УДК 519.61

Д. В. ВАСИЛЬЕВ, А. С. КУДИН

О ВЫБОРЕ ПОЛИНОМОВ ДЛЯ АЛГОРИТМА РЕШЕТА ЧИСЛОВОГО ПОЛЯ

Институт математики НАН Беларуси

(Поступила в редакцию 08.08.2014)

Введение. В настоящее время широко распространены криптосистемы, стойкость которых основана на высокой сложности задач факторизации больших целых чисел и дискретного логарифмирования в простом конечном поле. Например, такими являются криптосистема RSA [1], схемы Эль – Гамаля [2, гл. 11], схема Диффи – Хеллмана [3]. На сегодняшний день асимптотически и практически наиболее эффективными алгоритмами решения задач факторизации и дискретного логарифмирования в случае больших параметров являются различные версии алгоритма решета числового поля [4, 5, 6]. В то же время этот алгоритм является и наиболее сложным для программной реализации.

Алгоритмы решета числового поля состоят из нескольких основных трудоемких стадий (выбор полиномов, просеивание по простым числам из интервала, решение системы линейных сравнений), каждая из котоых допускает различные варианты реализаций с использованием тех или иных оптимизаций. Одной из основных стадий является стадия выбора двух или нескольких полиномов, задающих некоторое алгебраическое расширение поля $\mathbb{Q}$. Проблема оптимального выбора полиномов важна, так как более качественные полиномы могут значительно сократить время выполнения последующей стадии просеивания в алгоритме решета числового поля.

Одной из существенных проблем является вопрос, связанный с оценкой эффективности выбора полиномов. Чтобы понять насколько удачно выбраны полиномы, используются два подхода: первый заключается в вычислении теоретического рейтинга Б. Мерфи для пары полиномов, а второй – в выполнении непосредственного просеивания. Вычисление рейтинга дает довольно грубую оценку эффективности полинома. Второй подход дает точное представление об эффективности полинома, но просеивание полного интервала простых требует больших затрат времени.

В данной работе предлагается методика оценки эффективности выбора полиномов для алгоритма решета числового поля, основанная на разбиении всего интервала просеивания и частичном просеивании по коротким интервалам. На основе этой методики проведено экспериментальное исследование и сравнение эффективности различных вариантов стадии выбора полиномов, задающих алгебраические расширения, а именно метода разложения по основанию m (base-m метод) [7], методов Мерфи [7] и Кляйнюнга [8].

Методы выбора полиномов и рейтинг. Сформулируем основную задачу, решаемую на стадии выбора полиномов для алгоритмов решета числового поля. В ходе стадии выбора алгебраических расширений выбираются полиномы $f_1(x), f_2(x) \in \mathbb{Z}[x]$, обладающие следующими свойствами:

1) $f_1(x)$ и $f_2(x)$ неприводимы в $\mathbb{Z}[x]$;

2) существует такое целое $0 < m < n$, что $f_1(m) \equiv f_2(m) \equiv 0 \pmod{n}$.

В случае дискретного логарифма n заменяется на простое число p. На полиномы $f_1(x), f_2(x)$ могут накладываться дополнительные условия в зависимости от конкретной реализации алгоритма решета числового поля.

За стадией выбора алгебраических расширений следует другая трудоемкая стадия – просеивание [7, с. 6]. Пусть регион просеивания имеет вид $S = [-A, A] \times [1, B] \cap \mathbb{Z}^2$. Введем величину,

называемую скошенностью региона просеивания, равную $s = A / B$ [7, с. 77]. Обозначим однородные полиномы, соответствующие f_1 и f_2, как $F_i(x,y) = f_i\left(\frac{x}{y}\right) y^{\deg f_i},\ i = 1,2.$

Обозначим через U множество взаимно простых пар $(a,b) \in S$, для которых значения $F_1(a,b)$ и $F_2(a,b)$ являются одновременно гладкими с некоторыми границами гладкости B_1 и B_2 соответственно, выбранными заранее.

В ходе стадии просеивания нужно найти достаточно большое число пар из U. Ясно, что для ускорения стадии просеивания необходимо, чтобы значения на регионе просеивания $F_i(x,y)$ принимали как можно больше B_i-гладких значений. В настоящее время наиболее эффективно стадия просеивания реализуется с помощью метода решеточного просеивания с использованием специальных простых [5].

