Arithmetic algorithms for cryptology
2 October 2018, Paris

Attacks on RSA and discrete log cryptosystems

Razvan Barbulescu
CNRS and IMJ-PRG

cnrs dépasser les frontières

IMJ-PRG
Institut de Mathématiques
de
Jussieu-Paris Rive Gauche

Starting point

Notations

- q prime
- g a generator of $(\mathbb{F}_q)^*$
- X a (secret) integer less than q
- $Y = g^X \bmod q$.

Quote

"Computing X from Y, on the other hand can be computed much more difficult and, for certain carefully chosen values of q, requires on the order of $q^{1/2}$ operations [...]"

Diffie and Hellman 1976

Indeed, the best known algorithm in 1976 was Baby step giant step.

The DLP was as hard for $(\mathbb{F}_q)^*$ as for any other group (e.g. elliptic curves).

The L notation

For any integer $Q = e^n$, $L_Q(\alpha, c) = \exp\left(cn^{\alpha}(\log n)^{1-\alpha}\right)$.
When c is not specified we write $L_Q(\alpha)$.

Example

- $L_Q(1, \frac{1}{2}) = \exp\left(\frac{1}{2}n\right) = \sqrt{\exp(n)} = \sqrt{Q}$ (exponential algorithm).
- $L_Q(0, 3) = \exp\left(3 \log n\right) = n^3$ (cubic algorithm)
- $L_Q(1/2, 1) = \exp\left(\sqrt{n}\sqrt{\log n}\right) \approx \exp\left(\sqrt{n}\right) = e^{\sqrt{n}}$ (sub-exponential algorithm).

Exercice

- $L_{L_x(\alpha)}(\beta) = L_x(\alpha\beta)$.
- $L(\alpha)L(\beta) = L(\max(\alpha, \beta))^{1+o(1)}$.

History

- One year after the introduction of DLP in cryptography, a subexponential algorithm was proposed by Adleman (complexity $L(1/2)$).
- In 1978 when RSA was proposed, it was known that the continued fractions method of factorization was very fast. In early 80s Dixon and Pomerance proved that there are algorithms of complexity $L(1/2)$.
- Ideas traveled from factorization to discrete logarithm in finite fields and vice-versa.

Chronology

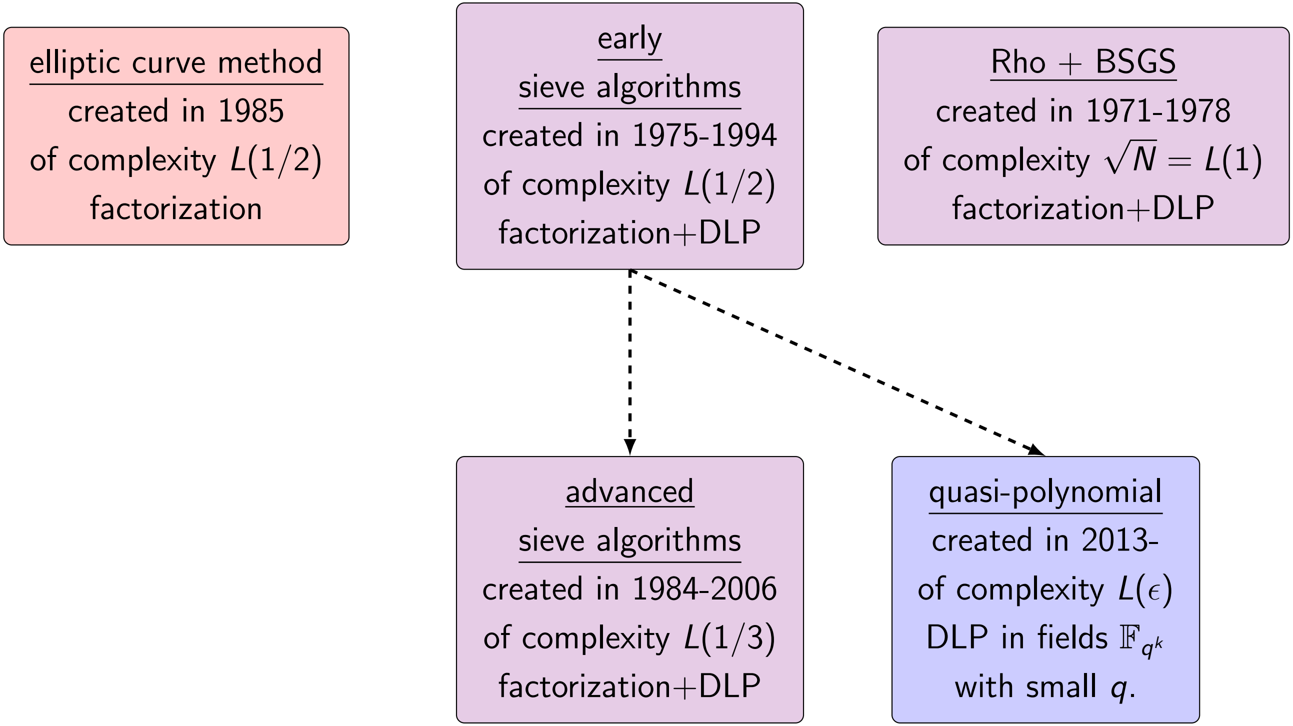
Dates below give the publication year of the first algorithm of each class. For discrete logarithm, there are several cases, which were solved by different algorithms.

complexity	factorization	DLP in finite fields
$L_Q(1/2)$	1970[a]	1979 – 1994
$L_Q(1/3)$	1989	1984 – 2006
$L_Q(\epsilon)$	––	2013–

[a] complexity unknown until 1980 (after the introduction of RSA)

Algorithms families

Let us use red for factoring, blue for DLP and violet for both.



elliptic curve method

created in 1985

of complexity $L(1/2)$

factorization

early sieve algorithms

created in 1975-1994

of complexity $L(1/2)$

factorization+DLP

Rho + BSGS

created in 1971-1978

of complexity $\sqrt{N} = L(1)$

factorization+DLP

advanced sieve algorithms

created in 1984-2006

of complexity $L(1/3)$

factorization+DLP

quasi-polynomial

created in 2013-

of complexity $L(\epsilon)$

DLP in fields $\mathbb{F}_{q^k}$

with small q.

