Chapter 8
Factorising and Finding Discrete Logarithms Simple Techniques

February 15, 2010

8

The security of many cryptographic techniques rests on:

1. The infeasibility of factorising large integers (Example: RSA, Rabin Encryption)

2. The infeasibility of finding discrete logarithms (DL) (Example: RSA, DSA Diffie-Hellman, etc.)

For example RSA relies on both these difficulties for safety: if $e = public\,key$, $n = modulus$, $secret\ key\ d = e^{-1}\ mod\ \phi(n)$ and $\phi(n)$ can only be found by factorising n; similarly if $m = a\ plaintext\ message$ and $c = cyphertext$ then we know $m = c^d\ mod\ n$ so if an attacker knows $m,\ d,\ n$ and can solve the DL problem he can find d.

Chapter 9 looks at sophisticated techniques which an attacker may use in these two cases; they are based on the use of a Factor Base. This module looks at some simpler techniques (a selection from many).

8.1 Simple Factorisation

Factorisation of n.

1. **Fermat's method**

   Assume n is odd, otherwise divide by two. Seek for solutions $x,\ y$ to $x^2 - y^2 = n$. If we find one we have $(x + y)(x - y) = n$, so $(x + y)$ or $(x - y)$ must have a common factor with n, and this can be found using Euclid. (Note: a useless solution are is $(x + y) = n,\ (x - y) = 1$)

Example $n = 35,\ x = 6,\ y = 1,\ (x + y) = 7,\ (x - y) = 5$

Fermat's method starts seeking $x,\ y$ with $x = \lfloor\sqrt{n}\rfloor + 1 = integral\ part\ of(\sqrt{n}) + 1$ and $y = 1$ and runs as follows.
$x = \lfloor\sqrt{n}\rfloor + 1,\ y = 1$

r=x$^2 - y^2 - n$
↓
(X) $r = 0?$ *Yes* ⟶ x$^2 - y^2 = n = (x + y)(x - y) = possible\ solution$
No
↓
$r > 0?$ *Yes* ⟶ r=r-2y=1
y=y+1
↓
Go to **(X)**
No
↓
r=r+2x+1
x=x+1
↓
Go to(**X**)

An Example factorise 7313

x	86	86	86	86	86	86	86	86	86	86	87	87	87	87	87	87	87
y	1	2	3	4	5	6	7	8	9	10	10	11	12	13	14	15	16
r	82	79	74	67	58	47	34	19	2	-17	156	135	112	87	60	31	0

$x = 87,\ y = 16,\ (x + y) = 103,\ (x - y) = 71,\ 7313 = 103 \times 71$

The process can be speeded up by noting that $x,\ y$ can never be both odd or both even since n is odd.

$x = \sqrt{n} + 1,\ y = 0$ if x is odd, $y = 1$ if x is even.
r=x$^2 - y^2 - n$
↓
(X) $r = 0?$ *Yes* ⟶ *Success*
No
↓
$r > 0?$ *No* ⟶ r=r+2(x+y), x=x+1, y=y-1 *Go to* **(X)**
Yes
↓
r=r-4y-4
y=y+2
↓
Go to(**X**)

This cuts out half the number of steps.

The method is efficient if n has a factor of approximately size $\sqrt{n}$. Otherwise it is very slow.

2. **The (p-1) Method of Factorisation**

   If p_i is the i^{th} prime and n is the integer to be factorised evaluate the Highest Common Factor (HCF) of n and $M = \prod_i p_i \; mod \; n$ where the product runs up to p_k some large prime appropriate to the scale of the problem (size of n) i.e. Find $HCF(M,\ n)$. Rather than calculating $\prod_i p_i$ which is slow we evaluate $M = (a^A - 1) \; mod \; n$ with $A = \prod_j p_j^{e_j}$ with p_j being successive primes and e_j being powers $e_j \sim \frac{K}{ln\ p_j}$ so that all components $p_j^{e_j}$ are of size K (which is appropriate to the scale of the problem). The idea is that A will be divisible by the LCM of $(p_i - 1)$, giving $M = (a^A - 1)$ divisible by some p_i. See Theorem 1.10 in Chapter 3.

   The range of the products $\prod p_i$ determined by k and or K can be determined by remembering that a random integer x has largest prime factor x^α with $E(\alpha) = 0.62$, and second largest prime factor x^β with $E(\beta) = 0.26$ approximately (see Knuth ACP Vol 24.5.4). An approach could be to choose $k,\ K$ so that all primes up to n^β are included in a^A, and then seek HCFs for a^{Ap} where p represents a succession of larger primes, to extend past n^α if necessary.

