Prime Interest

(Secrets and Primes)

November 29, 2016

Róbert Freud

Classical cryptosystems: before sending the message we replace every letter by another one, and we agree about it in advance with our partner. E.g.

$E\downarrow$	a	b	c	d	e	f	$\ldots$	$\uparrow D$
	s	c	m	t	r	h	$\ldots$	

The cryptosystem is a table: reading it downwards we obtain the enciphering or encrypting key E, reading it upwards we obtain the deciphering or decrypting key D.

E and D are functions which are the inverse functions of each other.

In general: A and B agree in an encrypting key E and its inverse, the decrypting key D. Instead of the plaintext message u its encrypted version $v = E(u)$ is sent to B by A. Then B applies D to the received ciphertext v and obtains the original plaintext $D(v) = u$.

The keys act on enormous – say, 50 digit — numbers into which the very long sequences of letters are transformed. The entire encrypting, transmitting, and decrypting procedure is processed electronically by computers.

Only B can understand the message from A, and no third party can fabricate false messages in the name of A.

Difficulties: It is clumsy and dangerous to produce the keys in advance; one cannot handle the eventual disputes between A and B; in multilateral (e.g. business) communication every relation requires a separate key.

Diffie and Hellman (1975): Make E public and keep only D secret.

But if one knows E, then (s)he knows also D!

In principle for sure. And in practice?

If the ciphertext is, say, 2016, what could be the plaintext?

In other words, find $D(2016)$, i.e. which u satisfies $E(u) = 2016$?

Let's try: $E(1) = 2016$?, $E(2) = 2016$? etc. Sooner or later we shall find it out!

Well, rather later than sooner, perhaps after several billions of years.

Can you use a Hungarian-English dictionary also as an English-Hungarian dictionary?

What is the Hungarian word for the English word "water"?

Also keys E and D are dictionaries: from plain to cipher and from cipher to plain, resp.; it is not enough to buy one of them!

The pair of keys for A is E_A and D_A, and for B these are E_B and D_B where E_A and E_B are public, but D_A is known only by A, D_B is known only by B.

Now A sends $v = E_B(u)$ instead of u to B. Only B can decipher it: $u = D_B(v)$.

But some C can send a false message in the name of A! This is an anonymous letter without a signature.

Therefore A signs the message first: $D_A(u)$, and only then puts it into an envelop, and sends $w = E_B(D_A(u))$ to B.

B (and only B) can decipher it: $u = E_A(D_B(w))$.

There is no need for preliminary key exchange, no controversy can occur between A and B, and for more parties each participant can use his/her own pair of keys. Such systems function well in business, this is how the safety of credit cards is guaranteed etc.

But how can we construct such a pair of keys E and D so that making E public D is known only by the key owner?

Obviously he/she must have some private extra information.

The RSA scheme invented by Rivest, Shamir, and Adleman in 1976 is based on the fact that according to our present knowledge we are unable to factor a very big composite number.

Thus if somebody multiplies two big prime numbers, then (s)he will be the only person who can factor this product, this will be his/her private top secret.

The encrypting function in RSA is raising the numbers to a suitable exponent e and taking their remainders when divided by the product N of two big primes. Here the quick implementation of the decrypting function (i.e. taking e-th root of the remainder) requires the factors of N which are available only for the key owner.

The prime numbers are those integers greater than one, which have no other positive divisors than 1 and themselves. Every integer greater than 1 has a unique decomposition into the product of primes. Already Euclid proved (around 300 BC) that there are infinitely many primes, but there are thousands of innocent looking unsolved questions concerning the primes which — as Erdős said — can be easily understood by a small child, but cannot be answered even by the best mathematicians in the world.

But why would it be difficult to factor a big number. We have just to test which smaller integers greater than one divide it.

E.g. $2016 = 2 \cdot 1008 = 2 \cdot 2 \cdot 504 = \ldots = 2^5 \cdot 3^2 \cdot 7$.

And 2017 is prime, since we do not find such a divisor. In fact, it is sufficient to check the (prime) numbers less than 45 as $45 \cdot 45 = 2025 > 2017$.

Let's consider now, say, a 50 digit number.

It is less than 10^{50}, thus its smallest (non trivial) divisor must be less than 10^{25}.

We have to divide our number with these potential divisors (it is sufficient to check just the odd ones, or just the primes, if we had a prime table, but even these do not accelarete the process significantly).

In 1 second we can perform, say, 10^{10} divisions.

During 1 day=86400 seconds we can do less than 10^{15} checks.

In 1 year we accomplish less than 10^{18} tests.

Thus in a general case it may take more than a million years to decide whether that single given number is prime or composite. And for a 200 digit integer we shall probably be not done in the lifetime of our universe.

Gauss (1801): The problem of distinguishing prime numbers from composites, and of resolving composite numbers into their prime factors, is one of the most important and useful in all of arithmetic. The dignity of science seems to demand that every aid to the solution of such an elegant and celebrated problem be zealously cultivated.

Lenstra: Suppose that the cleaning lady discarded the numbers p and q by mistake, but the product pq was saved. How can we recover the factors? We must feel as a defeat for mathematics that the most promising way is to rake the garbage dump and to apply mnemotechnical methods.

Well, then how can we decide it quickly whether a big positive integer is prime or composite?