Plan du cours

- Introduction
- Index Calculus
- Quadratic sieve (QS/MPQS)
- Sieving

Smoothness

Integers

- definition An integer is B-smooth if all its prime factors are less than B.
- computation One finds small prime divisors with ECM, which
  - is probabilistic;
  - relies on a conjecture of analytic number theory;
  - given an integer x, it finds all its factors less than B in average time $L_B(1/2, \sqrt{2})^{1+o(1)} \log(x)^4$. In practice, the dependency in $\log x$ is quadratic.

Polynomials

- definition A polynomial in $\mathbb{F}_q[t]$ is m-smooth if all its irreducible factors have degree less than or equal to m.
- computation One tests if a polynomial $P(t)$ is m-smooth by one of the two methods below:
  - by factoring it (correctness is trivial, probabilistic, slow);
  - by taking gcd with $P'(t) \cdot (t^{q^m} - t)$ (prove it!, deterministic, faster).

Smoothness

Integers

- definition An integer is B-smooth if all its prime factors are less than B.
- computation One finds small prime divisors with ECM, which
  - is probabilistic;
  - relies on a conjecture of analytic number theory;
  - given an integer x, it finds all its factors less than B in average time $L_B(1/2, \sqrt{2})^{1+o(1)} \log(x)^4$. In practice, the dependency in $\log x$ is quadratic.

Polynomials

- definition A polynomial in $\mathbb{F}_q[t]$ is m-smooth if all its irreducible factors have degree less than or equal to m.
- computation One tests if a polynomial $P(t)$ is m-smooth by one of the two methods below:
  - by factoring it (correctness is trivial, probabilistic, slow);
  - by taking gcd with $P'(t) \cdot (t^{q^m} - t)$ (prove it!, deterministic, faster).

It is not known how to define smooth elements on an elliptic curves with a fast smoothness test.

DLP: an example (1)

Parameters

- $p = 12101$
- $g = 7$ is a generator of $G = (\mathbb{Z}/p\mathbb{Z})^*$
- $\ell = 11$ is a prime factor of $(p-1) = \#G$
- $B = 10$ is the smoothness bound
- factor base $2, 3, 5, 7$

Finding relations among logs

$$7^{5} \bmod p = 4706 = 2 \cdot 13 \cdot 181$$

DLP: an example (1)

Parameters

- $p = 12101$
- $g = 7$ is a generator of $G = (\mathbb{Z}/p\mathbb{Z})^*$
- $\ell = 11$ is a prime factor of $(p-1) = \#G$
- $B = 10$ is the smoothness bound
- factor base $2, 3, 5, 7$

Finding relations among logs

$$\begin{aligned} 7^5 \bmod p &= 4706 = 2 \cdot 13 \cdot 181 \\ 7^6 \bmod p &= 8740 = 2^2 \cdot 5 \cdot 19 \cdot 23 \end{aligned}$$

DLP: an example (1)

Parameters

- $p = 12101$
- $g = 7$ is a generator of $G = (\mathbb{Z}/p\mathbb{Z})^*$
- $\ell = 11$ is a prime factor of $(p-1) = \#G$
- $B = 10$ is the smoothness bound
- factor base $2, 3, 5, 7$

Finding relations among logs

$$\begin{aligned}
7^5 \bmod p &= 4706 = 2 \cdot 13 \cdot 181 \\
7^6 \bmod p &= 8740 = 2^2 \cdot 5 \cdot 19 \cdot 23 \\
7^7 \bmod p &= 675 = 3^3 \cdot 5^2
\end{aligned}$$

DLP: an example (1)

Parameters

- $p = 12101$
- $g = 7$ is a generator of $G = (\mathbb{Z}/p\mathbb{Z})^*$
- $\ell = 11$ is a prime factor of $(p-1) = \#G$
- $B = 10$ is the smoothness bound
- factor base $2, 3, 5, 7$

Finding relations among logs

$$\begin{aligned} 7^5 \bmod p &= 4706 = 2 \cdot 13 \cdot 181 \\ 7^6 \bmod p &= 8740 = 2^2 \cdot 5 \cdot 19 \cdot 23 \\ 7^7 \bmod p &= 675 = 3^3 \cdot 5^2 \end{aligned}$$

The last relation gives:

$$7 = 3 \log_7 3 + 2 \log_7 5$$

DLP: an example (1)

Parameters

- $p = 12101$
- $g = 7$ is a generator of $G = (\mathbb{Z}/p\mathbb{Z})^*$
- $\ell = 11$ is a prime factor of $(p-1) = \#G$
- $B = 10$ is the smoothness bound
- factor base $2, 3, 5, 7$

Finding relations among logs

$$\begin{aligned}
7^5 \bmod p &= 4706 = 2 \cdot 13 \cdot 181 \\
7^6 \bmod p &= 8740 = 2^2 \cdot 5 \cdot 19 \cdot 23 \\
7^7 \bmod p &= 675 = 3^3 \cdot 5^2 \\
7^8 \bmod p &= \dots
\end{aligned}$$

The last relation gives:

$$\begin{aligned}
7 &= 3 \log_7 3 + 2 \log_7 5 \\
25 &= 8 \log_7 2 + 1 \log_7 3 \\
42 &= 6 \log_7 2 + 2 \log_7 5.
\end{aligned}$$

DLP: an example (2)

Thanks to the Pohlig-Hellman reduction

we do the linear algebra computations modulo $\ell = 11$.

