Annales Univ. Sci. Budapest., Sect. Comp. 40 (2013) 215–222

SIEVING FOR LARGE CUNNINGHAM CHAINS OF LENGTH 3 OF THE FIRST KIND

Gábor Farkas and Emil Vatai

(Budapest, Hungary)

Dedicated to Professors Zoltán Daróczy and Imre Kátai on their 75th anniversary

Communicated by Antal Járai

(Received April 02, 2013; accepted July 25, 2013)

Abstract. In this paper we study the details of sieving for Cunningham chains of the first kind of length 3. To find such prime triplets larger than the ones already known, we have to investigate the primality of 2^{37} numbers, each in the magnitude of 2^{34944} (more than 10 500 decimal digits). This would not be feasible if it weren't for the sieving process which reduces the estimated time of completion to only a few weeks on a grid or a supercomputer with multiple cores.

1. Introduction

The goal of our research group led by Antal Járai is to set new prime records, by developing effective computer programs which are able to find just a few probable primes and prove their primality. Usually we look for special prime combinations, e. g. twin primes, Sophie Germain primes and now we are looking for Cunningham chains. Since 2005 we set six world records, more

Key words and phrases: Cunningham chain, sieving.
2010 Mathematics Subject Classification: 11Y11.
The first author has been supported by the Hungarian and Vietnamese TET (grant agreement no. TET 10-1-2011-0645).

precisely, three by finding the largest known twin prime pairs [5, 7, 9] and three by finding largest known Sophie Germain primes [8, 10, 11].

A prime-searching project always starts with the process of sieving. We start with a large set of integers, and remove most of them by sieving out the ones with small prime factors. The remaining integers are more likely to be prime, and the primality tests are only executed on them. That is why the initial numbers are called "candidates".

A Cunningham chain of the first kind of length k is a sequence $p_1, p_2, \ldots, p_k$ of prime numbers, where $p = p_1$ and $p_{i+1} = 2p_i + 1$, for $1 \leq i < k$, that is:

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

For example $\{2, 5, 11, 23, 47\}$ is a Cunningham chain of the first kind of length 5. Let us observe that a Sophie Germain prime is the first member of a Cunningham chain of the first kind of length 2.

In this paper we focus our attention on the Cunningham chains of the first kind of length 3. In March 2013 the largest known prime combination of this kind has 10 477 decimal digits, more precisely the third element of this chain is $914546877 \cdot 2^{34774} - 1$. Our goal is to set a new world record by finding even larger primes, so we have to produce candidates with at least 10 500 decimal digits. Since we are looking for prime-triplets, we use the triple sieve method described below.

2. The generator polynomials

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

$$\begin{aligned} f_1(x) &= (h_0 + c \cdot x) \cdot 2^e - 1, \\ f_2(x) &= (h_0 + c \cdot x) \cdot 2^{e+1} - 1, \\ f_3(x) &= (h_0 + c \cdot x) \cdot 2^{e+2} - 1. \end{aligned}$$

Using f_1, f_2, f_3 we generate three series of positive integers from the set $\{0, 1, 2, \ldots H - 1\}$. Let us observe that

$$f_2 = 2f_1 + 1 \text{ and } f_3 = 2f_2 + 1 = 4f_1 + 3.$$

This means if $f_1(h)$, $f_2(h)$, $f_3(h)$ are simultaneously prime for a natural number h, than we get a Cunningham chain of length 3 of the first kind.

Since we do not want to sieve with primes 2, 3, 5, 7, 11, 13, we should have $c = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 = 30030$. Let us consider the following linear congruence:

$$f_1(x) = (h_0 + c \cdot x) \cdot 2^e - 1 \equiv 0 \pmod{p}, \tag{2.1}$$

for $p > 13$ prime. From this we get that

$$c \cdot x \cdot 2^e \equiv -h_0 \cdot 2^e + 1 \pmod{p}.$$

Since $(c \cdot 2^e, p) = 1$, the above congruence has exactly one solution, the exceptions being of course $p = 2, 3, 5, 7, 11, 13$ for which there are no solutions. We can carry out the same calculations for f_2 and f_3 too.