   Typically one takes $a = 2$. Note that raising an integer to a power $mod\ n$ can be speeded up by successive squarings and by using techniques such as Montgomery's

   (Trivial) Example ((p-1) method)
   Factorise 1309

   It is sufficient to find the second largest factor, so take $1309^{0.26} \sim 6$, with $A = 2^3 \times 3^2 \times 5$ say. Start off with

$$\begin{aligned} \text{A} &= 2^2 \times 3 \\ 2^{2^2 \times 3} &= 4096 \bmod 1309 = 169 \\ 2^A - 1 &= 168 \qquad \text{HCF}(168,\ 1309) = 7 \\ \textit{So the solution is already found } 1309 &= 7 \times (11 \times 17) \end{aligned}$$

   If we progress further we get

A	$2^2 \times 3$	$2^3 \times 3$	$2^3 \times 3^2$	$2 \times 3 \times 5$	2×5	3×5	etc.
$HCF(2^A - 1,\ 1309)$	7	17	17	11	11	7	etc.

3. **Pollard's Monte Carlo Method**
   The name (due to Pollard himself) arises because the method employs

random integers *mod* n - and looks for a common factor between them and n, not unlike the $(p-1)$ method.

To understand the method we need the results of the *Birthday Paradox* (see appendix). Effectively this states that if we pick random integers x_i *mod* n we shall find an $x_j = x_i$ with $j > i$ where $i = 1,\ 2,\ 3,\ .,\ .,\ .$ for the first time when $j \approx \sqrt{\frac{\pi n}{2}}$. (If $n = 365,\ j \approx 23.94$)

We consider $x_{i+1} = (x_i^2 - 1) mod\ n$ as the pseudo-random sequence, and n is to be factorised. Let x_j, where $j = t + c$, be the first x_j to equal a previous x_i, and let $i = t$.

$$\begin{aligned} \text{E(t+c)} &= \sqrt{\tfrac{\pi n}{2}} \quad by\ the\ Paradox \\ and\ \text{E(t)} &= \sqrt{\tfrac{\pi n}{8}} \quad (half - way\ to\ t + c) \end{aligned}$$

After x_t the series repeats with period c, $x_{t+j} = x_{t+c+j}$. Consider $(x_{2j} - x_j)$ and $Q_i = \prod_{j=1,i}(x_{2j} - x_j)$. Let $r = lowest\ j$ such that $x_{2j} - x_j = 0\ mod\ n$ then:

(a) If $t = 0$ then $r = c$

(b) If $t > 0$ then $\quad r \geq t$, $\ r = 0 \bmod c$, $\ r < t + c$

Exercise These constraints should be checked by the student.

$Q_i = 0$ for $i \geq r$.

Now consider $p|n$ and imagine $x'_i = x_i\ mod\ p$. Let $Q'_i = Q_i\ mod\ p$. We want to find p.

$$\begin{aligned} Q'_i &= 0 \bmod p\ for\ i \geq r' \text{ with } t' \leq r' \leq t' + c' \\ Q_i &= Q'_i + Kp \text{ for some } K \\ But\ Q'_i &= 0\ for\ i \geq r' \end{aligned}$$

Therefore look for $HCF(Q_i,\ n)$ and p should be found after $\sqrt{\frac{\pi p}{2}}$ steps.

Example $n = 1073,\ x_0 = 2,\ x_1 = 3,\ 8,\ 63,\ 749,\ 894,\ 923,\ 1039,\ 82$

$$\begin{aligned} Q_i &= (x_2 - x_1) \times (x_4 - x_2) \times (x_6 - x_3) \times (x_8 - x_4) \\ &= (5) \times (741) \times (860) \times (406) \\ Partial\ products &= \textstyle\prod Q_i = 5,\ 486,\ 563,\ 29 \\ \text{HCF}(29,\ 1073) &= 29 \end{aligned}$$

We can see how this works by considering the $x'_i\ mod\ 29$, $x'_i = 3,\ 8,\ 5,\ 24,\ 24$ i.e. first repetition on the fifth step. Note: $\sqrt{\frac{\pi 29}{2}} = 6.7$.

Clearly the Monte Carlo method is best for factorising an integer with a smallish prime factor.

Simple DL Attacks
The Discrete Logarithm (DL) problem is, given $n,\ y,\ \alpha$. Find x such that $y = \alpha^x \ mod\ n$. Typically n is a prime p, because little extra security is gained by having it otherwise.

4. **The Order of the Group Attack**

   If $\theta(\alpha) = rs$ the product of two integers $r,\ s$ we know $rs|(\phi(n))$ or if $n = p\ prime\ rs|(p-1)$. We can form $y_1 = y^r = (\alpha^r)^{x_1}\ mod\ p = \alpha_1^{x_1}\ mod\ p$ where $x_1 = x\ mod\ s,\ \alpha_1 = \alpha^r\ mod\ p$.

