The Evolution of Algorithms to find Prime Numbers

Maya Kaczorowski

Department of Mathematics and Statistics, McGill University, Burnside Hall, 1005 805, Sherbrooke Street West, Montreal, Quebec, Canada, H3A 2K6
Received: January 4, 2009 - Accepted: February 5, 2009

Prime numbers are what is left when you have taken all the patterns away. I think prime numbers are like life. They are very logical but you could never work out the rules, even if you spent all your time thinking about them.
- Mark Haddon, *The Curious Incident of the Dog in the Night-time*

Introduction

Prime numbers, and the deterministic formulas used to find them, have garnered considerable attention from mathematicians, professionals and amateurs alike. A prime number is a positive integer, excluding 1, whose only divisors are 1 and itself. For example, 23 is a prime number as it can only be divided by 1 and 23. A number that is not prime is called a composite number.

While prime numbers under 100 are fairly abundant, they become less frequent and difficult to find in a systematic manner as the digits in the number increase since they do not appear to follow a predictable distribution. So why do researchers keep studying them? For over 150 years, mathematicians have attempted to uncover a deterministic formula to identify prime numbers. If such a formula existed, all numbers could be factored relatively quickly using computers. Paradoxically, much of electronic data today is encrypted by taking advantage of the fact that it is difficult and time consuming for a computer program to factor a large composite number. A formula to find all prime numbers would be a significant breakthrough in mathematics, but severely detrimental to data security.

A simple algorithm to find prime numbers: The Sieve of Eratosthenes

As early as 200 BCE, in their efforts to determine the first few prime numbers, Greek mathematicians developed an algorithm requiring relatively easy calculations. All integer numbers greater than 1 can be uniquely factored as a product of prime numbers; this is the Fundamental Theorem of Arithmetic (Andrews, 1994). Consequently, it is a corollary that any composite number must have at least one factor smaller than or equal to its square root. For example, consider the factorization of 118: $118=2\cdot59$, and the factor 2 is less than $10.86=\sqrt{118}$.

Since all composite numbers have prime factors smaller than or equal to their square roots, it follows that prime numbers, which cannot be factored, do not. This idea prompted the Ancient Greek mathematician Eratosthenes to conceive of the Sieve of Eratosthenes to find small prime numbers (Ore, 1988). In order to find prime numbers less than 100, for instance, Eratosthenes would remove all factors of 2; then all factors of 3; then since 4 is not a prime, having already been removed as factors of 2, remove all factors of 5; etc. as seen in Figure 1.

The Sieve of Eratosthenes is an example of a deterministic algorithm used to unearth all prime numbers, but is only practical for "small" prime numbers, those less than 10,000,000 (Ore, 1988). Beyond that boundary, it is too resource-consuming for a computer to perform such a calculation.

The sporadic, but never-ending, primes: the Prime Number Theorem and the Infinity of Primes

Even though the Sieve of Eratosthenes offers an effective algorithm for finding small prime numbers, it gives little insight into the distribution of prime numbers. Carl Friedrich Gauss was the first to notice the only clear distributive property of prime numbers: they get scarcer as numbers get larger. Among the first 10 integers, 40% are prime; among the first 100, 1 in 4 is prime. This pattern continues, such that in the first 100,000 integers, 1 in 10.4 is prime (Peterson, 1996). In fact, Gauss wrote that "this frequency is on the average inversely proportional to the [natural] logarithm" (Tschinkel, 2006), so the approximate number of primes below a number *n* follows Equation(1):

$$\int \frac{dn}{\ln(n)}$$

Equation 1: Gauss' equation of the distribution of prime numbers

The French mathematician Adrien Marie Legendre independently developed a similar equation just a few years later. The result is known as the Prime Number Theorem, which while giving no definitive equation to find prime numbers, provides an approximation of the distance between prime numbers within any given interval. In fact, it states that the average distance between two consecutive primes near some number n is close to the natural logarithm of n (Peterson, 1996). For example, since ln(1000)=6.91, near 1000, approximately every seventh number should be prime.

As the density of prime numbers decreases, it might be expected that eventually prime numbers get so scarce that there exists a single largest prime number. However, the infinitude of prime numbers has been known since 300 BCE when it was established in Euclid's *Elements*. Euclid's proof hinges on the Fundamental Theorem of Arithmetic: if there is a single largest prime number, there would be a finite set of prime numbers. This theorem implies that all composite numbers could then be factored into these prime numbers. However, Euclid found a number that could not be divided by any of these prime numbers, thus necessitating the existence of another prime number (Ore, 1988). By induction, prime numbers are thus infinite. For a concise proof and examples, see Figure 2.