Рейтинги полиномов Мерфи. В работе [7] Б. Мерфи исследует вероятность появления гладких значений полинома $F_i(x,y)$ на регионе просеивания S и свойства, влияющие на нее: свойства размера и корней.

Под свойствами размера понимается величина значений, принимаемых полиномом на регионе просеивания. Влияние свойств размера на вероятность гладкости значений полинома очевидно, так как при фиксированной границе гладкости вероятность появления гладких значений полинома быстро падает с их ростом.

Под свойствами корней понимается распределение корней $F_i(x,y)$ по модулям степеней малых простых чисел. Их влияние на вероятность гладкости менее очевидно: если у $F_i(x,y)$ много корней по модулям степеней малых простых чисел, то его значения ведут себя при просеивании так, как если бы они были меньше.

Для оценки свойств размера полинома $F_i(x,y)$ на регионе просеивания $S = [-A, A] \times [1, B]$ со скошенностью $s = A / B$ Б. Мерфи вводит следующую величину:

$$I(F_i, S) = \frac{1}{2} \ln\left(\int_1^B \int_{-A}^A F_i^2(x,y)\, dx\, dy \right).$$

Скошенностью полинома F будем называть скошенность региона просеивания s, которая минимизирует $I(F,S)$ среди регионов просеивания одинаковой площади.

Для оценки свойств корня полиномов Б. Мерфи вводит величину

$$\alpha(F_i) = \sum_{p \le B_i} \left(1 - \frac{q_p p}{p+1}\right) \frac{\ln p}{p-1},$$

где q_p – число корней $F_i(x,y)$ по модулю p, т. е. q_p включает корни x / y многочлена $F(x,1)$ и проективные корни. Суммирование здесь ведется по простым, не делящим дискриминант f. Б. Мерфи высказывает и экспериментально обосновывает эвристическое предположение о том, что значения $F_i(x,y)$, имеющие порядок v, при просеивании ведут себя так же, как случайные числа порядка $ve^{\alpha(F_i)}$, и являются B_i-гладкими с той же вероятностью. Это предположение позволяет ввести для полинома F_i понятие рейтинга:

$$I(F_i, S) + \alpha(F_i).$$

Чем ниже этот рейтинг, тем выше вероятность B_i-гладкости чисел из $F_i(S)$.

Введем обозначение

$$u_{f_i}(\theta) = \frac{\ln\left| F_i\left(\sqrt{s}\cos\theta, \frac{1}{\sqrt{s}}\sin\theta\right)\right| + \alpha(F_i)}{\ln B_i}.$$

В работе [7] также вводится рейтинг пары полиномов, некоторым образом характеризующий вероятность одновременной гладкости $F_1(a,b)$ и $F_2(a,b)$ при $(a,b) \in S$:

$$E(f_1, f_2) = \sum_{k=1}^{K} \rho(u_{f_1}(\theta_k))\rho(u_{f_2}(\theta_k)),$$

где ρ – функция Дикмана, а θ_k лежат посредине $K = 1000$ равных интервалов, делящих $[0, \pi]$. Чем выше этот рейтинг, тем выше вероятность того, что $F_1(a,b)$ и $F_2(a,b)$ будут одновременно гладкими при $(a,b) \in S$.

Метод разложения по основанию m**.** В общем случае этот метод при заданных d и n строит число m и пару полиномов, удовлетворяющих свойствам 1 и 2, причем нелинейный полином f_1 – не обязательно унитарный:

$$f_1(x) = a_d x^d + a_{d-1} x^{d-1} + \ldots + a_1 x + a_0,$$
$$f_2(x) = x - m.$$

На первом шаге метода выбирается число $m \in (\sqrt[d+1]{n}, \sqrt[d]{n})$. Такой выбор m гарантирует существование разложения n из $d+1$ цифры в системе счисления с основанием m, цифры $a_d, \ldots, a_0$ которого находятся на втором шаге и полагаются равными коэффициентам многочлена f_1. Также мы могли бы разлагать по основанию m числа cn для небольших $c \in Z$, однако не ясно, какие преимущества это дает. Для полученных a_i справедливо $0 \le a_i < m$. Третьим шагом можно добиться, чтобы $|a_i| \le m/2$ при $i = 0, \ldots, d-1$. Для этого достаточно для тех $0 \le i < d$, при которых $a_i > m/2$, выполнять замену $a_i := a_i - m$ и $a_{i+1} := a_{i+1} + 1$. Эта замена сохраняет значение $f_1(m)$.