Linear algebra computations

We have to find the unknown $\log_7 2$, $\log_7 3$ and $\lg_7 5$ in the equation

$$\begin{pmatrix} 0 & 3 & 2 \\ 8 & 1 & 0 \\ 6 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} \log_7 2 \\ \log_7 3 \\ \log_7 5 \end{pmatrix} \equiv \begin{pmatrix} 7 \\ 25 \\ 42 \end{pmatrix} \mod 11.$$

Conjecture

The matrix obtained by the technique above has maximal rank.

We can drop all conjectures by modifying the algorithm, but this variant is fast and, even if the matrix has smaller rank we can find logs.

Solution

We solve to obtain $\log_7 2 \equiv 0 \mod 11$; $\log_7 3 \equiv 3 \mod 11$ and $\log_7 5 \equiv 10 \mod 11$. For this small example we can also use Pollard's rho method and obtain that

$$\log_7 3 = 8869 \equiv 3 \mod 11.$$

DLP: an example (3)

At this point, we know discrete logarithms of the factor base and of smooth numbers:

$$\log_7(10) = \log_7 2 + \log_7 5 \equiv 10 \mod 11.$$

DLP: an example (3)

At this point, we know discrete logarithms of the factor base and of smooth numbers:

$$\log_7(10) = \log_7 2 + \log_7 5 \equiv 10 \mod 11.$$

Smoothing by randomization

Consider a residue modulo p which is not 10-smooth, e.g. $h = 151$. We take random exponents a and test is $(g^a h) \bmod p$ is B-smooth.

$$7^{3} 151 \bmod p \quad = \quad 3389$$

DLP: an example (3)

At this point, we know discrete logarithms of the factor base and of smooth numbers:

$$\log_7(10) = \log_7 2 + \log_7 5 \equiv 10 \mod 11.$$

Smoothing by randomization

Consider a residue modulo p which is not 10-smooth, e.g. $h = 151$. We take random exponents a and test is $(g^a h)$ mod p is B-smooth.

$$\begin{aligned} 7^3 151 \bmod p &= 3389 \\ 7^4 151 \bmod p &= 11622 = 2 \cdot 3 \cdot 13 \cdot 149 \end{aligned}$$

DLP: an example (3)

At this point, we know discrete logarithms of the factor base and of smooth numbers:

$$\log_7(10) = \log_7 2 + \log_7 5 \equiv 10 \mod 11.$$

Smoothing by randomization

Consider a residue modulo p which is not 10-smooth, e.g. $h = 151$. We take random exponents a and test is $(g^a h) \bmod p$ is B-smooth.

$$\begin{aligned} 7^3 151 \bmod p &= 3389 \\ 7^4 151 \bmod p &= 11622 = 2 \cdot 3 \cdot 13 \cdot 149 \\ 7^5 151 \bmod p &= 8748 = 2^2 \cdot 3^7 \end{aligned}$$

DLP: an example (3)

At this point, we know discrete logarithms of the factor base and of smooth numbers:

$$\log_7(10) = \log_7 2 + \log_7 5 \equiv 10 \mod 11.$$

Smoothing by randomization

Consider a residue modulo p which is not 10-smooth, e.g. $h = 151$. We take random exponents a and test is $(g^a h) \bmod p$ is B-smooth.

$$\begin{aligned} 7^3 151 \bmod p &= 3389 \\ 7^4 151 \bmod p &= 11622 = 2 \cdot 3 \cdot 13 \cdot 149 \\ 7^5 151 \bmod p &= 8748 = 2^2 \cdot 3^7 \end{aligned}$$

The discrete logarithms of the two members are equal:

$$5 + \log_7(151) = 2 \log_7 2 + 7 \log_7 3.$$

We find $\log_7(151) \equiv 3 \mod 11$.

Remark

This part of the computations is independent of the relation collection and linear algebra stages. It is called individual logarithm stage.

Index Calculus

Input: p prime, g generator of $(\mathbb{Z}/p\mathbb{Z})^*$, ℓ prime divisor of $(p-1)$
h integer less than p

Output: $\log_g h \mod \ell$

```
Set $B$ to its optimal value
Make the list $\mathcal{F}$ of the primes less than $B$ (factor base)        ▷ Initialization
repeat
    $a \leftarrow$Random([1,p-1])
    if $(g^a \bmod p)$ is $B$-smooth then
        relations=relations $\bigcup \{a\}$                                   ▷ Relations collection
    end if
until $\#$relations $\geq \#\mathcal{F}$
Construct the matrix $M = (m_{a,q})$, $a$ in relations, $q \in \mathcal{F}$ as follows
        $m_{a,q} = \mathrm{val}_q\left(g^a \bmod p\right).$                  ▷ Linear algebra
Solve the linear system        $Mx = (a)_{a \text{ in relations}}.$
repeat
    $b \leftarrow$Random([1,p-1])
until $(g^b h \bmod p)$ is $B$-smooth
Factor $(g^b h \bmod p) = \prod q_i^{e_i}$                                   ▷ Individual logarithm
return $x = \sum e_i \log_g(q_i) - b$
```

Complexity analysis of Index Calculus (1)

Notation

For two integers x and y, call $P(x, y)$ the proportion of integers less than x which are y-smooth.

