Some Considerations concerning the Selection of RSA Moduli

Klaus Huber
Deutsche Bundespost TELEKOM
Research Institute
Am Kavalleriesand 3
P.O. Box 10 00 03
6100 Darmstadt
Germany

Abstract

In this contribution two conditions are stated which safe RSA moduli $n = p \cdot q$ must fulfill. Otherwise the factors of n can be found. First we consider the cycle-lengths of the recursion $c \leftarrow c^{\varphi(n)-1}+1 \bmod n$ which leads to a condition in terms of Fibonacci numbers. The second condition involves a property of Euler's function. We introduce a number-theoretic distance measure – the power-of-two distance (ptd) – which may be useful for evaluating the security of RSA moduli against 'number-theoretic integration'. The ptd of an RSA prime p must not be too small.

1 Introduction

The factorization of large numbers is a very old mathematical problem, which has found increasing interest since the advent of the RSA public-key cryptosystem (see [1]), whose security essentially relies on the difficulty of factoring. Nowadays – for security reasons – every user gets a different RSA modulus $n = p \cdot q$. As result thereof, a great amount of "secure" RSA moduli have to be generated. Hence there will not be enough (CPU-) time and resources to try all known factorization algorithms for a sufficiently long time on each RSA modulus n. For this reason it is of interest to have criteria telling whether a particular modulus n may be insecure

D.W. Davies (Ed.): Advances in Cryptology - EUROCRYPT '91, LNCS 547, pp. 294-301, 1991.

(e.g. it is well known that $p-1$ and $q-1$ must contain at least one large prime factor). In this contribution we state two conditions which secure RSA moduli must fulfill.

2 The first Condition

We start by giving a 'small' example. The "RSA number"

$$n = 525169521992627614583344195951527749$$

can be factored easily by assigning c the initial value 1 and then using the iteration

$$c \leftarrow c^{-1} + 1 \bmod n \tag{1}$$

until the greatest common divisor of c and n is greater than 1. c^{-1} denotes the inverse of c modulo n. A program implementing (1) on a powerful computer rapidly (much less than a second) finds the factor $p = 8242065050061761$. Hence the second factor is $q = 63718196690123467909$.

Now neither $p-1$ nor $q-1$ do have only 'small' prime divisors:

$$\begin{aligned} p-1 &= 2^6 \cdot 5 \cdot 53 \cdot 107 \cdot 109 \cdot 41667737 \\ q-1 &= 2^2 \cdot 3 \cdot 13 \cdot 67 \cdot 6096268340042429 \ , \end{aligned}$$

and the primes dividing $p+1$ and $q+1$ are not all 'small' either:

$$\begin{aligned} p+1 &= 2 \cdot 3^3 \cdot 17 \cdot 577 \cdot 15560284867 \\ q+1 &= 2 \cdot 5 \cdot 11 \cdot 59 \cdot 431 \cdot 449 \cdot 9677 \cdot 5242693 \ . \end{aligned}$$

The reason why the program handles this number so well is that the index u_p of the smallest Fibonacci number F_{u_p} which contains p as a factor is quite small, we have $u_p = 107$ (for Fibonacci numbers see the appendix), and, clearly, for initial value 1 the recursion (1) computes the convergents F_{j+1}/F_j. Before we study the cycle-lengths of (1) for arbitrary initial value a, we state the first condition:

Condition 1 *A prime p selected as factor of an RSA modulus n must have a large index u_p, where F_{u_p} is the smallest Fibonacci number which contains p as a divisor.*

To find the cycle-lengths of (1) we use continued fractions and set

$$\begin{aligned} a &= 1 + \cfrac{1}{1 + \ldots \frac{1}{1+\frac{1}{a}}} \\ &= [1, 1, 1, \ldots, 1, a] = [b_0, b_1, b_2, \ldots, b_l, a] \ , \end{aligned}$$

where computations are done modulo n. Note that if we set $c^{-1} \equiv c^{\varphi(n)-1} \bmod n$ the recursion (1) does make sense even if $\gcd(n, c) > 1$.

