

Cryptanalysis of RSA with smooth prime sum

Meryem Cherkaoui Semmouni *

Information Communication and Embedded Systems

Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes

Mohammed V University in Rabat

Rabat

Morocco

Abderrahmane Nitaj [†]

UNICAEN, CNRS, LMNO

Normandie University

14000 Caen

France

Mostafa Belkasmi [§]

Information Communication and Embedded Systems

Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes

Mohammed V University in Rabat

Rabat

Morocco

Abstract

Let $N = pq$ be an RSA modulus with balanced prime factors, that is $q < p < 2q$. There exist infinitely many integers x , y and z such that $ex - \phi(N)y = (p + q - 1)z$. We show that if the prime sum $p + q - 1$ has only small prime factors and e satisfies an equation of the form $ex - \phi(N)y = (p + q - 1)z$ with suitably small integers x , y and z , then one can factor the RSA modulus in polynomial time. In addition we show that the number of such RSA moduli, as well as the number of exponents that are vulnerable to our attack is non-negligible.

Subject Classification: 94A60.

Keywords: RSA, Factorization, Continued fractions, Elliptic curve factorization method (ECM).

* E-mail: cher.meryem@gmail.com (Corresponding Author)

[†] E-mail: abderrahmane.nitaj@unicaen.fr

[§] E-mail: mostafa.belkasmi@um5.ac.ma

1. Introduction

Invented in 1978 by Rivest, Shamir and Adleman [21], RSA has become the most widely known and widely used public-key cryptosystem. The public parameters in the RSA cryptosystem are the modulus N and the public exponent e . The RSA modulus $N = pq$ is the product of two large primes p and q . The prime numbers p and q are balanced if they are of the same bit size, that is $q < p < 2q$. Let e be an integer satisfying $\gcd(e, \phi(N)) = 1$ where $\phi(N) = (p-1)(q-1) = N+1-(p+q)$ is the Euler totient function. The private exponent in RSA is an integer d satisfying $ed \equiv 1 \pmod{\phi(N)}$. The public key (N, e) is used for the encryption of a message and the private exponent d is used for the decryption.

The security of RSA is based on the difficulty of factoring its modulus. Despite the hardness of the integer factorization problem, it is well known that RSA is vulnerable when the private exponent is too small. Wiener's continued fraction attack [26] can be used to factor N when the private exponent d is smaller than $\frac{1}{3}N^{\frac{1}{4}}$. Wiener's attack is based on the convergents of the continued fraction expansion of $\frac{e}{N}$. Boneh and Durfee's lattice-based attack [4] improved the bound on d up to $N^{0.292}$. Boneh and Durfee's attack is based on Coppersmith's method [7] for finding small roots of modular polynomial equations. Using specific properties of the prime factors, several attacks have been presented that improve the former bounds. In 2002, de Weger presented an attack [25] on RSA when the prime difference $\Delta = |p - q|$ is small, that is when p and q share an amount of their most significant bits. In 2007, Zhao and Qi presented an attack [27] on RSA when p and q share an amount of their least significant bits, that is when the prime difference $\Delta = |p - q|$ is such that $\Delta = 2^m \Delta'$ for some known integer m . The bounds were later improved by Sun et al. [23]. In 2008, Sarkar et al. [22] presented an attack on RSA and improved the bound on d when an amount of the most significant bits of p are known. Also, there are various attacks on RSA when a fraction of the most significant or least significant bits of d are known (see [5,1,9]). More attacks on RSA can be found in [3,11,18].

Further attacks on RSA have been presented when the exponent e satisfies variant equations of the form $ex - \phi(N)y = z$ with small unknown parameters x , y and z . In [2], Blömer and May showed how to find the prime factors p and q when p and q are of the same bitsize with a large prime difference $|p - q|$, under the conditions $1 \leq x \leq \frac{1}{3}N^{\frac{1}{4}}$ and $|z| < N^{-\frac{3}{4}}ex$. Their method consists in computing x and y by the

continued fraction expansion of $\frac{e}{N}$ which gives the approximation $N - \frac{ex}{y}$ of $p+q$ up to an error term less than $2N^{\frac{1}{4}}$. Then, by applying a result of Coppersmith [7], this leads to the exact value of $p+q$ from which p and q can be retrieved. Some variants of the RSA cryptosystem are based on elliptic curves such as the KMOV [12] and the Demytko [8] systems. In KMOV, the key equation is in the form $ed - k(p+1)(q+1)$ and was studied for vulnerabilities in [19], while the key equation in the Demytko system is in the form $ed - k(p-s)(q-r) = 1$, and was also studied for vulnerabilities in [20]. A recent survey by Mumtaz and Ping [16] gives more details on various attacks on RSA.

