

Public Key Cryptography

Lecture 4

Factorization Methods

- 1 The Problem
- 2 Fermat's Method
- 3 Pollard's $p - 1$ Method
- 4 Pollard's ρ Method
- 5 Continued Fraction Method

Factorization: the problem

Fundamental theorem of arithmetics

Every natural number has a factorization into primes, unique up to the order of factors.

Problem

Find a prime factor of a given large number n .

- In general the primality tests do not offer a prime factor of n , but only the information that n is composite.
- Out of the mentioned primality tests, only the slowest one (the trial division), gives us a prime factor of n .

A bit of history of factorization methods

- trial division (to determine small prime factors)
- Fermat's method (for numbers having factors relatively close one to each other)
- Pollard's $p - 1$ method (1974; to determine specific types of prime factors)
- Pollard's ρ method (1975; to determine relatively small prime factors)
- continued fraction method (Morrison and Brillhart 1975)
- quadratic sieve method (Pomerance 1981; the most effective for numbers having at most 100 digits)
- general number field sieve (1990's; the most effective for numbers having more than 100 digits)
- elliptic curve method (Lenstra 1987; the most effective to find divisors having 20-25 digits)

Remark. All of them are exponential-time algorithms!

Fermat's Method

- efficient factorization method for an $n = a \cdot b$ with $a \approx b$
- based on the following result:

Theorem

There is a bijective correspondence between the factorizations of n of the form $n = ab$, $a \geq b > 0$ and the representations of n of the form $n = t^2 - s^2$, $s, t \in \mathbb{N}$.

Proof.

- $n = ab \Rightarrow n = \left(\frac{a+b}{2}\right)^2 - \left(\frac{a-b}{2}\right)^2$.
- $n = t^2 - s^2 \Rightarrow n = (t+s)(t-s)$.
- If $n = ab$ and $a \approx b$, then $s = \frac{a-b}{2}$ is small and t is just a little greater than \sqrt{n} .

Fermat's Method (cont.)

- **Idea:** try for t all values starting with $\lceil\sqrt{n}\rceil + 1$, until $t^2 - n$ is a square, that will be exactly s^2 , and then determine a, b .
- Assume that n is not a square in order to avoid trivial exceptions.

Fermat's Algorithm

- Input: an odd composite number n (which is not a square), and a suitable bound B .
- Output: a non-trivial factor of n .
- Algorithm:
Let $t_0 = \lceil\sqrt{n}\rceil$.
For $t = t_0 + 1, \dots, t_0 + B$ do
 If $t^2 - n$ is a square s^2 , then $s^2 = t^2 - n$,
 $n = (t - s)(t + s)$, and STOP.

Fermat's Method (cont.)

Example. Let us factorize $n = 200819$.

We have $t_0 = \lfloor \sqrt{n} \rfloor = 448$.

For $t = 449$: $t^2 - n = 782$ is not a square.

For $t = 450$: $t^2 - n = 1681 = 41^2 = s^2$.

Hence $n = (t + s)(t - s) = 491 \cdot 409$.

Example. Let us factorize $n = 141467$.

We have $t_0 = \lfloor \sqrt{n} \rfloor = 376$.

For $t = 377$: $t^2 - n = 662$ is not a square.

For $t = 378$: $t^2 - n = 1417$ is not a square.

For $t = 377$: $t^2 - n = 2174$ is not a square.

...

For $t = 413$: $t^2 - n = 29102$ is not a square.

For $t = 414$: $t^2 - n = 29929 = 173^2 = s^2$ is a square.

Hence $n = (t + s)(t - s) = 587 \cdot 241$.

Generalized Fermat's Method (cont.)

Example. Let us factorize again $n = 141467$.

We take $t_0 = \lceil \sqrt{3n} \rceil = 651$.

For $t = t_0 + 1, t_0 + 2$ etc. we check if $t^2 - 3n$ is a square.

For $t = 655$: $t^2 - 3n = 4624 = 68^2 = s^2$.

