# Lenstra's Elliptic Curve Factoring Method
Connecticut Number Theory Summer School
May, 2018

Jeremy Teitelbaum

# The problem at hand

**Problem**

*Given a positive composite integer $N$, find a proper prime divisor of $N$.*

# Trial division is impractical

The 'grade school' method to solve the factoring problem by systematically trying integers less than $N$ (or prime numbers less than $N$) and checking to see if you find a factor requires, in the worst case, on the order of $\sqrt{N}$ divisions.

If a division takes, say, $10^{-12}$ seconds on some miracle computer, then factoring a 100 digit number would require $10^{38}$ seconds or more than $10^{30}$ years. (The universe is about $10^{10}$ years old.)

A different approach is needed.

# Overview of factoring methods

Modern methods of factoring fall into two categories:

- Methods based on algebraic groups, such as the $p-1$ method, the elliptic curve method, and generalizations.
- Sieve methods, such as the quadratic and number field sieves.

# Overview of factoring

Typically, the algebraic group methods are used first to identify "small factors" of large numbers $N$; and once those are found, or ruled out, the sieve methods are used.

In the best case these algorithms are believed to be sub-exponential, meaning that their running times grow more slowly than exponential in the number of digits of $N$; but they are far from polynomial time.

The complexity of factoring is not known.

There is a polynomial time algorithm for a "quantum computer."

# First make sure your number is composite

## Theorem (Fermat)

*Suppose that $N$ and $a$ are integers with $(a, N) = 1$. If*

$$a^{N-1} \not\equiv 1 \pmod{N}.$$

*then $N$ is composite.*

Fermat's theorem allows for a quick test of compositeness.

## The Fermat Test

Given $N$ (large), pick a random small $a$ and compute $a^{N-1}$ mod $N$.
If the result isn't 1, $N$ is composite.

## Definition

If $a^{N-1} \equiv 1 \pmod{N}$, then $N$ is called a pseudoprime to base $a$.

# Pseudoprimes are rare

There are 21853 pseudoprimes to base 2 less than $25 \times 10^9$.

If a number passes the Fermat test for a bunch of random bases, then spend your time trying to prove it prime rather than trying to factor it.

There are refinements to the Fermat test that are even more effective.

# Efficient Modular Exponentiation

Applying the Fermat Test requires computing $a^x$ mod $N$ where $x$ is large; and similar calculations are needed in the ECM method as well.

## Proposition

*$a^x$ mod $N$ can be computed in time $O(\log x)$ for fixed $N$ and $a$.*

## Algorithm

```
Set m=1 and s=a.
While x>0:
   if x is odd, set m=(m*s mod N)
   set s=s*s
   set x=x/2, rounding off
return m as your answer
```

# The $p-1$ algorithm

Suppose $N$ is (odd and) composite. Then the multiplicative group of units $(\mathbf{Z}/N\mathbf{Z})^*$ is not cyclic, so it is a product of cyclic groups by the fundamental theorem of abelian groups.

The strategy of the $p-1$ method is to

1. pick a base $a$ (like 2);
2. try to find an exponent $M$ so that $a^M \equiv 1$ in one of the cyclic factors of $(\mathbf{Z}/N\mathbf{Z})^*$ but not all of them.
3. then $(a^M - 1, N)$ will be a non-trivial factor of $N$.

If $p$ is an odd prime factor of $N$, then we can try $M = K(p-1)$ for various $K$. But how to find this $M$ if we don't know $p$?

# Smoothness

## Definition

An integer $N$ is called $B$-smooth if all the prime factors of $N$ are at most $B$. It is called $B$-powersmooth if all the prime powers dividing $N$ are at most $B$.

For example, the number

$$N = 334525266131638071081700620534407516651520000000000$$

is 41-smooth. (It is 41!). It is divisible by $2^{164}$ and all the other prime powers dividing 41! so 41! is $2^{164} + 1$ powersmooth.

# The $p-1$ method, 2

The hope for the $p-1$ method is that if $p$ is one of the prime divisiors of our integer $N$ then $p$ has the property that $p-1$ is $B$-powersmooth for some not too big $B$.