In this paper, we present a new attack on RSA. Since $\gcd(e, \phi(N)) = 1$, then the equation $ex - \phi(N)y = (p+q-1)z$ has infinitely many integer solutions. We use this relation to mount an attack on RSA when the following two conditions are fulfilled.

1. The first condition : The prime sum $p+q-1$ is P_0 -smooth, that is all prime factors of $p+q-1$ are less than P_0 , where P_0 is a bound up to which Lenstra's elliptic curve factorization method (ECM) [13] is efficient to find all prime factors. The cost of factoring a P_0 -smooth integer n is subexponential in P_0 and polynomial in $\log n$. Recent results shows that ECM can efficiently find all prime factors up to $P_0 = 2^{100}$ of any integer [28]. Moreover, the property that the sum $p+q-1$ is P_0 -smooth is not too rare. For a positive integer n , we show that the number of prime numbers p and q with $2^{n-1} < q < p < 2^n$

such that $p+q-1$ is P_0 -smooth is approximately $2^{n-1} \left(\frac{\log P_0}{n \log 2} \right)^{\frac{n \log 2}{\log P_0}}$.

With $n = 1024$ and $P_0 = 2^{100}$, this number is approximately 2^{988} .

2. The second condition: The solution (x, y, z) of the equation $ex - \phi(N)y = (p+q-1)z$ satisfies $|y - z|x < \frac{\sqrt{2}}{6}\sqrt{N}$. Similarly, such solutions are not rare. Indeed, we show that, in the interval $[1, N]$, the number of exponents e satisfying $ex - \phi(N)y = (p+q-1)z$ with $|y - z|x < \frac{\sqrt{2}}{6}\sqrt{N}$, and hence are weak to our attack is at least $\mathcal{O}(\sqrt{N})$.

Our attack is based on applying the continued fraction algorithm to the rational number $\frac{e}{N}$ to find $\frac{y}{x}$ amongst its convergents. By using x and y , we compute $(p+q-1)(z-y) = ex - (N+1)y$, and, by applying ECM, we can find the P_0 -smooth divisors of $ex - (N+1)y$. Then $p+q-1$ is one of these divisors, which leads to retrieving p and q . We notice that the average number of divisors of $ex - (N+1)y$ is approximately $\log |ex - (N+1)y|$, and, that the run time of our attack is equivalent to the run time of ECM to find the prime factors of $ex - (N+1)y$ that are less than P_0 .

The paper is organized as follows. Section 2 summarizes the preliminaries on the continued fraction algorithm, the elliptic curve factorization method and two lemmas that are used in the paper. Section 3 provides a study of the RSA moduli in an interval with a smooth prime sum. In Section 4, we present the new attack. The number of the keys that are vulnerable to our attack is studied in Section 5. In Section 6, we provide a comparison of our attack with former attacks on RSA. Finally, Section 7 draws the conclusion.

2. Preliminaries

In this paper, the notation $F(x) \sim g(x)$ means that $\lim_{x \rightarrow \infty} \frac{F(x)}{g(x)}$ is some finite constant. In this section, we present a few facts that will be used in the rest of the paper.

2.1 The continued fraction algorithm :

In this section, we briefly describe the continued fraction algorithm. Let $\xi \neq 0$ be a real number. Set $\xi_0 = \xi$, and $t_0 = \lfloor \xi_0 \rfloor$. For $n \geq 1$, if $\xi_{n-1} \neq t_{n-1}$, we set

$$\xi_n = \frac{1}{\xi_{n-1} - t_{n-1}}, t_n = \lfloor \xi_n \rfloor.$$

For $n \geq 1$, since $0 < \xi_{n-1} - t_{n-1} < 1$, then $\xi_n > 1$ and $t_n \geq 1$. This process is called the continued fraction algorithm and yields an expression of the form:

$$\xi = t_0 + \frac{1}{t_1 + \frac{1}{t_2 + \frac{1}{\dots}}}$$

which is called the continued fraction expansion of ξ . For $i \geq 0$, the irreducible fractions $\frac{p_i}{q_i} = [t_0, t_1, \dots, t_i]$ are called the convergents of the continued fractions expansion of ξ . If ξ is a rational number $\frac{a}{b}$, then its continued fraction expansion is finite and the algorithm terminates with complexity $\mathcal{O}((\log b)^2)$.