Thus $3n = (t + s)(t - s) = 723 \cdot 587$, whence $n = 241 \cdot 587$.

Note that b is close to $3a$.

Generalized Fermat's Algorithm

- Input: an odd composite number n (which is not a square), and a suitable bound B .
- Output: a non-trivial factor of n .
- Algorithm:

For $k = 1, 2, \dots$ do

Let $t_0 = \lceil \sqrt{kn} \rceil$.

For $t = t_0 + 1, \dots, t_0 + B$ do

If $t^2 - kn$ is a square s^2 , then $s^2 = t^2 - kn$,

$n = \frac{1}{k}(t - s)(t + s)$, and STOP.

Pollard's $p - 1$ Method

- used to efficiently find any prime factor p of an odd composite number n for which $p - 1$ has only small prime divisors.
- then we are able to find a multiple k of $p - 1$ without knowing $p - 1$, as a product of powers of small primes.
- **Idea:** By Fermat's Little Theorem, $a^k \equiv 1 \pmod{p}$, $\forall a \in \mathbb{Z}$ with $p \nmid a$. Then $p \mid a^k - 1$. If $n \nmid a^k - 1$, then $d = (a^k - 1, n)$ is a non-trivial divisor of n .
- The situation $d = n$, in which case the algorithm fails, occurs with a negligible probability.
- As candidates for k , the $p - 1$ method considers

$$k = \prod \{q^i \mid q \text{ prime}, i \in \mathbb{N}^*, q^i \leq B\}$$

or even $k = \text{lcm}\{1, \dots, B\}$. If the primes dividing $p - 1$ are smaller than B , then k is a multiple of $p - 1$.

Pollard's $p - 1$ Method (cont.)

Pollard's $p - 1$ Algorithm

- Input: an odd composite number n , and a bound B .
- Output: a non-trivial factor d of n .
- Algorithm:
 1. Let $k = \prod \{q^i \mid q \text{ prime}, i \in \mathbb{N}^*, q^i \leq B\}$ or $k := \text{lcm}\{1, \dots, B\}$.
 2. Randomly choose $1 < a < n - 1$.
 3. $a := a^k \bmod n$.
 4. $d := (a - 1, n)$.
 5. If $d = 1$ or $d = n$ then output FAILURE
else output d .

Remark. If the algorithm ends with a failure, it is repeated for another value $1 < a < n - 1$ or for another bound B .

Example. Let us factorize $n = 1241143$ using $a = 2$ and $B = 13$.

Version 1. We choose

$$k = \prod \{q^i \mid q \text{ prime}, i \in \mathbb{N}^*, q^i \leq B\} = 2^3 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 13 = 360360.$$

Then $(a^k - 1, n) = 547$, hence 547 is a factor of $n = 547 \cdot 2269$.

Version 2. We choose

$$k := \text{lcm}\{1, \dots, B\} = \text{lcm}\{1, \dots, 13\} = 360360.$$

Then $(a^k - 1, n) = 547$, hence 547 is a factor of $n = 547 \cdot 2269$.

- Pollard (1975)
- the simplest factorization algorithm that is substantially faster than trial division
- generally used to determine relatively small prime factors
- based on Floyd's algorithm for finding a cycle and on the remark (“birthday paradox” type) that t random numbers x_1, x_2, \dots, x_t from the interval $[1, n]$ contain a repetition with probability $P > 0.5$ if $t > 1.177n^{1/2}$.
- the birthday paradox concerns the probability that some pair of people out of randomly chosen n people have the same birthday. Probability 0.999 is reached with 70 people, and probability 0.5 with 23 people.

Auxiliary Problem

Let S be a finite set with n elements, let $f : S \rightarrow S$ be a random map and randomly choose $x_0 \in S$. Consider the sequence:

$$x_{j+1} = f(x_j), \quad j \in \mathbb{N}.$$

The sequence has a cycle (S is finite), which we would like to find.