A potential decrease in efficiency of the sieve would be the existence of a sieving prime $p > 13$, which divides more than one of the numbers $f_1(h)$, $f_2(h)$ and $f_3(h)$ for some integer $h \in [0, H)$. It is easy to prove, that this situation can not arise, because if there exists an integer $h \in [0, H)$ that satisfies $f_1(x) \equiv 0 \pmod{p}$ and $f_2(x) \equiv 0 \pmod{p}$ simultaneously, the following would be also true:

$$p \mid f_2(h) - f_1(h) = (h_0 + c \cdot h) \cdot 2^e.$$

But (2.1) implies that $p|1$, which is a contradiction. Extending this idea to the other polynomials it can be shown, that for every prime $p > 13$, the three congruences $f_1(x) \equiv 0 \pmod{p}$, $f_2(x) \equiv 0 \pmod{p}$ and $f_3(x) \equiv 0 \pmod{p}$ always have three different solution. This observation is important because sieving with one prime can always eliminate three different candidates which are composite, thus increasing the effectiveness of the sieve.

3. Theoretical background

In this section we describe how we chose the parameters for the sieve. First of all we compute the expected value of the number of Cunningham chains. Then we determine the compression factor of the triple sieve, and this provides us with an approximate value for the number of candidates which are tested for primality.

3.1. The expected value of the number of Cunningham chains

During the calculations we use the Bateman-Horn conjecture and Riesel theorem which can be found in [13], so in this paper we do not give the full presentation of these notions.

In the last seven years we have been searching for prime pairs (twin primes and Sophie Germain primes), and now we would like to find one or more prime-triplets which are even more infrequent then prime pairs. As a result, we have to increase the initial value of H (the number of candidates) and this causes some problems for the sieving algorithm. The solution for these problems is discussed in section 4. Naturally the number of primality tests to be executed also increase. In this project we choose $H = 2^{37}$.

In accordance with the Riesel test, we have $h_0 = 5\,775$. The parameter e sets the magnitude of the candidates, so having $e = 34\ 944$ means that the numbers investigated have more than 10 500 decimal digits.

In this case we can estimate the probability of $f_1(h)$, $f_2(h)$ and $f_3(h)$ being simultaneously prime with the following formula:

$$\frac{C_{f_1,f_2,f_3}}{\ln(f_1(h))\ln(f_2(h))\ln(f_3(h))}.$$

Since we carried out the sieving with linear polynomials, we can compute C_{f_1,f_2,f_3} easily from

$$C_3 = \prod_{p>3} \frac{1-3/p}{(1-1/p)^3} \sim 0.6351664804.$$

Let us observe that if the first sieving prime is 17, then for

$$C = \prod_{17>p\in\mathbb{P}} \left(1-\frac{1}{p}\right)^{-3}$$

we get that

$$C_{f_1,f_2,f_3} = C \cdot \prod_{17\le p\in\mathbb{P}} \frac{1-3/p}{(1-1/p)^3},$$

which implies that

$$C_{f_1,f_2,f_3} = \frac{C_3}{\left(1-\frac{1}{2}\right)^3 \cdot \left(1-\frac{1}{3}\right)^3 \cdot \left(1-\frac{3}{5}\right) \cdot \left(1-\frac{3}{7}\right) \cdot \left(1-\frac{3}{11}\right) \cdot \left(1-\frac{3}{13}\right)}.$$

Finally we get that

$$C_{f_1,f_2,f_3} \sim 134.1144099.$$

Now we can compute the expected value $\pi_{f_1,f_2,f_3}(a,b)$ of the number of integers $h \in [a,b)$ for which $f_1(h)$, $f_2(h)$ and $f_3(h)$ are simultaneously prime:

$$\pi_{f_1,f_2,f_3}(a,b) \sim C_{f_1,f_2.f_3} \cdot \int_a^b \frac{du}{\ln(f_1(h))\ln(f_2(h))\ln(f_3(h))}.$$