The main result from the theory of continued fractions that we use in this paper is the following (see [10], Theorem 184).

Theorem 1 : *Let ξ be a positive number. Suppose that $\gcd(r, s) = 1$ and*

$$\left| \xi - \frac{r}{s} \right| < \frac{1}{2s^2}.$$

Then $\frac{r}{s}$ is a convergent of the continued fraction expansion of ξ .

2.2 The elliptic curve factorization method :

The elliptic curve factorization method (ECM) invented by H. W. Lenstra [13] in 1985 is one of the asymptotically fastest known methods for finding relatively large prime factors of an integer.

Let N be a composite integer with a factor p . ECM is a sub-exponential time algorithm with expected run time of $\mathcal{O}(e^{\sqrt{(2+o(1)) \log p \log \log p}} M(N))$ where p is the prime factor to find, and $M(N)$ is an upper bound for the cost of a multiplication modulo N . To find a prime factor of N , the ECM algorithm starts by choosing a random elliptic curve with an equation $y^2 = x^3 + ax + b$ over the ring $\mathbb{Z} / N\mathbb{Z}$ and a point P on the curve. Then ECM applies a first algorithm, called phase 1 to find a prime factor of N . If phase 1 is not successful, then ECM applies a second algorithm, called phase 2.

In Phase 1, one computes kP for

$$k = \prod_{\substack{q=2 \\ q \text{ prime}}}^{B_1-1} q^{e_q}, \quad e_q = \left\lfloor \frac{\log B_1}{\log q} \right\rfloor,$$

where B_1 is the upper bound for the prime numbers under which Phase 1 is expected to succeed. If the computation of kP does not fail, then either try a new curve, or increase k . If this does fail, then there is a denominator d that satisfies $\gcd(d, N) \neq 1$. If $\gcd(d, N) < N$, this gives a factor of N , else, one should try another a new curve or proceed with Phase 2.

In Phase 2, one computes qP on the curve for the prime numbers q satisfying $B_1 < q < B_2$, where B_2 is the upper bound for the prime numbers

under which Phase 2 is expected to succeed. If the computation of qP does not fail, then try another q . If this does fail, then there is a denominator d that satisfies $\gcd(d, N) \neq 1$. If $\gcd(d, N) < N$, this gives a factor of N .

A complete description of Phase 1 and Phase 2 with detailed algorithms is presented in [29,15]. In [28], large prime factors, up to $10^{83} \approx 2^{275}$, of specific large integers are listed. We make the following assumption which will be essential to the heuristic part in our computations.

Assumption 1 : *The elliptic curve factorization method (ECM) finds all prime factors of an integer N with $P(N) \leq P_0$ with $P_0 = 2^{100}$ where $P(N)$ is the largest prime factor of N .*

2.3 Miscellaneous results

Let $N = pq$ be an RSA modulus with balanced prime factors p and q , typically, $q < p < 2q$. The following result gives bounds for p , q , and $p+q$ in terms of N (See [17]).

Lemma 1 : *Let $N = pq$ be the product of two unknown integers such that $q < p < 2q$. Then*

$$\frac{\sqrt{2}}{2}\sqrt{N} < q < \sqrt{N} < p < \sqrt{2}\sqrt{N}, \quad 2\sqrt{N} < p+q < \frac{3\sqrt{2}\sqrt{N}}{2}.$$

The following result concerns the solutions of the equation $ex - \phi(N)y = (p+q-1)z$. Let $N = pq$ be an RSA modulus.

Lemma 2 : *Let e be a public exponent such that $\gcd(e, \phi(N)) = 1$. Then there exist infinitely many integers x , y and z such that*

$$ex - \phi(N)y = (p+q-1)z.$$

Proof: Let z be an integer. Since $\gcd(e, \phi(N)) = 1$, then there exists an integer x such that $x \equiv \frac{(p+q-1)z}{e} \pmod{\phi(N)}$, that is $ex \equiv (p+q-1)z \pmod{\phi(N)}$.

Hence, there exists an integer y such that $ex - \phi(N)y = (p+q-1)z$. Since there are infinitely many choices for z , then there are infinitely many solutions for the equation.

3. RSA Moduli

In this section we study the existence and the number of the RSA moduli $N = pq$ in an interval with the property that the prime sum $p + q$ is composed by prime factors less than the ECM bound P_0 .

