Primality and factorization - a computer challenge

S. P. Khare
Ram Bilas Misra

Abstract

The second author had a long association with Satgur Prasad Khare, as his student both in graduate and postgraduate classes at the University of Allahabad, Prayagraj (India) for over five decades. Khare worked in the Theory of Numbers under the supervision of a celebrated Indian Number Theorist (Prof. H. Gupta). Even after his retirement, he continued his researches in Number Theory and, recently, he developed an interest in cryptography. Due to the lack of senior workers in his field, he trusted his own teacher and often used to send his researches to the author for review. He met a tragic and untimely end of his life on 30 April 2021 leaving a lot of his unpublished work that needs to be highlighted. He was completing a research monograph on Number Theory before his sudden cardiac arrest. The present article deals with the *primality and factorization* of RSA numbers including the latest test called the *AKS Primality Test*. Results of number theory are comprehensible by even the non-specialists but their proofs are most challenging.

Keywords: mathematical computation, primality and factorization, RSA numbers.

§ 1. Preliminaries

Let N, Z and P denote the sets of natural numbers, integers and prime numbers:

$$\mathrm{N} = \{ 1, 2, 3, \ldots\}, \qquad \mathrm{Z} = \{ 0, \pm 1, \pm 2, \pm 3, \ldots \},$$

$$\mathrm{P} = \{2, 3, 5, 7, 11, \ldots \}.$$

Unless stated otherwise, small (Roman) letters shall denote positive integers. If $n = a\,b$, $a, b \in \mathrm{N}$, $a \neq 1$, $b \neq 1$, then a and b are called *proper divisors* of n. So every natural number has either one, two or more divisors. Numbers having only two divisors are known as *prime numbers* (or briefly primes), while others as *composite numbers*. As per the *Fundamental theorem of arithmetic*, every natural number is expressible as a product of (distinct powers of) primes uniquely:

$$n = (p_1)^{a_1}.(p_2)^{a_2}.(p_3)^{a_3}\ldots(p_k)^{a_k},$$

also called the canonical form of n.

1.1. Some concerning results/conjectures about primes

(i) There are two forms of primes viz. $4k + 1$ and $4k - 1$;

(ii) Number of primes is infinite;

(iii) There are infinite pairs of the form $(p, p + 2)$;

(iv) There are infinitely many primes of the form $(p, p + 2, p + 4)$ and $(p, p + 4, p + 6)$;

(v) Does there exist any (simple) general formula for n^{th} prime?

(vi) Is there any formula satisfying $p_{n+1} = (p_n)^2 + 2$?

(vii) Is there a rule determining another prime greater than the given prime p?

(viii) How many primes may exist, which are not exceeding a given number x?

(ix) If $n \geq 6$, then n is the sum of two odd primes which still requires a proof – called *Goldbach conjecture*.

(x) There is always a prime between n^2 and $(n + 1)^2$ but it requires proof.

Note 1.1. Though number theory is classical, problems are easy to understand but really challenging to prove.

Given two numbers a and b, Euclid found two more numbers: q and r such that

$$a = q\,b + r; \qquad 0 \leq r < b\,.$$

This is known as *Euclidean Algorithm* – The first algorithm is known in Number theory (Computer Science). In particular, it is an algorithm for finding the *greatest common divisor* of two numbers a and b. Euclid devoted part of his life to the prime numbers and indivisibility topics that belong unambiguously to the Theory of Numbers and are basic results. He also gave the first proof of *infinitude of primes*.

1.2. Some basic primality results

(i) If the unit place of a number is either 0, 2, 4, 6 or 8, it is divisible 2;

(ii) If the first two digits (counted from right) are divisible by 4, the whole number is divisible 4;

(iii) More generally if first k digits number of a given number is divisible by 2^k or 5^k the whole number is divisible by 2^k or 5^k;

(iv) If sum of digits is divisible by 3 or 9, number is divisible by 3 or 9;

(v) If difference of sum of odd digits and sum of even digits is divisible by 11 the whole number is divisible by 11;

(vi) Divisibility by 7 can be checked by taking modulus of 1001. For example,

$$598236779 \equiv 598 - 236 + 779 = 1141 \equiv 0 \pmod 7.$$

Hence 598236779 is divisible by 7.