   If we can solve the reduced DL problem $y_1 = \alpha_1^{x_1}\ mod\ p$ for x_1 then we can try to solve also $y_2 = y^s = (\alpha^s)^{x_2}\ mod\ p = \alpha_2^{x_2}\ mod\ p$ with $x_2 = x\ mod\ r$ and $\alpha_2 = \alpha^s\ mod\ p$ for x_2.

   If we succeed we have $\begin{array}{lcl} \mathrm{x} & = & \mathrm{x}_1\ mod\ s \\ \mathrm{x} & = & \mathrm{x}_2\ mod\ r \end{array}$

   which can be solved for x by the Chinese Remainder Theorem (CRT).

   Example $7 = 2^x\ mod\ 113$ what is x?
   $\phi(113) = 112 = 7 \times 16$
   $Order\ (2) = 7 \times 4 = 28$ (Find by experiment)

   (a) $7^4 = (2^4)\ or\ 28 = 16^x\ mod\ 113$
   $16^2 = 30,\ 16^3 = 28,\ therefore\ x = 3\ mod\ 7$

   (b) $7^7 = (2^7)^x\ mod\ 113$ or $112 = 15^x\ mod\ 113$ therefore $x = 2\ mod\ 4$
   Using CRT $x = 10\ mod\ 28,\ 7 = 2^{10}\ mod\ 113$

   Therefore, a secure DL usage must not have small factors such as $r,\ s$ of the Order of $\alpha\ mod\ p$.

   This can often be arranged. Choose α wisely, as in the Diffie-Hellman technique.

   However in other cases this is not so, for example in RSA encryption where $x = d$ the secret key exponent; and $\alpha = c$ cypher-text. $y = m\ plaintext$. Again in RSA digital signatures $\alpha = H(m)$ hash of

message for signing y = *the signature*. In such cases we want the probability of $\theta(\alpha)$ to be large. If the modulus is p then this can be achieved by ensuring that $(p-1)$ has a large prime factor p', so that $(p-1) = p'r$.

By doing this we assure that there is a high probability than $\theta(\alpha)$ is divisible by p' because there exists a primitive β, say, *mod* p and an arbitrary element $m = \beta^k \ mod \ p$ for some k.

$Order\ (m) = k' = \frac{(p-1)}{HCF\ (k,\ (p-1))}$ See Chapter 3, Theorem 3.9

and the number of elements of *Order* $k' = \phi(k')$. If k is prime to $(p-1)$ we have $\phi(p-1) = \phi(p'r)$ elements of *order* $(p'r)$ and if $k = r$ we have $k' = p'$ so $\phi(p')$ elements of order (p'): In all $\phi(p'r) + \phi(p')$ elements whose order is divisible by p'.

The proportion over all $(p-1)$ elements is $\frac{(p'-1)(\phi(r)+1)}{p-1}$ and this is the probability that *Order* (m), m being an arbitrary integer, is divisible by p'. The larger p' the larger the probability.

Example $p = 23,\ (p-1) = 22 = 2 \times 11$
$\theta(\alpha)|22,\ p' = 11,\ r = 2$
giving probability $\theta(\alpha)$ divisible by n is $\frac{(p'-1)(\phi(r)+1)}{p-1} = \frac{10.2}{22} = \frac{10}{11}$
Specifically elements of order 22 are 5, 10, 20, 17, 11, 21, 19, 15, 7, 14.
Specifically elements of order 11 are 2, 4, 8, 16, 9, 18, 13, 3, 6, 12.
i.e. 20 out of 22.

5. **Shanks' Method**

This method of finding a digital logarithm is applicable after any simplification of the problem has been achieved by , for instance, an order of the group attack.

To solve $y = \alpha^x \ mod \ n$ for x with $y,\ \alpha,\ n$ known. Suppose $x = a\sqrt{n} + b,\ a, b = 0, \sqrt{n}$. Then $y = \alpha^{a\sqrt{n}}.\alpha^b \ mod \ n$ or $y(\alpha^{-\sqrt{n}})^a = \alpha^b \ mod \ n$. Tabulate $y,\ y\alpha^{-\sqrt{n}},\ y\alpha^{-2\sqrt{n}},\ \ldots\ y\alpha^{-(\sqrt{n}-1)(\sqrt{n})} \ mod \ n$ and put the results in ascending order in column A so that a match may be found rapidly.