Fermat: If n is prime, then $c^n - c$ is divisible by n for every integer c. E.g. for $n = 7$, $c = 3$ we have $3^7 - 3 = 2184$, and $2184 = 7 \cdot 312$, indeed.

Therefore if e.g. $2^n - 2$ is **NOT** divisible by n, then n can **NOT** be a prime. Thus we have proved its compositeness without findig a factor.

And what can we say if $2^n - 2$ is divisible by n?

Chinese mathematicians who were familiar with Fermat's "little" theorem already more than 1000(!) years ago, believed also the converse to be true, i.e. if this divisibility holds, then n must be a prime.

Unfortunately, this is not the case, e.g. $341 = 31 \cdot 11$ and still $2^{341} - 2$ is divisible by 341.

341 is a pseudoprime with base two.

We can try, of course, also $3^{341} - 3$ and will find that this is not divisible by 341, and thus revealed the compositeness of 341.

But there exist even composite numbers which pretend to be primes for every base, e.g. 1729 satisfies $1729 \mid c^{1729} - c$ for every c.

$$1729 = 7 \cdot 13 \cdot 19 = 12^3 + 1^3 = 10^3 + 9^3$$

And there are infinitely many such universal pseudoprimes. Does it mean that the Fermat test is useless?

It can be proven that the pseudoprimes occur very rarely compared to the primes. E.g. up to 10^9 there are 50847534 primes, 5597 pseudoprimes with base 2, and only 646 universal pseudoprimes. Hence it can happen extremely seldom that we "declare" a composite number to be prime erroneously if it passes the Fermat test with many bases.

And the Fermat test can be improved in several ways so that there do not exist universal pseudoprimes anymore against these versions (Solovay–Strassen and Miller–Rabin–Lenstra primality tests). For practical purposes these are really "bomb-proof", but from a theoretical point of view even these are surpassed by the AKS primality test developed by Agrawal, Kayal, and Saxena in 2002 which too is based in a certain sense on Fermat's Little Theorem.

We still need to be able to raise a number to a high exponent quickly.

In how many steps can we compute the remainder of, say, 5^{1000} when divided by 73?

If we always multiply by 5, then we need 1000 multiplications which is — for a 50 digit exponent instead of 1000 — completely hopeless. (Computing $2^n - 2$ this way is slower than to find a divisor of n.)

Rather, let's perform repeated squarings: $5^2 = 25$, $5^4 = 25^2 = 625 = 8 \cdot 73 + 41$, and for determining the remainder of 5^8 we can square 41 (instead of 625) etc. till in the ninth step we obtain the remainder of 5^{512}.

Finally $5^{1000} = 5^{512+256+128+64+32+8} = 5^{512} \cdot 5^{256} \cdot 5^{128} \cdot 5^{64} \cdot 5^{32} \cdot 5^8$,

thus we needed altogether only $9 + 5 = 14$ multiplications!

On October 22, 2009 a 100000 dollar prize of the Electronic Frontier Foundation was given

for a prime number!

The prize was due for finding explicitly the first prime having at least ten million decimal digits.

$$2^{43112609} - 1$$

It consists of 12978189 digits.

It was found by GIMPS (Great Internet Mersenne Prime Search) on August 23, 2008. http://www.mersenne.org

For exhibiting the first prime with at least one hundred million digits you can earn 150000 dollars, you can start to work on this project with GIMPS.

The present record is $2^{74207281} - 1$ consisiting of 22338618 decimal digits.

It is an unsolved problem whether there exist infinitely many primes of the form 2^k-1: $2^2-1=3, 2^3-1=7, 2^5-1=31, 2^7-1=127, \ldots$.

Erdős Pál: This is probably the most difficult though not the most vital problem mankind is facing.

$2^{100}-1$

$a^2-b^2=(a+b)(a-b)$; $2^{100}-1=(2^{50})^2-1^2=(2^{50}-1)(2^{50}+1)$

$2^{rs}-1=(2^r)^s-1^s$ is divisible by 2^r-1, therefore 2^k-1 cannot be prime unless the exponent k is a prime, too.

$2^{11}-1=2047=23\cdot 89$.

Mersenne's list (1644): $2^k - 1$ is prime, if $k = 2, 3, 5, 7, 13, 17, 19, 31, 67, 127, 257$, and is composite for every other k less than 257.

Mersenne (1644): A lifetime is not sufficient to determine whether a 15 or 20 digit number is prime or composite.

Lucas (1876): $2^{67} - 1$ is composite!

Cole (1903):

$$193\,707\,721 \cdot 761\,838\,257\,287$$

Also this illustrates well that primality testing is much easier than factoring.

For testing Mersenne numbers today we still use an improved version of Lucas' algorithm which made him possible to find the first error in Mersenne's list.

Mersenne primes are related to perfect numbers: a positive integer is perfect if it equals the sum of its positive divisors except itself.

E.g.: $6 = 1 + 2 + 3$, $28 = 1 + 2 + 4 + 7 + 14$.

Euclid's formula: If $2^k - 1$ is prime, then $2^{k-1}(2^k - 1)$ is perfect. We obtain 6 for $k = 2$, 28 for $k = 3$. The next perfect number is 496 when $k = 5$.

Euler proved more than 2000 years later that all even perfect numbers obey this rule.

The oldest unsolved problems of mathematics are whether there exist infinitely many perfect numbers, and are there odd perfect numbers at all. And even a child can understand these simple questions!