Метод разложения по основанию m имеет тот недостаток, что мы должны выбирать m как можно меньшим, если хотим получить небольшие коэффициенты f_1 и f_2. В общем случае могут существовать полиномы f_1 и f_2 с малыми коэффициентами и большим общим корнем m по модулю n, которые нельзя получить этим методом.

Метод Мерфи. Метод Мерфи, описываемый в работе [7], при заданном d и n позволяет найти пару полиномов вида

$$f_1(x) = a_d x^d + a_{d-1} x^{d-1} + \ldots + a_1 x + a_0,$$
$$f_2(x) = x - m,$$

удовлетворяющую свойствам 1 и 2 и имеющую наилучший рейтинг, путем последовательного улучшения рейтинга начальной пары полиномов, полученной методом разложения по основанию m. Предлагается улучшать рейтинг данной пары полиномов (f_1, f_2) следующими преобразованиями, применяемыми несколько раз:

1) уменьшая значение $I(F_1, S)$, варьируя скошенность региона просеивания;

2) сдвигая f_1 и f_2 на некоторый $t \in Z$, т. е. переходя к новой паре полиномов $(\overline{f}_1(x), \overline{f}_2(x)) = (f_1(x-t), f_2(x-t))$ и к новому $\overline{m} = m + t$;

3) прибавляя к $f_1(x)$ некоторое кратное $f_2(x)$, т. е. переходя к новой паре $(\overline{f}_1(x), \overline{f}_2(x)) = (f_1(x) + f_2(x)P(x), f_2(x))$ при некотором $P(x) \in \mathbb{Z}[x]$.

Первое и второе преобразования не изменяют свойства корней, но могут улучшить свойства размера. Третье преобразование может изменять и свойства корней, и свойства размера. Стоит отметить, что третье преобразование эффективно реализуется как поиск полиномов вида $f_{1,j_0,j_1}(x) = f_1(x) + (j_1 x - j_0)(x - m)$, имеющих наилучшие свойства корней, путем просеивания по j_0, j_1. На последнем шаге метода Мерфи выбирается пара полиномов, имеющая наиболее высокий рейтинг $E(f_1, f_2)$.

Метод Кляйнюнга. В работе [8] Т. Кляйнюнг предложил улучшенный и обобщенный вариант первого шага метода Мерфи (разложения по основанию m), а именно метод эффективного выбора полиномов вида

$$f_1(x) = a_d x^d + a_{d-1} x^{d-1} + \ldots + a_0,$$
$$f_2(x) = px - m, \tag{1}$$

где $p, m \in \mathbb{Z}$, $(p, m) = 1$ со следующими свойствами:

1) $F_1(m, p) = N$;

2) a_d имеет много делителей p^k для малых простых p;

3) мала норма полинома $\sup(f_1)$, определяемая выражением $\sup(f) = \min\limits_{s>0} \max\limits_{i} \left| a_i s^{i-\frac{\deg(f)}{2}} \right|$.

Важно отметить, что в работе [8] подробно описан метод Кляйнюнга только для случая, когда $\deg f_1 = 5$.

Алгоритм экспериментального сравнения эффективности выбора полиномов. В данном разделе исследуется проблема выбора метода из предложенных выше, генерирующего наилучшие пары полиномов, т. е. пары, минимизирующие время просеивания. Для выбора наилучшей пары полиномов можно пытаться использовать рейтинги, предложенные ранее, вычисление которых достаточно просто. Однако эти рейтинги имеют чисто эвристическое обоснование, а теоретическое их обоснование крайне сложно. Также они не учитывают многие аспекты поведения полиномов на регионе просеивания. Поэтому в данном разделе описывается экспериментальный метод сравнения качества пар полиномов. Стоит отметить, что его можно использовать для оценки точности различных рейтингов пар полиномов.