For estimating the above integral we can use Simpson's rule because the values of $f_1(h)$, $f_2(h)$ and $f_3(h)$ are large, so their logarithms are almost constant. Let

$$g(h) = \frac{1}{\ln(f_1(h))\ln(f_2(h))\ln(f_3(h))}.$$

Then we get that

$$\pi_{f_1,f_2,f_3}(a,b) \sim C_{f_1,f_2,f_3} \cdot \frac{H}{6} \cdot \left(g(a) + 4g\left(\frac{a+b}{2}\right) + g(b) \right).$$

Substituting the concrete values we can compute the number of Cunningham chains of the first kind of length 3 we can expect.

$$\pi_{f_1,f_2,f_3}(0,H) \sim 134.1144099 \cdot \frac{2^{37}}{6} \cdot \\ \cdot (0.7029147908 + 4 \cdot 0.7006047545 + 0.7005446987) \cdot 10^{-13} \sim \\ \sim 1.292084021.$$

This means that the expected value of the number of prime-triplets found is more than one.

3.2. The compression factor of the sieve

We can compute that before sieving every j-th number is probably a member of a Cunningham chain of length 3, for

$$j = \frac{H}{\pi_{f_1,f_2,f_3}(0,H)} = 106\ 369\ 981\ 600\,.$$

Let us investigate how can the Cunningham chains density in the set of candidates be increased by sieving. For this we use the method described in [13], so we get the compression factor from the following formula:

$$q_3^{a,b} = \prod_{a \le p < b} \frac{1}{1 - \frac{3}{p}}\,. \tag{3.1}$$

We compute the exact value of the product up to the prime $L = 1\,000\,003$, and from L to the largest sieving prime ($P_L \sim 2^{48}$) we use the approximation

$$\prod_{L \le p < P_L} \frac{1}{1 - \frac{3}{p}} \sim \left(\frac{\ln(P_L)}{\ln(L)} \right)^3.$$

So we get that the compression factor of the triple sieve is

$$q_3^{17,P_L} = \prod_{17\le p<P_L} \frac{1}{1-\frac{3}{p}} \sim \prod_{17\le p<L} \frac{1}{1-\frac{3}{p}} \cdot \left(\frac{\ln(P_L)}{\ln(L)}\right)^3 . \tag{3.2}$$

After the substitution the result is

$$q_3^{17,P_L} \sim 1551.743160.$$

Thus after sieving, the number of candidates would be reduced from H to

$$\frac{H}{q_3^{17,P_L}} \sim 88\ 570\ 684,$$

and divided by the expected value calculated above, we get that after the triple sieve the number of expected candidates for Cunningham chains of the first kind of length 3 is

$$\frac{H}{q_3^{17,P_L} \cdot \pi_{f_1,f_2,f_3}(0,H)} \sim 68\ 548\ 703.$$

4. Sieveing details

The sieving of the candidates is done in two phases. As mentioned earlier we need primes up to $P_L \sim 2^{48}$ and producing these primes is the first phase of the process. This is done using the usual sieve of Eratosthenes, which produces the small primes ($p < 2^{24}$) and the large primes ($2^{24} < p < 2^{48}$). There are numerous ways of optimizing this sieve (see [12]), but it is not as important as the second phase.

The second phase is sieving the candidates, eliminating the $h \in [0, H)$ integers for which $f_i(h)$ yields a composite. Up until now, in similar projects, the sieve table containing the candidates was much smaller, and the sieving approach was to put the large primes into so called prime sieve tables, which can be generated on separate computers (nodes). The first node would sieve the candidates with the small primes, and send a list of the remaining candidates to the other nodes. Each node would sieve this list with the primes from its prime sieve table. After all nodes finished, the partial sieving results would be merged into the list of potential primes meant for the Fermat test.

But now we have a set of $H = 2^{37}$ candidates, which is represented by 128 Gbits, that is 32Gbytes. This amount of RAM can not be found in computers at our disposal and transmitting such amount of data would waste a lot of time. There were different approaches to compress the sieve table e. g. switching

from a bit table representation to an array of integer values, but this would still be quite slow due to the vast amount of candidates. However, this is the scenario where the inverse sieve [14] can be beneficial.

