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 it on the RSA-numbers defined by the RSA challenge. We analyse the time spent on factoring to check the speed and efficiency of the 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 cryptographical 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.

CASE I

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. We will refer to this paper written by on solving conic sections.

Note: Since this one is a conic section and to find the values of h,k we will need to factorise N using SIQS or ECM, hence this one will not have an algorithm.

CASE II

Given $f(h,k,m)$: $100k^2+20hk+h^2+148k-252h-132=m^2$ Let $10k+h+m=34$ if we solve it, 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.

Let $M > x$

Note: This one will have an algorithm. Name it ALGORITHM 1.

CASE III

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, $(B+xk-h)^2-4(A-k)(C+xh)=m^2$

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

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 $2m + o = 72$, $10k + h + n = 126$

For $k = 2$

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

For $k = 3$

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

Note: ALGORITHM 2

CASE IV

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.

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

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

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

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

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

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

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.

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

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

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.

The solution to this problem we have two options.

OPTION 1

We assume that the reason why some of the values did not give us deterministic results was due to the fact that both a and c must have been fractions, hence by Diophantine solving were skipped. If we take that both a and c are of the form $\frac{i}{\theta}$ and since we want to get complete integer solutions we get the general solution:

$-4ac + \theta \times 28a + \theta \times 24c + \theta^2(b^2 - 12b - m^2 - 132) = 0$

In which we continuously iterate theta until we find p=gcd(f(x), N)

Note: ALGORITHM 3

OPTION 2

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

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

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:

$20301 * 840 = 29 \times 729$

$10302 * 2808 = 23 \times 570$

$5304 * 1320 = 23 \times 288$

$2808 * -720 = 29 \times 72$

$840 * -150 = 23 \times 30$

Therefore p=gcd(e, N)   ALGORITHM 4

OPTION 3

Combine option 1 and 2. Leave this one tho.