3.1 Decomposition of an even integer as the sum of two primes

In number theory, Goldbach's conjecture is a well-known unsolved problem. It states that every even integer $S_0 \geq 4$ is the sum of two primes p and q , that is $S_0 = p + q$. The conjecture was formulated by Goldbach in 1742, and no counter-example has been found to date. The following result gives an explicit approximation of the number of ways to represent an even integer $S_0 = S + 1$ as the sum of two prime numbers in an interval.

Lemma 3 : *Let n be a positive integer. Let S be an odd integer such that $2^n < S < 2^{n+1}$. Then the set*

$$E(S) = \{(p, q) \mid 2^{n-1} < q < p < 2^n, p + q - 1 = S, p, q \text{ primes}\}.$$

has cardinality

$$\#E(S) \sim \frac{2^{n-1}}{(n \log 2)^2}.$$

Proof: Let x be a positive integer. The Prime Number Theorem asserts that the number $\pi(x)$ of primes $p \leq x$ satisfies (Theorem 6 of [10])

$$\pi(x) \sim \frac{x}{\log(x)}.$$

This implies that an integer x selected at random is prime with probability

$$\frac{\pi(x) - \pi(x-1)}{x - (x-1)} \sim \frac{x}{\log x} - \frac{x-1}{\log(x-1)} \sim \frac{x}{\log x} - \frac{x-1}{\log x} = \frac{1}{\log x}.$$

Observe that if $S = p + q - 1$ with $q \leq p$, then $2q \leq S + 1$ and $q \leq \frac{S+1}{2}$. For an even integer $S + 1$ such that $2^n < S + 1 < 2^{n+1}$, the number of representations of $S + 1$ in the form $S + 1 = p + q$ with two prime integers satisfying $2^{n-1} < q < p < 2^n$ is then

$$\#E(S) \sim \sum_{q=2^{n-1}}^{\frac{S+1}{2}} \frac{1}{\log q} \cdot \frac{1}{\log(S+1-q)}.$$

Since $\log q \approx \log S$ and $\log(S+1-q) = \log p \approx \log S$, then

$$\#E(S) \sim \sum_{q=2^{n-1}}^{\frac{S+1}{2}} \frac{1}{(\log S)^2} \sim \frac{S-2^n}{2(\log S)^2} \sim \frac{2^{n+1}-2^n}{2(\log S)^2} \sim \frac{2^{n-1}}{(n \log 2)^2},$$

which gives the result.

3.2 The number of RSA moduli in an interval

The following result gives a very conservative estimation of the number of RSA moduli of the form $N = pq$ with $2^{n-1} < q < p < 2^n$.

Lemma 4 : *Let n be an integer. The set of the RSA moduli*

$$\mathcal{R}(n) = \{N = pq \mid 2^{n-1} < q < p < 2^n, p, q \text{ primes}\},$$

has cardinality

$$\#\mathcal{R}(n) \sim \frac{2^{2n-3}}{(n \log 2)^2}.$$

Proof : We have

$$\#\mathcal{R}(n) = \sum_{\substack{p=2^{n-1} \\ \text{prime}}}^{2^n} \left(\sum_{\substack{q=2^{n-1} \\ q \text{ prime}}}^{p-1} 1 \right).$$

Let $p_0, p_1, \dots, p_k$ be the list of the prime numbers satisfying

$$2^{n-1} < p_0 < p_1 < \dots < p_k < 2^n.$$

Then

$$\#\mathcal{R}(n) = \sum_{\substack{p=p_0 \\ p \text{ prime}}}^{p_k} \left(\sum_{\substack{q=2^{n-1} \\ q \text{ prime}}}^{p-1} 1 \right) = \sum_{i=1}^{k-1} i = \frac{k(k-1)}{2} \sim \frac{1}{2}k^2.$$

Using the Prime Number Theorem, we get

$$k = \pi(2^n) - \pi(2^{n-1}) \sim \frac{2^n}{n \log 2} - \frac{2^{n-1}}{(n-1) \log 2} \sim \frac{(n-2)2^{n-1}}{n(n-1) \log 2} \sim \frac{2^{n-1}}{n \log 2}.$$

Then

$$\# \mathcal{R}(n) \sim \frac{1}{2} \left(\frac{2^{n-1}}{n \log 2} \right)^2 = \frac{2^{2n-3}}{(n \log 2)^2},$$

which gives the result.