Stages complexity

- Relations collection The integers ($g^a \bmod p$) are random integers less than p. The expectancy of the number of trials before finding a relation is $1/P(p, B)$. Hence, $\text{cost(relations collection)} = B/P(p, B)$.
- Linear algebra We have to invert a square matrix of size B. With the Gaussian method it requires B^3 operations. The Wiedemann algorithm (1986) takes time $B^2\lambda$, where λ is the average number of non-zero entries per row. In our case λ is the number of divisors of the integer $g^a \bmod p$, which is less than p. Since an integer less than p has less than $\log_2 p$ prime factors, $\lambda \leq \log p$. Hence we have $\text{cost(linear algebra)} = B^{2+o(1)}$.
- Individual logarithm The integers ($g^a h \bmod p$) are random among integers less than p. Hence, on average we need $1/P(p, B)$ trials: $\text{cost(individual logarithm)} = 1/P(p, B)$.

Complexity analysis of Index Calculus (2)

We have:

$$\begin{aligned}\text{cost(Index Calculus)} &= B/P(p,B) + B^{2+o(1)} + 1/P(p,B).\\ &= B^{1+o(1)}\left(1/P(p,B)+B\right)\\ &\leq 2B^{1+o(1)}\max\left(1/P(p,B),B\right).\end{aligned}$$

When we increase B, $1/P(p,B)$ decreases, so the optimal value of $\max\left(1/P(p,B),B\right)$ is obtained when

$$B = 1/P(p,B).$$

Theorem (Canfield,Erdös,Pomerance 1983, simplified statement)

Assume that x and u are two real numbers such that u is between $(\log x)^{c_1}$ and $(\log x)^{c_2}$ for some fixed constants c_1 and c_2, $0 < c_1 < c_2 < 1$. Then, we have

$$P(x, x^{1/u}) = 1/u^{u(1+o(1))},$$

where the $o(1)$ depends only on x.

Conclusion

The solution of the equation $B = 1/P(p,B)$ corresponds to $B = p^{1/u}$ for $u = 1/\sqrt{2}(\log p)^{1/2}/(\log\log p)^{1/2}$. Then the complexity is equal to B^2 and we obtain

$$\text{cost}\left(\text{Index Calculus}\right) = L_p(1/2,\sqrt{2})^{1+o(1)}.$$

DLP in real life

Comparing algorithms

- The Index Calculus “evolved” into algorithms which inherited its characteristics: factor base, relations collection and linear algebra.
- The L notation allows to compare algorithms: Index calculus $L_p(1/2, \sqrt{2})$, Gaussian Integers $L_p(1/2, 1)$, and the Number Field Sieve (NFS) $L_p(1/3, \sqrt[3]{64/9})$.
- Records are published on the NMBRTHRY mailing list.

Records allow to find the crossing point

year	algorithm	size of p in decimal digits	author	cost in GIPS[a] years
1991	Gaussian Integers	58	Lammachia, Odlyzko	0.01
1996	Gaussian Integers	85	Weber	0.10
1998	NFS	85	Weber	0.05
1999	NFS	100	Lercier, Joux	0.05
2000	NFS	110	Lercier, Joux	0.20
2005	NFS	130	Lercier, Joux	1.5
2007	NFS	160	Franke et al.	55
2014	NFS	180	Caramel	260
2016	NFS	232	Kleinjung et al	8000

[a] Giga instructions per second

Plan du cours

- Introduction
- Index Calculus
- Quadratic sieve (QS/MPQS)
- Sieving

Fermat's idea

Idea

Fermat (XVII century) computed solutions of the equation

$$X^2 \equiv Y^2 \mod N. \tag{1}$$

It became a classical idea for factoring, e.g. mechanical machines were built in France in early XX century to solve the above equation.[a]

[a] "Discovery of a lost factoring machine", Shallit, Wiliams, Morain

Lemma

If $N = pq$, Equation (1) has four solutions Y for each $X \neq 0$.

Proof.

Using the identity $X^2 - Y^2 = (X - Y)(X + Y)$ we have $Y \equiv \pm X \mod p$ and $Y \equiv \pm X \mod q$. We call X' the unique integer less than N which satisfies the system

$$\begin{aligned} Y &\equiv -X \mod p \\ Y &\equiv X \mod q. \end{aligned}$$

Then the solutions of Equation (1) are $Y = X$, $Y = -X$, $Y = X'$ and $Y = -X'$. □

Fermat's idea

Idea

Fermat (XVII century) computed solutions of the equation

$$X^2 \equiv Y^2 \mod N. \tag{1}$$

It became a classical idea for factoring, e.g. mechanical machines were built in France in early XX century to solve the above equation.[a]

[a] "Discovery of a lost factoring machine", Shallit, Wiliams, Morain

Lemma

If $N = pq$, Equation (1) has four solutions Y for each $X \neq 0$.

Proof.

