A NEW APPROACH TO FACTORING SEMI-PRIMES

It is well known that the factoring of semi-primes \( N=pq \) is a difficult task when \( N \) becomes large and exceeds one-hundred or so digit length such as encountered in public key cryptography. Even with the fastest supercomputers, it is still a lengthy and time-consuming process to factor such semi-primes. This stems from the fact that it can take up to \( 2\sqrt{N}/\ln(N) \) divisions to factor such numbers. Looking at things in terms of entropy, we can say that \( p \) and \( q \) are the low entropy state and \( N \) is the state of high entropy. It is always more difficult to go from a state of high entropy to one of low as so well suggested by the nursery rhyme-

Humpty Dumpty sat on a wall,  
Humpty Dumpty had a great fall.  
All the king's horses and all the king's men  
Couldn't put Humpty together again.

We want here to introduce a new technique for factoring semi-primes which keeps in mind the entropy reversal one is using to factor such numbers. The basic idea is to start with the lower entropy number \( A \) which is the nearest integer to \( \sqrt{N} \) and then make the expansion-

\[
p = A - a \quad \text{and} \quad q = A + b
\]

where \( a \) and \( b \) are positive integers. Taking the product one finds the linear Diophantine Equation-

\[
A(b-a) - ab = N - A^2
\]

On letting \( x = ab \) and \( y = b - a \) be unknown integers, we have-

\[
Ay - x = N - A^2
\]

with the known solutions-

\[
x = (ab) = C + k \quad \text{and} \quad y = (b - a) = \frac{N - A^2 + C + k}{A}
\]

where \( k = 0, \pm 1, \pm 2, \ldots \)

Here \( C \) is an integer smaller than \( A \) and hence of lower entropy. Eliminating \( b \) from these last two equations and setting \( C = A^2 - N \) produces the quadratic equation-

\[
a^2 + a\left(\frac{k}{A}\right) - (C + k) = 0
\]
This solves as-
\[ a = -\left( \frac{k}{2A} \right) + \sqrt{\left( \frac{k}{2A} \right)^2 + (C + k)} \]

But we know that \( k/(2A) \) must be an integer, say \( n \). This produces the desired result-
\[ a = -n + \sqrt{n^2 + 2An + C} = -n + R \]

Here we must find the integer value of \( n \) which makes the radical \( R \) an integer. Once this \( n \) has been found the problem is solved with-

\[
\begin{align*}
p &= (A + n) - \sqrt{n^2 + 2An + C} \\
\text{and} \\
q &= (A + n) + \sqrt{n^2 + 2An + C}
\end{align*}
\]

The constants \( A \) and \( C \) are known once \( N \) has been specified. To get an idea of what the value of \( n \) might be we start with the square of \((A+n-a)\) to get-
\[ (A - p + n)^2 = n^2 + 2An + C \]

Next solving for \( n \) we have the identity-
\[ n = \frac{\{(A - p)^2 - C\}}{2p} = \frac{(N - \alpha A^2)(2 - \alpha)}{2\alpha A} \approx \frac{(1 - \alpha)^2}{2\alpha} \]

Here \( \alpha \) is an unknown constant with a value in the range \( 0 < \alpha < 1 \). Since we don’t know \( \alpha \) beforehand one does not know the exact value of \( n \) either. Usually one can start the search for integer \( R \) assuming \( \alpha = 1 \). This will work well if \( a \) and \( b \) are small. However, often it may be more of an advantage to let \( \alpha = 3/4 \) or even \( 1/2 \).

Inverting the problem, we can always find \( \alpha \) via the quadratic equation-
\[ \alpha^2 A^2 - 2\alpha A (A + n) + N = 0 \]

once \( N \), \( A \), and \( n \) are known. The equation says that \( \alpha = 0.8260 \) once we set \( N = 551 \), \( A = 23 \), and \( n = 1 \). This shows that \( p = A - a = \alpha A = 19 \) for this case.
Let us now demonstrate this new semi-prime factoring procedure for several specific cases. Start with –

\[ N = 12319 \quad \text{for which } \sqrt{N} = 110.9901 \ldots, \text{so that } A = 111, \text{ and } C = A^2 - N = 2 \]

We start the search with \( n = 0 \). Already at \( n = 1 \) we find the integer value \( R = 15 \). So we have the answer -

\[ p = 112 - 15 = 97 \quad \text{and} \quad q = 112 + 15 = 127 \]

This required very little effort. The reason is that \( n \) remained small here was because \( C = (A^2 - N) \) was just 2 and the ratio \( p/q \) is only a little less than one allowing us to say \( \alpha \approx 1 \).