Then for $b = 0$ *to* $\sqrt{n}$ evaluate α^b and seek a match in column A. When that occurs we have the solution

$$\begin{aligned} \text{y } \alpha^{-a\sqrt{n}} &= \alpha^b \ mod \ n \\ \text{x} &= \text{a}\sqrt{n} + b \end{aligned}$$

Example *The same problem as before*

$$\begin{aligned} 7 &= 2^x \ mod\ 113 \\ \text{Try} \quad x &= 11a+b \\ \text{So} \quad 7\times 2^{-11a} &= 2^b \end{aligned}$$

Now $\theta(2) = 28$ so $2^{11} \times 2^{17} = 1\ mod\ 113$ and $2^{-11} = 2^{17}$. But $2^{14} = -1$ therefore $2^{17} = -8 = 2^{-11}$.

a	0	1	2	3	4	5	6	7	8	9	10
7×2^{-11a}	7	57	109	32	83	14	1	105	64	53	28

Sorting this we get

7×2^{-11a}	1	7	14	28	32	53	57	83	105	109
a	6	0	5	10	3	9	1	4	7	2

$$\begin{aligned} 2^b\ mod\ 113 \text{ for } b &= 0,\ 1,\ 2,\ \text{to } 11 \\ gives\ 2^0 &= 1\ and\ we\ get\ a\ match\ immediately \\ \text{a=1} \quad \text{b=0}\ gives & \\ 7\times 2^{-11\times 6} &= 2^0\ mod\ 113 \\ so\ 7 &= 2^{66}\ mod\ 113 \\ &= 2^{10} \text{ since } 2^{56} = 1 \end{aligned}$$

$$\begin{aligned} We\ can\ also\ see\ that\ \text{b} &= 5\ gives \\ 2^b &= 32{=}2^{-11.3} \times 7 \\ or\ 2^5 &= 2^{-33} \times 7 \\ 2^{38} &= 7 \text{ mod } 113 \\ i.e.\ 2^{10} &= 7\ again \end{aligned}$$

Appendix X
Birthday Paradox

Let x_i be a series of random integers $0 \le x_i \le n$, $x_1,\ x_2,\ x_3,\ \ldots$ *etc.* Let x_{j+1} be the first of the series which has the same value as a preceding integer: $[x_{j+1} = x_i$ for some $i < j$ and all x_i for $i \le j$ are distinct].

$$\begin{aligned} \text{Let } P_j &= \text{Probability that first repeat occurs at (j+1)} \\ &= j_n \times [\tfrac{n!}{(n-j)!}] \times \tfrac{1}{n^j} \\ &= j_n \times [Prob\ (no\ repeats\ for\ i \le j)] \end{aligned}$$

Example $n = 365$, then $Prob(no\ repeats\ for\ i \le j) = \frac{1}{365^j} \times \frac{365!}{(365-j)!} < 0.5$ when $j \ge 23$

Let $S = \sum_{k=0}^{n} \frac{n^k}{k!}$. The expected value for j is

$$\begin{aligned}
\text{E(j)}=\textstyle\sum_{j=1,n} jP_j &= (\text{n!})/(\text{n}^{n+1}) \textstyle\sum_{1,n} j^2 \frac{n^{n-j}}{(n-j)!} = (n!)/(n^{n+1})) \textstyle\sum (n-k)^2 \frac{n^k}{k!} \\
&= (\text{n!})/(\text{n}^{n+1}))[n^2 S - 2n^2(S - \tfrac{n^n}{n!})) + \textstyle\sum \frac{k(k-1)n^k}{k!} + \sum k \frac{n^k}{k!}] \\
&= (\text{n!})/(\text{n}^{n+!})[2n^2 \tfrac{n^n}{n!} - n^2 S + n^2(S - \tfrac{n^{n-1}}{(n-1)!} - \tfrac{n^n}{n!}) + n(S - \tfrac{n^2}{n!})] \\
&= (\text{n!})/(\text{n}^{n+1})[nS - \tfrac{nn^n}{n!}] \\
&= (\text{n!})/(\text{n}^{n})[S - \tfrac{n^n}{n!}] \\
\text{We take } \textstyle\sum_{i=0,\ n-1} \frac{n^k}{k!} &= \text{S-}((\text{n}^n)/(n!)) \approx (\text{e}^n)/(2) \ \textit{for largish } n \\
\text{Therefore E(j)} &= ((\text{e}^n)/(2)) \times ((\text{n!})/(\text{n}^n)) \\
&= ((\text{e}^n)/(2)) \times ((\sqrt{2\pi n})/(e^n)) \text{ by Stirling's Formula} \\
\text{Therefore E(j)} &= \sqrt{\tfrac{\pi n}{2}}
\end{aligned}$$

Example $n = 365,\ E(j) = 23.9$