§ 2. Pierre de Fermat (1601 – 65 A.D.)

The celebrated mathematician never published his works, especially those on Number Theory. His works were found in his letters to mathematicians and private marginal notes. He wrote down nearly no proof in Number Theory. He also had no models in this area. He repeatedly made use of Mathematical Induction, introducing the method of *infinite descent*.

Fermat's first interest was in Perfect Numbers of the form $2^h (2^{h+1} - 1)$, where $2^{h+1} - 1$ is a prime. The largest known perfect number is

$$2^{57885160}. (2^{57885161} - 1),$$

which is a 34 million digits long number. Till now no odd perfect number is known. No proof is found in known correspondences made during 1636 A.D. onward. From 1643 A.D., his interests shifted to Diophantine Problems.

Fermat defined some numbers of the form, called Fermat's numbers:

$$F_n = 2^{(2^n)} + 1.$$

Particularly,

$$F_o = 3, \quad F_1 = 5, \quad F_2 = 17, \quad F_3 = 257,$$

$$F_4 = 65537, \quad F_5 = 4294967297.$$

He conjectured that these numbers are primes. In 1732 A.D., Euler proved that F_5 is composite and factorized it as

$$F_5 = 641 \times 6700417.$$

In 1880 A.D., Landry proved that F_6 is also not prime. It is factorized as

$$F_6 = 274177 \times 67280421310721.$$

No prime has been found beyond F_4.

Fermat's numbers satisfy the recurrence relations:

(i) For $n \geq 1$, $F_n = (F_{n-1} - 1)^2 + 1$;

(ii) For $n \geq 2$, $F_n = (F_{n-1})^2 - 2 (F_{n-2} - 1)^2$;

(iii) $F_{n-1} = 2^{(2^n)}. F_o \ldots F_{n-2}$;

(iv) $F_n = 7 \pmod{10}$, for all $n \geq 2$.

Remark. It is very likely that Fermat's Primes are finite but this requires proof.

2.1. Another big challenge in the theory of numbers / Computer Science

Considering the above statement to be true then Fermat' Numbers are composite and have the divisor of the form $k \cdot 2^{n+2} + 1$, where k is some positive integer.

The numbers of the form $2^p - 1$ (called *Mersenne primes*) are primes for

p = 2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2203, 2281, 3217, 4253, 4423, 9689, 9941, 11213, 19937, 21701, ….

The number $2^{21701} - 1$ is a prime having 6533 digits.

On 18th February 2005, the German medical doctor (Dr Martin Nowak) discovered $2^{25964951} - 1$, having 7816230 digits, as the largest prime number accounted till then.

Next, in March 2007, $2^{32582657} - 1$ was known as the largest known prime number having 98,08,358 digits. Assuming a page containing 60 lines each with 8 characters about 2000 pages were needed to print it. Thereafter, the Electronic Frontier Foundation (U.S.A.) offered $100,000 USD prize to discover still the larger prime. It was found by Dr Cooper, Department of Computer Science, University of Missouri that $2^{57885161} - 1$ is a prime of 1,74,25,170 digits. It is the largest known prime till now worked out at Global Gross Root Super Computer performing 150 trillion operations per second.

Fermat biggest achievements are his theorem (of 1640 A.D.):

$$\text{If } p \text{ is a prime then } a^{p-1} \equiv 1 \pmod{p},$$

for any a relatively prime to p, i.e. $(a, p) = 1$; but its converse is false, i.e. If $a^{n-1} \equiv 1 \pmod{n}$, $(a, n) = 1$, then n need not be prime.

Example. $2^{560} \equiv 1 \pmod{561}$, while 561 is not a prime.

Hence, Fermat Theorem does not guarantee for a number n to be a prime.