Let us first consider the recursion modulo a prime p. We set c equal to an initial value a and start the recursion

$$c \leftarrow c^{p-2} + 1 \bmod p \,. \tag{2}$$

The question we ask is how many calls of (2) do we need until we come back to the initial value a.

The answer to this question can be obtained by considering the jth convergents $f_j = A_j/B_j$ of the continued fraction $[1, 1, \ldots, a]$. We get

$$\begin{aligned} A_j &= A_{j-1} + A_{j-2} \quad , \quad 1 \le j \le l; \quad \text{where} \quad A_{-1} = A_0 = 1 \\ B_j &= B_{j-1} + B_{j-2} \quad , \quad 1 \le j \le l; \quad \text{where} \quad B_{-1} = 0, B_0 = 1 \\ &\Rightarrow \quad A_j = F_{j+2}, \quad B_j = F_{j+1} \,. \\ \text{Hence} \quad a &= f_{l+1} = \frac{aF_{l+2} + F_{l+1}}{aF_{l+1} + F_l} \,. \end{aligned}$$

If $a \neq -F_j/F_{j+1}$ we get $(a^2 - a - 1)F_{l+1} \equiv 0 \bmod p$ and the cycle-length is u_p or 1, otherwise the cycle-length equals $u_p - 1$. To summarize, the possible cycle-lengths of (2) are

$$\begin{array}{ll} 1 & \text{for } a = \frac{1 \pm \sqrt{5}}{2} \\ u_p - 1 & \text{for } a = -\frac{F_j}{F_{j+1}} \\ u_p & \text{else} \end{array} \tag{3}$$

The cycle-length 1 occurs if 5 is a quadratic residue of p, i.e. if the last digit of p is 1 or 9. From the properties of the Fibonacci numbers we find that there is only one cycle of length $u_p - 1$. The remaining $\frac{p-(5/p)-u_p}{u_p}$ cycles all have length u_p.

Example: $p = 101$, $u_{101} = 50$

a	cycle-length
1	49
4	50
23	1
79	1

Having this result for primes we infer that for RSA-moduli n the possible cycle-lengths of (1) are given by the least common multiples of the cycle-lengths of its prime-factors. Let us consider two examples:

First: $n = 77 = 7 \cdot 11$, $u_7 = 8$, $u_{11} = 10$

a	cycle-length
1	$63 = \text{lcm}(7, 9) = \text{lcm}(u_7 - 1, u_{11} - 1)$
4	$7 = \text{lcm}(7, 1) = \text{lcm}(u_7 - 1, 1)$
8	$7 = \text{lcm}(7, 1) = \text{lcm}(u_7 - 1, 1)$

Second: $n = 611 = 13 \cdot 47$, $u_{13} = 7$, $u_{47} = 16$

a	cycle-length
1	$30 = \text{lcm}(6, 15) = \text{lcm}(u_{13} - 1, u_{47} - 1)$
3	$105 = \text{lcm}(7, 15) = \text{lcm}(u_{13}, u_{47} - 1)$
4	$112 = \text{lcm}(7, 16) = \text{lcm}(u_{13}, u_{47})$
6	$48 = \text{lcm}(6, 16) = \text{lcm}(u_{13} - 1, u_{47})$
$\vdots$	$\vdots$

We now give a simple combined test for primality of p and large u_p. We use the well-known matrix description of Fibonacci numbers which follows immediately from the definition in eqn. (12):

$$\begin{pmatrix} F_{j+1} & F_j \\ F_j & F_{j-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^j . \tag{4}$$

By 'square and multiply' any $F_j \bmod n$ can be computed with $O(\log n)$ operations. For a prime p we get (see appendix)

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{p-(\frac{5}{p})} \equiv \begin{pmatrix} (\frac{5}{p}) & 0 \\ 0 & (\frac{5}{p}) \end{pmatrix} \bmod p . \tag{5}$$