Then we take a integer $M$ that is divisible by powers of the primes less than $B$ hoping to get a multiple of $p-1$. For example, take:

$$M = \prod_{p \leq B} p^{[\log_p(B)]}.$$

Then compute $(a^M - 1, N)$ and see what happens. If you don't find anything, make $B$ bigger.

# A simple example

Suppose $N = F_5 = 2^{2^5} + 1$ is the fifth Fermat number. We can't use $a = 2$ because clearly high powers of $a$ are going to be $-1$ mod $N$. Let's try $a = 5$ instead. Take $M = 10!$.

$$5^M - 1 \equiv 1869036133 \pmod{F_5}$$

and $(1869036133, F_5) = 641$.

# The Elliptic Curve Method

For the $p-1$ method to work, we have to be lucky enough to have a prime factor $p$ of $N$ so that $p-1$ is $B$-powersmooth for a relatively small $B$.

If the number $N$ we are trying to factor doesn't have this property, then the $p-1$ method won't work.

The elliptic curve method opens the door to more situations in which we can apply the idea of the $p-1$ method.

# ECM, cont'd

Suppose $N = UV$ where $U$ and $V$ are proper factors. Let $E$ be an elliptic curve over $\mathbf{Z}$. Then

$$E(\mathbf{Z}/N\mathbf{Z}) = E(\mathbf{Z}/U\mathbf{Z}) \times E(\mathbf{Z}/V\mathbf{Z}).$$

Suppose that we can find a point $P$ on this curve mod $N$ so that a multiple $K$ of $P$ is zero in the first factor but not the second.

If we were to write $E$ in Weierstrass form, and the point $P$ in (reduced) homogeneous coordinates $[x(P) : y(P) : z(P)]$, then this condition would mean that $z(KP)$ is divisible by $U$ but not by $V$.

In other words, $(z(KP), N)$ would give us a proper factor of $N$.

# ECM,3

If we were fortunate enough that (say) the order of the first of the two factor groups $n = |E(\mathbf{Z}/U\mathbf{Z})|$ were $B$-powersmooth for a (relatively) small $B$, Then we could use the trick of the $p-1$ method and choose our $K$ to hopefully be divisible by $n$.

The Riemann hypothesis for elliptic curves over finite fields tells us that if $U$ is prime then $n$ is roughly $p$.

It seems reasonable to assume that these group orders are essentially random numbers of size roughly $p$. So the chance that $n$ is $B$-powersmooth is the same order as $p-1$ having that property.

But there are many elliptic curves!

# ECM: An example

Consider the $7^{th}$ Fermat number

$$F_7 = 2^{128} + 1 = 340282366920938463463374607431768211457.$$

A Fermat test to base 3 confirms that $N$ is composite:

$$3^{N-1} \equiv 47511664169441434718291075092691853899 \not\equiv 1 \pmod{N}$$

# ECM: An example, cont'd

Try many elliptic curves $E$ and a large but not hopelessly large $B$ so that, if $U$ is a factor of $N$, then $|E(\mathbf{Z}/U\mathbf{Z})|$ is $B$-powersmooth.

Consider the family of elliptic curves $E_a : y^2 = x^3 + ax + 1$ which has the obvious point $P = (0, 1)$ on it.

Consider $a$ as integers (for example in the range $[-100, 100]$) and try $B = 10000$. Let

$$B_4 = \prod_{p<10000} p^{[\log(10000)/log(p)]}$$

For each $a$, compute $[B_4](P)$ in $E_a(\mathbf{Z}/N\mathbf{Z})$. Use the same idea as in modular exponentiation to compute $[B_4](P)$ (repeated doubling of the point).

In computing this, you have to do modular inversion mod $N$. If that fails: you've found a factor of $N$.

# ECM: an example, cont'd

A very naive search finds that

$$E_{-91} : y^2 = x^3 - 91x + 1$$

gives us our factorization and finds the factor

$$N = UV = (59649589127497217)(5704689200685129054721)$$

Some quick pseudoprime tests suggests that this is prime (it is) and $U - 1$ has a prime divisor with 15 digits and $V - 1$ has a prime factor with 12 digits, so this factorization is unlikely to have been found by the $p - 1$ method.

# Lenstra's ECM for theoretical purposes

Let $N$ be an integer that is not a prime power and is relatively prime to 6.

