MORE ON FACTORING LARGE SEMI-PRIMES

Recently (see the earlier note on “Factoring N=pq”) we looked at a process for factoring large composite semi-primes N=pq using a non-elliptic curve factoring approach. It involved a graphical method using a new integer K whose value can be chosen to lie near sqrt(N). At the time we concluded that such an approach, although quite handy for smaller N, becomes impractical if N has 100 digit length or longer. Here we re-look at the problem from a slightly different angle using the basic definition of the tangent of the angle of a right triangle having a hypotenuse of (p+q)/2 and two legs (p-q)/2 and sqrt(N). The approach works a bit more efficiently than brute force division of N by all odd numbers up to about sqrt(N), but still requires prodigious computer efforts to factor large semi-primes.

We will show below that the factoring of N involves solving the Diophantine equation:

\[ y(x) = \frac{(K + x)^2 - N}{K + x} \]

where x=p-K and y=p-q are integer solutions with x running over the range 1 to approximately sqrt(N).

Let us indicate how we arrive at this result. One starts with the basic definition of a semi-prime:

\[ N = pq = \left(\frac{p+q}{2}\right)^2 - \left(\frac{p-q}{2}\right)^2 \text{ where } q < \sqrt{N} < p \]

One can recast this definition into a N=pq triangle as shown:
Taking the tangent, one finds that:

\[
\tan(\theta) = \frac{(p - q)}{2\sqrt{N}} = \frac{(p^2 - N)}{2p\sqrt{N}}
\]

If we now choose a new integer \( K \), close to the non-integer value of \( \sqrt{N} \), we can define:

\[
p = K + \alpha \quad \text{and} \quad q = K - \beta
\]

where \( \alpha \) and \( \beta \) are unknown integers. Multiplying \( \tan(\theta) \) by \( 2\sqrt{N} \) and replacing \( \alpha = x \), we find:

\[y(x) = p - q = 2\sqrt{N} \tan(\theta) = \frac{[(K + x)^2 - N]}{(K + x)} \quad \text{with} \quad 0 < x \approx \sqrt{N}
\]

This last result represents essentially a non-linear Diophantine equation which has to be solved for its integer solution pair \([x, y] \). That is, we need to find the integer pair \([x, y] \) for which:

\[y(x) = \frac{[(K + x)^2 - N]}{(K + x)}
\]

is satisfied. Note that this function represents a monotonically increasing function of \( x \) as indicated by the form of its first derivative:

\[
\frac{dy}{dx} = 1 + \frac{N}{(K + x)^2}
\]

It’s nth derivative equals:

\[
\frac{d^n y}{dx^n} = (-1)^{n+1} \frac{n! N}{(K + x)^{n+1}}
\]

There is also be no inflection point for finite \( x \). The function has zero value when \( x \) has the non-integer value of \( \sqrt{N} - K \).
Let us next look at two specific examples. Take first the trivial case of N=143, where we can see by inspection that p=13 and q=11. For K we can take 12 which is close to sqrt(143)=11.95826... We thus are interested in solving:

\[ y(x) = \frac{x^2 + 24x + 1}{(12 + x)} \]

This equation clearly is satisfied by the integer pair \( [x, y] = [1, 2] \) so that \( p=12+1=13 \) and \( q=p-2=11 \).

Let us next look at a more complicated problem involving the five digit semi-prime number \( N=89893 \). Here \( \sqrt{N}=299.8216... \) so that we can choose \( K=300 \). This produces the non-linear algebraic equation:

\[ y(x) = \frac{[(300 + x)^2 - 89893]}{(300 + x)} \]

Plotting \( y(x) \) in the range \(-100 < x < 100\) yields the result shown:
We have indicated the two values of x for which the equation y(x) has integer values. These values are derived below. Note that the value of y(73)=y(-59). To find the value of x corresponding to p=300+x, we can use the one-line MAPLE program –

\[
\text{for x from 70 to 80 do } \{x, \text{evalf}(y)\}\text{od;}
\]

with the result-

\[
\{70, 127.0459459\}
\{71, 128.7008086\}
\{72, 130.3521505\}
\{73, 132\}
\{74, 133.6443850\}
\{75, 135.2853333\}
\{76, 136.9228723\}
\]

\(\text{Solution } y(73)=132\)

Thus p=300+73=373 and q=p-y=373-132=241. Hence

\[89893 = 241 \times 373\]

Notice that the integer values of y=p-q will be an even number while p=K+x will be odd.

One can extend the above procedure to still larger values of N but the process becomes more involved requiring an ever larger number of multiplications and divisions. According to the prime number theorem this could require some \(2 \sqrt{N}/ \ln(N)\) operations. For a hundred digits semi-prime this could require some \(10^{48}\) computer operations which could take years to carry out. Existing alternative factoring procedures such as the elliptic curve method allow for fewer mathematical operations but involve more intricate and involved evaluation procedures compared to the present approach. At this time no one has succeeded in factoring semi-primes above about 300 digit length. That is why public keys in cryptography based on N values of this length or greater are considered secure.

Finally, we point out that the present search procedure applies equally to non-semi-prime numbers including even numbers. Thus N=728 using K=27 produces \([x,y]=[1,2]\) among several other possibilities. That is, one factor product for N=728 is \(pq=(27+1)x(28-2)=28x26\).

November 2011