Annales Univ. Sci. Budapest., Sect. Comp. 38 (2012) 117-128

SIEVING FOR LARGE TWIN PRIMES AND CUNNINGHAM CHAINS OF LENGTH 2 OF THE SECOND KIND

T. Csajbók, G. Farkas and J. Kasza

(Budapest, Hungary)

Dedicated to Dr. Bui Minh Phong on his 60th birthday

Communicated by A. Járai

(Received November 20, 2012)

Abstract. In this paper we study the details of a triple sieving method for Cunningham chains of length 2 of the second kind and twin primes. The theoretical background is described, and also concrete computational results are published. The magnitude of the investigated numbers is 2^{253824} (more than 76 000 decimal digits).

1. Introduction

The sieving methods play an important role in prime-searching processes. Our goal is to develop effective programmes for finding large prime number combinations. Since 2005 we reached six world records, we found a twin prime pair and a Sophie Germain prime [6], [7], [9], [8], [10], [11]. Naturally "to find

The second author has been supported by the Hungarian and Vietnamese TET (grant agreement no. TET 10-1-2011-0645).

Mathematics Subject Classification: 11Y11

a large prime" means "to prove its primality". The function of the sieving method in a prime-searching project is to produce numbers prepared for a primality test. It means they do not have small prime factors and they are in a special form suitable for the probabilistic and deterministic tests. A number like this is called candidate.

A sequence of prime numbers

$$\{p, 2p + d, 4p + 3d, 8p + 7d, ..., 2^{k-1}p + (2^{k-1} - 1)d\},$$

where $k \geq 2$, is a Cunningham chain of length k of the first kind if $d = 1$ and of second kind if $d = -1$. For example $\{89, 179, 359, 719, 1439, 2879\}$ is a Cunningham chain of length 6 of the first kind and $\{2, 3, 5\}$ is a Cunningham chain of length 3 of the second kind.

In this paper we focus our attention on Cunningham chains of length 2 of the second kind. As of January 2013 the largest known prime combination has 44 652 decimal digits. Our goal was to produce candidates with more than 75 000 decimal digits of this type for which there is a good chance for being a Cunningham chain. In order to increase the effectiveness of the process we carried out so called triple sieving. This means that the sieving was performed for not only Cunningham chains but also for twin primes simultaneously. We can expect at least two world recorder Cunningham chains of length 2 of the second kind and two bronze medalist twin primes.

2. The generating of the candidates

Let us consider the following three linear polynomials with positive integer coefficients:

$$f_1(x) = (h_0 + c \cdot x) \cdot 2^e - 1,$$

$$f_2(x) = (h_0 + c \cdot x) \cdot 2^e + 1,$$

$$f_3(x) = (h_0 + c \cdot x) \cdot 2^{e+1} + 1.$$

By the help of f_1, f_2, f_3 we generate three series of positive integers from the set of candidates $\{0, 1, 2, \ldots H - 1\}$. Let us observe that if $f_2(h)$ is prime for a natural number $h > 2$, then we get a Cunningham chain of length 2 of the second kind if $f_3(h)$ is a prime, and a twin prime pair if $f_1(h)$ is a prime. Since the primality checking of $f_2(h)$ for every $h < H$ takes a very long time, we sieve the series decreasing the number of candidates.

After sieving we use the well-known Fermat primality test for the remaining candidates. If we find an integer $f_2(h)$ which is probably prime for some $h < H$, we check $f_1(h)$ and $f_3(h)$ too. If the result of Fermat test is "maybe prime" for $f_3(h)$, we have a good chance that we have found a Cunningham chain [and for $f_1(h)$ a twin prime pair]. In the case $f_2(h)$ and $f_3(h)$, the proof of primality is carried out with the Proth test. It is based on the following assertion published by a self educated French farmer Francois Proth in 1878:

Theorem 2.1. (Proth's theorem) *Let $N = k \cdot 2^n + 1$, where k odd and $2^n > k$. If there exists an integer a such that*

$$a^{(N-1)/2} = -1 \pmod{N},$$

then N is prime.