2.2. In 1640 A.D., Fermat made two very important statements:

(i) If $(a, b) = 1$, then $a^2 + b^2$ is not divisible by any prime congruent to -1 (modulo 4);

(ii) but it is divisible by any prime congruent to 1 (modulo 4).

Fermat also made a conjecture in 1637 A.D. (known as Fermat's last Theorem):

$$x^n + y^n = z^n \text{ ; has no integral solution if } n \geq 3.$$

He claimed (on the last page of his book) to have discovered truly marvellous proof of his above theorem but did not detail it because of the narrow margin therein. Did Fermat really know it or did he just mislead the mathematical community? Some mathematicians claim that Fermat never knew it. The unsolved problem had been the most notable theorem in the history of mathematics. It simulated the development of algebraic number theory in the 19th century and its proof in the 20th century. Prior to its proof, it was classified as an open problem in the "Guinness World Records". The reason was the largest number of unsuccessful proofs. Proof of Fermat's last theorem has been, finally, settled down after 358 years of hard work and continued efforts of mathematicians. Ultimately, Sir Andrew John Wiles, FRS (born on 11th April 1953), a British Mathematician and Royal Society Research Professor at the University of Oxford, a specialist in Number Theory proved it. His proof is most notable for which he received the 2016 *Abel Prize*.

§ 3. Computation in Number Theory algorithm

An algorithm is a specific procedure for solving a well-defined computational problem. Such methods are as old as any recognizable mathematics. The oldest algorithm is *Euclidean algorithm* for finding the *greatest common divisor* of two numbers. In the 5^{th} and 6^{th} century A.D. an algorithm called *kuttaka* (pulverizer), i.e. finding the quantities whose existence is ensured with proof or correctness.

There arise two questions:

(i) Can we compute it, (ii) Can we compute it rapidly.

For example, can anyone test whether a number is prime or not if it is not split into two prime factors? Doing so rapidly matters that there are fast methods for testing primality but truly no fast algorithm.

However, on a different side, something may not be computable at all. For example, there is no Turing machine, which can solve all Diophantine Equations.

In view of the above considerations, *factorization* and *primality testing* are two important concepts in Computer Science / Mathematics. From a purely academic motivation, it is an intriguing question to ask.

How can we determine whether a given number is a prime or not? The next logical question is if a number is composite can we calculate its factors? These two questions are invariably related. If we can factorize a number, obviously it is not a prime. If we cannot, then it is a prime. Hence, the definition of primality is very much derivable from *divisibility* or *factorability*. When we program through known and developed primality test and factorization algorithm, it will be clear that primality and factorization are intertwined. They occupy two levels of computational difficulty.

The study and development of a *test for primality* have been focused on the fast and deterministic algorithm. While primality and factorization are studied in many branches of mathematics, much of the recent study has been sparred by advances in cryptography and Number Theory. This application has motivated the recent interest in efficient primality and factorization problems. Primality and Factorization theory has a greater impact on modern encryption methods and public-key systems. The computational complexity of these two problems is different. Primality is an efficient problem as 20^{th} Century saw the biggest revelation in the area of cryptography when asymmetric encryption systems were conceived independently in 1976 by Diffie and Hellman. It was considered that the public key cryptosystem would completely outdate the *symmetric system* but it was turned out to be a false perception. This has weaknesses and computing power of computer was increasing day by day.

3.1. RSA algorithm: In 1977, Ron *Rivest*, Adi *Shamir* and Leonard *Adleman* gave RSA algorithm based on the product of two large primes. The algorithm was assumed to be secure. Martin Gardener wrote an article 'A new kind of cipher that would take millions of years to break' in his *Mathematical Game* column in *Scientific American*. This is how cryptography works. Gardener issued a challenge to his readers. He printed a ciphertext and also provided the public key that had been used to encrypt it.

The chosen number was

N = 114 381 625 757 888 867 669 235 779 976 146 612 010 218 296 721 242 362 562 561 842 935 706 935 245 733 897 830 597 123 563 958 705 058 989 075 147 599 290 026 879 543 541.