Мы сравнивали пары полиномов (f_1, f_2) степеней (5, 1), сгенерированные методами разложения по основанию m, Мерфи и Кляйнюнга. Так как сравниваемые пары полиномов имеют одинаковые степени (5, 1), правомерно предположить, что оптимальные размеры региона просеивания $S = [-A, A] \times [1, B] \cap \mathbb{Z}^2$, границы факторных баз B_1 и B_2, и границы больших простых $\overline{B}_1$ и $\overline{B}_2$ для них одинаковы. Также правомерно считать одинаковым количество пар из U, которое нужно найти для составления системы уравнений.

Как показывают экспериментальные данные (табл. 2), время полного просеивания при фиксированных B_1, B_2, $\overline{B}_1$, $\overline{B}_2$ практически не зависит от пары полиномов. Поэтому лучшей парой полиномов является та, которая позволит набрать необходимое число пар из U быстрее.

Для оценки числа пар из U, даваемых парой полиномов, можно было бы использовать полное просеивание. Однако для факторизации чисел размером более 100 десятичных знаков стадия просеивания занимает несколько часов на одном персональном компьютере, поэтому полное просеивание не выполнялось. Вместо этого интервал больших простых $[\overline{B}_1, B_1]$ полинома степени 5 покрывался $K + 1$ интервалами $[l_t, l_{t+1})$ одинаковой длины так, чтобы $l_0 = \overline{B}_1$, $l_K = B_1$, т. е. чтобы последний интервал начинался в точке B_1. И с помощью алгоритма 1 для каждой пары полиномов выполнялось решеточное просеивание для полинома степени 5 со специальными простыми [5] из начала каждого интервала $[l_t, l_{t+1})$, т. е. со спецальными простыми из $[l_t, l_t + h_t)$, где $h_t = L \,|\, l_{t+1} - l_t \,|$, для какого-то $0 < L < 1$.

Алгоритм 1.

1. Зная $\overline{B}_1, B_1$, L и K, вычислить координаты интервалов $[l_t, l_{t+1})$ и $h_t = L \,|\, l_{t+1} - l_t \,|$.

2. С помощью просеивателя вычислить $\hat{\chi}_t$ и χ_t – количество идеалов из факторной базы на интервалах $[l_t, l_{t+1})$ и $[l_t, l_t + h_t)$ соответственно.

3. С помощью просеивателя выполнить частичные просеивания на интервалах специальных простых $[l_t, l_t + h_t)$, измеряя δ_t – время просеивания интервала $[l_t, l_t + h_t)$ и ε_t – число пар из U на интервале $[l_t, l_t + h_t)$.

4. Зная δ_t, ε_t, χ_t, $\hat{\chi}_t$, вычислить $\bar{\delta}_t = \frac{\hat{\chi}_t}{\chi_t}\delta_t$ – экстраполированное время просеивания интервала $[l_t, l_{t+1})$ и $\bar{\varepsilon}_t = \frac{\hat{\chi}_t}{\chi_t}\varepsilon_t$ – экстраполированное число пар из U на интервале $[l_t, l_{t+1})$.

5. Далее вычислить $\overline{\varepsilon}_t = \frac{\bar{\varepsilon}_{t+1} + \bar{\varepsilon}_t}{2}$, $0 \le t < K$, – оценочное время просеивания интервала $[l_t, l_{t+1})$ и $\overline{\delta}_t = \frac{\bar{\delta}_{t+1} + \bar{\delta}_t}{2}$, $0 \le t < K$, – оценочное число пар из U на интервале $[l_t, l_{t+1})$.

В дальнейшем обозначим как $\hat{\delta}_t$ и $\hat{\varepsilon}_t$ соответственно время полного просеивания и число пар из U для интервала $[l_t, l_{t+1})$.