In the case $f_1(h)$ the Riesel test is used for proving primality (it is described in [5]).

2.1. The choice of the parameters

First of all we have to choose the parameters of the searching project: h_0, e, c. The magnitude of the candidates depend on the parameter e. Since we chose $e = 253824$, $h_0 = 17325$ and $c = 30030$, the candidates have at least 76 423 decimal digits. We wanted 17 to be the first sieving prime. Thus we had to use such h_0 and c for which none of the integers $f_1(h), f_2(h), f_3(h)$ is divisible by $2, 3, 5, 7, 11, 13$ for any natural number $h < H$. We can observe that $c = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13$, so $c \cdot x \equiv 0 \pmod{p}$, if p is a factor of c. The choice of h_0 is more complicated, because it has to satisfy the lucasian criteria for the primality of an integer $N = k \cdot 2^n - 1$ which is the base of the Riesel primality test. It means that there exists a sequence with the following properties:

$$v_0 \text{ is given, } \quad v_i = v_i^2 - 2 \quad \text{and}$$

$$v_{n-2} \equiv 0 \pmod{N} \quad \text{if and only if } N \text{ is prime.}$$

This test was published by H. Riesel in [5]. In order to demonstrate his assertion let us consider a $\mathbb{Q}(\sqrt{D})$ real quadratic field, where D is a squarefree natural number. Then

$$\mathbb{Q}(\sqrt{D}) = \{u + v\sqrt{D} \mid u, v \in \mathbb{Q}\}.$$

The conjugate of a number $\alpha = u + v\sqrt{D}$ is $\overline{\alpha} = u - v\sqrt{D}$, and its norm is $N(\alpha) = \alpha\overline{\alpha} = u^2 - v^2 D$. Let us denote the set of algebraic integers of $\mathbb{Q}(\sqrt{D})$ by I_D, then in case $D \equiv 2, 3 \pmod{4}$

$$I_D = \{m + n\sqrt{D} \mid m, n \in \mathbb{Z}\},$$

in case $D \equiv 1 \pmod 4$

$$I_D = \{m + n\omega \mid m, n \in \mathbb{Z}\}, \quad \omega = \frac{1+\sqrt{D}}{2}.$$

There are infinitely many units in I_D, and there exists an ε, so called fundamental unit, for which all units ζ are given by $\zeta = \pm\varepsilon^s$, for $s \in \mathbb{Z}$, i. e. the set of units is

$$U_D = \{\pm\varepsilon^m \mid m = 0, \pm 1, \pm 2, \ldots\}.$$

The absolute value of the norm of an arbitrary unit equals 1. H. Riesel proved the following assertion:

Theorem 2.2. *Suppose that $n \geq 2$, k is an odd integer, $k < 2^n$, $N = k \cdot 2^n - 1$, $r = \mid a^2 - b^2 D \mid$, where $a, b \in \mathbb{Z}$,*

$$\alpha = \frac{\left(a + b\sqrt{D}\right)}{r},$$

$$\left(\frac{D}{N}\right) = -1, \quad \textit{and}$$

$$\left(\frac{r}{N}\right) \frac{\left(a^2 - b^2 D\right)}{r} = -1,$$

where $\left(\frac{D}{N}\right)$ means the Jacobi symbol. Then N is a prime if and only if

$$v_{n-2} \equiv 0 \pmod N,$$

if $v_i = v_i^2 - 2$ with $v_0 = \alpha^k + \alpha^{-k}$.

In our prime-searching project we use $k = h_0 + c \cdot h$ and the quadratic field $\mathbb{Q}(\sqrt{13})$ for the exact primality test. With respect to the above mentioned theorem we found that the appropriate residue classes can be represented by $1, 0, 0, 0, 0, 9$ modulo $2, 3, 5, 7, 11, 13$ (more solutions are possible). Finally, we can get the value of h_0 by the Chinese remainder theorem.