Since every sum $S+1 = p+q$ of two prime numbers p and q satisfying $2^{n-1} < q < p < 2^n$ corresponds to an RSA modulus $N = pq$, then a direct consequence of Lemma (4) is the following result.

Lemma 5 : *Let n be an integer. The set*

$$\mathcal{S}(n) = \{S = p+q-1 \mid 2^{n-1} < q < p < 2^n, p, q \text{ primes}\},$$

has cardinality

$$\# \mathcal{S}(n) \sim 2^{n-2}.$$

Proof : Let S be an odd integer with $2^n < S+1 < 2^{n+1}$. Then, by Lemma (3), the number of representations of $S+1$ as a sum $S+1 = p+q$ with $2^{n-1} < q < p < 2^n$ is $\#E(S) \sim \frac{2^{n-1}}{(n \log 2)^2}$. Then, combining with Lemma (4), we get that the cardinality of the set

$$\mathcal{S}(n) = \{S = p+q-1 \mid 2^{n-1} < q < p < 2^n, p, q \text{ primes}\},$$

is

$$\# \mathcal{S}(n) = \frac{\# \mathcal{R}(n)}{\# E(S)} \sim \frac{\frac{2^{2n-3}}{(n \log 2)^2}}{\frac{2^{n-1}}{(n \log 2)^2}} = 2^{n-2}.$$

This terminates the proof.

We notice that $\# \mathcal{S}(n)$ is smaller than $\# \mathcal{R}(n)$. This is accurate since for a fixed odd integer S , the equation $S = p+q-1$ has many solutions (p, q) with $2^{n-1} < q < p < 2^n$ in accordance with Lemma (3), while, for a fixed integer N , the equation $N = pq$ has only one solution in the same interval.

3.3 Divisors and smooth numbers

Let n be a positive integer. The largest prime factor of n is denoted $P(n)$, the number of distinct prime divisors of n is denoted $\omega(n)$, and the number of divisors of n is denoted $d(n)$. The average order of $\omega(n)$ and $d(n)$ are given by the following result (see Theorem 430 and Theorem 319 of [10]).

Theorem 2 : *Let n be a positive integer. The average order of $\omega(n)$ and $d(n)$ are*

$$\omega(n) \sim \log \log n, \quad d(n) \sim \log n.$$

Definition 1 : Let y be a positive integer. An integer n such that $P(n) \leq y$ is called y -smooth. Let x and y be two positive integers. The probability that a randomly chosen integer m in the interval $[1, x]$ has all its prime factors less than y can be estimated by $\frac{\Psi(x, y)}{x}$ where

$$\Psi(x, y) = \#\{m \mid 1 \leq m \leq x, P(m) \leq y\}.$$

Let $u = \frac{\log x}{\log y}$ so that $y = x^{\frac{1}{u}}$. Then, for large x and u , the Canfield-Erdős-Pomerance Theorem [6] (see also [24]) states that

$$\Psi(x, x^{\frac{1}{u}}) = xu^{-u+o(u)},$$

holds for $u < (1 - \varepsilon) \frac{\log x}{\log \log x}$. Equivalently, using y , we get the following result

Theorem 3 : *Let x and y be two positive integers. Then the number of positive integers $m \leq x$ such that $P(m) \leq y$ is*

$$\Psi(x, y) = x \left(\frac{\log y}{\log x} \right)^{\frac{\log x}{\log y} + o\left(\frac{\log x}{\log y}\right)},$$

for $y > (\log x)^{1+\varepsilon}$.

3.4 RSA moduli with smooth sum of primes :

Let x be a positive integer, and $P(x)$ be the largest prime factor of x . For positive integers n and y , we consider the number of RSA moduli $N = pq$ with prime factors p and q satisfying $2^{n-1} < q < p < 2^n$ and $P(p+q-1) < y$.

Lemma 6 : *Let n and y be positive integers such that $y > ((n+1)\log 2)^{1+\varepsilon}$.*

Then the set

$$\mathcal{F}(n, y) = \{S = p + q - 1 \mid p, q \text{ primes}, 2^{n-1} < q < p < 2^n, P(p + q - 1) \leq y\},$$

has cardinality

$$\#\mathcal{F}(n, y) \sim 2^{n-1} \left(\frac{\log y}{n \log 2} \right)^{\frac{n \log 2}{\log y}}.$$

