International Journal of Mathematics and Statistics Studies

Vol. 1, No.1, March 2013, pp. 39-44

Published by European Centre for Research Training and Development, UK (www.ea-journals.org)

FACTORIZATION OF LARGE COMPOSITE INTEGER

Aldrin W. Wanambisi[1*] Shem Aywa[2], Cleophas Maende[3], Geoffrey Muchiri Muketha[4]

1. School of Pure and Applied Science, Mount Kenya University, P.O box 553-50100, Kakamega, Kenya.
2. Dept of Mathematics, MasindeMuliro University of Science and Technology, P.O Box 190-50100, Kakamega, Kenya.
3. School of Post graduate studies, Mount Kenya University, P.O box 553-50100, Kakaemega, Kenya.
4. Dept of Computer Science, Meru University of Science and Technology, P.O Box 672 - 60200, Meru, Kenya.

Abstract: *In this paper we practically deal with the problem of factorizing large integers. The various algorithms that have been proposed are not efficient that is they do not run in polynomial time. We use the algebraic approach proposed by Wanambisi et al [1]. We define a large integer based on the number of digits and seek to decompose these numbers based on place values.*

Keywords: Integer factorization, Composite integer, Algebraic approach, Primes, Algorithms, Prime differences

1.0 Introduction

This section discusses the concept of a large composite, the basic concepts, the trends in differences in primes and the basics of the algebraic approach.

1.1 Large composite integers

The question of what is and what is not a large integer cannot be addressed at once. We pose the question how large is a large integer. Considering the RSA-2048 which is a 232- digit is a large enough integer to be referred to as a large integer? The answer is no! What about if we added to it just one more decimal digit then clearly we get a larger number and adding a digit still gives us still larger integers. The first number with more than a thousand digits known to be prime was M_{4253}. The largest number on that list was found on 2003-Nov-17. This number has 6, 320, 430 digits.

Now large composite integers are hard to factor or simply there is no known algorithm that can factor them in polynomial time. This is of course the basis of most cryptosystem. So this problem acts as a padlock that locks up most of the world's money and secrets. Since no one break down the integer problem in the background then the information in the background cannot be accessed.

Is there therefore no answer to the questions: what is the largest integer known? Or is there a limit to the size of a large integer which can be used in cryptographic primitives?

Definition 1: In this paper a composite integer is a number $m = pq$ where p and q are large primes.

Definition 2: A large prime p is any number divisible only by one and itself with number of decimal digits (l) greater than or equal to 5.

Definition 3: A large composite integer is a composite integer $m = pq$ where both *p*, *q* are of decimal digit length $l \geq 5$.

Definition 4: Large composite integer factorization is the process of factorizing a large composite integer.

2.0 The Basic concepts

2.1 Theorem (Fundamental Theorem of Arithmetic)

Every number greater than 1 factors into a product of primes $n = p_1 p_2 \dots p_s$. Further, writing the primes in ascending order $p_1 \leq p_2 \leq \cdots \leq p_s$ makes the factorization unique.

We will break the proof of the Fundamental Theorem into a sequence of Lemmas [5]

2.2 Lemma (Euclid's Lemma)

If p is a prime and $p|ab$, then $p|a$ or $p|b$.

Proof

Assume that $p|ab$. If $p|a$ then we are done, so suppose that it does not. Let *c* = gcd (*p, a*). Note that *c* > 0, and that $c|p$ and $c|a$. Since $c|p$ we have that *c* = 1 or *c* = *p*. If *c* = *p* then $p|a$, which we assumed was not true. So we must have *c* = 1. Hence gcd (*p, a*) = 1 and $p|ab$. Thus $p|b$ [5].

2.3 Lemma

Let p be prime. Let $a_1, a_2, \dots, a_n, n \geq 1$, be integers. If $p|a_1 a_2 \dots a_n$, then $p|a_i$ for at least one $i \in \{1,2,..n\}$.

Proof

We use induction on *n*. For the $n = 1$ base case the result is clear. For the inductive step, assume the inductive hypothesis: that the lemma holds for *n* such that $1 \leq k \leq n$. We must show that it holds for $n = k + 1$. Assume that p is prime and that $p|a_1 a_2 \dots a_k a_{k+1}$. Write $a_1 a_2 \dots a_k$ as *a*, and a_{k+1} as *b*. Then $p|a$ or $p|b$. If $p|a = a_1 \dots a_k$ then by the induction hypothesis, $p|a_i$ for some $i \in \{1,2,..k\}$. If $p \mid b$ then $p|a_{k+1}$. So we can say that $p|a_i$ for some $i \in \{1,2,..k+1\}$. This verifies the lemma for $n = k + 1$. Hence by mathematical induction, it holds for all $n \geq 1$[5].

2.4 Lemma (Fundamental Theorem, Existence)

If *n* > 1 then there exist primes $p_1, \dots, p_s$, where $s \geq 1$, such that $n = p_1 p_2 \dots p_s$ and $p_1 \leq p_2 \leq \cdots \leq p_s$.