2.2. The sieving method in details

As a matter of fact the sieving is a simple version of the generalized sieving with three linear polynomials. It means that if $17 \leq p \leq P_S$ is a prime and $f_1(x)$, $f_2(x)$, or $f_3(x)$ is congruent 0 modulo p, than we remove the set $\{h, h+p, h+2p, ...\}$ from $\{0, 1, 2, \ldots, H-1\}$. First we produce so called "small

primes" up to a suitable bound P_S. P_S depends on the implementation and the capacity of the available hardware. In our case $P_S = 16777213$.

We define an array B such that the remaining elements of $\{0, 1, \ldots, H-1\}$ are the indexes of B. Every value in B is initially 1. If we find that $f_1(h)$, $f_2(h)$, or $f_3(h)$ has a prime factor $17 \leq p \leq P_S$ for an $h < H$, we change the values $B[h], B[h+p], B[h+2p], \ldots, B[h+qp]$ for 0, for all $h+pq < H$. After the small prime sieving we are able to continue our computations on lower performance computers. We store the indices of these elements in an array A. Due to the sieving the size of A is significantly smaller than that of B. Therefore, we can take advantage of the cache memory and we can make the parallelization of the further sieving even with lower-capacity processors.

Now let us assume that we have a multiprocessor computer or a grid with n processors and let us consider the set of "large primes", i. e. $\{p \in \mathbb{P} \wedge P_S < p < P_L\}$, where $\mathbb{P}$ denotes the set of prime numbers and P_L is the largest sieving prime. In our case $P_L = 281537253736433 \sim 2^{48.00031915}$. Every processor gets a copy of A, a lower and an upper bound from the interval (P_S, P_L). Let us denote by A_i the copy of A received by the ith processor. Then every processor performs a sieving process from the given lower bound to upper bound producing an array filled primes, so called prime sieve table, and sieves its own copy of A with the primes found in its own prime sieve table. An $h \in A$ is removed from A, if $f_1(h)$, $f_2(h)$, or $f_3(h)$ has a prime factor $p < L_P$ in the following way.

For large primes we do not sieve directly. It means that we do not eliminate the appropriate h values from A_i, but we store them in an array A_i^* if $h < H$. As a matter of fact A_i^* includes such elements $h < H$ which will be removed later. When the cardinality of A_i^* reaches a bound ($\sim 10^6$) we sort A_i^*. Then we try to find every element of A_i^* in A_i with binary searching. If an element of A_i^* can be found in A_i, we store it in the array A_i^{**}. If the cardinality of A_i^{**} reaches a suitable bound ($\sim 10^6$), than we sort A_i^{**} and remove its elements from A_i.

If the ith processor runs out of the large primes stored in its prime sieve table, A will be renewed. It means that only such elements of A will not be removed which are present in every A_i, $1 \leq i \leq n$. Then if necessary the prime sieve table will be refilled by large primes and the sieve of the renewed A_i may restart. The larger the sieving primes are the slower the sieve is. To speed up the process we can use different sieving methods, e.g. cache optimized linear sieve [12], and the inverse sieve [13]. We keep on sieving until the it is faster than the probabilistic primality test.

3. Theoretical base

In this section we study the following questions:

I. How can we choose the value of H to have a good chance of finding twin primes or Cunningham chains, without the sieving and primality tests taking too much time?

II. How can we set the parameters of sieve to decrease the number of candidates, without the triple sieving eliminating probable twin primes or Cunningham chains?

To answer the first question let us consider the fundamental "prime number theorem". It states that

$$\pi(x) \sim \frac{x}{\ln(x)},$$

from which we get that the probability for a number to be prime $n \in \mathbb{N}$ is:

$$\frac{1}{\ln(n)}\ .$$

However, for an integer $0 \leq h < H$ the events that $f_1(h)$ and $f_2(h)$ are both primes are not independent. Therefore, the probability that $f_1(h)$ and $f_2(h)$ are twin primes is not

$$\frac{1}{\ln(f_1(h))\ln(f_2(h))}.$$

By the help of a result published in 1962 we can improve the preciseness of the above estimation. Although the Bateman-Horn conjecture is not proved, it is very useful in practice.