Proof : If p and q are prime numbers satisfying $2^{n-1} < q < p < 2^n$ and $S = p + q - 1$, then S is odd and $2^n < S < 2^{n+1}$. Setting $x_1 = 2^n$, $x_2 = 2^{n+1}$, and using $\Psi(x_1, y)$ and $\Psi(x_2, y)$ to count y -smooth integers as in Theorem (3), we get for $y > (\log x_2)^{1+\varepsilon}$,

$$\begin{aligned} \#\mathcal{F}(n, y) &= \frac{1}{2}(\Psi(x_2, y) - \Psi(x_1, y)) \\ &= \frac{1}{2}x_2 \left(\frac{\log y}{\log x_2} \right)^{\frac{\log x_2}{\log y} + o\left(\frac{\log x_2}{\log y}\right)} - \frac{1}{2}x_1 \left(\frac{\log y}{\log x_1} \right)^{\frac{\log x_1}{\log y} + o\left(\frac{\log x_1}{\log y}\right)}. \end{aligned}$$

Since x_1 and x_2 are much larger than y and $\frac{\log x_2}{\log x_1} \sim 1$, then,

$$\frac{\log x_2}{\log y} \sim \frac{\log x_1}{\log y}, \quad \frac{\log y}{\log x_2} \sim \frac{\log y}{\log x_1}.$$

Hence, $\#\mathcal{F}(n, y)$ transforms to

$$\#\mathcal{F}(n, y) \sim \frac{x_2 - x_1}{2} \left(\frac{\log y}{\log x_1} \right)^{\frac{\log x_1}{\log y}} \sim 2^{n-1} \left(\frac{\log y}{n \log 2} \right)^{\frac{n \log 2}{\log y}}.$$

This terminates the proof.

4. The new attack

In this section, we present the new attack on RSA when the prime factors p and q are such that $p + q - 1$ is P_0 -smooth where P_0 is the bound for ECM in Assumption (1). The attack combines the continued fraction algorithm and the elliptic curve factorization method ECM.

The first step of the attack is the following result which is based on Theorem (1).

Lemma 7 : Let $N = pq$ be an RSA modulus with $q < p < 2q$. Let e be a public exponent satisfying the equation $ex - \phi(N)y = (p+q-1)z$ with $\gcd(x, y) = 1$. If

$$|y - z| x < \frac{\sqrt{2}}{6} \sqrt{N},$$

then $\frac{y}{x}$ is a convergent of the continued fraction expansion of $\frac{e}{N+1}$.

Proof : By Lemma (2), the exponent e satisfies an equation of the form $ex - \phi(N)y = (p+q-1)z$. Then $ex - Ny = (p+q-1)(z-y)$ and

$$\left| \frac{e}{N} - \frac{y}{x} \right| = \frac{(p+q-1)|y-z|}{Nx}.$$

Using Lemma (1), we get

$$\left| \frac{e}{N} - \frac{y}{x} \right| \leq \frac{3\sqrt{2}\sqrt{N}|y-z|}{2Nx}.$$

Hence, if $|y - z| x < \frac{\sqrt{2}}{6} \sqrt{N}$, then

$$\left| \frac{e}{N} - \frac{y}{x} \right| < \frac{1}{2x^2}.$$

Hence, by Theorem (1), it follows that $\frac{y}{x}$ is a convergent of the continued fraction expansion of $\frac{e}{N}$.

Let P_0 be the bound for ECM in Assumption (1). The next step of the attack is to show how to factor the modulus $N = pq$ when the parameters in the equation $ex - \phi(N)y = (p+q-1)z$ satisfy the conditions of Lemma (7) and the sum $p+q-1$ or the difference $|y-z|$ is P_0 -smooth.

Theorem 4 : Let $N = pq$ be an RSA modulus. Let e be a public exponent satisfying the equation $ex - \phi(N)y = (p+q-1)z$ with $\gcd(x, y) = 1$ and $|y - z| x < \frac{\sqrt{2}}{6} \sqrt{N}$. If

$$P(p+q-1) \leq P_0, \quad \text{or} \quad P(|y-z|) < P_0,$$

then, under Assumption (1), one can factor N in the same run time than ECM.