Using the identity $X^2 - Y^2 = (X - Y)(X + Y)$ we have $Y \equiv \pm X \mod p$ and $Y \equiv \pm X \mod q$. We call X' the unique integer less than N which satisfies the system

$$\begin{aligned} Y &\equiv -X \mod p \\ Y &\equiv X \mod q. \end{aligned}$$

Then the solutions of Equation (1) are $Y = X$, $Y = -X$, $Y = X'$ and $Y = -X'$. □

$$\boxed{\text{50\% of the solutions, i.e. } X' \text{ and } -X'\text{, give } \gcd(X - Y, N) = p \text{ or } q.}$$

Factoring: an example (1)

Not squares but smooth numbers

Let us factor $N = 2041$. We search integers a such that $a^2 - N$ is a square. In order to keep $a^2 - N$ small, we take a approximately equal to $\sqrt{N}$: 46, 47, ... Squares seem to be rare! Kraitchik (1922) proposed to collect integers which are 10-smooth.

We call factor base the set of primes less than 10: 2, 3, 5 and 7.

Collecting relations

$$46^2 - N \; = \; 75 \; = 3 \cdot 5^2$$

Factoring: an example (1)

Not squares but smooth numbers

Let us factor $N = 2041$. We search integers a such that $a^2 - N$ is a square. In order to keep $a^2 - N$ small, we take a approximately equal to $\sqrt{N}$: 46, 47, ... Squares seem to be rare! Kraitchik (1922) proposed to collect integers which are 10-smooth.

We call factor base the set of primes less than 10: 2, 3, 5 and 7.

Collecting relations

$$\begin{aligned} 46^2 - N &= 75 = 3 \cdot 5^2 \\ 47^2 - N &= 168 = 2^3 \cdot 3 \cdot 7 \end{aligned}$$

Factoring: an example (1)

Not squares but smooth numbers

Let us factor $N = 2041$. We search integers a such that $a^2 - N$ is a square. In order to keep $a^2 - N$ small, we take a approximately equal to $\sqrt{N}$: 46, 47, ... Squares seem to be rare! Kraitchik (1922) proposed to collect integers which are 10-smooth.

We call factor base the set of primes less than 10: 2, 3, 5 and 7.

Collecting relations

$$\begin{aligned} 46^2 - N &= 75 \;\; = 3 \cdot 5^2 \\ 47^2 - N &= 168 = 2^3 \cdot 3 \cdot 7 \\ 48^2 - N &= 263 = 263^1 \end{aligned}$$

Factoring: an example (1)

Not squares but smooth numbers

Let us factor $N = 2041$. We search integers a such that $a^2 - N$ is a square. In order to keep $a^2 - N$ small, we take a approximately equal to $\sqrt{N}$: 46, 47, ... Squares seem to be rare! Kraitchik (1922) proposed to collect integers which are 10-smooth.

We call factor base the set of primes less than 10: 2, 3, 5 and 7.

Collecting relations

$$\begin{aligned}
46^2 - N &= 75 = 3 \cdot 5^2 \\
47^2 - N &= 168 = 2^3 \cdot 3 \cdot 7 \\
48^2 - N &= 263 = 263^1 \\
49^2 - N &= 360 = 2^3 \cdot 3^2 \cdot 5
\end{aligned}$$

Factoring: an example (1)

Not squares but smooth numbers

Let us factor $N = 2041$. We search integers a such that $a^2 - N$ is a square. In order to keep $a^2 - N$ small, we take a approximately equal to $\sqrt{N}$: 46, 47, ... Squares seem to be rare! Kraitchik (1922) proposed to collect integers which are 10-smooth.

We call factor base the set of primes less than 10: 2, 3, 5 and 7.

Collecting relations

$$\begin{aligned}
46^2 - N &= 75 = 3 \cdot 5^2 \\
47^2 - N &= 168 = 2^3 \cdot 3 \cdot 7 \\
48^2 - N &= 263 = 263^1 \\
49^2 - N &= 360 = 2^3 \cdot 3^2 \cdot 5 \\
50^2 - N &= 459 = 3^3 \cdot 17
\end{aligned}$$

Factoring: an example (1)

Not squares but smooth numbers

Let us factor $N = 2041$. We search integers a such that $a^2 - N$ is a square. In order to keep $a^2 - N$ small, we take a approximately equal to $\sqrt{N}$: 46, 47, ... Squares seem to be rare! Kraitchik (1922) proposed to collect integers which are 10-smooth.

We call factor base the set of primes less than 10: 2, 3, 5 and 7.

Collecting relations

$$\begin{aligned}
46^2 - N &= 75 = 3 \cdot 5^2 \\
47^2 - N &= 168 = 2^3 \cdot 3 \cdot 7 \\
48^2 - N &= 263 = 263^1 \\
49^2 - N &= 360 = 2^3 \cdot 3^2 \cdot 5 \\
50^2 - N &= 459 = 3^3 \cdot 17 \\
51^2 - N &= 560 = 2^4 \cdot 5 \cdot 7
\end{aligned}$$

Factoring: an example (2)

Combining relations

With the previous relations we have, for all non-negative integers u_{46}, u_{47}, u_{49}, u_{51}:

$$\left(46^{2u_{46}}47^{2u_{47}}49^{2u_{49}}51^{2u_{51}}\right) \equiv 2^{3u_{47}+3u_{49}+4u_{51}}3^{u_{46}+u_{47}+2u_{49}}5^{2u_{46}+u_{49}+u_{51}}7^{u_{47}+u_{51}} \mod N$$

Linear algebra stage

We find u_{46}, u_{47}, u_{49}, u_{51} in $\mathbb{Z}/2\mathbb{Z}$ satisfying

$$\begin{aligned}
u_{47}+3u_{49}+4u_{51} &\equiv 0 \mod 2\\
u_{46}+u_{47}+2u_{49} &\equiv 0 \mod 2\\
2u_{46}+u_{49}+u_{51} &\equiv 0 \mod 2\\
u_{47}+u_{51} &\equiv 0 \mod 2.
\end{aligned}$$

We obtain $u_{46}=u_{47}=u_{49}=u_{51}=1$.