In general, $S = \mathbb{Z}_n$ and $f : \mathbb{Z}_n \rightarrow \mathbb{Z}_n$ is a polynomial map (but not linear, bijective, $f(x) = x^2$ or $f(x) = x^2 - 2, \dots$), and usually it is chosen to be $f(x) = x^2 + 1$.

Pollard's ρ Method (cont.)

Reduction: The problem is to find two indexes j and k , say $j < k$, such that $x_j = x_k$. Then we get a cycle of length $l = k - j$.

Floyd's method: Start with the pair (x_1, x_2) and successively computes (x_i, x_{2i}) from the previous pair $(x_{i-1}, x_{2(i-1)})$ until $x_m = x_{2m}$ for some m .

There is such a value m , for instance let m be the least multiple of l greater than or equal to j , say $m = ls$. Then

$$x_m = x_{ls} = x_{ls+l} = x_{l(s+1)} = x_{l(s+2)} = \cdots = x_{l \cdot 2s} = x_{2m}.$$

Pollard's ρ Algorithm

- Input: an odd composite number n and a suitable random polynomial map f (implicitly, $f(x) = x^2 + 1$).
- Output: a non-trivial factor d of n .
- Algorithm:

Let $x_0 = 2$.

For $j = 1, 2, \dots$ compute the sequence:

$$x_j = f(x_{j-1}) \bmod n$$

and $d = (|x_{2j} - x_j|, n)$.

- If $1 < d < n$, then STOP and d is a non-trivial factor of n .
- Else, continue with the next value of j .

Pollard's ρ Algorithm (cont.)

Example. Let us factorize $n = 4087$ using $f(x) = x^2 + x + 1$ and $x_0 = 2$.

We have modulo n :

$$x_1 = f(x_0) = 7; x_2 = f(x_1) = 57;$$

$$(|x_2 - x_1|, n) = (50, 4087) = 1;$$

$$x_3 = f(x_2) = 3307; x_4 = f(x_3) = 2745;$$

$$(|x_4 - x_2|, n) = (2688, 4087) = 1;$$

$$x_5 = f(x_4) = 1343; x_6 = f(x_5) = 2626;$$

$$(|x_6 - x_3|, n) = (681, 4087) = 1;$$

$$x_7 = f(x_6) = 3734; x_8 = f(x_7) = 1647;$$

$$(|x_8 - x_4|, n) = (1098, 4087) = 61.$$

Hence a factor of $n = 4087$ is 61 and thus $4087 = 61 \cdot 67$.

Continued Fraction Method

Idea (Fermat): if we obtain a congruence

$$t^2 = s^2 \pmod{n} \text{ with } t \not\equiv \pm s \pmod{n},$$

then $n \mid t^2 - s^2 = (t + s)(t - s)$, and so $a = (t + s, n)$ or $a = (t - s, n)$ is a non-trivial factor of n .

Definition

- By the *least absolute residue* of a number a modulo n we mean the integer in the interval $[-\frac{n}{2}, \frac{n}{2}]$ to which a is congruent modulo n .
- A *factor base* is a set $B = \{p_1, p_2, \dots, p_h\}$ of primes, where p_1 may be also -1 .

For $b \in \mathbb{Z}$, b^2 is a *B-number* for a given n if the least absolute residue $b^2 \bmod n$ can be written as a product of numbers from B .

Continued Fraction Method (cont.)

Consider now \mathbb{Z}_2^h , which is a vector space over \mathbb{Z}_2 .

We associate to each B -number a vector

$$v = (x_1, \dots, x_h) \in \mathbb{Z}_2^h$$

as follows: we write

$$b^2 \bmod n = p_1^{r_1} \dots p_h^{r_h}$$

and we put

$$x_j = r_j \bmod 2 \text{ for } j = 1, \dots, h.$$

Example. Let $n = 4633$ and $B = \{-1, 2, 3\}$. Then 67^2 , 68^2 , 69^2 are B -numbers because

$$67^2 \bmod n = -144 = (-1) \cdot 2^4 \cdot 3^2$$

$$68^2 \bmod n = -9 = (-1) \cdot 3^2$$

$$69^2 \bmod n = 128 = 2^7$$

Hence the vectors from \mathbb{Z}_2^3 corresponding to our B -numbers are

$$v_1 = (1, 0, 0), v_2 = (1, 0, 0), v_3 = (0, 1, 0).$$

Continued Fraction Method (cont.)