Результаты эксперимента. Эксперименты проводились с использованием пакетов программ с открытым исходным кодом *ggnfs* [9] и *gdlog* [10], предназначенных для факторизации и логарифмирования алгоритмом решета числового поля соответственно. Как оказалось, просеивание с двумя неунитарными полиномами f_1 и f_2 (которые генерируются методом Кляйнюнга) поддерживает только программа *gdlg_sieve* из пакета *gdlog*, поэтому для экспериментального просеивания использовалась именно она. Так как нам не нужно полное решение задачи логарифмирования, то для экспериментального просеивания было достаточно задать только простые числа p и $q \mid p-1$, как описано в работе [6].

Для заданной пары p и q алгоритм проведения эксперимента следующий.

А л г о р и т м 2.

1. Используя программы из пакета *ggnfs*, сгенерировать параметры просеивания для факторизации чисел размера p, а именно, границы факторных баз B_1 и B_2, границы больших простых $\overline{B}_1$ и $\overline{B}_2$, и границы регионов просеивания A и B.

2. Сгенерировать пару полиномов методом разложения по основанию m с помощью программы, созданной авторами.

3. Извлечь параметры *ggnfs* по умолчанию для факторизации чисел размера p и в дальнейшем использовать их при вызове программ *polyselect*, *pol*51*m*0*b* и *pol*51*opt*.

4. Сгенерировать полиномы по методу Мерфи программой *polyselect* из пакета *ggnfs*.

5. Сгенерировать полиномы по методу Кляйнюнга программами *pol*51*m*0*b* и *pol*51*opt* из пакета *ggnfs*.

6. С помощью просеивателя *gdlg_sieve* и алгоритма 1 выполнить частичное просеивание в каждом регионе $[l_t, l_{t+1})$ специальных простых полинома степени 5, и вычислить $\overline{\overline{\varepsilon}}_i$ и $\overline{\overline{\delta}}_i$.

Все эксперименты проводились при $L = \frac{1}{40}$, и $[\overline{B}_1, B_1]$ разбивался на $K = 16$ интервалов. Для оценки вносимой погрешности из-за просеивания более короткого промежутка специальных простых было выполнено полное просеивание трех промежутков специальных простых при p размером 100 десятичных знаков. Получившиеся погрешности представлены в табл. 1.

Таблица 1. **Ошибки экспериментальной оценки**

Номер региона $[l_t, l_{t+1}), t$	Относительная ошибка оценки времени просеивания, $\frac{\lvert\overline{\overline{\varepsilon}}_t - \widehat{\varepsilon}_t\rvert}{\widehat{\varepsilon}_t}$			Относительная ошибка оценки числа найденных пар из U, $\frac{\lvert\overline{\overline{\delta}}_t - \widehat{\delta}_t\rvert}{\widehat{\delta}_t}$		
	base-*m*	Мерфи	Кляйнюнг	base-*m*	Мерфи	Кляйнюнг
0	0,0273	0,0099	0,0121	0,0591	0,0274	0,0121
8	0,0047	0,0056	0,0160	0,0285	0,0022	0,0059
15	0,0104	0,0069	0,0140	0,0498	0,0384	0,0052

Из представленных данных вытекает, что выбранные $L = \frac{1}{40}$ и $K = 16$ позволяют получать адекватные оценки времени просеивания и числа найденных пар из U.

В табл. 2 преставлены зависимости оценок времени просеивания и числа найденных пар из U от области специальных простых $[l_t, l_{t+1})$ для p размером 100 десятичных знаков.

Из табл. 2 можно сделать вывод, что время просеивания незначительно зависит от метода выбора полиномов, а именно, максимальное время просеивания отличается от минимального не более чем в 1,12 раза. Небольшую разницу во времени просеивания можно объяснить дополнительными временными затратами на обработку большего числа пар из U. Видно, что метод Кляйнюнга дает значительно больше пар, чем все остальные, и в 2,2–2,5 раза больше пар, чем метод Мерфи. Очевидно, что метод разложения по основанию m дает значитаельно меньше пар, чем остальные (не менее чем в 3,7 раза), поэтому в дальнейших результатах этот метод не присутствует.

Для p размером 100, 104, 108, 112, 116, 120 десятичных знаков был проведен эксперимент, описанный выше, и в табл. 3 даны рейтинги полученных пар полиномов и границы отношения числа пар из U для метода Кляйнюнга к числу пар для метода Мерфи.