Equation (5) can be used as primality test (for a composite number $(\frac{5}{n})$ denotes Jacobi's symbol). There are 7 odd composite numbers smaller than 50000 which fulfill equation (5), namely $\{4181, 5777, 6721, 10877, 13201, 15251, 34561\}$. Even if a composite odd number fulfills eqn. (5) its compositeness can sometimes be established within the test, e.g. for $n = 6721$ we have $n - (5/n) = 2^6 \cdot 105$ and

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{2^4 \cdot 105} \equiv \begin{pmatrix} 6579 & 0 \\ 0 & 6579 \end{pmatrix} \bmod n$$

$$\Rightarrow 6579^2 \equiv (-1)^{2^4 \cdot 105} \equiv 1 \bmod n \Rightarrow \gcd(n, 6579 \pm 1) = \begin{cases} 47 \\ 143 \end{cases} .$$

As $F_p \equiv 5^{\frac{p-1}{2}} \bmod p$ we can extend eqn. (5) to give the following *primality test*:

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{n-(\frac{5}{n})} \stackrel{?}{\equiv} \begin{pmatrix} (\frac{5}{n}) & 0 \\ 0 & (\frac{5}{n}) \end{pmatrix} \stackrel{?}{\equiv} \begin{pmatrix} 5^{\frac{n-1}{2}} & 0 \\ 0 & 5^{\frac{n-1}{2}} \end{pmatrix} \bmod n \ . \tag{6}$$

None of the seven odd composite numbers given above passes this test.

Eventually the number u_p can be found among the divisors of $p-(5/p)$. This follows from eqn. (5), for further details see e.g. ([4] or [5]).

To increase confidence in the primality of a number, we may still run e.g. Miller's test for one or two bases. Essentially, however, the combined test for primality and large u_p given above does not increase the cost of selecting an RSA-prime, if the primes are generated from the bottom up (as in [1] p.124), i.e. if the factors of $p-(5/p)$ are known.

To summarize. for an RSA number $n = p \cdot q$ we must demand that u_p and u_q are both large. Otherwise either the recursion (1) with initial value $a = 1$ or – faster –

$$\gcd(n, F_i \bmod n) \qquad i = 1, 2, 3, \ldots \tag{7}$$

will factor n with $\min\{u_p, u_q\}$ steps. (The only case that $\gcd(n, F_{u_p})$ does not factor n is if $u_p = u_n$. Note that (7) may also be useful to extract small factors from large composite numbers.)

To get very long cycles of (1), we may demand – according to the cycle-lengths which do occur – that $\gcd(u_p, u_q)$, $\gcd(u_p - 1, u_q)$, $\gcd(u_p, u_q - 1)$, $\gcd(u_p - 1, u_q - 1)$ are small. $\gcd(u_p, u_q)$ is small if $\gcd(p-(5/p), q-(5/q))$ is small, this can be checked even if the factors of $p-(5/p)$ and $q-(5/q)$ are not known. To be sure against other ideas exploiting Fibonacci numbers (e.g. the Iteration Theorem in [5]), it is reasonable to demand that u_p and u_q contain a large prime. Finally note that safe primes of the form $p = 2p' + 1$, where p' is also prime, do ensure a very large u_p if $(5/p) = 1$, but not if $(5/p) = -1$.

3 The second Condition

In this section a number-theoretic distance function – the power-of-two distance (ptd) – is introduced. The ptd of RSA moduli should not be too small in order to give security against cryptanalytic attacks.