Despite providing insight into the distribution of prime numbers over the real number line, the Prime Number Theorem did not contribute to creating a definitive formula to find prime numbers.

Finding a faster algorithm: Euler's formula, the Riemann hypothesis, and a polynomial-time algorithm

Although the Sieve of Eratosthenes is a foolproof method to find prime numbers, this primitive algorithm is very time consuming, and mathematicians have devoted their efforts to finding a faster method.

Leonhard Euler spent many years working on a deterministic formula for finding prime numbers and eventually developed the equation seen in Equation 2. However, this equation only works for restricted inputs and does not determine consecutive prime numbers, meaning that its use as a test of primality is limited.

$$f(x) = x^2 + x + 41 \quad \text{for} \quad 0 \le x \le 39$$
$$\text{For example, } f(4) = 4^2 + 4 + 41 = 61 = \text{prime}$$

Equation 2: Euler's prime generating function

Multiples of 2 are underlined, of 3 are overlined, of 5 are **bolded**, and of 7 are *italicized*.
The primes are boxed.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	**25**	26	27	28	29	30
31	32	33	34	**35**	36	37	38	39	40
41	42	43	44	45	46	47	48	*49*	50
51	52	53	54	**55**	56	57	58	59	60
61	62	63	64	**65**	66	67	68	69	70
71	72	73	74	75	76	*77*	78	79	80
81	82	83	84	**85**	86	87	88	89	90
91	92	93	94	**95**	96	97	98	99	100

Figure 1: The Sieve of Eratosthenes

For primes p_i, $(p_1p_2p_3...p_n + 1)$ has remainder 1 when divided by all primes p_i, so cannot be decomposed by the finite set of primes. This contradicts the Fundamental Theorem of Arithmetic. Then another prime must exist, either $(p_1p_2p_3...p_n + 1)$ or a smaller number which divides it.

For example, this clearly holds on small primes:
$2 \cdot 3 + 1 = 7 = prime$
$2 \cdot 3 \cdot 5 + 1 = 31 = prime$
$2 \cdot 3 \cdot 5 \cdot 7 + 1 = 211 = prime$
$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 + 1 = 2311 = prime$
etc.

Figure 2: Short proof of the infinitude of prime numbers

Euler also developed the zeta function, relating a sum of fractions to a product of prime numbers, shown in Figure 3. Bernhard Riemann extended it into what is now known as the Riemann Zeta function (Ivic, 2003). In 1859 Riemann published his results and hypothesized that a function that has a zero root uniquely defines a prime number. Recently, in 2004, Gourdon and Sebah verified the Riemann hypothesis for the first 10 trillion zeroes (Crandall and Pomerance, 2005); nevertheless, the hypothesis remains unproven.

The discovery of prime numbers is simplified by computer programs, where the main challenge is finding a more rapid algorithm. Computer algorithms are usually compared using runtime analysis, which determines the worse case runtime given an input of length n. The runtime of a program is a function of the length of the input, and can be a polynomial, logarithmic or exponential equation. As the input size increases, an exponential runtime will always be longer than a polynomial runtime, which in turn will be longer than a logarithmic runtime.

The Sieve of Eratosthenes is an exponential algorithm to find prime numbers, which renders it ineffective in finding exceptionally large prime numbers. Since Eratosthenes, all deterministic algorithms to find primes have been exponential, so it was remarkable when a relatively simple deterministic polynomial algorithm was finally discovered in 2002 by Agrawal, Kayal and Saxena (AKS). The AKS algorithm, based on Fermat's Little Theorem and other proven mathematical assumptions, is an improvement but may still have an extremely long runtime, rendering it impractical. The search for an expeditious yet deterministic formula to find prime numbers is still underway, and there is no doubt that computers will continue to provide mathematicians with the ability to make further improvements.

$$\begin{aligned}
\zeta(s) &= 1 + \frac{1}{2^s} + \frac{1}{3^s} + \dots \\
&= \sum_{n=1}^{\infty} \frac{1}{n^s} \\
&= \prod_{primes\ p} \frac{1}{1 - p^{-s}} \\
&= \prod_{primes\ p} \frac{p^s}{p^s - 1} \\
&= \left(\frac{2^s}{2^s - 1}\right)\left(\frac{3^s}{3^s - 1}\right)\left(\frac{5^s}{5^s - 1}\right)\dots
\end{aligned}$$

