FBA

Abstract

One of the most difficult challenges in number theory and cryptography is the problem of factoring large integers since factoring large numbers more than 200 digits remain difficult. This difficulty has been useful in proving the strength of the RSA encryption scheme, which uses two large numbers of similar size. In this work, we introduce factorisation by bases algorithms (FBA) and implement them on the RSA-numbers defined by the RSA challenge. We analyse the time spent on factoring to check the speed and efficiency of each algorithm.

Introduction

Integer factorization is the task of computing the divisors of natural numbers. It is a problem with a long and fascinating history, and it is certainly among the most influential in algorithmic number theory. While there is a variety of algorithms significantly faster than the brute-force search for divisors, it is still an open problem to construct a technique that efficiently factors general numbers with hundreds to thousands of digits. The hardness of this problem is fundamental for the security of widely used cryptography schemes, most prominently the RSA cryptosystem. Nevertheless, there is no proof for its hardness besides the fact that decades of efforts have failed to construct a more efficient technique. Quite regularly, there are set new records1 concerning the factorization of numbers of certain size, mostly due to improved implementations of the best available algorithms and advances in the hardware and computing power. In addition, the bound for the deterministic integer factorization problem has been improved multiple times in recent years ([11], [12], [10], [14]). On the other hand, there has only been little progress in the development of new techniques for practical integer factorization since the invention of the Number Field Sieve ([19]) in the 1990s. One of the earlier algorithms with sub-exponential runtime was by Dixon ([7]) in 1981.

Background

Let N be the number we want to factorize. We will always assume that N is odd, composite and not a perfect power of another number. We will also assume that N is made up of 2 prime factors, p and q. We will explain properties of factorisation by bases techniques using one number in base 10.

Lemma: N can be converted into a polynomial given a base.

Example: 667 in base 10 is $6x^2 + 6x + 7$

Now 667 is built by two prime factors which are 23 and 29. If both of them are converted into base 10 they become, $2x + 3$ and $2x + 9$ respectively.

$(2x + 3) \times (2x + 9) = 4x^2 + 24x + 27$ which is different from the first form of N gotten. (A1)

From this derivation we notice that $6x^2 + 6x + 7$ cannot be factorised to give us p and q because the discriminant was imaginary. Furthermore, we also notice that for any converted form of N we get we cannot factorise if also the discriminant is imaginary. From (A1) we noticed that a change of

the form $(6-k)x^2+(6+10k-h)x+(7+10h)$ happened with h=2 and k=2 (A2) which results in, $f(x)$: $6x^2+6x+7\equiv 0(mod\ 667)$. This becomes the obtained equation.

If we say $(6-k)x^2+(6+10k-h)x+(7+10h)=0$

We obtain the discriminant: $(6+10k-h)^2-4(6-k)(7+10h)=m^2$ which reduces to

$f(h,k,m)$: $100k^2+20hk+h^2+148k-252h-132=m^2$ (A3)

To solve such a problem is quite complex. We want to reduce the equation to 2 variables which can be easily solved either as Diophantine equations or through some ingenious method. There are many ways to do this.

First given $f(h,k,m)$: $100k^2+20hk+h^2+148k-252h-132=m^2$ Here if we insert a value for m for example, 12 we get: $f(h,k,m)$: $100k^2+20hk+h^2+148k-252h-276=0$ which is a conic section. Methods of solving conic sections such as SIQS and ECM, involve the factorisation of a number to its prime factors of which this is return back to not solving the problem.

Another choice is that of given $f(h,k,m)$: $100k^2+20hk+h^2+148k-252h-132=m^2$ Let $10k+h+m=M$ if we say M=34, we get $k=\frac{-2(m-41)}{29}$, $h=\frac{-(9m-166)}{29}$ We will now only be left with finding m of which by Diophantine solutions, we know that m is of a specific parametric solution. However, this method is difficult since predicting the exact values of M which will give us prime solutions is difficult and iterating all possible solutions is quite a slow method.

To solve all these problems, we present three case scenarios, each with a different perspective.