Suppose now that we have a set of B -numbers $b_i^2 \bmod n$, $i = 1, \dots, k$ such that

$$v_1 + v_2 + \dots + v_k = 0 \in \mathbb{Z}_2^h.$$

Then the product of the least absolute residues of b_i^2 is equal to the product of some even powers of the primes p_j from B . Denote by a_i the least absolute residue of $b_i^2 \bmod n$. If for $i = 1, \dots, k$ we write $a_i = p_1^{r_{i1}} \dots p_h^{r_{ih}}$, then

$$a_1 \dots a_k = p_1^{r_{11} + \dots + r_{k1}} \dots p_h^{r_{1h} + \dots + r_{kh}},$$

where the exponent of each p_j is even.

Hence the right hand side is the square of $p_1^{\gamma_1} \dots p_h^{\gamma_h}$, where

$$\gamma_j = \frac{1}{2}(r_{1j} + \dots + r_{kj})$$

for $j = 1, \dots, h$.

Continued Fraction Method (cont.)

Let c be the least absolute residue of $p_1^{\gamma_1} \dots p_h^{\gamma_h} \bmod n$ and b be the least absolute residue of $b_1 \dots b_k \bmod n$.

Then we have $b^2 = c^2 \bmod n$ by construction.

- If $b = \pm c \bmod n$, then we need to consider another subset of B -numbers that have the sum of the corresponding vectors equal to 0.
- Since n is composite, randomly choosing b_i 's, the probability that $b = \pm c \pmod n$ is at most $1/2$. As previously seen, when we find b, c such that $b^2 = c^2 \pmod n$, but $b \not\equiv \pm c \pmod n$, we immediately have a proper factor of n , namely $(b + c, n)$ or $(b - c, n)$. The probability that the process to find b, c with the above properties takes more than l steps is at most 2^{-l} .

How to choose B and the b_i 's in practice?

Continued Fractions

Definition

Let $x \in \mathbb{R}$. For every $i \geq 1$ define

$$a_0 = [x], \quad x_0 = x - a_0,$$
$$a_i = \left[\frac{1}{x_{i-1}} \right], \quad x_i = \frac{1}{x_{i-1}} - a_i.$$

Remarks. (i) The process ends when and if $x_i = 0$.

(ii) Note that the process ends $\Leftrightarrow x \in \mathbb{Q}$.

By the construction of a_0, a_1, \dots, a_i , we can write for each i

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots \frac{1}{a_i + x_i}}} \stackrel{\text{not.}}{=} a_0 + \frac{1}{a_1 +} \frac{1}{a_2 +} \dots \frac{1}{a_i + x_i}.$$

Continued Fractions (cont.)

Suppose that $x \in \mathbb{R}$ is irrational. Then the rational number

$$\frac{b_i}{c_i} = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots \frac{1}{a_i}}}$$

is called *the i -th convergent* of the continued fraction x .

Theorem

$$(i) \quad \frac{b_0}{c_0} = \frac{a_0}{1}, \quad \frac{b_1}{c_1} = \frac{a_0 a_1 + 1}{a_1},$$

$$\frac{b_i}{c_i} = \frac{a_i b_{i-1} + b_{i-2}}{a_i c_{i-1} + c_{i-2}}, \quad \forall i \geq 2.$$

$$(ii) \quad b_i c_{i-1} - b_{i-1} c_i = (-1)^{i-1}, \quad \forall i \geq 1.$$

(iii) If $b_i = a_i b_{i-1} + b_{i-2}$ and $c_i = a_i c_{i-1} + c_{i-2}$, then $(b_i, c_i) = 1$.