If the function has a root, then

$$\begin{aligned}
0 = \zeta(s) &= \left(\frac{2^s}{2^s - 1}\right)\left(\frac{3^s}{3^s - 1}\right)\left(\frac{5^s}{5^s - 1}\right)\dots \\
0 &= \frac{p^s}{p^s - 1}
\end{aligned}$$

for some prime p

Figure 1: The Riemann-Zeta function

The use of prime numbers in Cryptography and the consequences of a deterministic formula for finding prime numbers

Cryptography, the science of encrypting and decrypting messages for transmission between a sender and an intended recipient, may at first glance seem unrelated to the discovery of prime numbers. Beginning in Roman times, information was frequently encrypted using a private-key, meaning that the sender and the recipient had to define a decryption codec that would allow the recipient to decode the message. This practice became impractical as technology evolved, since there was often no secure way to communicate the private-key, especially over considerable distances. To deal with these challenges, public-key cryptography was developed. Presently, the most common encryption methods in use rely on the difficulty of efficiently finding prime numbers.

Whereas prime numbers are known up to several thousands of digits, it is much harder to factor composite numbers with several thousands of digits, especially if they are composed of large prime numbers; a deterministic exponential time algorithm could take centuries to factor the composite number. This is the basis of public-key cryptography using the RSA encryption method created in 1978 by Rivest, Shamir and Adleman. Two large prime numbers p and q decide a large composite number N=pq as well as the encryption key e using an equation. From p, q and e, the decryption key d is determined by the same equation. An individual who wants to receive information securely makes public N and e, allowing anyone to send them the information. If the information is intercepted, then knowledge of d, which can only be determined if p, q and e are known, is required for decryption. This means that in order to decrypt the intercepted information, the individual, who only knows N and e, must factor N back into p and q, which is a restrictively time consuming process. On the other hand, if prime numbers could be found quickly, then composite numbers could be factored much more swiftly, and the RSA method would fail, rendering electronic public-key encryption insecure.

A similar threat to secure encryption is quantum computing, which is based on the principle that quantum properties could be used to represent data and perform operations as a traditional computer does. If quantum computing evolves beyond the experimental stages where it currently is, it presents the possibility of performing computations in record time. The same algorithms could be used, but would be executed much faster. For example, while it may take centuries to break an RSA code using a traditional computer, a quantum computer could take just seconds or minutes.

Conclusion

The field of prime numbers is ever changing, with new prime numbers discovered every few years. Early mathematicians dealt with the distribution and infinitude of prime numbers, whereas modern mathematicians aspire to find a deterministic formula to identify all prime numbers. Algorithms that quickly generate small prime numbers already exist, and the recent AKS polynomial time algorithm will even allow large prime numbers to be found quickly. Although finding prime numbers with ease would be a seminal accomplishment in mathematics, it would also create new challenges for the safety of electronic data encryption. Are the benefits of a deterministic prime generator to mathematics worth the destruction of the most common form of data encryption in computer science? Only time will tell.

References

1. Agrawal. Manindra. Kayal. Neeraj. Saxena. Nitin. 2004. PRIMES is in P. Annals of Mathematics 160: 781-793.

2. Andrews, G. E. 1994. Number Theory. Mineola, NY: Courier Dover Publications.

3. Crandall, R. E. and Pomerance, C. 2005. Prime Numbers. New York, NY: Springer.

4. Haddon, M. 2003. The Curious Incident of the Dog in the Night-time. London: Vintage Random House.

5. Ivic, A. 2003. The Riemann Zeta-Function. Mineola, NY: Courier Dover Publications.

6. Ore, O. 1988. Number Theory and Its History. Mineola, NY: Courier Dover Publications.

7. Peterson, I. 1996. Prime Theorem of the Century. Math-Trek, 1 (41). December

8. 14, 2008. http://www.maa.org/mathland/mathland_12_23.html

9. Rivest, R.L. Shamir, A. Adleman, L. 1978. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. Communications of the ACM 21(2):120-126.

10. Tschinkel, Y. 2006. About the cover: On the distribution of primes – Gauss' tables. Bulletin of the American Mathematical Society 43:89-91.