CASE I

If we are to go back to the original form with only variables, we notice that $N=Ax^2+Bx+C$ and $f(x)$: $(A-k)x^2+(B+xk-h)x+(C+xh)\ \equiv 0\ (mod\ N)$. For the discriminant we get, $(B+xk-h)^2-4(A-k)(C+xh)=m^2$

$\Rightarrow\ x^2k^2+h^2-m^2+B^2+2xkh+2Bxk-2Bh-4AC-4Axh+4Ck=0$

Making unknowns subjects of the knowns, for example, making subject of B, we get.

$B^2+(2xk-2h)B+(x^2k^2+h^2-m^2+2xkh-4AC-4Axh+4Ck)=0$

$\Longrightarrow(2xk-2h)^2-4(x^2k^2+h^2-m^2+2xkh-4AC-4Axh+4Ck)=o^2$

$\Longrightarrow 4m^2-16xkh+16AC+16Axh-16Ck=\ o^2$

And for $N=667$ in base 10, we obtain.

$f(h,k,m)$: $4m^2-160kh+672+960h-112k=\ o^2$

Well, this means that we now have 2 equations of $f(h,k,m)$ which we can use to find solutions. We now have four variables of which we need a third equation to reduce all the variables to one variable which can easily be solved with.

If we have,

$f(h,k,m)$: $100k^2+20hk+h^2+148k-252h-132=m^2$ (T1)

$f(h,k,m)$: $4m^2-160kh+672+960h-112k=\ o^2$ (T2)

And then substitute (T1) into (T2) to eliminate m, we have,

$f(h,k,m)$: $400k^2 - 80hk + 4h^2 + 480k - 48h + 144 = o^2$ (T4)

$\Rightarrow f(h,k,m)$: $4(10k-h)^2 + 48(10k-h) + 144 - o^2 = 0$

Which means, if we put a value for $10k - h$ we find o and conversely.

But since m is the holy grail of this solution, eliminating it makes it harder for us to calculate it, rather let's eliminate a value such as k, and that must give us a desirable solution.

Given we have (T1) and (T2), the third equation is establishing a relationship between o and m, such of their squares, for example: $2m + o = R$. If we insert any value for R and then iterate for any value of k, we get deterministic results for all the variables. Examples are shown below.

Let (T1), (T2) and $2m + o = 72$, $10k + h + n = 126$

The value of $10k + h + n$ was found from calculating the discriminant of $f(x) - N$.

$\Rightarrow 6x^2 + 6x - 660: D = 12x + 6 = 126$

For $k = 2$

$h = 2,\ m = 12,\ o = 48,\ n = 104$ $\quad h = \frac{-410}{11},\ m = \frac{1092}{11},\ o = \frac{-1392}{11},\ n = \frac{1576}{11}$

For $k = 3$

$h = \frac{-331}{4},\ m = \frac{619}{4},\ o = \frac{-475}{2},\ n = \frac{715}{4}$ $\quad h = \frac{101}{16},\ m = \frac{101}{16},\ o = \frac{475}{8},\ n = \frac{1435}{16}$

This is because if we look closely at the situation, we have

$f(h,k,m)$: $100k^2 + 20hk + h^2 + 148k - 252h - 132 = m^2$ (T1)

$f(h,k,m)$: $4m^2 - 160kh + 672 + 960h - 112k = o^2$ (T2)

$2m + o = R$

If we put a value for k, for example, k=2, we get:

$h = \frac{R^2+104R+448}{4(R-160)}$, $m = \frac{R^2-320R-17088}{4(R-160)}$, $o = \frac{R^2+17088}{2R-320}$

Hence if we put any value for R we get h and m values which will give us integer solutions. However, most of these solutions will be N of which we want to find p or q, making it difficult.

To conquer this difficulty we substitute the gotten values of h, k and m, into $f(x)$ and (T1), meaning we have to just solve:

$(6-k)x^2 + (6 + 10k - h)x + (7 + 10h) = 0$

$f(h,k,m)$: $100k^2 + 20hk + h^2 + 148k - 252h - 132 = m^2$ (T1)

And we get:

$x = \frac{R}{16}$ or $\frac{2(5R+534)}{R-160}$ and we know that one of the values of R is $R = -72$

The challenge we face here is that of finding the gcd of numerator and the denominator since we should first find x in its simplest form so that we can later equate the product of p generated from x to N. but we know that the gcd must be of the factors of 16, however, even if we predict gcd to be 8, we find ourselves in an interesting situation when trying to solve for the other x, since this will imply that both the numerator and the denominator must be divisible by the gcd.

