Industrial Strength Factorization

Lawren Smithline
Cornell University
lawren@math.cornell.edu
http://www.math.cornell.edu/~lawren

Industrial Strength Factorization

Given an integer N,

determine the prime divisors of N.

Definition. An integer p is prime means:

$1/p$ is not an integer, and

for integers a, b,

if $p \mid ab$, then $p \mid a$ or $p \mid b$.

Industrial Strength Factorization

Application:
break certain public key cryprography.

Metaproblem I: How much does it cost to factor a d digit number N?

Metaproblem II: Given d, design a strategy for picking N which maximizes the cost to factor.

Metaproblem III: Given a determined *adversary* who has solved Metaproblem II and who picks N, how much does it cost to factor N?

Fast and Slow

Definition (for today) A fast algorithm is one which runs in polynomial time in the length of its input.

Example. Multiplication of two numbers N_1 and N_2 each with at most d digits.

about $2d^2$ operations, depending how you count.

$$
\begin{array}{rrrrr}
 & & 1 & 3 & 7 \\
 & \times & & 3 & 7 \\
\hline
 & & & 4 & 9 \\
 & & 2 & 1 & \\
 & & 2 & 1 & \\
 & & 9 & & \\
 & & 7 & & \\
+ & 3 & & & \\
\hline
 & 5 & 0 & 6 & 9
\end{array}
$$

Fast and Slow

Fast examples:

Let N be a d digit integer.

Write N on the board: $O(d)$.

For $a < N$, compute Na: $O(d^2)$.

For $a < N$, compute $GCD(N, a)$: $O(d^3)$.

Fast and Slow

What is slow?

Factor an odd integer N by trial division.

Example $N = 209$.

try 3
try 5
try 7
try 9
try 11 !

Fast and Slow

To factor N with d digits

by trial division, we may have to go up to $\sqrt{N}$, which has about $d/2$ digits.

Slow example: $O(\sqrt{N}) = O(\exp(d/2))$.

For $d = 150$, at speed of 10^9 trials / second, trial division potentially takes 10^{66} seconds.

One year is $\pi \cdot 10^7$ seconds.

$\pi = \sqrt{10}$.

$\pi \cdot 10^{58}$ years is a long time.

Trial Division

Trial division is great for finding small factors.

The adversary picks N with exactly two prime divisors, both large.

What if we magically knew prime numbers?

We wouldn't have to try 9 or 77.

Definition. $\pi(x)$ is the number of primes between 1 and x.

Now check only $\pi(\sqrt{N})$ things.

Trial Division

Definition. $\pi(x)$ is the number of primes between 1 and x.

Chebyshev's Theorem $\pi(x) > x/2\log x$.

Prime Number Theorem (1896) $\pi(x) \sim x/\log x$.

conjectured by Gauss, proved by Hadamard and de la Vallée-Poussin.

Conclusion. $O(\pi(\sqrt{N}))$ is still slow.

Fun fact:

$$\prod_{s<p\leq s+100}^{p \text{ prime}} \approx e^{100}.$$

Summary of Fast and Slow

Definition. $L_N[v, \lambda]$ is

$$O(\exp(\lambda(\log N)^v(\log\log N)^{1-v})).$$

$L_N[v]$ includes $L_N[v, \lambda]$ for every λ.

$L_N[1]$ is slow. e.g. trial division.

$L_N[0]$ is fast, e.g. $GCD(N, a)$.

Avoid embarrassment!
Check whether N is composite.

Answering "Is N composite?" is $L[0]$.
New result in '02. Previous best was

$$O((\log N)^{c \log\log\log N}).$$

Analytical Estimate

Let $\Psi(x, y)$
be number of integers between 1 and x
with all prime divisors less than y.

Theorem For u sufficiently large and $x > 1$,

$$\Psi(x, x^{1/u}) > x/u^{u(1+o(1))}.$$

Lemma (Canfield-Erdős-Pomerance)
There is a constant c such that
for $u > c$ and all $x > 1$,
if $u > (\log x)^{3/8}$, then

$$\Psi(x, x^{1/u}) > x/u^{3u}.$$

Analytical Estimate

Lemma (Canfield-Erdős-Pomerance)
There is a constant c such that
for $u > c$ and all $x > 1$,
if $u > (\log x)^{3/8}$, then

$$\Psi(x, x^{1/u}) > x/u^{3u}.$$

Proof. If $x < u^{3u}$, trivial.

Suppose $x \geq u^{3u} \geq c^{3c} \geq c^3$.

By Chebyshev, $\pi(x^{1/u}) > ux^{1/u}/2\log x$.

Let $m = \lfloor u \rfloor$; $u = m + \theta$.

Let $\pi'(y) = \max(1, \pi(y))$.

Analytical Estimate

Lemma (Canfield-Erdős-Pomerance)
There is a constant c such that
for $u > c$ and all $x > 1$,
if $u > (\log x)^{3/8}$, then

$$\Psi(x, x^{1/u}) > x/u^{3u}.$$

Proof (continued).