(iv) Let $x \in \mathbb{R}$. Then the sequence of convergents of x is convergent and has limit x .

Continued Fractions (cont.)

Lemma

Let $x \in \mathbb{R}$, $x > 1$, with the i -th convergent $\frac{b_i}{c_i}$. Then for every i ,

$$|b_i^2 - x^2 c_i^2| < 2x.$$

Theorem

Let $n \in \mathbb{N}$, which is not a square. Let $\frac{b_i}{c_i}$ be the i -th convergent of the writing of \sqrt{n} as a continued fraction. Then the least absolute residue of $b_i^2 \pmod n$ is less than $2\sqrt{n}$.

Proof. Apply the previous lemma for $x = \sqrt{n}$. Then $b_i^2 = b_i^2 - nc_i^2 \pmod n$ and the last integer is less than $2\sqrt{n}$ in absolute value. \square

Remark. This theorem is the key of the continued fraction method.

Continued Fraction Algorithm

All computations will be done modulo n , the sums and products being reduced modulo n to the least positive residue (or to the least absolute residue in Step 5.).

Continued Fraction Algorithm

- Input: a composite number n .
- Output: a non-trivial factor of n .
- Algorithm:
 1. Let $b_{-1} = 1$, $b_0 = a_0 = [\sqrt{n}]$ and $x_0 = \sqrt{n} - a_0$.
 2. Compute $b_0^2 \bmod n$ (that will be $b_0^2 - n$).
 3. Let $a_i = [\frac{1}{x_{i-1}}]$. Then $x_i = \frac{1}{x_{i-1}} - a_i$.
 4. Let $b_i = a_i b_{i-1} + b_{i-2}$ (reduced modulo n).
 5. Compute $b_i^2 \bmod n$ for several i 's.
 6. Choose out of these numbers those that factorize in absolute value in small primes.

Continued Fraction Algorithm (cont.)

Continued Fraction Algorithm (cont.)

7. Choose the factor base $B = \{p_1, \dots, p_h\}$ as consisting of -1 and the primes appearing in more than one element $b_i^2 \pmod n$ (or that appear with an even power in a single element).
8. Write all numbers $b_i^2 \pmod n = p_1^{r_{i1}} \dots p_h^{r_{ih}}$ that are B -numbers and their associated vectors $v_i \in \mathbb{Z}_2^h$.
9. Find a subset of vectors v_i with the sum $0 \in \mathbb{Z}_2^h$.
10. Let $b = \prod b_i$, where everything is done modulo n and the product is taken for those b_i 's for which $\sum v_i = 0$. Let $c = \prod p_j^{\gamma_j}$, where the p_j 's are the elements of B except for -1 and $\gamma_j = \frac{1}{2} \sum r_{ij}$, the sum being done after the same indexes i 's.
11. If $b \not\equiv \pm c \pmod n$, then $(b + c, n)$ or $(b - c, n)$ is a non-trivial factor of n . If $b \equiv \pm c \pmod n$, then we look for another subset of indexes i 's with the above properties. If this is not possible, we compute more values a_i , b_i and $b_i^2 \pmod n$, enlarging the factor base B .

Continued Fraction Algorithm (cont.)

Example. Let us factorize $n = 9073$. We make a table of values $a_i, b_i, b_i^2 \bmod n$:

i	0	1	2	3	4
a_i	95	3	1	26	2
b_i	95	286	381	1119	2619
$b_i^2 \bmod n$	-48	139	-7	87	-27

Note that the last row contains least absolute residues. Their factorizations are as follows:

- ❶ $i = 0$: $-48 = (-1) \cdot 2^4 \cdot 3$
- ❷ $i = 1$: 139
- ❸ $i = 2$: $-7 = (-1) \cdot 7$
- ❹ $i = 3$: $87 = 3 \cdot 29$
- ❺ $i = 4$: $-27 = (-1) \cdot 3^3$

Analyzing them, we decide that the primes 29 and 139 are too large, and we choose $B = \{-1, 2, 3, 7\}$.

