FACTORING LARGE SEMI-PRIMES

It is well known that it is difficult to factor a large semi-prime number N into its two prime components. We look more into this problem here and show ways to factor such numbers making use of the Goldbach Conjecture. Our starting point is the formula-

$$N = p \cdot q$$

where p and q are the two prime numbers whose product equals N. To aid in our analysis we apply the Goldbach Conjecture that the sum of two primes will always be equal to an even number. That is-

$$p + q = 2n$$

This means that n is the average integer value of p and q. Next eliminating q from these two equations, yields the quadratic form-

$$p^2 - 2pn + N = 0$$

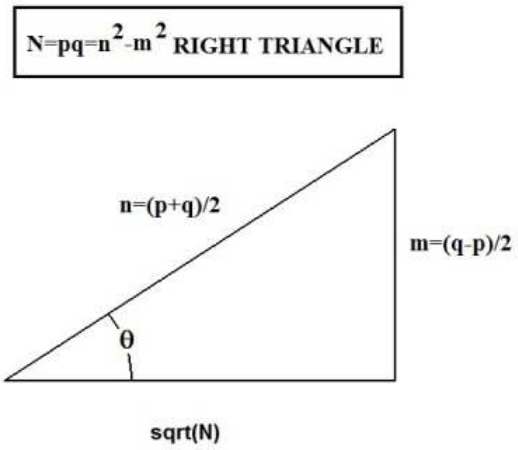
Solving this algebraic expression yields the solutions-

$$p = n - \sqrt{n^2 - N} \quad and \quad q = n + \sqrt{n^2 - N}$$

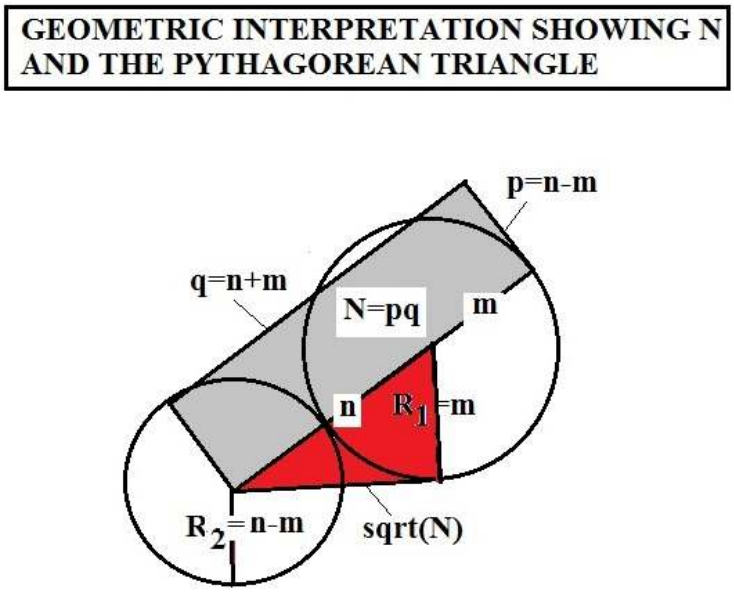
provided that p<sqrt(N)<q. One can also rewrite things as the Diophantine Equation–

$$n^2 - m^2 = N$$

with m=(q-p)/2 and n and m both integers. Fermat first used these integers in his factoring approach to semi-primes. One can also think of n, m and N as parts of a Pythagorean triple represented by the following right triangle-

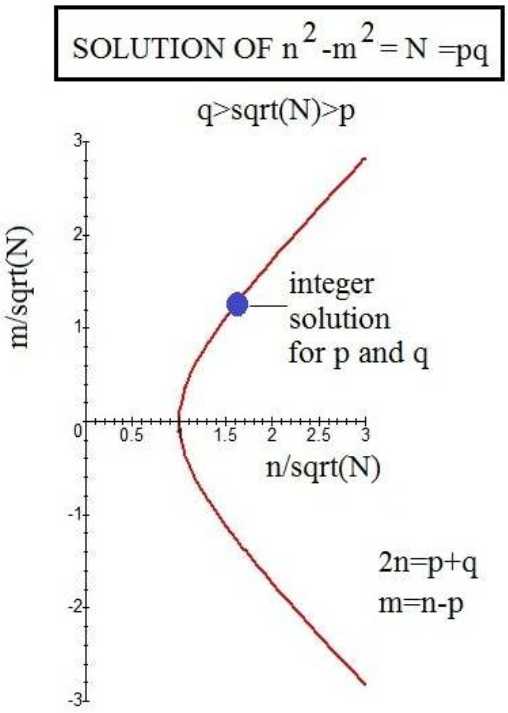


Comparing this triangle with the solution rectangle N=pq=(n-m)(n+m) yields the geometric figure-



The ratio of the grey area to the red area is $2\text{sqrt}(N)/\text{sqrt}(n^2-N)$.