Conjecture (Bateman-Horn). *Let $f_1(x), f_2(x), \ldots, f_s(x)$ be irreducible polynomials, with integral coefficients and positive leading coefficients. If $\pi(r)$ denotes the number of integers $1 < n < r$, such that*

$$f_1(n), f_2(n), \ldots, f_s(n)$$

are all primes, then

$$\pi(r) \approx C_{f_1,\ldots,f_s} \cdot \frac{1}{\deg(f_1(x)) \cdots \deg(f_s(x))} \cdot \sum_{n=2}^{r} \frac{1}{(\ln(n))^s},$$

where

$$C_{f_1,\dots,f_s} = \prod_{p\in P}\left(1-\frac{w(p)}{p}\right)\cdot\left(1-\frac{1}{p}\right)^{-s}, \tag{3.1}$$

where $w(p)$ denotes the number of solutions x of the congruence

$$f_1(x)\cdots f_s(x)\equiv 0 \pmod{p}.$$

In the definition (3.1) the number $1-w(p)/p$ means the chance that none of the integers $f_1(n), f_2(n), \dots, f_s(n)$ are divisible by p and $(1-1/p)^s$ is the chance that none of the members of an s-tuple are divisible by p. Thus in our case we can estimate the probability that $f_1(h)$ and $f_2(h)$ are simultaneously primes with the following formula:

$$\frac{C_{f_1,f_2}}{\ln(f_1(h))\ln(f_2(h))}\ . \tag{3.2}$$

Since we carried out the sieving with linear polynomials, we can compute C_{f_1,f_2} easily from

$$C_s = \prod_{p>s}\frac{1-s/p}{(1-1/p)^s}.$$

Consequently, we can estimate how many twin primes and Cunningham chains of length 2 can be expected at the beginning of our project. For this let us consider the following computation, where the first sieving prime is 17. Let

$$C =$$

$$= \left(1-\frac{1}{2}\right)^{-2}\cdot\left(1-\frac{1}{3}\right)^{-2}\cdot\left(1-\frac{1}{5}\right)^{-2}\cdot\left(1-\frac{1}{7}\right)^{-2}\cdot\left(1-\frac{1}{11}\right)^{-2}\cdot\left(1-\frac{1}{13}\right)^{-2},$$

then

$$C_{f_1,f_2} = C\cdot\prod_{17\le p\in P}\frac{1-2/p}{(1-1/p)^2}.$$

From this we get

$$C_{f_1,f_2} = \left(1-\frac{1}{2}\right)^{-2}\times$$

$$\times\left(\left(1-\frac{2}{3}\right)\cdot\left(1-\frac{2}{5}\right)\cdot\left(1-\frac{2}{7}\right)\cdot\left(1-\frac{2}{11}\right)\cdot\left(1-\frac{2}{13}\right)\right)^{-1}\cdot C_2,$$

where $C_2 = 0.660161815846869573927812110014 5...$ is naturally the twin prime constant, thus

$$C_{f_1,f_2} \sim 26.69987788.$$

Then we get that the expected number $\pi_{f_1,f_2}(a,b)$ of such n's $\in [a,b)$ for which $f_1(h)$ and $f_2(h)$ are primes simultaneously

$$\pi_{f_1,f_2}(a,b) \sim C_{f_1,f_2} \cdot \int_a^b \frac{du}{\ln(f_1(h))\ln(f_2(h))}.$$

For estimating the above integral we can use the Simpson's rule since the values of f_1 and f_2 are large, so their logarithms are almost constant. Thus we obtain

$$\pi_{f_1,f_2}(a,b) \sim C_{f_1,f_2} \cdot \frac{H}{6} \cdot$$
$$\cdot \left(\frac{1}{\ln(f_1(a))\ln(f_2(a))} + \frac{4}{\ln(f_1(\frac{a+b}{2}))\ln(f_2(\frac{a+b}{2}))} + \frac{1}{\ln(f_1(b))\ln(f_2(b))} \right).$$

We can compute the expected value of the number of Cunningham chains in the same way. Simply we use f_3 instead of f_1.