Because of the failed attempt, we can still solve it using another way. In this case we use, (T2). Because m, h, k are always strict, we assume that x is varying to find o.

If we have that $m = 12, h = 2, k = 2$, we get this sample of data between x and o.

x	0	10	24	42	64	90
o	32	48	64	80	96	112

And if we have $m = 6, k = 5, h = 24$, we also get.

x	0	2	10	16	32	42
0	16	32	64	80	112	128

This means that we can use the values of x to obtain the required solutions for $h, k, m.$ to test our hypotheses, we will use the first scenario in which m was 12.

$4m^2 - 16x_1kh + 672 + 96x_1h - 112k = 32^2$ (T8)

$4m^2 - 16x_2kh + 672 + 96x_2h - 112k = 48^2$ (T9)

(T9) – (T8) gives: $-16x_2kh + 96x_2h + 16x_1kh - 96x_1h = 48^2 - 32^2$

$\Rightarrow 16h(x_1 - x_2)(k - 6) = (48 - 32)(48 + 32)$

$\Rightarrow h(x_1 - x_2)(k - 6) = 80$ (T10)

Meaning that we just need to guess $(x_1 - x_2)$ and then we can solve for h,k. of which we know that $(x_1 - x_2)$ must be a factor of 80. Testing for $(x_1 - x_2) = -10$ it gave us these solutions:

$(k, h) = (-2, 1)\ (14, -1)\ (2, 2)\ (10, -2)\ (4, 4)\ (8, -4)\ (5, 8)\ (7, -8)$

Note: ALGORITHM 1, a breaker, which might give us deterministic solutions. Contains the ability to prove that P=NP.

But does that mean if we can find the RHS value, then the solution will be deterministic. Let's test this by trying for a larger number, probably a twelve-digit number made by Mersenne primes.

CASE II

Let's have $f(x)$: $(A-a)x^2+(B-b)x+(C-c)\equiv 0(mod\ N)$ then it shows that for the discriminant we have $(B-b)^2-4(A-a)(C-c)=m^2$

$\Longrightarrow\ -4ac+4aC+b^2-2bB+4cA-4AC+B^2=\ m^2$

Here we can insert any value for b and m and get deterministic results.

For example, for N = 667, x = 10, we have: $b^2-4ac+28a-12b+24c-m^2-132=0$

If $b=30$ and $m=12$ we get: $-4ac+28a+24c+264=0$. The solution set for 6 values is shown below.

$(a,b,c)\ (-26,10,8)\ \Longrightarrow 32x^2-4x-1\quad \therefore x=\frac{1}{4}\ \ or\ -\frac{1}{8}$

$(a,b,c)\ (-10,10,9)\ \Longrightarrow 16x^2-4x-2\quad \therefore x=\frac{1}{2}\ \ or\ -\frac{1}{4}$

$(a,b,c)\ (-2,10,11)\ \Longrightarrow 8x^2-4x-3\quad \therefore x=\frac{1}{1}\ \ or\ -\frac{1}{2}$

$(a,b,c)\ (2,10,15)\ \Longrightarrow 4x^2-4x-8\quad \therefore x=\frac{2}{1}\ \ or\ -\frac{1}{1}$

$(a,b,c)\ (4,10,23)\ \Longrightarrow 2x^2-4x-16\quad \therefore x=\frac{4}{1}\ \ or\ -\frac{2}{1}$

$(a,b,c)\ (5,10,39)\ \Longrightarrow x^2-4x-32\quad \therefore x=\frac{8}{1}\ \ or\ -\frac{4}{1}$

Such solutions are a closed set in which x and the complement of x are all solutions.

If we substitute x with 10 for all the solutions, we will find none equating to 667 and its multiples. However, if we try to find solutions for x such that $ax^2+bx+c\equiv 0(mod\ N)$ we get the prime solutions, for example.

