FINDING LARGE SEMI-PRIMES

It is well known that any positive integer \( N \) may be decomposed into the product of one or more prime numbers. For example \( 111 = 3 \times 37 \) and \( 6479 = 11 \times 19 \times 31 \). When the number of factors is just one term (1 and \( N \) excluded) it is a prime number and when it factors in the product of two primes it is referred to as a semi-prime. An example the number 1237 is a prime while 2501 = 41 \times 61 \) is a semi-prime. A number with multiple factors such as 85085 = 5 \times 7 \times 11 \times 13 \times 17 \) is a composite. We want here to identify large semi-primes \( N = p \cdot q \) since these play a critical role in public key cryptography.

Our starting point is our earlier defined number fraction –

\[
f = \frac{\sigma(N) - (N+1)}{N}
\]

where \( \sigma(N) \) is the sigma function of number theory. The number fraction has zero value when \( N \) is a prime and a value greater than one when \( N \) is a composite number with multiple factors. A semi-prime will have a value near zero, but not zero. Typically if \( N \) is a \( k \) digit long semi-prime one can expect the value of \( f \) to equal about \( 10^{-\frac{k}{2}} \). Furthermore, a semi-prime \( N \) must have its mod(6) operation yield a value of 1 or -1 since we know that all primes above 3 have the form \( 6n \pm 1 \). In terms of our earlier defined hexagonal integer spiral we have the following diagram for the semi-prime \( N = 77 \) showing its components \( p = 11 \) and \( q = 7 \).

Performing a \( \text{mod}(6) \) operation on \( N = p \cdot q \) produces \((-1) = (-1)(+1)\). This shows why \( p \) and \( q \) lie along different diagonals in the graph. The \( f \) value for \( N = 77 \) is \( 18/77 = 0.2337 \), and so is small but not zero.
We have run through a set of numbers \( N \) which are semi-primes and find the results shown in the following table:

<table>
<thead>
<tr>
<th>( N = pq )</th>
<th>( N \mod(6) )</th>
<th>( f )</th>
<th>( p )</th>
<th>( p \mod(6) )</th>
<th>( q )</th>
<th>( q \mod(6) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>-1</td>
<td>0.2337</td>
<td>11</td>
<td>-1</td>
<td>7</td>
<td>+1</td>
</tr>
<tr>
<td>391</td>
<td>+1</td>
<td>0.10230</td>
<td>23</td>
<td>-1</td>
<td>17</td>
<td>-1</td>
</tr>
<tr>
<td>12091</td>
<td>+1</td>
<td>0.018195</td>
<td>113</td>
<td>-1</td>
<td>107</td>
<td>-1</td>
</tr>
<tr>
<td>765469</td>
<td>+1</td>
<td>0.0026780</td>
<td>1559</td>
<td>-1</td>
<td>491</td>
<td>-1</td>
</tr>
<tr>
<td>8932479</td>
<td>+1</td>
<td>0.0008037</td>
<td>70949</td>
<td>-1</td>
<td>1259</td>
<td>-1</td>
</tr>
<tr>
<td>123456763</td>
<td>+1</td>
<td>0.000281264</td>
<td>30703</td>
<td>+1</td>
<td>4021</td>
<td>+1</td>
</tr>
<tr>
<td>3000000089</td>
<td>-1</td>
<td>0.0000471499</td>
<td>115469</td>
<td>-1</td>
<td>25981</td>
<td>+1</td>
</tr>
</tbody>
</table>

We picked these semi-primes by finding the lowest (non-zero) value of \( f \) for a given range of \( N \). Once such a value for \( f \) had been found, we next applied an `ifactor(N)` operation to get the components. Note that \( N, p \) and \( q \) always lie on either the \( 6n+1 \) or \( 6n-1 \) diagonal in the hexagonal integer plane. The \( \mod(6) \) values for \( N \) show that \( N \mod(6)=+1 \) implies that both \( p \) and \( q \) lie on the same diagonal while a \( N \mod(6)=-1 \) result says that \( p \) and \( q \) must lie on different diagonals. The values of \( f \) for these semi-primes decrease rapidly in value as the number of digits in \( N \) increases. We can estimate the value of \( f \) for a semi-prime by noting that:

\[
f(N) = \frac{p+q}{N} \approx \frac{2\sqrt{N}}{N} = \frac{2}{\sqrt{N}}
\]

since we know that \( q < \sqrt{N} < p \).

That is, the value of \( f \) for a semi-prime is approximately \( 2/\sqrt{N} \). For the ten digit number 3000000089 we have a number fraction estimate \( f \approx 0.0000365 \). This is close to the exact value of \( f \) given in the above table. If \( f \) lies much above the approximation \( 2/\sqrt{N} \) then we know we are dealing not with a semi-prime but rather with one having three or more prime factors. For example, \( N=453583 \) factors into four prime components \( 13 \times 23 \times 37 \times 41 \) and has a value \( f=0.182264 \). If this where a semi-prime then the value of \( f \) should be near \( f=0.002969 \) and not 61 times larger.

We can summarize the above observations by noting that a semi-prime \( N=p \cdot q \) has the following properties:

1. It must be of the form \( 6n+1 \) or \( 6n-1 \), where \( n=1,2,3,.. \)
2. The value of its number fraction should lie near \( f=2/\sqrt{N} \).
3. If \( N \mod(6)=1 \) than both \( p \) and \( q \) must lie along the same diagonal. If \( N \mod(6)=-1 \) then \( p \) and \( q \) lie along different diagonals in the hexagonal spiral plane.

To test things out using these criteria for a large semi-prime consider the 12 digit long number \( N=460969682477 \). It has \( N \mod(6)=-1 \) and an \( f \) estimate of \( 2/\sqrt{N}=0.000029457 \). Evaluating the actual \( f \) value yields \( f=[\text{sigma}(N)-(N+1)]/N =0.00000554281 \) and so we are close and one can conclude that \( N \) is a semi-prime. We can then proceed, using the technique discussed in earlier notes, to factor this number. It says essentially that –
\[ p = \frac{N}{2} + \sqrt{\left(\frac{N}{2}\right)^2 - N} \]

Using the above values of \( N \) and \( f \) then yields \( p = 2359603 \) and \( q = \frac{N}{p} = 195359 \), so that-

\[ 460969682477 = 2359603 \times 195359 \]

The secret to a precise evaluation is to be able to find an accurate value of \( f = \frac{(p+q)}{N} \) quickly. This will no longer be possible when \( N \) has digit lengths above 50 or so. Note here that \( p = 6(393267) + 1 \) while \( q = 6(32560) - 1 \), so that \( p \) and \( q \) do indeed lie along different diagonals as predicted by (3).

June 1, 2013