Continued Fraction Algorithm (cont.)

Then $b_i^2 \bmod n$ is a B -number for $i = 0, 2, 4$.

The associated vectors v_i are:

$$v_0 = (1, 4, 1, 0), \quad v_2 = (1, 0, 0, 1), \quad v_4 = (1, 0, 3, 0).$$

Then we have

$$v_0 + v_4 = 0 \pmod{2}.$$

Hence

$$b = b_0 \cdot b_4 = 95 \cdot 2619 = 3834 \pmod{n},$$

$$c = (-1)^{\frac{1+1}{2}} \cdot 2^{\frac{4+0}{2}} \cdot 3^{\frac{1+3}{2}} \cdot 7^{\frac{0+0}{2}} = -2^2 \cdot 3^2 = -36.$$

By construction we always have $b^2 = c^2 \pmod{n}$.

Since $b \not\equiv \pm c \pmod{n}$, a factor of n is $(3834 + 36, 9073) = 43$ or $(3834 - 36, 9073) = 211$. Thus $n = 43 \cdot 211$.

Continued Fraction Algorithm (cont.)

Example. Let us factorize $n = 17873$. We make a table as in the previous example.

i	0	1	2	3	4	5
a_i	133	1	2	4	2	3
b_i	133	134	401	1738	3877	13369
$b_i^2 \bmod n$	-184	83	-56	107	-64	161

We choose $B = \{-1, 2, 7, 23\}$. Then $b_i^2 \bmod n$ is a B -number for $i = 0, 2, 4, 5$. The associated vectors v_i are:

$$v_0 = (1, 3, 0, 1), v_2 = (1, 3, 1, 0), v_4 = (1, 6, 0, 0), v_5 = (0, 0, 1, 1).$$

Then $v_0 + v_2 + v_5 = 0 \pmod{2}$. It follows that

$$b = b_0 \cdot b_2 \cdot b_5 = 133 \cdot 401 \cdot 13369 = 1288 \pmod{n}$$

$$c = 2^3 \cdot 7 \cdot 23 = 1288.$$

Continued Fraction Algorithm (cont.)

We have $b = c \pmod{n}$, so we need to generate more values.

i	6	7	8
a_i	1	2	1
b_i	17246	12115	11488
$b_i^2 \pmod{n}$	-77	149	-88

We choose now $B = \{-1, 2, 7, 11, 23\}$. Then $b_i^2 \pmod{n}$ is a B -number for $i = 0, 2, 4, 5, 6, 8$. The associated vectors v_i are:

$$v_0 = (1, 3, 0, 0, 1), v_2 = (1, 3, 1, 0, 0), v_4 = (1, 6, 0, 0, 0),$$




$$v_5 = (0, 0, 1, 0, 1), v_6 = (1, 0, 1, 1, 0), v_8 = (1, 3, 0, 1, 0)$$

Then $v_2 + v_4 + v_6 + v_8 = 0 \pmod{2}$, whence

$$b = b_2 \cdot b_4 \cdot b_6 \cdot b_8 = 7272 \pmod{n}, \quad c = 2^6 \cdot 7 \cdot 11 = 4928.$$

Since $b \not\equiv \pm c \pmod{n}$, a factor of n is $(7272 + 4928, 17873) = 61$ or $(7272 - 4928, 17873) = 293$. Thus $n = 61 \cdot 293$.

Selective Bibliography

-  N. Koblitz, *A Course in Number Theory and Cryptography*, Springer, 1994.
-  A.J. Menezes, P.C. van Oorschot, S.A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1997.
[<http://www.cacr.math.uwaterloo.ca/hac>]
-  M. Cozzens, S.J. Miller, *The Mathematics of Encryption: An Elementary Introduction*, American Mathematical Society, 2013.