The challenge was to factorize N into two prime numbers p and q and use them to find the plain text. The prize was USD 100. Gardener' challenge took seventeen (17 years) before the cipher would be broken. On April 26, 1994, a team of 600 volunteers announced the factors of

$$N = p\,q,$$

where

$$p = 32\ 769\ 132\ 993\ 266\ 709\ 549\ 961\ 988\ 190\ 834\ 461\ 413\ 177\ 642\ 967\ 992\ 942\ 539\ 798\ 288\ 533,$$

and

$$q = 3\ 490\ 529\ 510\ 847\ 650\ 949\ 147\ 849\ 619\ 903\ 898\ 133\ 417\ 764\ \ 638\ 493\ 387\ 843\ 990\ 820\ 577.$$

Using these values as private keys they were able to decipher the message which was a series of numbers. When converted into text it yielded the magic words "Squeamish Ossifrage", i.e. a large number.

3.2. Another RSA challenge was the factorization of 155 digits number. A group of sixteen researchers, in four months, on a cluster of 292 computers factorized 155 digits number into primes p and q each of 78 digits:

p = 102 639 592 829 741 105 772 054 196 573 991 675 900 716 567 808 038 066 803 341 933 521 790 711 307 779,

q = 106 603 488 380 168 454 820 927 220 360 012 878 679 207 958 575 989 291 522 270 608 237 193 062 808 643.

There was another USD 10,000 bounty on factorizing 174 digit number

N = 188 198 812 920 607 963 838 697 239 461 650 439 807 163 563 379 417 382 700 763 356 422 988 859 715 234 665 485 319

060 606 504 743 045 317 388 011 303 396 716 199 692 321 205 734 031 879 550 656 996 221 305 168 759 307 650 257 059.

This number is known as RSA 576 for having 576 digits - when written in binary form. This number was factorized at German Federal Agency for Information Technology Security in 2003. The two prime factors p and q are

$$p = 398\ 075\ 086\ 424\ 064\ 937\ 397\ 125\ 500\ 550\ 386\ 491\ 199\ 064\ 362\ 342\ 526\ 708\ 406\ 385\ 189\ 575\ 946\ 388\ 957\ 261\ 768\ 583\ 317,$$

and

$$q = 472\ 772\ 146\ 107\ 435\ 302\ 536\ 223\ 071\ 973\ 048\ 224\ 632\ 914\ 695\ 302\ 097\ 116\ 459\ 852\ 171\ 130\ 520\ 711\ 256\ 363\ 590\ 397\ 527.$$

On 2nd November 2005, RSA 640 with 193 digits carrying a challenge of USD 20,000, was factorized by F. Bahr, M. Boehm, J. Franke, T. Kleinjun. It took five months. The number is

$$N = 310\ 741\ 824\ 049\ 004\ 372\ 135\ 075\ 003\ 588\ 856\ 793\ 003\ 734\ 602\ 284\ 272\ 754\ 572\ 016\ 194\ 882\ 320\ 644\ 051\ 808\ 150\ 455\ 634\ 682\ 967\ 172\ 328\ 678\ 243\ 791\ 627\ 283\ 803\ 341\ 547\ 107\ 310\ 850\ 191\ 954\ 852\ 900\ 733\ 772\ 482\ 278\ 352\ 574\ 238\ 645\ 401\ 469\ 173\ 660\ 247\ 765\ 234\ 660\ 9.$$

One of the prime factors of the above number is

$$p = 163\ 473\ 364\ 580\ 925\ 384\ 844\ 313\ 388\ 386\ 509\ 085\ 984\ 178\ 367\ 003\ 309\ 231\ 218\ 111\ 085\ 238\ 933\ 310\ 010\ 450\ 815\ 121\ 211\ 816\ 751\ 157\ 9.$$

The above team also factorized 663 bits RSA challenge integer with 207 digits. The factorizations were done by algorithm number field sieve based on Elliptic Curve Factorization.