Factoring: an example (3)

Computing X

We multiply the left sides of all the relations to find

$$\begin{aligned} X &= 46^{u_{46}}47^{u_{47}}49^{u_{49}}51^{u_{51}} \mod N \\ &= 46 \cdot 47 \cdot 49 \cdot 51 \mod N \\ &= 311. \end{aligned}$$

Computing Y

We multiply the right sides of all the relations to find

$$\begin{aligned} Y &= \left(2^{3u_{47}+3u_{49}+4u_{51}}3^{u_{46}+u_{47}+2u_{49}}5^{2u_{46}+u_{49}+u_{51}}7^{u_{47}+u_{51}}\right)^{1/2} \mod N \\ &= 2^5 \cdot 3^2 \cdot 5^2 \cdot 7 \mod N \\ &= 1416. \end{aligned}$$

Euclid gives the factorization!

Since $X \not\equiv \pm Y \mod N$, we succeed. We compute

$$\gcd(Y - X, N) = \gcd(1416 - 311, 2041) = 13.$$

The factorization is $2041 = 13 \cdot 157$.

Quadratic sieve

Input: integer $N = pq$ for two primes p and q
Output: p and q

```
Set $B$ to its optimal value
Make the list $\mathcal{F}$ of the primes less than $B$ (factor base)        ▷ Initialization
$a \leftarrow \lfloor \sqrt{N} \rfloor$
repeat$a \leftarrow a + 1$
    if $(a^2 - N)$ is $B$-smooth then
        relations=relations $\bigcup \{a\}$        ▷ Relations collection
    end if
until  #relations $\geq \#\mathcal{F}$
Construct the matrix $M = (m_{a,q})$, $a$ in relations, $q \in \mathcal{F}$ as follows
        $m_{a,q} = \mathrm{val}_q\left(a^2 - N\right)$ .        ▷ Linear algebra
Solve the linear system        transpose$(x)M \equiv 0 \mod 2$.
Compute $X = \prod_{a \text{ in relations}} a^{x_a}$.
Compute $Y = \prod_{q \in \mathcal{F}} q^{(\sum_a x_a)/2}$.
Compute $g = \gcd(X - Y, N)$        ▷ Square root
if $g \neq 1$ or $N$ then
    return $p = g$, $q = N/g$
else
    Find more relations and do the linear algebra again.
end if
```

Correctness and complexity analysis

Correctness

The algorithm relies on two heuristics:

1. The proportion of integers $a^2 - N$ which are B-smooth is the same as the proportion of integers of the same size which are B-smooth.(proven by Pomerance in the simples form, but not in the variants of the algorithm)
2. Each solution of $X^2 \equiv Y^2 \mod N$ found by the algorithm has 50% chances to give a factor. The block version of the Wiedemann algorithm allows to compute 32 or 64 solutions of our linear system with no extra cost. Hence, we only redo the post-processing.

Complexity analysis

We write $a = \lfloor\sqrt{N}\rfloor + x$. Then

$$a^2 - N = (\lfloor\sqrt{N}\rfloor + x)^2 - N = 2x\lfloor\sqrt{N}\rfloor + (x^2 + \lfloor\sqrt{N}\rfloor^2 - N)$$

We consider the case where $x = N^{o(1)}$ trials are enough to collect $\#\mathcal{F}$ relations. After all computations are done, we prove that this choice is possible. Then we have

$$|a^2 - N| = N^{1/2+o(1)}.$$

We repeat the complexity analysis of Index Calculus with $P(\sqrt{N}, B)$ instead of $P(p, B)$. We obtain

$$\text{cost}(QS) = L_N(1/2, 1)^{1+o(1)}$$

Factoring in real life

Comparing algorithms

The quadratic sieve was improved by proposing new variants of the same complexity, like the Multiple Polynomial Quadratic Sieve (MPQS). Later, a factorization variant of the number field sieve (NFS) was invented of complexity $L_N(1/3, \sqrt[3]{64/9})$.

The RSA company offered prizes to whoever factored RSA keys

year	digit size	algorithm	author
1991	100	QS	Lenstra et al
1992	110	QS	Lenstra et al
1993	120	QS	Denny et al
1996	130	QS	Lenstra et al
1999	140	NFS	Franke et al
2003	160	NFS	Franke et al
2005	200	NFS	Franke et al
2009	232	NFS	international

The computation of the present record, RSA 768 bits took 2000 CPU years.

You can factor at home with CADO `http://cado-nfs.gforge.inria.fr`

Plan du cours

- Introduction
- Index Calculus
- Quadratic sieve (QS/MPQS)
- Sieving

The idea of sieving

What we need

In QS, we collect integers $a = \lceil \sqrt{N} \rceil + x$, where x is a small integer, such that $a^2 - N$ is B-smooth.

We need to find the smooth values of $Q(x)$, when $Q(x) = \left(\lceil \sqrt{N} \rceil + x \right)^2 - N$.

Eratosthenes sieve

Given a polynomial $Q(x) \in \mathbb{Z}[x]$, one can compute the values x in an interval $[E_1, E_2]$ such that $Q(x)$ is prime. One marks with a line every value of x which is divisible by two, then by three and so on. The values of x which have no marks correspond to prime values of Q.