$(a,b,c)\ (4,10,23)\ \Longrightarrow 2x^2-4x-16=667\times 2\quad \therefore x=\ 27\ \ or\ -25$

For $(a,b,c)\ (4,10,23)\ \Longrightarrow 2x^2-4x-16\quad \therefore x=\frac{4}{1}\ \ or\ -\frac{2}{1}$

Using $x=27$, for the solutions of x to find p and q, we get:

$p=27\times 1-4=23,\quad q=27\times 1+2=29$

But finding solutions such that $ax^2+bx+c\equiv 0(mod\ N)$ takes time especially when resolving large numbers, since current methods require us to factorise N first.

Here we put C to be C-N for f(x).

$f(x)$: $(A-a)x^2+(B-b)x+(C-N-c)\equiv 0(mod\ N)$

Using N=667, x=10, we get.

$b^2-4ac-2640a-12b+24c-m^2+126^2=0$

Let $b=-24$ and $m=10$, we get: $-4ac-2640a+24c+16640=0$

As for the solution sets, we have:

$(a,b,c)\ (206,-24,-659) \implies -200x^2+30x-1 \quad \therefore x=\frac{1}{20}\ \ or\ \ \frac{1}{10}\ ,\quad f(10)=-19701$

$(a,b,c)\ (-194,-24,-661) \implies 200x^2+30x+1 \quad \therefore x=-\frac{1}{20}\ \ or\ \ -\frac{1}{10}\ ,\quad f(10)=20301$

And if we take all the solutions of $f(10)$ we have:

$$S=\begin{Bmatrix}-9702,10302,-4704,5304,-2208,2808,-240,840,-720,1320,-1710,2310,\\ -3705,4305,-525,1125,-150{,}750,600,-19701,20301\end{Bmatrix}$$

We had no $f(10)=667$ or its multiples.

Assuming S is a group and using the preliminary laws of algebra, a group is said to be associative, commutative and contains and identity element as well as an inverse. In our case, let the identity be the prime factor we want to find. Then we have $f(x)*f'(x)=e$ whereby * is addition.

Examples:

$20301*840=29\times729$ $\qquad$ $10302*2808=23\times570$

$5304*1320=23\times288$ $\qquad$ $2808*-720=29\times72$

$840*-150=23\times30$

Therefore $p=gcd(e,N)$ $\qquad$ ALGORITHM 2

If for $f(10)*f(10)=e\ \mid gcd(e,N)=p$ was found, then we can get the same solutions for $p(10)*p(10)=e$.

Let's take the situation in (A9) whereby.

If $b=30$ and $m=12$ we get: $-4ac+28a+24c+264=0$.

$(a,b,c)\ (-26,10,8) \implies 32x^2-4x-1 \quad \therefore x=\frac{1}{4}\ \ or\ \ -\frac{1}{8}\quad ,\quad p(10)=39\ \ or\ \ 81$

$(a,b,c)\ (-10,10,9) \implies 16x^2-4x-2 \quad \therefore x=\frac{1}{2}\ \ or\ \ -\frac{1}{4}\quad ,\quad p(10)=19\ \ or\ \ 41$

$(a,b,c)\ (-2,10,11) \implies 8x^2-4x-3 \quad \therefore x=\frac{1}{1}\ \ or\ \ -\frac{1}{2}\quad ,\quad p(10)=9\ \ or\ \ 21$

$(a,b,c)\ (2,10,15) \implies 4x^2-4x-8 \quad \therefore x=\frac{2}{1}\ \ or\ \ -\frac{1}{1}\quad ,\quad p(10)=8\ \ or\ \ 11$

$(a,b,c)\ (4,10,23) \implies 2x^2-4x-16 \quad \therefore x=\frac{4}{1}\ \ or\ \ -\frac{2}{1}\quad ,\quad p(10)=6\ \ or\ \ 12$

$(a,b,c)\ (5,10,39) \implies x^2-4x-32 \quad \therefore x=\frac{8}{1}\ \ or\ \ -\frac{4}{1}\quad ,\quad p(10)=2\ \ or\ \ 14$

Then for $p(10)*p(10)=e$ where * is addition we have:

$39*19=58\ =29\times2$ $\qquad$ $81*11=92\ =23\times4$

$2*21=23\ =23\times1$ $\qquad$ $14*9=23\ =23\times1$

Even if we insert any pair of (b, m) to find (a, c), it's still not enough to give us some desired solutions. Some solutions will be omitted from the mix. So to find the whole solutions, firstly we insert the pair (b, m) such that they are small enough. After this we obtain the pairs of (a, c). Then using any pair of (a, c), we substitute into the original equation so that we find the pairs of (b, m). We can then combine them to find the pairs of (a, b, c).