Proof : Suppose that e satisfies the equation $ex - \phi(N)y = (p+q-1)z$ with $\gcd(x, y) = 1$ and $|y - z| x < \frac{\sqrt{2}}{6} \sqrt{N}$. Then, by Lemma (7), one can find x

and y using the continued fraction expansion of $\frac{e}{N}$. Using x and y , we can compute $M = |ex - Ny|$. Then $p + q - 1$ satisfies

$$(p + q - 1) |z - y| = M,$$

where $|z - y|$ is an unknown factor of M . In the next, we suppose that the largest prime factor of $|z - y|$ or $p + q - 1$ is less than the bound P_0 of the ECM method for integer factorization as in Assumption(1). By applying ECM to M , one gets $M = M_0 \cdot M_1$, where all prime factors of M_0 are bounded by P_0 , that is

$$M_0 = \prod_{i=1}^k p_i^{\alpha_i}, \quad p_i \leq P_0, \quad 1 \leq \alpha_i \leq \left\lfloor \frac{\log M}{\log p_i} \right\rfloor,$$

where k is the number of prime divisors of M_0 . By Theorem 2, we know that $k = \omega(n) \sim \log \log M_0 < \log \log M$. Then, the divisors of M_0 are of the form

$$g = \prod_{i=1}^k p_i^{\beta_i}, \quad 0 \leq \beta_i \leq \alpha_i,$$

and there are $\prod_{i=1}^k (\alpha_i + 1)$ such divisors. Nevertheless, by Theorem 2, the average number of divisors of M_0 is

$$d(M_0) \sim \log M_0 \leq \log M = \log((p + q - 1) |z - y|).$$

We have $|z - y| \leq |y - z| x < \frac{\sqrt{2}}{6} \sqrt{N}$ by assumption, and $p + q - 1 < \frac{3\sqrt{2}\sqrt{N}}{2}$ by Theorem 1. Plugging this in $d(M_0)$, we get

$$d(M_0) < \log \left(\frac{\sqrt{2}}{6} \sqrt{N} \cdot \frac{3\sqrt{2}\sqrt{N}}{2} \right) = \log \left(\frac{1}{2} N \right) < \log N.$$

First, suppose that $P(p + q - 1) \leq P_0$. For each divisor g of M_0 , we compute the solution of the system of equations

$$p_0 + q_0 - 1 = g, \quad p_0 q_0 = N, \quad p_0 \geq q_0 \geq 2.$$

If the solution (p_0, q_0) is a couple of integers, then $p = p_0$ and $q = q_0$ give the factorization of N .

Next, suppose that $P(|z - y|) \leq P_0$. For each divisor g of M_0 , we compute the solution of the system of equations

$$p_0 + q_0 - 1 = \frac{M}{g}, \quad p_0 q_0 = N, \quad p_0 \geq q_0 \geq 2.$$

If the solution (p_0, q_0) is a couple of integers, then $p = p_0$ and $q_0 = q$ give the factorization of N .

Since the continued fraction algorithm of a rational is a polynomial time algorithm, and there are less than $\log N$ divisors to process, then the factorization of N is achieved in the same run time algorithm than ECM. This terminates the proof.

5. The Number of Weak Keys

Let $N = pq$ be an RSA modulus where the prime sum $p + q - 1$ is P_0 -smooth. In this section, we give an explicit family of exponents e that are vulnerable to our attack, and give a very conservative estimation of the number of such weak exponents. The attack of Theorem (4) is based on two main conditions :

- The sum $S = p + q - 1$ is P_0 -smooth where P_0 is the bound in Assumption (1).
- The public exponent satisfies an equation $ex - \phi(N)y = (p + q - 1)z$ where the parameters are such that $\gcd(x, y) = 1$ and $|y - z| x < \frac{\sqrt{2}}{6} \sqrt{N}$.

The RSA modulus $N = pq$ is the product of two random prime numbers p and q with the same bit-size, typically $2^{n-1} < q < p < 2^n$. The sum $S = p + q - 1$ seems also random. Nevertheless, Lemma (5) and Lemma (6) show that the probability that S is P_0 -smooth is

$$\frac{\#\mathcal{F}(n, P_0)}{\#\mathcal{S}(n)} \sim \frac{2^{n-1} \left(\frac{\log P_0}{n \log 2} \right)^{\frac{n \log 2}{\log P_0}}}{2^{n-2}} = 2 \left(\frac{\log P_0}{n \log 2} \right)^{\frac{n \log 2}{\log P_0}}.$$

This is twice the probability that a random integer less than 2^{n+1} is P_0 -smooth.

