Factoring Algorithms
The $p - 1$ Method and Quadratic Sieve

November 17, 2008
Fermat’s factoring method

- Fermat made the observation that if \( n \) has two factors that are near one another (and hence near the square root of \( n \)) then one can find them by searching the sequence \( n + y^2 \) for \( y = 0, 1, 2, 3, \ldots \) until finding a perfect square \( x^2 \).

- Then \( n + y^2 = x^2 \), so

\[
    n = x^2 - y^2 = (x - y)(x + y)
\]

is a factorization of \( n \). If \( n = pq \) is a product of two primes that are near to one another, this finds the factors fairly quickly.
The $p - 1$ method

- Due to Pollard 1974.
- Assume that $n$ has a prime factor $p$ such that all the prime factors of $p - 1$ are fairly small, then we can find a nontrivial factor of $n$ by computing $b \equiv a^{B!} \pmod{n}$ for some chosen $B$. This computation can be done quickly so long as $B$ is not chosen too large.
- If $p - 1$ has only small prime factors, then for $B$ sufficiently large $p - 1$ will divide $B!$, so $B! = (p - 1)k$ for some integer $k$. Hence
  \[
  b = a^{B!} = a^{(p-1)k} = (a^{p-1})^k \equiv 1^k \equiv 1 \pmod{p}
  \]
  by Fermat’s Little Theorem, so $p$ is a factor of $b - 1$ and $n$. Thus, $\gcd(b - 1, n)$ will have $p$ as a factor, so by computing $\gcd(b - 1, n)$ we have found a non-trivial factor of $n$. 
The $p − 1$ method

- The $p − 1$ method works so long as $B$ is *big enough* so that all the prime factors (with their multiplicity) of $p − 1$ occur in $B!$, but *not so big* that computing $b ≡ a^{B!} \pmod{n}$ is prohibitively time consuming.

- Note that it is well known that the sequence $B!$ of factorials is of exponential growth rate, so no matter how fast your computer, there will be some value of $B$ such that the computation will take more than your lifetime to finish.

- To emphasize this last point, let’s record the first few terms of the sequence $n!$ below:

  $1, 2, 6, 24, 120, 720, 5040, 40320, 362880, 3628800, 39916800, 479001600, 6227020800, \ldots$

  Obviously it is growing pretty fast. In fact, this sequence eventually grows faster than $a^n$ for any given base $a$. 
The $p-1$ method

- MORAL: In choosing $p, q$ for an RSA system, it is important that both $p$ and $q$ are chosen so that each of $p-1$ and $q-1$ has at least one large prime factor. Otherwise, a clever attacker just might get lucky with the $p-1$ method, factor $n$, and decode the message.

- This is easy to do. Choose a large prime $p_0$ at random, say with 40 or so decimal digits. Now look in the sequence of numbers of the form $kp_0 + 1$, for $k = 10^{60} + 1, 10^{60} + 2, 10^{60} + 3, \ldots$ until you find some prime $p$ or roughly 100 decimal digits such that $p = kp_0 + 1$. Then $p-1$ has a large prime factor, namely $p_0$, by construction. Repeat the method to find $q$.

- It should also be noted that $p, q$ should not be too close together, or else someone might find the factors of $n$ using Fermat’s approach. For this reason, it is important to choose $p, q$ to NOT have exactly the same number of decimal digits.
The Quadratic Sieve

**Theorem (Basic Principle)**

Let \( n \) be a positive integer. Suppose there exist integers \( x, y \) such that \( x^2 \equiv y^2 \pmod{n} \) but \( x \not\equiv \pm y \pmod{n} \). Then \( \gcd(x - y, n) \) gives a non-trivial factor of \( n \).

**Proof.**

Set \( d = \gcd(x - y, n) \). Then \( d \) is a divisor of \( n \) so \( 1 \leq d \leq n \). If \( d = n \) then \( n \mid (x - y) \) so \( x \equiv y \pmod{n} \), which is contrary to the hypothesis. If \( d = 1 \) then \( n \) does not divide \( x - y \). But \( n \) divides \( x^2 - y^2 = (x - y)(x + y) \) by hypothesis, so \( n \) must therefore divide the second factor \( x + y \), by Euclid’s Lemma. In other words, \( x \equiv -y \pmod{n} \), which is again contrary to hypothesis.