Example:

$N = 667, \ \ x = 10$

$f(x) = \ b^2 - 4ac + 28a - 12b + 24c - m^2 - 132$

Let $b = 30, \quad m = 12$

$\Rightarrow -4ac + 28a + 24c + 264 = 0$

$(a, c) = \ (10, 34)(-30, 4)(-6, -2)(8, 61)(-102, 6)(12, 25)(33, 11)(114, 8)(18, 16)(-21, 3)$

$(2, -20)(4, -47)(0, -11)(60, 9)(42, 10)(5, -101)(24, 13)(15, 19)(9, 43)(-3, -5)(3, -29)$

$(7, 115)\ (-48, 5)\ (-12, 1)$

Using $(a, c) \ = \ (10, 34)$

$\Rightarrow \ b^2 - 12b - m^2 - 396 = 0$

$(b, m) = (30, \pm 12)\ (45, \pm 33)\ (-103, \pm 107)\ (27, \pm 3)\ (37, \pm 23)\ (-50, \pm 52)\ (-15, \pm 3)$

$(115, \pm 107)(62, \pm 52)\ (-18, \pm 12)(-33, \pm 33)(-25, \pm 23)$

CASE III

Here we try to solve in a different manner from the rest. Previously, the solutions were based on the derivative of a function but here we solve using integration instead. Rather than using roots to solve a function, we propose that we can find similar solutions even if we use the area under a cave, as long as it is bound within some roots.

Let's take for example, four situations in which $f(x)$ gave us prime solutions.

$f(x) = 4x^2 + 24x + 27 : \ \ x = \frac{-3}{2} \ \ or \ \ \frac{-9}{2}$ (V1)

$f(x) = 20x^2 - 169x + 357 : \ \ x = \frac{17}{4} \ \ or \ \ \frac{21}{5}$ (V2)

$f(x) = x^2 + 32x + 247 : \ \ x = \frac{-13}{1} \ \ or \ \ \frac{-19}{1}$ (V3)

$f(x) = 16x^2 - 112x + 187 : \ \ x = \frac{17}{4} \ \ or \ \ \frac{11}{4}$ (V4)

From previous examples, we noticed that for most functions in which we got p,q = 1 or N, if we added or subtracted such functions from that in which we get p,q not equal to N or 1, we would derive deterministic prime solutions. Let's now apply this for the integration, ignoring the constant. Let A_n be the area between one of the deterministic roots against that when x is the base value.

Using V1

$\int 4x^2 + 24x + 27 = \frac{4}{3}x^3 + \frac{24}{2}x^2 + 27x$

Meaning the area can be between $x = \frac{-3}{2}$, $x = 10$ or $x = \frac{-9}{2}$, $x = 10$

For the area bound between $x = \frac{-3}{2}$, $x = 10$ we get: $A_1 = \frac{8464}{3}$

For the area bound between $x = \frac{-9}{2}$, $x = 10$ we get: $A_2 = \frac{8410}{3}$

We notice that $gcd(8464, 667) = 23$ and $gcd(8410, 667) = 29$

And using V3

$\int x^2 + 32x + 247 = \frac{1}{3}x^3 + \frac{32}{2}x^2 + 247x$

Meaning the area can be between $x = \frac{-13}{1}$, $x = 10$ or $x = \frac{-19}{1}$, $x = 10$

For the area bound between $x = \frac{-13}{1}$, $x = 10$ we get: $A_1 = \frac{16928}{3}$

For the area bound between $x = \frac{-19}{1}$, $x = 10$ we get: $A_2 = \frac{16820}{3}$

We notice that $gcd(16928, 667) = 23$ and $gcd(16820, 667) = 29$

And using V4

$\int 16x^2 - 112x + 187 = \frac{16}{3}x^3 - \frac{112}{2}x^2 + 187x$

Meaning the area can be between $x = \frac{17}{4}$, $x = 10$ or $x = \frac{11}{4}$, $x = 10$

For the area bound between $x = \frac{17}{4}$, $x = 10$ we get: $A_1 = \frac{4232}{3}$

For the area bound between $x = \frac{11}{4}$, $x = 10$ we get: $A_2 = \frac{4205}{3}$

We notice that $gcd(4232, 667) = 23$ and $gcd(4205, 667) = 29$

Using the above scenarios, we notice that the area bound between x=base and x=root will give a prime solution which is factor of N.