The solution we are looking for is shown by the blue dot on the following graph-



One can now attempt to evaluate the expression for p for a specified N by treating the average value n as a running variable starting with the integer nearest to the sqrt(N). Thus , if we are given N=3763 where sqrt(N)=61.343.., we start our evaluation with n=62 followed by n=63 etc. In this case the answer will already follow from the first operation, namely,-

$$p = 62 - \sqrt{3844 - 3763} = 62 - 9 = 53$$

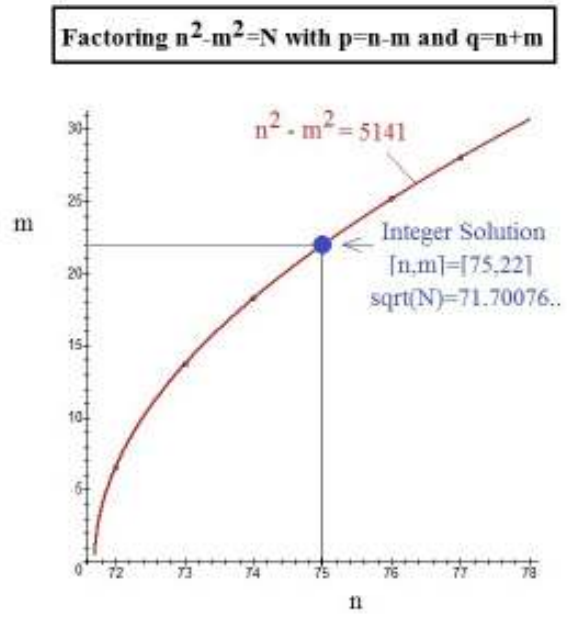
The value of q follows from 2n-p=N/p=71. For most Ns, especially when they become larger, the analysis will typically require a much larger group of ns before the radical sqrt(n^2-N) is found to have an integer value. In such cases one can use several approaches. One is to simply carry out an evaluation of the p expression for an n value near the sqrt(N) and just keep operating with ever larger ns stepped off in unit increments. Such a procedure can be nicely implemented on a computer. Take the case of N=5141 where sqrt(N)=71.70076... The one line program we run for this case using MAPLE is-

for n from 72 to 80 do {n, evalf(n-sqrt(n^2-N))}od;

At n=75 it spits out the integer answer-

[n,m]=[75,22]

It thus took just four simple operations to determine the value p=n-m=53 and q=n+m=97. This should be compared with the 16 divisions required by the brute force approach of dividing N in turn by the primes 3 ,5, 7, 11,…through 16 . A picture of the hyperbola for $5141=n^2-m^2$ and the integer solution [75,22] follows-



It should be pointed out that in most cases simple evaluations as those shown above will no longer be possible, in particular for large Ns such as encountered in cryptography. Under those cases the value of n is likely to be far removed from sqrt(N). This means that if a stepping procedure is to be used then one needs to start with some larger values of n. How can such values be determined? One possible way is to note that $\varepsilon=N/n^2=4pq/(p+q)^2$ is often a small parameter suggesting a series expansion of the above hyperbolic function $n^2-m^2=N$ in powers of ε . Carrying out such an expansion we find-

$$m = n\left\{1-\frac{\varepsilon}{2}\left[1+\frac{\varepsilon}{2\cdot\ 2!}+\frac{(1\cdot 3)\varepsilon^2}{2^2\cdot 3!}+\frac{(1\cdot 3\cdot 5)\varepsilon^3}{2^3\cdot 4!}+\cdots\right]\right\}$$

