PROPERTIES OF SEMI-PRIMES

One of the more important quantities encountered in number-theory is the semi-prime N=pq, where p and q are prime numbers. This combination plays a critical role in modern day cryptography because it is extremely hard to factor when N gets large but very simple to construct when the prime components are given. Let us demonstrate this fact for the semi-prime N=1373 x 3217=4416941. The multiplication of the two primes readily produces the seven digit long semi-prime. However to find its factors is much more difficult. We can always take p to be less than sqrt(N)=2101.6519 so that q will be somewhat greater than sqrt(N). Thus one can write –

p=2102-a and q=2102+b

, with a and b being unknown integers . This is equivalent to saying-

a=[2102b+1460]/[b+2102]

Next by varying b from b=0 on up, we need 1115 trials until an all integer solution is found. The solution reads-

[a,b]=[729,1115]

This result then yields the answer-

p=2102-729=1373 and q=2102+1115=3217

It took 1115 trial calculations moving one unit at a time to get this answer. Typically one can expect the trials to increase to about sqrt(N) before an integer solution for a and b is found. For large semi-primes of 100 digit or larger size used in public key cryptography it presently becomes impractical to test for integer p and q for such large semi-primes thus making modern day cyber security safe to use.

It is our purpose here to find additional properties of semi-primes and then to use this information to speed up the factorization process from the brute force approach demonstrated above.

We begin by noting that the average value of p and q must equal-

S=(p+q}/2=[σ(N)-1-N]/2

, where the sigma function σ(N) for the semi-prime N reads σ(N)=1+p+q+N. One can next write-

p=S-R and q=S+R

, with R given by the radical-

R=sqrt[S^2-N]

Here R represents half the distance between p and q. Once one knows the value of σ(N) and hence S, the radical will be known with the prime factors becoming-

$$[p,q]=S\mp sqrt[S\hat{}2-N]$$

Fortunately the value of sigma(N) is given by most advanced mathematics programs for Ns up to about forty digit length. Thus , for the semi-prime discussed above , where N=4416941, our computer yields in a split second that σ(N)=4421532, S=2295 and R=922. Thus we have the factors-

$$p=1373 \quad \text{and} \quad q=3217$$

This result is obtained at only a very small fraction of the time it takes to find p and q by the above brute force search approach.

We notice that any semi-prime N, when its factors are both greater than three, must have the form-

$$N=6s\pm1 \text{ with } s=1,2,3,\ldots$$

This means that N mod(6)=1 or 5 without exception. The seven digit long semi-prime N=4416941, discussed above, has N mod(6)=5. This fact allows us to state that p=6n±1 and q=6m∓1 . It agrees with the actual results p=1373=6(229)-1 and q=3217=6(536)+1. These results also mean that all prime numbers greater than three must have the form 6n±1.

In the past decade or so we have come up with several new properties concerning semi-primes . One of these is the Number Fraction which for semi-primes reads-

$$f(N)=[\sigma(N)-N-1]/N =(p+q)/pq$$

It has the interesting property that f(p)=f(q)=0 for he primes p and q. For N=4416941 we find f(N) equal to-

$$f(N)=4590/4416941]=0.00103918$$

For the small semi-prime N=77, where σ(77)=1+7+11+77=96 , we find f(77)=(96-78)/77 =18/77=0.233766 .

One also has the identities-

$$S=(p+q)/2=Nf(N)/2 \qquad \text{and} \qquad \sigma(N)=1+N+Nf(N)$$

One can estimate the value of σ(N) as 1+N+2sqrt(N). For the N=77 case this estimate is 1+77+18=96 which is the exact value while for N=4416941 we get the estimate-

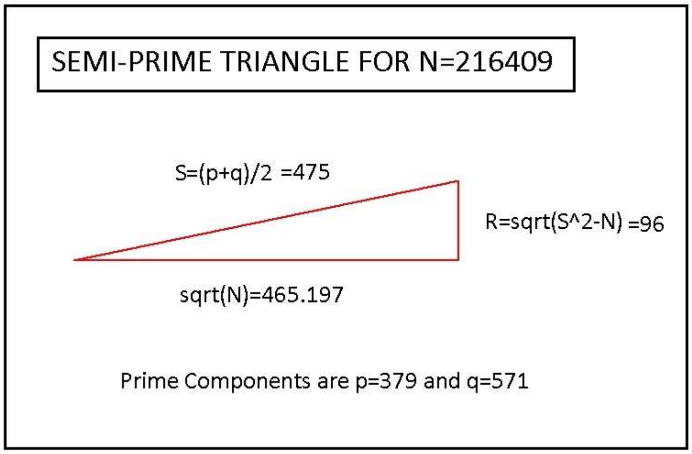
$$\sigma(N)\approx 1+N+2\text{sqrt}(N)=1+4416941+4203=4472114$$

Here the exact value is σ(N)=1+1373+3217+4416941=4421532 and thus about one percent in error.

Here is a list of five representative semi-primes and their properties-

Semi-prime N	p	q	S	σ(N)	R
77	7	11	9	96	2
216409	379	571	475	217360	96
4416941	1373	3217	2295	4421532	922
110764567	8231	13457	10844	110786256	2613
28287742959451	3678923	7689137	5684030	28287754327512	2005107

We see that S always equals the mean value of p and q while R is half the distance between p and q. Also σ(N) lies only slightly above N with p=S-R and q=S+R. In addition we have N=S^2-R^2. The last allows for the existence of a right triangle with a hypotenuse of S and sides sqrt(N) and R. Here is this right triangle drawn for N=216409, p=379 and q=572-



Using the above formulas makes the factoring of semi-primes trivial for all Ns small enough for my computer to give a value for sigma(N) . With my MAPLE program, the largest N for which σ(N) is readily found has about fortyt digit length. More research is needed to find a way to determine the sigma function for Ns greater than this. Should such a search be successful, the security of public key approaches in public key cryptography will become obsolete. Note that if p and q are given, σ(N) always known from its semi-prime definition 1+p+q+N.

.

U.H.Kurzweg
June 8, 2021
Gainesville, Florida