The idea of sieving

What we need

In QS, we collect integers $a = \lceil \sqrt{N} \rceil + x$, where x is a small integer, such that $a^2 - N$ is B-smooth.

We need to find the smooth values of $Q(x)$, when $Q(x) = \left(\lceil \sqrt{N} \rceil + x \right)^2 - N$.

Eratosthenes sieve

Given a polynomial $Q(x) \in \mathbb{Z}[x]$, one can compute the values x in an interval $[E_1, E_2]$ such that $Q(x)$ is prime. One marks with a line every value of x which is divisible by two, then by three and so on. The values of x which have no marks correspond to prime values of Q.

$$\boxed{\text{Numbers which have many marks are smooth.}}$$

Sieving: an example

Problem

Find values a in the interval $[3,7]$ such that $Q(a) = a^2 + 1$ is prime, respectively 6-smooth.

Table of sieving

a	3	4	5	6	7
ticks					
$\log(a^2+1)$	$\log 10$	$\log 17$	$\log 26$	$\log 37$	$\log 50$

Computations

Sieving: an example

Problem

Find values a in the interval $[3,7]$ such that $Q(a) = a^2 + 1$ is prime, respectively 6-smooth.

Table of sieving

a	3	4	5	6	7
ticks					
$\log(a^2+1)$	$\log 10$	$\log 17$	$\log 26$	$\log 37$	$\log 50$

Computations

Consider primes less than 6 and their powers less than $\max\{Q(a) \mid a \in [2,7]\}$:

- $\underline{p=2}$, solutions of $a^2 + 1 \equiv 0 \mod 2$ are $\{1\}$;

Sieving: an example

Problem

Find values a in the interval $[3,7]$ such that $Q(a)=a^2+1$ is prime, respectively 6-smooth.

Table of sieving

a	3	4	5	6	7
ticks	$/$		$/$		$/$
$\log(a^2+1)$	$\log 5$	$\log 17$	$\log 13$	$\log 37$	$\log 25$

Computations

Consider primes less than 6 and their powers less than $\max\{Q(a) \mid a \in [2,7]\}$:

- $\underline{p=2}$, solutions of $a^2+1 \equiv 0 \mod 2$ are $\{1\}$;
- $\underline{q=2^2}$, solutions of $a^2+1 \equiv 0 \mod 4$ are $\varnothing$;

Sieving: an example

Problem

Find values a in the interval $[3,7]$ such that $Q(a) = a^2 + 1$ is prime, respectively 6-smooth.

Table of sieving

a	3	4	5	6	7
ticks	$/$		$/$		$/$
$\log(a^2+1)$	$\log 5$	$\log 17$	$\log 13$	$\log 37$	$\log 25$

Computations

Consider primes less than 6 and their powers less than $\max\{Q(a) \mid a \in [2,7]\}$:

- $\underline{p = 2}$, solutions of $a^2 + 1 \equiv 0 \mod 2$ are $\{1\}$;
- $\underline{q = 2^2}$, solutions of $a^2 + 1 \equiv 0 \mod 4$ are $\varnothing$;
- $\underline{p = 3}$, solutions of $a^2 + 1 \equiv 0 \mod 3$ are $\varnothing$;

Sieving: an example

Problem

Find values a in the interval $[3,7]$ such that $Q(a) = a^2 + 1$ is prime, respectively 6-smooth.

Table of sieving

a	3	4	5	6	7
ticks	/		/		/
$\log(a^2+1)$	$\log 5$	$\log 17$	$\log 13$	$\log 37$	$\log 25$

Computations

Consider primes less than 6 and their powers less than $\max\{Q(a) \mid a \in [2,7]\}$:

- $\underline{p = 2}$, solutions of $a^2 + 1 \equiv 0 \mod 2$ are $\{1\}$;
- $\underline{q = 2^2}$, solutions of $a^2 + 1 \equiv 0 \mod 4$ are $\varnothing$;
- $\underline{p = 3}$, solutions of $a^2 + 1 \equiv 0 \mod 3$ are $\varnothing$;
- $\underline{p = 5}$, solutions of $a^2 + 1 \equiv 0 \mod 5$ are $\{2,3\}$;

Sieving: an example

Problem

Find values a in the interval $[3,7]$ such that $Q(a) = a^2 + 1$ is prime, respectively 6-smooth.

Table of sieving

a	3	4	5	6	7
ticks	$//$		$/$		$//$
$\log(a^2+1)$	0	$\log 17$	$\log 13$	$\log 37$	$\log 5$

Computations

Consider primes less than 6 and their powers less than $\max\{Q(a) \mid a \in [2,7]\}$:

- $\underline{p = 2}$, solutions of $a^2 + 1 \equiv 0 \mod 2$ are $\{1\}$;
- $\underline{q = 2^2}$, solutions of $a^2 + 1 \equiv 0 \mod 4$ are $\varnothing$;
- $\underline{p = 3}$, solutions of $a^2 + 1 \equiv 0 \mod 3$ are $\varnothing$;
- $\underline{p = 5}$, solutions of $a^2 + 1 \equiv 0 \mod 5$ are $\{2,3\}$;
- $\underline{p = 5^2}$, solutions of $a^2 + 1 \equiv 0 \mod 25$ are $\{7,18\}$

Sieving: an example

Problem

Find values a in the interval $[3, 7]$ such that $Q(a) = a^2 + 1$ is prime, respectively 6-smooth.