Euler's function $\varphi(n)$, given in equation (8), plays a central role in the RSA cryptosystem and for factoring as well.

$$\varphi(n) \ = \ | \{i \mid gcd(i, n) = 1, 1 \leq i \leq n\} | \tag{8}$$

$$\varphi(n) \;=\; n \cdot \prod_{p|n} \frac{p-1}{p} \qquad \text{where } p \text{ means prime.} \tag{9}$$

Now let us define the iterated function $\varphi^{\{j\}}(n)$:

$$\varphi^{\{j\}}(n) = \varphi(\varphi^{\{j-1\}}(n)) \qquad j = 1, 2, \ldots \tag{10}$$

where $\varphi^{\{0\}}(n) = n$. By applying Euler's function φ repeatedly, we will in most cases arrive quite rapidly at a power of two. This follows immediately from eqn. (9), as in each step every prime factor is reduced by one and thus contributes at least one factor of two. For illustration consider the following example:

$$\begin{aligned}
n &= 98765432109876543210 \\
n = \varphi^{\{0\}}(n) &= 2 \cdot 3^2 \cdot 5 \cdot 17^2 \cdot 101 \cdot 3541 \cdot 27961 \cdot 379721 \\
\varphi^{\{1\}}(n) &= 2^{17} \cdot 3^3 \cdot 5^5 \cdot 11 \cdot 17 \cdot 59 \cdot 233 \cdot 863 \\
\varphi^{\{2\}}(n) &= 2^{29} \cdot 3^2 \cdot 5^5 \cdot 29^2 \cdot 431 \\
\varphi^{\{3\}}(n) &= 2^{34} \cdot 3 \cdot 5^5 \cdot 7 \cdot 29 \cdot 43 \\
\varphi^{\{4\}}(n) &= 2^{40} \cdot 3^2 \cdot 5^4 \cdot 7^2 \\
\varphi^{\{5\}}(n) &= 2^{43} \cdot 3^2 \cdot 5^3 \cdot 7 \\
\varphi^{\{6\}}(n) &= 2^{46} \cdot 3^2 \cdot 5^2 \\
\varphi^{\{7\}}(n) &= 2^{48} \cdot 3 \cdot 5 \\
\varphi^{\{8\}}(n) &= 2^{50} = 1125899906842624
\end{aligned}$$

In a way Euler's function behaves like a 'number-theoretic derivative' – it makes big primes small. Therefore the function $\varphi^{\{j\}}(n)$ is refered to as $j-$th (number-theoretic) derivative. For most numbers repeated application of Euler's function leads quite rapidly to a power of two. This simple observation leads to the following definition of the power-of-two distance:

Definition 1 *The power-of-two-distance (ptd) of a number $n > 1$ is defined by*

$$ptd(n) := \min\{j \mid \varphi^{\{j\}}(n) = 2^i, i = 1, 2, \ldots\} \; .$$

For example $\mathrm{ptd}(2^{16}) = 0$, $\mathrm{ptd}(2^{16}+1) = 1$, $\mathrm{ptd}(3977) = 2$. From the fact that almost all numbers around n have about $\ln\ln n$ prime divisors (see [3]), we obtain the following crude approximation for the average value of ptd for a randomly chosen number n:

$$E\{\mathrm{ptd}(n)\} \approx \frac{\log_2 n}{\ln\ln n} \; .$$

Our second condition is now:

Condition 2 *A prime p selected as factor of an RSA modulus n must have a sufficiently big value of ptd(p).*

Clearly, we have

$$\max\{\text{ptd}(a), \text{ptd}(b)\} \leq \text{ptd}(a \cdot b) \leq \text{ptd}(a) + \text{ptd}(b) \ .$$

The left half of the above inequality may be useful to bound the ptd if a number can be factored only partially.

In analogy to 'number-theoretic differentiation' we can refer to 'number-theoretic integration' of a number r as finding a number which belongs to the set Ψ_r, which is defined by

$$\Psi_r := \{i \mid \varphi(i) = r\} \ . \tag{11}$$

We refer to finding the whole set Ψ_r as complete integration of r. For example integrating $r = 100$ completely gives $\Psi_{100} = \{101, 125, 202, 250\}$.

If the factors of r are known, integration of r is an easy task. It can be done in a systematic way from the representations $r = \prod \varphi(p_i) \cdot p_i^{f_{p_i}}$. For integration note, that all odd numbers ≥ 3 are non-integrable, and if an odd number t belongs to Ψ_r then $2t$ is also in Ψ_r. Also the density of non-integrable even numbers increases the larger the numbers get. If the ptd(p) is too 'small', the factors of $n = p \cdot q$ can be found by repeated integration, starting from a power of 2 close to $\sqrt{n}$.