This shows that \( 1 < d < n \), so \( d \) is a nontrivial factor of \( n \). That’s what we needed to show.
The Quadratic Sieve

EXAMPLE. Suppose we would like to factor $n = 3837523$. We observe the following:

\[
\begin{align*}
9398^2 & \equiv 5^5 \cdot 19 \\
19095^2 & \equiv 2^2 \cdot 5 \cdot 11 \cdot 13 \cdot 19 \\
1964^2 & \equiv 3^2 \cdot 13^3 \\
17078^2 & \equiv 2^6 \cdot 3^2 \cdot 11
\end{align*}
\]

where all the congruences are mod $n$. By multiplying these together, we obtain the congruence

\[
(9398 \cdot 19095 \cdot 1964 \cdot 17078)^2 \equiv (2^4 \cdot 3^2 \cdot 5^3 \cdot 11 \cdot 13^2 \cdot 19)^2
\]

which simplifies mod $n$ to give

\[
2230387^2 \equiv 2586705^2 \pmod{3837523}.
\]
EXAMPLE, CONTINUED.

In this case, \(2230387 \not\equiv \pm 2586705 \pmod{3837523}\) so by computing \(\gcd(2230387 - 2586705, 3837523) = 1093\) we find a factor 1093 of \(n\). Then the other factor is \(n/1093 = 3511\), so \(n = 1093 \cdot 3511\).
The Quadratic Sieve

- The idea is to find several relations of the form

\[ x_i^2 \equiv \text{a product of small primes} \pmod{n}. \]

If you get enough relations of that form, then some of them can be combined to give a congruence \( x^2 \equiv y^2 \pmod{n} \).

- Sometimes you are unlucky, and \( x \equiv \pm y \pmod{n} \). If that happens, look for more relations of the indicated type and try again. Eventually, if you happen upon a case where \( x^2 \equiv y^2 \pmod{n} \) but \( x \not\equiv \pm y \pmod{n} \) then you have factored \( n \).

- This is how the RSA challenge (Scientific American 1977) was finally cracked in 1994, using a factor base of more than half a million primes and 1600 computers in parallel. The project took 7 months to complete.
The Quadratic Sieve

How does one find numbers $x_i$ such that

$$x_i^2 \equiv \text{a product of small primes (mod } n)$$

(The set of desirable small primes is called the factor base.)

Examine numbers of the form $\left[\sqrt{kn} + j\right]$ where $j$ is fairly small. Here $[r]$ means the integer part of a real number $r$. The square of such a number $x_i$ will be likely to have only small factors mod $n$, since its residue mod $n$ is fairly small relative to the size of $n$.

EXAMPLE. For $n = 3837523$ as before, we get $8077 = \left[\sqrt{17n} + 1\right]$ and $9398 = \left[\sqrt{23n} + 1\right]$. 
The Quadratic Sieve

- Let \( \{p_1, p_2, \ldots, p_t\} \) be a chosen factor base. Find numbers \( x_i \) such that \( x_i^2 \) is congruent to a product of primes from the factor base. So we can write

\[
x_i^2 \equiv p_1^{e_{i1}} p_2^{e_{i2}} \cdots p_t^{e_{it}} \pmod{n}
\]

for each \( i \). Here the exponents \( e_{ij} \) are non-negative integers. (Some of them might be zero.)

- The exponents \( e_{ij} \) produced above give a matrix with \( t \) columns. We want to find rows in that matrix such that the sum of the rows gives a row vector whose entries are all even, since then the product of the corresponding \( x_i^2 \) will be congruent to the square of a product of primes in the factor base.
**The Quadratic Sieve**

- **EXAMPLE.** One can find such a matrix for \( n = 3837523 \) by repeatedly examining numbers of the form \( x_i = \lfloor \sqrt{kn} + j \rfloor \) and factoring these numbers.

- This gives a matrix of the form

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>11</th>
<th>13</th>
<th>17</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>9398^2</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>19095^2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1964^2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17078^2</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8077^2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3397^2</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14262^2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>