Before the sieving the expected value of the number of the Cunningham chains was

$$\pi_{f_2,f_3}(0,H) \sim 26.69987788 \cdot \frac{2^{35}}{6} \cdot$$
$$\cdot (0.3230232755 + 4 \cdot 0.3229347404 + 0.3229321964) \cdot 10^{-10} .$$

Hence

$$\pi_{f_2,f_3}(0,2^{35}) \sim 29.62737432.$$

This means that at the very start every 1 159 729 445th $(H/\pi_{f_2,f_3}(0,H))$ number is probably a member of a Cunningham chain of length 2. For twin primes we received the following values:

$$\pi_{f_1,f_2}(0,2^{35}) \sim 29.62749103,$$

and

$$H/\pi_{f_1,f_2}(0,H) = 1\ 159\ 724\ 876.$$

Let us consider now question II. First of all we want to increase the twin prime and Cunningham chain density. The main problem is that if we make triple sieving we can eliminate twin primes or Cunningham chains. Therefore,

we have to investigate how many twin primes or Cunningham chains can we expect after the triple sieving.

Due the Bateman-Horn conjecture we can study the prime compressing effect of the sieving. Namely, we get from the above idea that if we use the sieve with primes $a \le p < b$, then the density of the prime s-tuples is increased by the factor

$$q^{a,b}_{f_1,\dots,f_s} = \prod_{a\le p<b} \frac{1}{1-\frac{w(p)}{p}}.$$

Naturally, the number of candidates decrease by this factor reducing the required time of the primality tests. Since we used linear polynomials, the above formula implies that, in our case, the compressing factor of the sieve is

$$q^{a,b}_{s} = \prod_{a\le p<b} \frac{1}{1-\frac{s}{p}}\ . \tag{3.3}$$

The value of the product (3.3) can be estimated in the following way: for $p < L \sim 1\ 000\ 000$ we perform the calculation accurately and for the remaining part of the product we use the approximation $(\ln(b)/\ln(L))^s$ which has a relative error below 0.1%.

If we carried out the sieve just for two primes, namely if we made double sieving for twin primes or Cunningham chains of length 2, the compression factor would be

$$q^{17,P_L}_{2} = \prod_{17\le p<P_L} \frac{1}{1-\frac{2}{p}} \sim \prod_{17\le p<L} \frac{1}{1-\frac{2}{p}} \cdot \left(\frac{\ln(P_L)}{\ln(L)}\right)^2\ . \tag{3.4}$$

In our case this value is

$$q^{17,P_L}_{2} = 131.5305598.$$

This means that after sieving, the amount of candidates will reduce from H to

$$H/q^{17,P_L}_{2} = 261\ 230\ 077,$$

i.e. every

$$H/\left(q^{17,P_L}_{2}\cdot \pi_{f_1,f_2}(0,H)\right) = 8817151$$

number would be a candidate for a twin prime, and

$$H/\left(q^{17,P_L}_{2}\cdot \pi_{f_2,f_3}(0,H)\right) = 8817186$$

for a Cunningham chain of length 2.

The compression factor of triple sieving is significantly better than that of the double one's. We can count this in the above mentioned way (3.4):

$$q_3^{17,P_L} = \prod_{17\le p<P_L} \frac{1}{1-\frac{3}{p}} \sim \prod_{17\le p<L} \frac{1}{1-\frac{3}{p}} \cdot \left(\frac{\ln(P_L)}{\ln(L)}\right)^3 . \tag{3.5}$$

Since in our case $P_L = 281537253736433$, we get that

$$q_3^{17,P_L} = 1551.774114.$$

This implies that the triple sieving reduces H to

$$H/q_3^{17,P_L} = 22\ 142\ 229,$$

i.e. every

$$H/\left(q_3^{17,P_L} \cdot \pi_{f_1,f_2}(0,H)\right) = 747\ 354$$

number is a candidate for a twin prime, and

$$H/\left(q_3^{17,P_L} \cdot \pi_{f_2,f_3}(0,H)\right) = 747\ 357$$

for a Cunningham chain of length 2.

