, y_1 = 5, \ldots, y_{n+1} = F(F(y_n))$$

и для всех n вычисляем $d = gcd(x_n, y_n)$. Если $d > 1$, то d – нетривиальный делитель n.

Эффективность алгоритма. Метод эффективен для нахождения небольших простых делителей числа n. За 87 шагов будут гарантированно найдены все делители, меньшие 1669, за 990 шагов – все делители, меньшие 81517, за 1860 шагов – меньшие 290 047, за 4961 шагов – меньшие 1 633 123, за 9568 шагов – меньшие 5 845 753, за 19546 шагов – меньшие 20 739 487, за 36286 шагов – меньшие 62 918 203. Делители большего размера тоже могут быть обнаружены, однако лишь с некоторой вероятностью. В таблице приведена вероятность (в процентах) обнаружения простого делителя с заданным количеством битов (1-й столбец) в зависимости от количества шагов (n). Подробнее: для заданного количества битов k выбиралась 1000

последовательных простых чисел, начиная с $4/3 \cdot 2^k$ и для каждого простого числа определялось, будет ли оно найдено алгоритмом в качестве простого делителя.

$nbits$	$n = 100$	$n = 1000$	$n = 2000$	$n = 5000$	$n = 10000$	$n = 20000$
16	9.3	99.9	100.0	100.0	100.0	100.0
17	5.3	97.5	100.0	100.0	100.0	100.0
18	2.4	85.6	99.9	100.0	100.0	100.0
19	1.0	65.4	97.7	100.0	100.0	100.0
20	0.8	43.5	87.5	100.0	100.0	100.0
21	0.3	24.1	67.7	99.6	100.0	100.0
22	0.0	13.3	46.5	95.5	100.0	100.0
23	0.0	7.6	25.1	81.2	99.0	100.0
24	0.0	3.3	14.1	59.5	95.2	100.0
25	0.0	1.5	6.6	37.0	81.2	99.5
26	0.0	1.3	3.7	21.7	58.9	96.0
27	0.0	0.2	1.2	9.8	33.1	81.0
28	0.0	0.0	0.3	5.7	22.8	57.5
29	0.0	0.0	0.3	2.7	11.1	39.3
30	0.0	0.0	0.4	1.8	6.6	22.5
31	0.0	0.0	0.0	0.8	2.6	10.5

Таким образом, этот алгоритм полностью перекрывает метод простого деления во всех случаях. Эксперименты показывают, что ρ-метод полностью (в пределах проведенных экспериментов, до 2^{64}) перекрывает и $(p-1)$-метод Полларда. Метод Ферма эффективен лишь в редком случае двух почти одинаковых простых делителей: если сомножители имеют размер нескольких миллиардов, их разность не должна превышать десятки миллионов. Таким образом, среди сравнительно несложных методов факторизации, ρ-метод можно считать наиболее универсальным и подходящим во всех случаях.

Выборочная проверка 5000 простых чисел, превышающих $3 \cdot 10^9$, показала, что для их обнаружения в качестве простых делителей достаточно 226 тысяч шагов ρ-метода.

4. $p \pm 1$-метод факторизации

Эти [1, 6, 4] методы эффективны в случаях, когда для некоторого простого сомножителя p числа n число $p-1$ (или $p+1$, соответственно) раскладывается на небольшие простые множители. Если s – наибольший простой множитель числа $p-1$ ($p+1$ соответственно), то для разложения этим методом требуется перебор всех простых чисел, меньших s, т. е. $\pi(s)$ шагов. На каждом шаге требуется модулярное возведение в степень по модулю n, т. е. довольно медленная операция. Для этих методов разработана так

называемая «вторая стадия», на которой для каждого последующего простого числа s требуется не возведение в степень, а лишь умножение, т. е. операция гораздо более быстрая.

Очень часто для произвольного n вполне может оказаться, что для p – одного из его простых делителей, либо $p-1$, либо $p+1$ раскладывается на небольшие простые множители. Но в худшем случае это далеко не так. Совсем не трудно подобрать простые числа p любого заданного размера так, чтобы $(p-1)/4$ и $(p+1)/6$ были простыми числами. Такими являются, например, 29, 43, 536 868 797. Таким образом, для нахождения простого делителя p этими методами в наихудшем случае потребуется

$$\pi(p/6) \approx \frac{p}{6\ln(p/6)}$$

шагов. Таким образом, этими методами гарантировано можно найти множители лишь в пределах нескольких десятков миллионов, максимум – до миллиарда.

В то же время эти алгоритмы заслуживают реализации. Возьмем в качестве условной границы число $M = 1\,000\,000$ и будем называть простое число p гладким [1, 6, 5], если одно из чисел $p \pm 1$ раскладывается на множители, меньшие M. Будем считать, что число n раскладывается на множители $(p \pm 1)$-методами Полларда, если оно имеет хотя бы один гладкий множитель.

Среди 50 847 531 простых чисел, меньших 10^9, не гладкими являются лишь 5 006 778. Поэтому доля составных чисел, разложимых $(p \pm 1)$-методами Полларда, весьма велика и оба метода можно считать достаточно эффективными.

Список литературы

1. *Ишмухаметов Ш. Т.* Методы факторизации натуральных чисел : учеб. пособие. Казань : Казанский университет, 2011. С. 53–55.
2. *Толстопятов А. А.* О возможности использования булевых уравнений для сжатия файлов // Вестник Иван. гос ун-та. 2003. Вып. 3. С. 82–84.
3. *Хашин С.И.* Small (till 2^{64}) prime numbers in C++. URL: http://math.ivanovo.ac.ru/dalgebra/Khashin/Eratosthenes/z64.html.
4. *Cohen H. A.* A Course in Computational Algebraic Number Theory. Berlin : Heildelberg ; New York ; Springer, 2000. 550 p.
5. *Crandall R. E.,Pomerance C.* Prime Numbers: a computational perspective, second edition. Springer ; New York, 2005. 597 p.
6. *Pollard J. M.* Theorems of Factorization and Primality Testing // Proceedings of the Cambridge Philosophical Society. 1974. V. 76(3). P. 521–528.
7. *Ribenboim P.* My numbers, my friends: popular lectures on number theory. Springer, 2000. 392 p.

Поступила в редакцию 26.11.2012.