AN ALTERNATE TECHNIQUE FOR FACTORING LARGE SEMI-PRIMES

There are several well known techniques for factoring large semi-primes N=pq including elliptic curve factorization and generalized sieve methods. These methods are able to handle the factorization of large semi-primes when the number of digits contained in the semi-prime N=pq does not exceed about 100 digit length. Above this number the required computer time becomes prohibitive so that most cryptic codes based on public keys of several hundred digit length are assumed to be secure. We consider here a related factorization method similar to the sieve method but unlike the sieve method(which typically requires the addition of many exponential vectors V_i,) we accomplish things by a modification of the Fermat method which starts with the variables n=(p+q)/2 and m=(q-p)/2 . This fact allows one to write pq=N in its equivalent form $m^2=n^2-N$. Taking the root we have –

$$n-\text{sqrt}(m)=p \text{ and } n+\text{sqrt}(m)=q$$

If the sqrt(m) is an integer then we can write directly that-

$$\gcd(n-\text{sqrt}(m)),N)=p$$

Unfortunately, in most instances m will not be a perfect square. However we can modify things slightly by introduction a random number n_1 usually near the value of sqrt(N) and then postulate that-

$$\gcd[(n_1n_2-\text{sqrt}(m_1m_2)) \bmod N, N)=p \text{ or } q$$

If we now let $m_1=m_2=n_1^2-N$ we get-

$$\gcd((n_1(n_2-n_1)+N) \bmod N, N)\rightarrow\gcd(n_1(n_2-n_1) \bmod N ,N)=p \text{ or } q$$

That is, for any given n_1 we vary n_2 over a range $-k<n_2<k$ until a point is hit where the gcd differs from unity. This value will then be either p or q. As we will demonstrate the process works very well for semi-primes less than about ten digits long but has a problem with finding a proper starting value of n_1 when the number N is much larger than this. Typically we will find that $|n_2|<<n_1$. Let us give some examples using this alternate method.

Take first the five digit semi-prime N=16351 for which sqrt(N)=127.87…. We pick as our starting value an even number $n_1=300$ and then search through the odd numbers $n_2=2n+1$ in the above gcd formula. We use the one line MAPLE search program-

```
for n from 0 to 30 do {n, gcd(300*(2*n-299),16351)}od;
```

for doing this. An evaluation produces a table of thirty values of gcd=1 plus one solution where gcd=83=p occurring at n_2=25. Since we know q=N/p, we have factored N to obtain the result-

16351=83 x 197

Notice we could have picked as our initial value for n_1 any other even number such as 200 or even 100. The latter will produce at n_2=8 the value p=83. So that the same factoring is accomplished. What is interesting about this greatest common divisor approach is that the only possible solutions are 1 or p or q. Thus, with enough patience, the factor p or q can always be found. It should be pointed out that n_1 and n_2 differ from the Fermat variables n and m given above. For the N=16351 case we have shown p can be produced by n_1=100 and n_2=8 while n=(p+q)/2=140 and m=(q-p)/2=57 in terms of the original numbers.

Moving on, we consider next the ten digit semi-prime N=1232895479 where sqrt(N)=35112.6114…Playing around with the choice for n_1, we find n_1=70002>sqrt(N) makes for an easy evaluation. One finds that at n_2=34 the expression-

gcd(70002(2*34-70001),N)=23311=p

We thus have that-

1232895479=23311 x 52889

Next let us look at one of the classic mathematics problems involving the factoring of the ten digit Fermat number $N=2^{32}+1$=4294967297 which Leonard Euler first showed to be a semi-prime. In this case sqrt(N)=65536.00001… so that we can choose as n_1 any integer somewhere around this value of 65537. Again testing several different values for n_1 we took n_1=80134 and obtained the expression-

$$\gcd(80134(2n_2-80133),2^{32}+1)$$

Running the calculations in the neighborhood of n_2=0, one finds that gcd=641=p at n_2=4 Thus we have with relatively little effort that-

$2^{32}+1=4294967297=641 \times 6700417$

It is amazing that Euler was able to accomplish the factoring of this number without the aid of electronic computer. He was probably aided by his prodigious number calculating abilities. His factorization was most likely accomplished by the brute force approach of dividing $N=2^{32}+1$ by successive primes. Since the number 641 represents the 116th prime, he could have accomplished his factorization with just 116 divisions. He was lucky that p in this case is very

small compared to the square root of N. Euler was Swiss by birth, spent most of his life at both the Prussian and the Russian Academy of Sciences and is buried at the Alexander Nevsky Monastery in St. Petersburg, Russia.

Finally let us complete our discussion by revisiting a semi-prime number I posed as a problem for someone to factor in an earlier note. The nineteen digit number reads N=1127451830576035879 and its root is sqrt(N)= 1061815346.74.. . So searching for n_2 starting with different random choices of n_1 above 1.07 billion and running $-100<n_2<100$ we were unable to find any solutions other than gcd=1. We know the unique values p=486100619 and q=2319379541 exist since we started with this in the original construction of N. This example brings out the shortcoming of the present approach. It is that the search for the right n_1 and n_2 increases dramatically with increasing digit size of N. We are presently in the process of finding a better than random selection approach for the starting value of n_1 which will minimize the number of calculations involving n_2. Some preliminary observations are that when n_1 is close to to an integer multiple of p then the number n_2 will be small. This is good to know but is not really of help unless one already has some idea of what the value of p is. Also one finds that $(n_1-2n_2-1)=p\cdot$int. for all even n_1. Thus for N=22261 and n_1=300 we find p=113 at n_2=-20. Thus (300+40-1)=339=113·3. Also we observe that one can modify the gcd search procedure by multiplying n_2 by a factor of 2^b , with b a small positive integer. As an example N=1232895479 where sqrt(N)=35112.61140… for n_1=26000 and b =7 produces-

$$\gcd(26000(128n_2-25999)\bmod N,N)=23311=p \text{ at } n_2=21$$

Another way of stating this last result is-

$$\gcd(626809479,1232895479)=23311$$

May 2012