The absolute number of candidates left after sieving with small primes is very big, however it is quite small compared to the sieve table. In other words the sieve table becomes very sparse after sieving with small primes because of the triple sieve (and the plethora of small primes). As a direct result, the probability of a large prime eliminating a candidate which hasn't already been eliminated is very small, and this is exploited by the inverse sieve.

The idea behind the inverse sieve, is to "smear" the bits left in the sieve table, and represent these smears in a compressed bit table. Of course, it is most efficient to shrink the table by a power of two e. g. shrinking the 32Gbyte table by 16 reduces it to a 2Gbyte table which can fit in most of today's computer's RAM. Shrinking by 16 is done, by dividing the original sieve table into 16 bit segments and each segment containing at least one potential prime would be represented by 1 (this is a smear) and each segment containing only composites would be represented by 0.*

After receiving the compressed smear table the nodes can sieve with their prime sieve table, i. e. large primes. The offset in the smear table can be easily obtained by a binary shift operation. If the bit in the compressed table at that offset is 0, then the actual offset would sieve in a segment which contains only 0's, that is candidates already eliminated by small primes. These offsets can be discarded†, and using this method the only offsets remaining are ones which could (potentially) eliminate some of the candidates from the sieve table. Hopefully most of the offsets are be discarded. The remaining offsets are sent back to the first node where they can be merged with the results from the other nodes. This method is especially useful when sieving with primes $2^{37} < p < 2^{48}$, because these primes produce at most one offset in the sieve table, and if it is not near a potential prime, it can be thrown away all together.

References

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

[2] **Indlekofer, K.-H. and A. Járai,** Largest Known Twin Primes and Sophie Germain Primes, *Mathematics of Computation*, **68** (1999), 1317–1324.

*The potential primes are usually represented by 1 and the composites by 0, so it comes naturally to represent "smears" with 1 and "blank" segments with 0.

†We eliminate primes and offsets instead of candidates, hence the name "inverse sieve".

[3] **Indlekofer, K.-H. and A. Járai,** Largest Known Twin Primes, Mathematics of Computation. **65** (1996), 427–428.
[4] **Csajbók, T., G. Farkas, A. Járai, Z. Járai and J. Kasza**, Report on the largest known twin primes, *Annales Univ. Sci. Budapest, Sect. Comp.*, **25** (2005) 247–248.
[5] **Csajbók, T., G. Farkas, A. Járai, Z. Járai and J. Kasza,** 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).
[6] **Csajbók, T., G. Farkas, A. Járai, Z. Járai and J. Kasza,** Report on the largest known Sophie Germain and twin primes, *Annales Univ. Sci. Budapest, Sect. Comp.*, **26** (2006) 181–183.
[7] **Csajbók, T., G. Farkas, A. Járai, Z. Járai and J. Kasza,** 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., G. Farkas, A. Járai, Z. Járai and J. Kasza,** 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., G. Farkas, A. Járai, Z. Járai and J. Kasza,** 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., G. Farkas, A. Járai, Z. Járai and J. Kasza,** 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., G. Farkas, A. Járai, Z. Járai and J. Kasza,** 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 E. Vatai,** Cache optimized linear sieve, *Acta Universitatis Sapientiae Informatica*, **3(2)** (2011), 205–223.
[13] **Csajbók, T., G. Farkas and J. Kasza,** Sieving for Large Twin Primes and Cunningham chains of length 2 of the second kind, *Annales Univ. Sci. Budapest, Sect. Comp.*, **38** (2012), 117–128.
[14] **Vatai, E.,** Inverse sieve, *Annales Univ. Sci. Budapest, Sect. Comp.*, **40** (2013),

G. Farkas and E. Vatai
Department of Computer Algebra
Eötvös Loránd University
H-1117 Budapest
Hungary
`farkasg@compalg.inf.elte.hu`
`emil.vatai@gmail.com`