Let us observe that the triple sieving eliminates some probable twin primes and Cunningham chains decreasing the expected value of them by its compression factor. So, a new question arises: how many prime pair will remain after sieving? Considering (3.4) and (3.5) we can get the answer by evaluating the following formulas:

$$\pi_{f_1,f_2}(0,H) \cdot \frac{q_2^{17,P_L}}{q_3^{17,P_L}} = 2.511267875\ , \tag{3.6}$$

for twin primes, and

$$\pi_{f_2,f_3}(0,H) \cdot \frac{q_2^{17,P_L}}{q_3^{17,P_L}} = 2.511257982\ , \tag{3.7}$$

for Cunningham chains of length 2.

Thus we get that after triple sieving the expected value of the number of twin primes and Cunningham chains of length 2 of the second kind is at least 2.5.

References

[1] **Indlekofer, K.-H. and Járai, A.,** Largest known twin primes and Sophie Germain primes, *Mathematics of Computation,* **68** (1999), 1317-1324.

[2] **Indlekofer, K.-H. and Járai, A.,** Largest known twin primes, *Mathematics of Computation,* **65** (1996), 427-428.

[3] **Csajbók, T., Farkas, G., Járai, A., Járai, Z. and Kasza, J.,** Report on the largest known twin primes, *Annales Univ. Sci. Budapest. Sect. Comp.,* **25** (2005), 247-248.

[4] **Csajbók, T., Farkas, G., Járai, A., Járai, Z. and Kasza, J.,** Report on the largest known Sophie Germain and twin primes, *Annales Univ. Sci. Budapest. Sect. Comp.,* **26** (2006), 181-183.

[5] **Riesel, H.,** Lucasian criteria for the primality of $N = h \cdot 2^n - 1$, *Math. Comp.,* **23** (1969), 869-875.

[6] **Csajbók, T., Farkas, G., Járai, A., Járai, Z. and Kasza, J.,** The largest known twin primes of the world $16869987339975 \cdot 2^{171960} \pm 1$ (51779 digits), `http://primes.utm.edu/top20/page.php?id=1`, 2005.

[7] **Csajbók, T., Farkas, G., Járai, A., Járai, Z. and Kasza, J.,** The largest known twin primes of the world $100314512544015 \cdot 2^{171960} \pm 1$ (51780 digits), `http://primes.utm.edu/top20/page.php?id=1`, 2006.

[8] **Csajbók, T., Farkas, G., Járai, A., Járai, Z. and Kasza, J.,** The largest known Sophie Germain prime of the world $137211941292195 \cdot 2^{171960} - 1$ (51780 digits), `http://primes.utm.edu/top20/page.php?id=2`, 2006.

[9] **Csajbók, T., Farkas, G., Járai, A., Járai, Z. and Kasza, J.,** The largest known twin primes of the world $194772106074315 \cdot 2^{171960} \pm 1$ (51780 digits), `http://primes.utm.edu/top20/page.php?id=1`, 2007.

[10] **Csajbók, T., Farkas, G., Járai, A., Járai, Z. and Kasza, J.,** The largest known Sophie Germain prime of the world $620366307356565 \cdot 2^{253824} - 1$ (76424 digits), `http://primes.utm.edu/top20/page.php?id=2`, 2009.

[11] **Csajbók, T., Farkas, G., Járai, A., Járai, Z. and Kasza, J.,** The largest known Sophie Germain prime of the world $648621027630345 \cdot 2^{253824} - 1$ (76424 digits), `http://primes.utm.edu/top20/page.php?id=2`, 2009.

[12] **Járai, A. and Vatai, E.,** Cache optimized linear sieve, *Acta Universitatis Sapientiae Informatica,* **3** (2) (2011), 205-223.

[13] **Vatai, E.,** Inverse sieve, *Annales Univ. Sci. Budapest. Sect. Comp.,* (submitted)

T. Csajbók, G. Farkas, J. Kasza
Department of Computer Algebra
Eötvös Loránd University
Pázmány Péter s. 1/C
H-1117 Budapest, Hungary
farkasg@compalg.inf.elte.hu