Remark. Smallest currently open challenge is RSA 704 worth bounty $30,000. It has 212 decimal digits and is definitely composite.

$$N = 740\ 375\ 634\ 795\ 617\ 128\ 280\ 467\ 960\ 974\ 295\ 731\ 425\ 931\ 888\ 892\ 312\ 890\ 849\ 362\ 326\ 389\ 727\ 650\ 340\ 282\ 662\ 768\ 919\ 964\ 196\ 251\ 178\ 439\ 958\ 943\ 305\ 021\ 275\ 853\ 701\ 189\ 680\ 982\ 867\ 331\ 732\ 731\ 089\ 309\ 005\ 525\ 051\ 168\ 770\ 632\ 990\ 723\ 963\ 807\ 867\ 100\ 860\ 969\ 625\ 379\ 346\ 505\ 637\ 963\ 59.$$

§ 4. AKS *primality* test

On August 6, 2002, Computer Scientist Prof. Manindra Agrawal of I.I.T., Kanpur (India) and his students Neeraj Kayal and Nitin Saxena achieved a remarkable result. They discovered an efficient and deterministic test for the primality of a number. The mathematical community is

fascinated by this result not only because it answers an open question of long-standing but also for the mathematics behind it is beautiful and easily accessible as compared to other modern research advances.

4.1. Primality over the fields $K(i)$, $K(\rho)$, $K(\sqrt{m})$ and modular arithmetic over Elliptic Curves are new areas of research leading to its vast application to communication theory. On the other hand, computation of primitive irreducible polynomials of degree > 33 and study of non-linear logic are burning issues of the secure communication of modern era to face the ongoing problems to be faced by the country.

Now we have the following challenges for computer scientists to work on the following problems, if we are thinking of up-gradation in technological advances in the country:

Challenge 1. To work on the problems concerning primes and their applications leading to modern cryptography.

Challenge 2. Users of RSA should pick up such a large value of *n* to secure important information that it may be impossible to break the encoded message on this planet even if a large number of supercomputers may be working on it. So is the case with the discrete *log* problem.

Because of the Application of Algorithm Number Field Sieve based on Elliptic Curve Factorization, factorization is not difficult as it was a couple of decades ago. That is why public-key systems are in use in the commercial Sectors of the Government as well as in Private Sectors. These are also used for authentication and recognition purposes in the public as well as private sectors.

Conclusively, this paper raises the question of whether our younger people working in this area may take up such challenges and utilize their knowledge in the development of indigenous cryptosystems so that academia may help the Government/PSUs in Secure Communications.

References

Cindy, T. (2010). *Fermat's Numbers*. University of Washington Math 414 Number Theory Notes.

Hardy, G. H. & Wright, E. M. (2008). *An introduction to the theory of numbers.* (7 ed.). David Burton Paperback.

Misra, R. B. (2010). *Basic mathematics at a glance*. Saarbrücken, Germany: Lambert Academic Publishers. ISBN 978-3-8433-8696-8.

Singh, S. (1999). *The code book: The science of secrecy from ancient Egypt to quantum cryptography.* Fourth Estate and Doubleday.

Stein, W. (2008). *Elementary number theory, a computational approach.* Springer.

Acknowledgement

The second author wishes to express his sincere thanks to the referee for his valuable comments and the editor of the journal for bringing such useful information to its inquisitive readers stimulating their interest to explore the research in this domain. The paper is dedicated to the memory of Prof. Peter K. Anderson who had been a senior colleague of the second author at the Divine Word University, Madang (PNG).

Authors

(Late) Dr. S. P. Khare
Ex Director, Joint Cipher Bureau, Govt. of India, New Delhi;

Prof. Dr. Ram Bilas Misra
Ex Vice-Chancellor, Avadh University, Ayodhya / *Faizabad* (*India*); *Research and Strategic Studies Centre, Lebanese French University, Erbil, KRG* (*Iraq*); 380 A, Gomti Nagar Extension, Sector 1, Lucknow – 226010, U.P. (India);
Email: rambilas.misra@gmail.com; misrarb1@rediffmail.com