Since the set $\Psi_{\varphi(q)}$ of a safe prime q contains only one even number – thus reducing its effective ptd – we should measure the ptd of primes which lead to long chains of safe primes by the ptd of the smallest prime in the chain (e.g. 2879, 1439, 719, 359, 179, 89 is a chain of length 5, hence it is ptd(89) = 3 which measures the security of $p = 2879$).

4 Appendix: Fibonacci numbers

To make this paper self-contained we recall the most important properties of Fibonacci numbers (see e.g. [2] pp.78-86, or [3] p.150). Fibonacci numbers are defined by the recursion

$$F_{j+2} = F_{j+1} + F_j \qquad\qquad j = 2, 3, \ldots \tag{12}$$

with initial values $F_0 = 0$ and $F_1 = 1$. From the roots of the characteristic equation $x^2 - x - 1 = 0$ we get the n-th Fibonacci number as $F_n = ((\frac{1+\sqrt{5}}{2})^n - (\frac{1-\sqrt{5}}{2})^n)/\sqrt{5}$.

This leads to $2^{n-1} F_n = n + \binom{n}{3} 5 + \binom{n}{5} 5^2 + \ldots + \begin{cases} 5^{\frac{n-1}{2}} & n \text{ odd} \\ n 5^{\frac{n}{2}-1} & n \text{ even} \end{cases}$. (13)

Thus for a prime p from the above equation and Euler's criterion we get

$$F_p \equiv 5^{\frac{p-1}{2}} \equiv (5/p) \bmod p \ , \tag{14}$$

where $(5/p)$ denotes Legendre's symbol. Using Gauß's law of quadratic reciprocity we find

$$(\frac{5}{p}) = (\frac{p}{5}) = \begin{cases} 1 & \text{for } p \equiv 1 \text{ or } 4 \bmod 5 \\ -1 & \text{for } p \equiv 2 \text{ or } 3 \bmod 5 \end{cases}$$

It is easily seen (e.g. from the determinant of the matrices of eqn. (4)) that $F_{n+1}F_{n-1} - F_n^2 = (-1)^n$ holds. Hence $F_{p+1}F_{p-1} \equiv 0 \bmod p$. It can be shown that either F_{p-1} or F_{p+1} contains p as divisor. More precisely, using (13) we find

$$\begin{aligned} F_{p-1} &\equiv 0 \bmod p \quad \text{for } p \equiv 1, 4 \bmod 5 \\ F_{p+1} &\equiv 0 \bmod p \quad \text{for } p \equiv 2, 3 \bmod 5 \ . \end{aligned}$$

By induction one can show that $F_{n+m} = F_m F_{n+1} + F_{m-1}F_n$, and an important divisibility property of Fibonacci numbers follows, namely F_a divides $F_{a \cdot b}$ where a, b are integers. Thus $F_{u_p} \equiv 0 \bmod p \Rightarrow F_{k \cdot u_p + 2} \equiv F_{k \cdot u_p + 1} \bmod p$.

References

[1] R.L.Rivest, A.Shamir, L.Adleman: "A Method for Obtaining Digital Signatures and Public Key Cryptosystems", Communications of the ACM, Vol.21 Nr.2, pp.120-126, Feb.1978.

[2] D.E.Knuth: "The Art of Computer Programming" Vol.1, Fundamental Algorithms, Reading, MA:Addison-Wesley, 1968.

[3] G.H.Hardy, E.M.Wright: "An Introduction to the Theory of Numbers", Oxford University Press: Oxford, Fifth edition 1979.

[4] K.Barner: "Zur Fibonacci-Folge modulo p", Monatshefte für Mathematik. Bd.69/2, pp.97-104, 1965.

[5] J.D.Fulton, W.L.Morris: "On Arithmetical functions related to the Fibonacci numbers", Acta Arithmetica, XVI, pp.105-110, 1969.