Table of sieving

a	3	4	5	6	7
ticks	$//$		$/$		$///$
$\log(a^2 + 1)$	0	$\log 17$	$\log 13$	$\log 37$	0

Computations

Consider primes less than 6 and their powers less than $\max\{Q(a) \mid a \in [2, 7]\}$:

- $\underline{p = 2}$, solutions of $a^2 + 1 \equiv 0 \mod 2$ are $\{1\}$;
- $\underline{q = 2^2}$, solutions of $a^2 + 1 \equiv 0 \mod 4$ are $\varnothing$;
- $\underline{p = 3}$, solutions of $a^2 + 1 \equiv 0 \mod 3$ are $\varnothing$;
- $\underline{p = 5}$, solutions of $a^2 + 1 \equiv 0 \mod 5$ are $\{2, 3\}$;
- $\underline{p = 5^2}$, solutions of $a^2 + 1 \equiv 0 \mod 25$ are $\{7, 18\}$

Sieving: an example

Problem

Find values a in the interval $[3,7]$ such that $Q(a)=a^2+1$ is prime, respectively 6-smooth.

Table of sieving

a	3	4	5	6	7
ticks	//		/		///
$\log(a^2+1)$	0	$\log 17$	$\log 13$	$\log 37$	0

Computations

Consider primes less than 6 and their powers less than $\max\{Q(a)\mid a\in[2,7]\}$:

- $\underline{p=2}$, solutions of $a^2+1\equiv 0 \mod 2$ are $\{1\}$;
- $\underline{q=2^2}$, solutions of $a^2+1\equiv 0 \mod 4$ are $\varnothing$;
- $\underline{p=3}$, solutions of $a^2+1\equiv 0 \mod 3$ are $\varnothing$;
- $\underline{p=5}$, solutions of $a^2+1\equiv 0 \mod 5$ are $\{2,3\}$;
- $\underline{p=5^2}$, solutions of $a^2+1\equiv 0 \mod 25$ are $\{7,18\}$

Conclusion

The prime values of Q are $Q(4)=17$ and $Q(6)=37$.
The 6-smooth values of Q are $Q(3)=10$ and $Q(7)=50$.

Algorithm for sieving

Algorithm

Input: a monic polynomial $Q(x)$ in $\mathbb{Z}[x]$ and parameters B, E_1, E_2;
Output: all the integers $x \in [E_1, E_2]$ for which $Q(x)$ is B-smooth.

```
Make a list $(p^k, r)$ of prime powers $p^k \leq \max\{|Q(x)|, x \in [E_1, E_2]\}$, with $p < B$,
  and integers $0 \leq r < p^k$ such that $Q(r) \equiv 0 \bmod p^k$
Define an array indexed by $x \in [E_1, E_2]$ and initialize it with $\log_2 |Q(x)|$
for all $(p^k, r)$ considered above do
    for $x$ in $[E_1, E_2]$ and $x \equiv r \bmod p^k$ do
        Subtract $\log_2 p$ from the entry of index $x$;
    end for
end for
Collect the indices $x$ where the array is close to 0 (numerical errors).
```

Cofactorization

In practice we sieve on primes smaller than a bound **fbb** $< B$ and we collect indices x whose value is smaller than a threshold. Then we test smoothness with ECM on the survivals in a step called cofactorization. In the literature, the smoothness bound B is called **lpb**, “large prime bound”, to distinguish from **fbb**, “factor base bound”.

Exercice

What is the condition on **fbb** and **lpb** such that ECM is not needed, i.e., we know that there is only one large prime.

Complexity

New complexity

The complexity of the sieve is

$$B(\log B)^{O(1)} + \left(\sum_{p \text{ prime}} \frac{1}{p-1}\right)(E_2 - E_1).$$

Using the formula $\sum_{p<x} \frac{1}{p} \sim \log\log x$, the complexity of the sieve is $(E_2 - E_1)^{1+o(1)}$. In first approximation, this is the cardinality of the sieved array.

Old complexity

One can show that

$$L_{L_x(\alpha)}(\beta) = L_x(\alpha\beta).$$

Hence, if the smoothness bound B is $L_p(\alpha)$ for some $\alpha < 1$, then one can simply enumerate values to test for smoothness, with a cost of $L(\alpha/2)$ per test. Then the complexity of the algorithm is

$$B^{O(1)} L_p(\alpha/2) = B^{O(1)}.$$

We conclude that the sieving procedure does not change complexity, at the first level of approximation.

Conclusion of first lecture

- DLP in finite fields and factoring have a common history;
- smoothness is the key notion for these problems;
- complexity is best expressed with the L notation;
- the new algorithms evolved from Index Calculus and Quadratic Sieve;
- sieving is an important practical improvement, but unseen in the first approximation of the complexity.

Exercices

1. What is the complexity of solving linear systems by Gaussian elimination.

2. If $f, g : \mathbb{R} \to \mathbb{R}$ are two functions then

$$\min_{x\in\mathbb{R}} \max(f, g) \leq \min_{x\in\mathbb{R}} f + g \leq 2 \min_{x\in\mathbb{R}} \max(f, g).$$

   If f is increasing and g is decreasing and both are continous then the the minimum value of $\max(f, g)$ is obtained when $f(x) = g(x)$.

3. Find $\alpha, c > 0$ so that $B = L_p(\alpha, c)$ is solution of $B = 1/P(p, B)$.

4. What is the speed-up obtained when using the sieve w.r.t. a direct smoothness test with ECM.