Proof

We will use induction on *n*. The base step is *n* = 2: in this case, since 2 is prime we can take *s* = 1 and $p_1 = 2$. For the inductive step, assume the hypothesis that the lemma holds for $2 \leq k \leq n$; we will show that it holds for $n = k + 1$. If $k + 1$is prime then *s* = 1 and $p_1 = k + 1$. If $k + 1$ is composite then write $k + 1 = ab$ where $1 < a < k + 1$ and $1 < b < k + 1$. By the induction hypothesis there are primes $p_1, \dots, p_u$ and $q_1, \dots, q_v$ such that $a = p_1 \dots p_u$and $b = q_1 \dots q_v$. This gives that *k* + 1 is a product of primes $k + 1 = ab = p_1 p_2 \dots p_u q_1 q_2 \dots q_v$, where *s* = *u* + *v*. Reorder the primes into ascending order, if necessary. The base step and the inductive step together give us that the statement is true for all *n* > 1 [5].

2.5 Lemma (Fundamental Theorem, Uniqueness)

If $n = p_1 p_2 \dots p_s$ for $s \geq 1$ with $p_1 \leq p_2 \leq \cdots \leq p_s$, and also $n = q_1 q_2 \dots q_t$ for $t \geq 1$ with $q_1 \leq q_2 \leq \cdots \leq q_t$, then *t* = *s*, and $p_i = q_i$ for all $i \in [1, s]$ [5].

This result can as well be proved by mathematical induction [5].

3.0 **The algebraic approach**

3.1 Twin primes

Twin primes have a difference of 2. Thus if p and q are any two consecutive twin primes then the product pqcan be given by $(n-1)(n+1)$. Now consider a composite integer $m = pq$ that is a product of twin primes. Then

$$(n-1)(n+1) = m$$
$$n^2 + 1 = m$$
$$n = \pm\sqrt{m+1}$$

Taking the appropriate value of *n* to be approximately *p* then *q* follows at once [1].

3.2 Blum integers

A Blum integer is a composite integer that is a product of two primes both congruent to 1 modulo 4. The difference between any two such consecutive primes is 4. Thus if p and q are two primes both congruent to 1 modulo 4, then

$$|p - q| = 4$$

Therefore if we let p to be say n-2 and q to be say n+2, then a Blum integer $m = pq$

$$(n-2)(n+2) = m$$
$$n^2 - 4 = m$$
$$n^2 = m + 4$$
$$n = \pm\sqrt{m+4}$$

An appropriate value of *n* gives us approximate p and q hence the prime factors of our composite integer [1].

3.3 Other composite integers

In this section we consider other composite integers which are products of primes with differences of 6, 8, 10, 12, 14, 16 e.t.c. now just like the case of Twin and Blum integers, if $m = pq$ is any composite integer, with difference of say 6, then clearly

$$|p - q| = 6$$

Thus if we let $p = n - 3$ and $q = n + 3$ then a composite integer $m = pq$

$$(n-3)(n+3) = m$$
$$n^2 - 9 = m$$
$$n^2 = m + 9$$

Therefore $n = \pm\sqrt{m+9}$

An appropriate value of *n* gives approximate values of the prime factors of the composite integer [1].

For any general composite integer $m = pq$, with the difference $|p - q|$ the prime factors are approximately

$$n = \pm\sqrt{m + \left(\frac{|p-q|}{2}\right)^2}$$

Using the relation (3), we can obtain prime factors of composite integers on condition that the integer is a product of two primes no matter how large. This method reduces the steps that lead to factorization of a large integer to say polynomial time. If for example we take the case of RSA cryptosystem which is based on the prime factorization problem in which the primes are relatively close, the steps taken to arrive at the prime factors are greatly reduced. This then can be done in polynomial time [1].

4.0Cryptographic security

The Rivest, Shamir, Adleman (RSA) cryptosystem is an example of a *public key cryptosystem*. RSA uses a *public key* to encrypt messages and decryption is performed using a corresponding *private key*. We can distribute our public keys, but for security reasons we should keep our private keys to ourselves. Just like the RSA, most of the existing cryptographic primitives draw their security from the hardness of composite integer factorization. Say for large integer $m = pq$, the choice of numeric values for p and q for the remainder of this paper, always bearing in mind that they have been chosen for illustrative purposes only. Refer [2], [3] and [4] for in-depth discussions on the security of RSA, or consult other specialized texts.

For RSA, we can compute the value $\varphi(m)$ for arbitrarily large prime numbers p and q, this can take an enormous amount of time. Indeed, the private key can be quickly deduced from the public key once you know $\varphi(m)$, so it is an important part of the security of the RSA cryptosystem that $\varphi(m)$ cannot be computed in a short time, if only m is known. On the other hand, if the private key or the factorization of n is available, we can compute $\varphi(m) = (p-1)(q-1)$ in a very short time [6].