Claim. $(2\log x)^{m+1} < u^{3u}$:

$$(2\log x)^{m+1} < (2\log x)^{u+1}.$$

For $u > 3$ and $u \geq \log^{3/8} x$,

$$(2\log x)^{u+1} \leq (u^3)^u.$$

Analytical Estimate

Lemma (Canfield-Erdős-Pomerance)
There is a constant c such that
for $u > c$ and all $x > 1$,
if $u > (\log x)^{3/8}$, then

$$\Psi(x, x^{1/u}) > x/u^{3u}.$$

Proof (concluded).

$$\begin{aligned}
\Psi(x, x^{1/u}) &> \pi(x^{1/u})^m \pi'(x^{\theta/u})/(m+1)! \\
&> \frac{(ux^{1/u})^m x^{\theta/u}}{2u^m \log^{m+1} x} \\
&= x/(2\log x)^{m+1} \\
&> x/u^{3u}.
\end{aligned}$$

Exeunt Canfield, Erdős, and Pomerance.
Enter Fermat.

Fermat's Method

Quick... factor 3599.

$3599 = 60^2 - 1^2$.

If $x^2 \equiv y^2\ (N)$, compute $GCD(x - y, N)$.

50% chance to get a factor of N, because

$$(x - y)(x + y) \equiv 0\ (N).$$

$3599 = 59 \cdot 61$.

Fermat's Method

Fermat's method: for each $x = 1, 2, ...,$
check whether $N + x^2$ is a square.

The adversary chooses $N = pq$ with
prime factors far away from $\sqrt{N}$.

$p \approx N^{1/e}$ will do.

Now, Fermat's method is $O(\sqrt{N})$, which is slow.

Let's use another idea of Fermat.

Fermat's Method

(This isn't the other useful Fermat idea.)

Factor 64027.

Yes, $x^3 + y^3 = (x+y)(x^2 - xy + y^2)$.

But cubes are rarer than squares.

Pollard's $p-1$ Method

Fermat's Little Theorem
If p is prime and $a \in \mathbf{Z}$, and $GCD(a,p) = 1$, then

$$a^{p-1} \equiv 1 \ (p).$$

Given N, if a kind oracle would tell us $p-1$ for p dividing N, then

$$a^{p-1} - 1 \equiv 0 \ (p).$$

Therefore, $GCD(a^{p-1} - 1, N)$ is p or N.

Pollard's method finds m such that $(p-1) \mid m$.

Thus, $GCD(a^m - 1, N)$ is p or N.

Pollard's $p-1$ Method

Let $m_j = LCM(1, 2, \ldots, j)$.

Given $N = pq$, for large enough j, $p-1$ divides m_j.

So: $a^{m_j} \equiv 1\ (p)$. Maybe $a^{m_j} \not\equiv 1\ (q)$.

Then $GCD(a^{m_j} - 1, N) = p$.

Here's an example for $N = 20701$.

Pollard's $p-1$ Method for $N = 20701$

Pick $a = 2$.
(If $GCD(a, N)$, that's an instant win.)

Let $m_j = LCM(1, 2, \ldots, j)$.

Let $X_j = 2^{m_j}$ reduced modulo N.

Let $F_j = GCD(X_j - 1, N)$.

j	m_j	X_j	F_j
2	2	4	1
3	6	64	1
4	12	4096	1
5	60	6493	1
6	60	6493	1
7	420	17273	127
8	840	13717	127
9	2520	6986	127

So $N = 127 \cdot 163$.

Next: another example.

Pollard's $p-1$ Method for $N = 5029$

j	m_j	X_j	F_j
2	2	4	1
3	6	64	1
4	12	4096	1
5	60	1156	1
6	60	6493	1
7	420	4153	1
8	840	2968	1
9	2520	625	1
11		4920	1
13		4327	1
16		4991	1
17		850	1
19		2077	1

Better luck with the next method!

Pollard's $p-1$ Method Performance

Pollard's method finds factors p such that prime power factors of $p-1$ are all small.

Typical size of largest prime power factor of $p-1$ is

$$p^{1-1/e} \approx N^{1/e-1/e^2}.$$

So: Pollard's method is slow against a smart adversary.

Pollard's $p-1$ Method Performance

The adversary already chooses:

$N = pq$, the product of exactly two primes,

p, q are not small.

p, q are not too close to $\sqrt{N}$

$(p \approx N^{1/e})$

Now add: $p-1, q-1$ each have at least some large prime power factor.

Rational Sieve Method

Given N, let's solve $\quad x^2 \equiv y^2 \ (N)$.

Method:

Find lots of integers a such that a and $N + a$ have only small prime factors.

Pick a subset of the $a's$ such that

$$\prod_i a_i/(N + a_i) = s^2/t^2.$$

Conclude: For $x = s$, $y = t$,

$$x^2 \equiv y^2 \ (N).$$

$GCD(x - y, N)$ has even odds to be a prime divisor of N.

Rational Sieve Vocabulary

Given N, we pick B, the *smoothness bound*.

We treat a prime p less than B as small.

An integer a is B-smooth means every prime p dividing a is less than B.

The set $\mathcal{B}$ of small primes is the *factor base*.

A pair $(a, N+a)$ of smooth numbers is a *relation* or *smooth relation*.