Таблица 2. **Оценки времени просеивания и числа пар из U для p размером 100 десятичных знаков**

Номер региона $[l_t, l_{t+1})$, t	Оценка времени просеивания, $\bar{\varepsilon}_t$, с			Оценка числа найденных пар из U, $\bar{\delta}_t$		
	base-m	Мерфи	Кляйнюнг	base-m	Мерфи	Кляйнюнг
0	3760	3992	4103	10618	39434	94378
1	3760	3892	4103	9008	37225	90180
2	3744	3997	4171	8807	38462	88316
3	3695	3887	4074	9056	37240	82141
4	3687	3934	4061	9169	36045	80416
5	3662	3866	4059	8310	33396	81412
6	3653	3802	3979	7421	31373	78229
7	3646	3817	4016	8062	31042	78336
8	3687	3837	3980	8304	32122	75062
9	3587	3741	3964	6971	30812	74037
10	3631	3702	3943	7280	29817	75102
11	3587	3794	3975	7491	30126	73130
12	3570	3726	3859	7399	28726	67947
13	3561	3762	3895	7306	28754	68055
14	3537	3657	3831	7110	28696	66750
15	3535	3758	3914	7069	29803	66265

Таблица 3. **Результаты эксперимента для p размером 100, 104, 108, 112, 116, 120 десятичных знаков**

Размер p	$E(f_1, f_2)$ для пары полиномов Кляйнюнга	$E(f_1, f_2)$ для пары полиномов Мерфи	Отношение рейтингов Кляйнюнг/Мерфи	Границы отношения числа пар из U Кляйнюнг/Мерфи
100	55,2856	39,1366	1,4126	2,2057–2,5236
104	44,0243	34,1833	1,2879	1,6303–1,7646
108	33,8394	22,309	1,5169	1,8115–2,0008
112	28,9876	20,9761	1,3819	1,5553–1,7810
116	24,7566	16,2829	1,5204	1,9949–2,3471

Это показывает, что полиномы метода Кляйнюнга дают приблизительно в 2 раза больше пар из U, чем полиномы метода Мерфи. Таким образом, алгоритм 1 дает методику оценки времени полного просеивания и числа пар из U, что позволяет существенно сократить время оценки качества полиномов для алгоритма решета числового поля. Для пар полиномов фиксированных степеней эта методика позволяет сравнить эффективность различных методов выбора полиномов.

Литература

1. *Rivest R., Shamir A., Adleman B.* // Communications of the ACM. 1978. Vol. 21, iss. 2.
2. *Vanstone S., Menezes A., Oorschot P. van.* Handbook of Applied Cryptography. London, 1996.
3. *Diffie W., Hellman M.* // Information Theory IEEE Transactions. 1976. Vol. 22 (2). P. 644–654.
4. *Buhler J. P., Lenstra H. W., Pomerance C.* // Springer Berlin / Heidelberg, Lecture Notes in Mathematics. 1993. Vol. 1554. P. 50–94.
5. *Pollard J. M.* // Springer Berlin / Heidelberg, Lecture Notes in Mathematics. 1993. Vol. 1554. P. 43–49.
6. *Lercier R., Joux A.* // Mathematics of computation. 1999. Vol. 72. P. 953–967.
7. *Murphy B. A.* Polynomial Selection for the Number Filed Sieve Integer Factorization Algorithm. PhD thesis. The Australian National University, 1999.
8. *Kleinjung Th.* // Math. Comp. 2006. Vol. 75. P. 2037–2047.
9. *Monico C.* GGNFS suite // sourceforge. net [Electronic resource]. 2012. Mode of access: http://sourceforge.net/projects/ggnfs/. Date of access: 17.09.2013.
10. General number filed sieve implementation // sourceforge. net [Electronic resource]. 2012. Mode of access: http://gdlog.sourceforge.net/. Date of access: 17.09.2013.

D. V. VASILYEV, A. S. KUDIN

ON THE CHOICE OF POLYNOMIALS FOR THE NUMBER FIELD SIEVE

Summary

In the paper an algorithm is suggested for quick estimate of sieving time and the number of relations for the number field sieve algorithm.