$\int (6-k)x^2 + (6 + 10k - h)x + (7 + 10h) = \frac{(6-k)}{3}x^3 + \frac{(6+10k-h)}{2}x^2 + (7 + 10h)x$

$A(10) - A(x_i) = \mu \ : \ \gcd(N, \mu) = p$

In this situation we try to solve using the simplest possible solutions. One way is to input the value of $A(x_i)$ then iterate. In one situation we found $A\left(\frac{-9}{2}\right) = 0$ for V1, hence we can assume that such a value must exist for some $A(x_i)$.

Let $f(x) = (6-k)x^2 + (6 + 10k - h)x + (7 + 10h)$

$\int f(x) = \frac{(6-k)}{3}\, x^3 + \frac{(6+10k-h)}{2}\, x^2 + (7+10h)\, x$ and if equal to zero we obtain.

$x\left[\frac{(6-k)}{3}\, x^2 + \frac{(6+10k-h)}{2}\, x + (7+10h)\right] = 0$ and to remove the fractions, we get.

$2(6-k)x^2 + 3(6+10k-h)x + 6(7+10h) = 0$ then as for the discriminant:

$3^2\,(6+10k-h)^2 - 4 \times 2 \times 6\,(6-k)(7+10h) = m^2$ of which m was found to be zero.

This implies that: $900k^2 + 9h^2 + 300kh + 1416k - 2988h - 1692 = 0$ of which solving it like this is more difficult as it resorts us to the first case scenario (CASE I). Hence, we combine this solution with the discriminant solution of $f(x)$ to reach a consensus.

$900k^2 + 9h^2 + 300kh + 1416k - 2988h - 1692 = 0$ (T16)

$100k^2 + h^2 + 20kh + 148k - 252h - 132 = 9m^2$ (T17)

$(T17) - 9(T16)$ results in: $-120kh - 84k + 720h + 504 = 9m^2$

where $gcd(-120, -84, 720, 504) = 12$

when $m = 12$ $(k, h) = (2, 2)$ $(42, -1)$

when $m = 2$ $(k, h) = (7, -1)$

when $m = 10$ $(k, h) = (31, -1)$

when $m = 16$ $(k, h) = (70, -1)$

Using the above cases, we notice that the value of h remained constant regardless of the value of m. And when letting $h = -1$, we got $36k - 216 = m^2$ which gave us a number of parametric solutions sets. One of them was $(k, m) = (144r^2 + 48r + 10, -72r - 12)$.

In the next scenario we notice an anomaly in the sense that if we get some value of k, m and h then substitute them in $f(x)$ we get an $f(x)$ which does not satisfy some of the requirements we stated should happen for a full factorization, but even so, if we continue with the formulas stated we return with the correct value for p.

Example:

Given $(k, m) = (144r^2 + 48r + 10, -72r - 12)$.

$h = \frac{9m^2+84k-504}{720-120k}$, $x_1 = \frac{-(6+10k-h)+m}{2(6-k)}$, $x_2 = \frac{-(6+10k-h)-m}{2(6-k)}$

When $r = 0, k = 10, m = -12, h = \frac{-17}{5}, x_1 = \frac{607}{40}, x_2 = \frac{487}{40}$

It results in $p_1 = 40 * 10 - 607 = -207$ such that $gcd(-207, 667) = 23$

And $p_2 = 40 * 10 - 487 = -87$ such that $gcd(-87, 667) = 29$

Also if we use $A(10) - A(x_2) = \mu$ we get $6\mu = -9039.27825$ and expanding the number, we find out that the $gcd(-903927825, 667) = 29$ which confirms our proposition.

ALGORITHM 3