5.0 Factoring large composite integers

This section contains results of this research. We shall use the terms factorization and decomposition interchangeably because we are dealing with numbers that products of only two primes.

5.1 Proposition 1

Let $m = pq$ be a large composite integer of decimal digit length $l \geq 5$ and difference $|p - q|$ denoted as $d \geq 0$, then the prime factors p and q are approximately $p = q = \pm\sqrt{m + \left(\frac{d}{2}\right)^2}$

Consider the integer $n = 21$, though not necessarily large as suggested in this paper but for purposes of illustrating the above proposition we have the prime factors as:

$$p = \pm\sqrt{21 + \left(\frac{4}{2}\right)^2}$$

$$p = \pm\sqrt{25} = \pm 5$$

Now taking the positive square root and adding or subtracting $\frac{d}{2}$ we obtain

$p = 3$ and $q = 7$.

5.2 Large integers that are used in cryptographic primitives

5.2.1 Mersenne primes

A prime number of the form $M_n = 2^n - 1, n \geq 2,$ is a Mersenne prime. Consider the RSA public key with $m = p.q$ where p and q are both Mersenne primes. With $m = 68718821377$. We want to find the prime factors of *m*, applying the proposition 1 we have

$$p = \pm\sqrt{68718821377 + \left(\frac{393216}{2}\right)^2}$$

Evaluating this and adding and subtracting the value $\left(\frac{393216}{2}\right)$ yields values approximately

$p = 524287$ and $q = 131071$

With the factorization determining $\varphi(m) = (p-1)(q-1)$ is easy.

5.2.2 Blum integers

Unlike Mersenne primes, Blum integers have predictable differences between consecutive primes, to be precise 4. Any other difference will be a multiple of 4. Applying proposition 1 on $m = pq$ such as $m = 62393801$

Here we check for the difference 4 and if it doesn't give us the solution, we return and pick the next multiple of 4; we do this until we obtain the difference that gives a solution. That is in this case 40. Substituting we obtain:

$$p = \pm\sqrt{62393801 + \left(\frac{40}{2}\right)^2}$$

Evaluating this equation we get the values of *p* and *q* as 7879 and 7919.

5.2.3 Twin primes

Just like the Blum integers, differences between consecutive Twin primes are predictable and factoring is similar to that of Blum integers replacing the difference 4 with 2.

6.0 Conclusion

Assuming that *m* is the product of two odd primes *p* and *q* between 1 and *N*, the Algebraic approach which utilizes the prime differences as presented in section 3.3 makes no more than $\frac{N}{2}$steps since there are $\frac{N}{2}$ differences. Hence takes $O\left(\frac{N}{2}\right)$steps. The algorithm can take even fewer steps the ones given above since the differences repeat themselves. A study I have carried out on the prime differences reveals that in the first 1000 primes, the maximum difference is 14; now taking the even numbers between 1 and 14 we have

only 7! This shows that if we are dealing with the products of consecutive primes then it will take a record maximum 7 steps to achieve the prime factors!
The complication comes in when the primes are not consecutive. This means that each of the differences has to have its multiples worked out and each of them tested to establish the difference between each prime. Since the multiples are even then all even numbers can be checked between 1 and *N*. This brings the number of steps to a maximum of $\frac{N}{2}$.

Authors' contributions

All authors contributed to the conceptualisation of the paper which is an excerpt of W.A.W. PhD dissertation. W.A.W. did the initial review, the selection of abstracts, and the identification of papers to be included in the final review. SA, CM and GMM contributed to the assessment of papers and reviewed the results of the analysis. W.A.W drafted the manuscript, and all authors contributed to its completion.

Acknowledgements

Thanks to those who have been instrumental in the success of this research: The Masinde Muliro University of Science and Technology, the adviser, for participating in this research study and for their support of this study.

References

Wanambisi A.W., Aywa S., Maende C. Muketha G. M., *Algebraic approach to composite integer factorization,* International Journal of Mathematics and Statistics Studies, March 2013

1. J. Menezes, P. C. van Oorschot, and S. A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, Boca Raton, FL, USA, 1996.
2. D. R. Stinson. *Cryptography: Theory and Practice*. Chapman & Hall/CRC, Boca Raton, USA, 3rd edition, 2006.
3. W. Trappe and L. C. Washington. *Introduction to Cryptography with Coding Theory*. Pearson Prentice Hall, Upper Saddle River, New Jersey, USA, 2nd edition, 2006.
4. J. Hefferon *Elementary Number Theory*, St Michael's College, 2003-Dec
5. Wesstein, Eric W. *RSA Encryption*. From Mathworld, an online encyclopedia. April, 2001. Available: http://mathworld.wolfram.com/RSAEncryption.html

* E-mail of the corresponding author: wawanambisi@gmail.com