Here is an example with $N = 5029$, $B = 20$.

Rational Sieve for $N = 5029$

Factor $N = 5029$.

Factor base $\mathcal{B} = \{2, 3, 5, 7, 11, 13, 17, 19\}$.

	a	$a+N$	2	3	5	7	11	13	17	19
A	11	5040	0	0	1	1	1	0	0	0
B	20	5049	0	1	1	0	1	0	1	0
C	25	5054	1	0	0	1	0	0	0	0
D	91	5120	0	0	1	1	0	1	0	0
E	119	5148	0	0	0	1	1	1	1	0
F	171	5200	0	0	0	0	0	1	0	1
G	196	5225	0	0	0	0	1	0	0	1
H	221	5250	1	1	1	1	0	1	1	0
I	275	5304	1	1	0	0	1	1	1	0

Eliminate 19 with $F + G$ replacing F, G.

Eliminate 2 and 3 with $B + C + H$ and $H + I$ replacing B, C, H, I.

Eliminate 19 with $F + G$.
Eliminate 2 and 3 with $B + C + H$ and $H + I$.

	5	7	11	13	17
A	1	1	1	0	0
D	1	1	0	1	0
E	0	1	1	1	1
$F+G$	0	0	1	1	0
$B+C+H$	0	0	1	1	0
$H+I$	1	1	1	0	0

Only row E has 17, so strike it.

	5	7	11	13
$H+I$	1	1	1	0
A	1	1	1	0
D	1	1	0	1
$F+G$	0	0	1	1
$B+C+H$	0	0	1	1

Relations:

$A + H + I$
$A + D + F + G$
$B + C + F + G + H$

Rational Sieve for $N = 5029$

Assembling relations:

	a	$a+N$	2	3	5	7	11	13	17	19
A	11	5040	0	0	1	1	1	0	0	0
H	221	5250	1	1	1	1	0	1	1	0
I	275	5304	1	1	0	0	1	1	1	0

Now compute over $\mathbf{Z}$, not mod 2:

	$a/(a+N)$	2	3	5	7	11	13	17	19
A	11/5040	−4	−2	−1	−1	1	0	0	0
H	221/5250	−1	−1	−3	−1	0	1	1	0
I	275/5304	−3	−1	2	0	1	−1	−1	0
	$A+H+I$	−8	−4	−2	−2	2	0	0	0

$$11^2 \equiv (2^4 \cdot 3^2 \cdot 5 \cdot 7)^2$$

For $x = 11, y = 5040$, $\quad x^2 \equiv y^2 \ (N)$.

Alas, $N = y - x$.

Rational Sieve for $N = 5029$

Assembling relations:

	a	$a+N$	2	3	5	7	11	13	17	19
A	11	5040	0	0	1	1	1	0	0	0
D	91	5120	0	0	1	1	0	1	0	0
F	171	5200	0	0	0	0	0	1	0	1
G	196	5225	0	0	0	0	1	0	0	1

Now compute over $\mathbf{Z}$, not mod 2:

	$a/(a+N)$	2	3	5	7	11	13	17	19
A	11/5040	-4	-2	-1	-1	1	0	0	0
D	91/5120	-10	0	-1	1	0	1	0	0
F	171/5200	-4	2	-2	0	0	-1	0	1
G	196/5225	2	0	-2	2	-1	0	0	-1
	total	-16	0	-6	2	0	0	0	0

For $x = 7, y = 2^8 5^3 = 32000, \quad x^2 \equiv y^2\ (N)$.

Hooray! GCD(N, 32007) = 47.

Rational Sieve Performance

How common are smooth numbers?

From a random sample of integers of size $L[v, \lambda]$, the fraction of $L[w, \mu]$ smooth ones is:

$$\Psi(L[v, \lambda], L[w, \mu])/L[v, \lambda].$$

The expected sample size to find one $L[w, \mu]$ smooth ones is:

$$L[v, \lambda]/\Psi(L[v, \lambda], L[w, \mu]) =$$

$$L[v - w, (v - w)\lambda/\mu].$$

The expected sample size to find $L[w, \mu]$ smooth ones is:

$$L[w, \mu]L[v - w, (v - w)\lambda/\mu].$$

Rational Sieve Performance

We choose a small prime bound B of size $L[w, \mu]$.

Factor base $\mathcal{B}$ has size $L[w, \mu]$.

Search for smooth pairs $a, N + a$;
a is B-smooth. $N + a$ has size $L[1, 1]$.

Search region size:

$$L[1 - w, (1 - w)/\mu]L[w, \mu].$$

Linear algebra problem:

$$L[w, \mu]^2 = L[w, 2\mu].$$

Rational Sieve Performance

Search region size:

$$L[1-w,(1-w)/\mu]L[w,\mu].$$

Minimize $\max(w, 1-w)$: $w = 1/2$.
Search region size is:

$$L[1/2, \mu + 1/2\mu].$$

Minimum at $\mu = \sqrt{2}/2$.
Search region size is:

$$L[1/2, \sqrt{2}].$$

Linear algebra is also

$$L[1/2, \sqrt{2}].$$