We next try factoring the semi-prime \( N = 455839 \). This number is often used to demonstrate the Lenstra Elliptic Curve Factorization Method. We have here -

\[ A = 675 \quad \text{and} \quad C = (675)^2 - 455839 = -214 \]

On varying \( n \) in the radical, we find \( R \) to have the integer value 81 for \( n = 5 \). So we find -

\[ p = 680 - 81 = 599, \quad \text{and} \quad q = 680 + 81 = 761 \]

Again this was accomplished with amazing speed (compared to the Lenstra Method) requiring only five divisions.

As a third specific example, consider the nine digit semi-prime-

\[ N = 453266657 \quad \text{where} \quad \sqrt{N} = 21290.06005. \]

Here \( A = 21290, \ A^2 - N = -2557, \) and the radical reads -

\[ R = \sqrt{n^2 + 2(21290)n - 2557} \]

We carry out a search for finding the value of integer \( n \) which produces a positive integer for the radical via the following one line computer program -

\[
\text{for n from 0 to 100 do } \{n, R\} \text{ od;}
\]

After 61 divisions, one finds \( R = 1612 \) at \( n = 61 \). So we have-

\[ p = 21351 - 1612 = 19739 \quad \text{and} \quad q = 21351 + 1612 = 22963 \]

that is, we have the factored form -
453266657 = 19739 x 22963

Because N was a larger number than in the earlier examples, the number of divisions increased. However, even the 61 divisions it took to factor this nine digit long semi-prime is still quite reasonable compared to a brute force factoring where one divides N by all prime numbers less than \( \sqrt{N} \). This would require 2391 divisions. We could also have speeded up the search by assuming \( \alpha = 0.93 \) which would tell us to start the search with \( n = 56 \) and would therefore require only five divisions. Of course we really did not know the value of \( \alpha \) beforehand. One could guess for \( \alpha \) and then use the following graph to try the corresponding n for the integer R search-

![Graph showing estimate for n given a value for alpha](image)

We complete our discussion by looking at the even and odd character of the most important terms in our solution procedure. For this purpose we have constructed the following table-

<table>
<thead>
<tr>
<th>N</th>
<th>A</th>
<th>C=A^2-N</th>
<th>n</th>
<th>R</th>
<th>p=A+n-R</th>
<th>q=A+n+R</th>
</tr>
</thead>
<tbody>
<tr>
<td>3403</td>
<td>58</td>
<td>-39</td>
<td>4</td>
<td>21</td>
<td>41</td>
<td>83</td>
</tr>
<tr>
<td>7663</td>
<td>88</td>
<td>81</td>
<td>0</td>
<td>9</td>
<td>79</td>
<td>97</td>
</tr>
<tr>
<td>12319</td>
<td>111</td>
<td>2</td>
<td>1</td>
<td>15</td>
<td>97</td>
<td>127</td>
</tr>
<tr>
<td>100651</td>
<td>317</td>
<td>-162</td>
<td>9</td>
<td>75</td>
<td>251</td>
<td>401</td>
</tr>
<tr>
<td>351163</td>
<td>593</td>
<td>486</td>
<td>29</td>
<td>189</td>
<td>433</td>
<td>811</td>
</tr>
<tr>
<td>9431047</td>
<td>3071</td>
<td>-6</td>
<td>53</td>
<td>573</td>
<td>2551</td>
<td>3697</td>
</tr>
<tr>
<td>32983151</td>
<td>5743</td>
<td>-1102</td>
<td>38</td>
<td>943</td>
<td>4877</td>
<td>6763</td>
</tr>
<tr>
<td>263100319</td>
<td>16220</td>
<td>-11919</td>
<td>1020</td>
<td>5841</td>
<td>11399</td>
<td>23081</td>
</tr>
</tbody>
</table>
Because both p and q will be odd numbers, we can say that A=p+a is even when a is odd and visa versa. Also if A is even then n +R is odd. If A is odd then n+R is even. Often one can halve the number of divisions required for a search by assuming n is either even or odd. In that case one has a fifty-fifty chance of requiring only half of the required divisions to establish an integer value for the radical R. The integers n and R also relate to each other via the equality:

\[ n = -A + \sqrt{N + R^2} \]

Thus if N=36557 we have A=191 and find n=10 and R=62. Thus-

\[ 10 = -191 + \sqrt{36557 + 62^2} = -191 + 201 \]

We note that n is typically smaller than R. We also observe that the values of n in the above table are all fairly small with exception of the last entry. To make the number of required divisions smaller for N=263100319 one can first apply the above formula for n when \( \alpha = 0.7 \). This tells us to search near n=1043.

Finally we provide the reader with a flow chart for factoring semi-primes by the present method. Here is the chart-

![Flow Chart for Factoring N=pq](image)

Here is a worked out example-
FLOW CHART FOR N=5496811

(1)-sqrt(N) = 2344.527...

(2)-so A = 2345

(3)-find C = A^2 - N = 2214

(4)-search for integer R near n=151 to get R=885 at n=161

(5)-p = 2345 + 161 - 885 = 1621 and q = 2345 + 161 + 885 = 3391

U.H.Kurzweg
January 2015