1. Choose integers $v$ and $w$ and set

$$k = \prod_{2}^{w} r^{e(r)}$$

   where $e(r)$ is maximal so that $r^{e(r)} \leq v + 2\sqrt{v} + 1$.

2. Draw three elements $a, x_0, y_0$ at random modulo $N$. Let $b = y_0^2 - x_0^3 - ax_0$. Then $P = (x_0, y_0)$ is a point on the curve $E$ with equation $y^2 = x^3 + ax + b$.

3. Try to compute $kP$. If you find a divisor, cheer! Otherwise, try again until you've tried $h$ times, then give up.

Here $v$ serves as bound on the smallest prime factor of $N$ and $w$ a bound on the largest prime divisor of the group of points mod that prime.

# Analysis of ECM

The method described above will find a divisor provided that the following conditions hold:

1. $N$ has a prime divisor $p \leq v$.
2. The elliptic curve chosen is non-singular mod $p$.
3. The order of the group of points on that curve is $w$-smooth.
4. For some *other* divisor $q$ of $N$, $E$ is non-singular mod $q$ and the order of $P$ mod $q$ is NOT divisible by the largest prime divisor of the order of $P$ mod $p$.

How likely is this to happen? There is a trade off between time spent on one curve $v, w$ and number of times you try $h$.

# Analysis of ECM, cont'd

Lenstra uses properties of elliptic curves to show the following. Let

$$u = |\{s : |s - p - 1| < \sqrt{p} \text{ and } s \text{ is } w\text{-smooth}\}|$$

So $u$ is the number of $w$-smooth numbers that could possibly be the order of an elliptic curve mod $p$. Then the chance that a single attempt at Lenstra's ECM method works (i.e. that a particular $(a, x, y)$ succeeds) is closely related to the chance of a smooth group order, or $u/(2\sqrt{p} + 1)$.

Conjecture: The chance that a randomly chosen integer in the range $(x - \sqrt{x} + 1, x + \sqrt{x} + 1)$ is $w$-smooth is the same as the chance that a randomly chosen integer of that size is $w$-smooth.

# Analysis of ECM, cont'd

**Theorem**

*Suppose $N$ has two prime divisors, that $p$ is the smallest prime divisor of $N$, and that $N$ is prime to 6. Let $f(w)$ be the probability that a randomly chosen integer in the range*

$$(p + 1 - \sqrt{p}, p + 1 + \sqrt{p})$$

*is $w$-smooth. Assuming $v \geq p$, then there is an explicitly computable constant $c$ so that the chance of success of the ECM with parameters $v, w, h$ is*

$$1 - c^{-hf(w)/\log(v)}.$$

# Analysis of ECM, cont'd

Meanwhile, the running time for the algorithm is

$$O(hw \log(v) M(N))$$

because you have to compute $kP$, and

$$\log k \leq w \log(v).$$

# Analysis of ECM, cont'd

Let

$$L(x) = e^{\sqrt{\log(x)\log\log(x)}}.$$

Lenstra invokes a famous result of Canfield, Erdös, and Pomerance that states that the probability that a random positive integer $s \leq x$ is $L(x)^{\alpha}$-smooth is (asymptotically)

$$L(x)^{\frac{1}{2\alpha}+o(1)}$$

# Analysis of ECM, cont'd

One way to think of the problem is: for a fixed chance of success, minimize running time. In other words, for a fixed value of $hf(w)$, minimize $hw$. This boils down to minimizing $w/f(w)$.

The CEP result says that

$$w/f(w) = O(L(p)^{\frac{1}{2\alpha}+\alpha+o(1)}$$

so the minimum value happens when $\alpha = \frac{1}{\sqrt{2}}$, and, in that case.

$$w/f(w) = O(L(p)^{\sqrt{2}}).$$

In other words, for a fixed chance of success, the running time behaves like

$$L(p)^{\sqrt{2}} = e^{\sqrt{((2+o(1))\log p \log\log p}}$$

where $p$ is the smallest prime factor of $N$.

# Further Discussion

To make ECM work in practice there are many refinements, including:

1. Using highly optimized elliptic curve operations and modular arithmetic;
2. Careful choice of parameters (smoothness bound, for example).

Bernstein's notes from the 2006 Winter School address some of these optimizations.