Thus, if ε is small, we have the approximation-

$$\frac{m}{n}\approx 1-\frac{N}{2n^2}$$

Although in general one won't know the value of n beforehand, one could make an assumption that n=k sqrt(N), where k is a constant with a value of one or greater. Let us demonstrate such an approach by looking at the famous Fermat Number-

$$N= 2^{32}+1=4294967297$$

Fermat thought it to be prime but Euler proved that it wasn't. Indeed, Euler showed that this number is composite with p=641 and q=6700417. Here n=(641+6700417)/2=3350529 so that N/n^2=0.0003825895458… is indeed small . We can thus estimate that an integer value for m equals approximately $m\approx n[1-(N/2n^2)]$. At n=3350529 we find m=3349888.0613… which is very close to the exact value of m=3349888. Searching for an integer solution near n=3350529 with the program-

```
for n from 3350527 to 3350532 do{n , evalf(n-sqrt(n^2-4294967297),10)}od;
```

produces the following-

n	p=n-sqrt(n^2-N)
3350527	641.00038269
3350528	641.00019135
3350529	**641 ←integer solution**
3350530	640.99980865
3350531	640.99961730

From this result we also have that q=2n-p=6701058-641=6700417. The value of k=n/sqrtN to use in this case is about 51 meaning we are a long way from n≈sqrt(N) for this semi-prime. If we had not known the factors p and q beforehand , a long search would have been required trying ns for many different values of the constant k. That is, the procedure is a little like playing the boardgame battleship where one aims for a point in the range sqrt(N)<n<(N+1)/2 .

When N has hundreds to several thousand digit length, the search using a large number of different ns becomes essentially impractical even with the use of the fastest existing electronic computers and application of improved factorization methods such as general number field sieving(GNFS) or elliptic curve factorization . I am still waiting for someone to break the following 1084 digit public key-

N=160731047637009729259688920385507056726966793490579598495689711866432421212774967
029895340327197901756096014299132623454583177072050452755510701340673282385647899694
0838813161946424174515704834663277821357305755648561855464870530344045600634336147

238364567902664574388316263755568541338669583498171727274624625164668984795744028410717039091380624565676245657842541015683784072422732076608920368697081906880333516015394016215765079648415972059527224877506709045229323287315306407064573821626447385388132471393154562134015866188205178235764270941251970012703500878782708897174454011457922316740989484168888682501435920269738539737851202170779517665469395775208972453921865472795724941776802915065785089627079348791249148808855007264396250330219367289492773901853990242765470359959156489381704156637573786372070113915380095968333541077371562730374947278583020286633662969439250086473487692720355322650480497098272751793812528986759655285106192583767791710305564828845357288129162166254301870395336686775280795441768976473034451536435253548174136508485447786906882010052744437176805938 99

which I constructed about a year ago using a new prime number generating formula. It took me about an hour to generate the primes p and q and then to construct N. It's time for NSA (or anyone else capable of factoring large semi-primes) to get busy on trying to crack this number. My bet is that they won't be able to, at least not anytime soon.

Finally I leave you with three semi-primes N=100160065 , N=3289192017, and N=455839 all of which are very easy to factor. For the first one the analysis starts with n=10008>sqrt(100160065)= 10007.99995. One finds at once via a single evaluation that p=10007 and from this follows q=10009. It is an example of having the values of p and q too close to each other for key security. In the second example a single division of 3289192017 by the lowest odd prime p=3 yields a q=1096397339. It is an example of having the value of p too small. The third number is a number often used to show how Lenstra Elliptic Curve Factorization works (see Wikepedia). For this example it is much faster and easier to simply use the present running n approach. Indeed we have here that sqrt(N)=675.158.. so that we start the evaluation of n-sqrt(n^2-N) using n=676. Already at n=680 we find the integer factor p=599 so that q=761. Good secure public keys require N in the several hundred digit range while at the same time having the ratio q/p neither near one or approaching very large values . Already an eleven digit semi-prime such as N=74197213369 requires 2203 operations using the n variation (also known in the literature as Fermat factorization) approach starting with n=272392 ≈ sqrt(N). The first integer solution pair found is [n,m]=[274595,34716] to yield the factor [p,q]=[239879,309311]. See if you can factor the 19 digit number-

N=1127451830576035879

into its prime factors -

p=486100619 and q=2319379541

using the n variation approach for a good choice for k.

You can go to-

http://www.alpertron.com.ar/QUAD.HTM

to find a calculator which solves $n^2-m^2=N$ for smaller semi-primes N out to about ten digit length. It is just necessary to set n=x, m=y, a=1, b=0, c=-1, d=0, e=0 and f=-N in the generalized Diophantine equation $ax^2+bxy+cy^2+dx+ey+f=0$. The program is based on the elliptic curve factorization method.

April 2012