Next, for an RSA modulus $N = pq$ with $2^{n-1} < q < p < 2^n$, suppose that $S = p + q - 1$ is P_0 -smooth. Consider the set

$$\mathcal{K} = \left\{ e \equiv \frac{(p+q-1)Z}{x} \pmod{N} \mid \gcd(x, Z) = 1, x \mid Z < \frac{\sqrt{2}}{6} \sqrt{N} \right\}.$$

If $e \in \mathcal{K}$, then there exists an integer y such that $ex - (p+q-1)Z = Ny$. Let $z = y + Z$. Then e satisfies the equation $ex - \phi(N)y = (p+q-1)z$ with $x \mid y - z \mid < \frac{\sqrt{2}}{6}\sqrt{N}$. These are the conditions of Theorem (7), and the exponent e is vulnerable to our attack under the condition $P(p+q-1) \leq P_0$ by Theorem (4). The following result shows that two different couples (x, Z) and (x', Z') produce two different exponents.

Lemma 8 : Let (x, Z) and (x', Z') satisfying $\gcd(x, Z) = 1$, $\gcd(x', Z') = 1$, $x \mid Z \mid < \frac{\sqrt{2}}{6}\sqrt{N}$ and $x' \mid Z' \mid < \frac{\sqrt{2}}{6}\sqrt{N}$. Define

$$e \equiv \frac{(p+q-1)Z}{x} \pmod{N}, \quad e' \equiv \frac{(p+q-1)Z'}{x'} \pmod{N}.$$

If $e = e'$, then $(x, Z) = (x', Z')$.

Proof : Suppose that $e \equiv \frac{(p+q-1)Z}{x} \pmod{N}$, $e' \equiv \frac{(p+q-1)Z'}{x'} \pmod{N}$, with $\gcd(x, Z) = 1$, $\gcd(x', Z') = 1$, $x \mid y - z \mid < \frac{\sqrt{2}}{6}\sqrt{N}$ and $x' \mid y' - z' \mid < \frac{\sqrt{2}}{6}\sqrt{N}$. If $e = e'$, then

$$\frac{e}{e'} = \frac{Zx'}{Z'x} \equiv 1 \pmod{N},$$

or equivalently $Zx' - Z'x \equiv 0 \pmod{N}$. We have

$$\mid xZ' - x'Z \mid \leq 2 \max(x \mid Z' \mid, x' \mid Z \mid) < 2 \cdot \left(\frac{\sqrt{2}}{6}\sqrt{N} \right)^2 = \frac{N}{9} < N.$$

Hence $xZ' - x'Z = 0$, and $\frac{Z}{x} = \frac{Z'}{x'}$. Since $\gcd(x, Z) = \gcd(x', Z')$, then $x = x'$, and $Z = Z'$. This terminates the proof.

The following result gives a lower bound for the size of the number of distinct exponents that are weak for our attack.

Proposition 1 : The number $\mathcal{N}$ of distinct exponents $e \in \mathcal{K}$ with $\gcd(e, \phi(N)) = 1$ that are weak to our attack satisfies

$$\mathcal{N} \geq \frac{\sqrt{2}}{6}\sqrt{N}.$$

Proof : Let $M = \left\lfloor \frac{\sqrt{2}}{6} \sqrt{N} \right\rfloor$. The number $\mathcal{N}$ of distinct exponents $e \in \mathcal{K}$ that are weak to our attack satisfies

$$\begin{aligned}
 \mathcal{N} &= \sum_{\substack{x=1 \\ \gcd(x, |Z|)=1, \gcd(e, \phi(N))=1}}^M \sum_{Z=-\frac{M}{x}}^{\frac{M}{x}} 1 \\
 &= 2 \sum_{\substack{x=1 \\ \gcd(x, |Z|, \gcd(e, \phi(N))=1)}}^M \sum_{Z=1}^{\frac{M}{x}} 1 \\
 &= 2M \sum_{\substack{x=1 \\ \gcd(x, |Z|, \gcd(e, \phi(N))=1)}}^M \frac{1}{x} \\
 &\geq 2M,
 \end{aligned}$$

and consequently $\mathcal{N} \geq \frac{\sqrt{2}}{6} \sqrt{N}$.

6. Comparison with former attacks

Many attacks on RSA, mainly based on the key equation $ed - k\phi(N) = 1$, have been presented such as the attack of Wiener [26] and the attack of Boneh-Durfee [4]. Since our equation is in the form $ex - \phi(N)y = (p+q-1)z$, we only compare our attack to the attack of Blömer-May [2] which is based on the variant equation $ex' - \phi(N)y' = z'$. The attack of Blömer-May is possible if the following conditions are fulfilled

$$1 \leq x' \leq \frac{1}{3} N^{\frac{1}{4}}, \quad |z'| < N^{-\frac{3}{4}} ex'. \quad (1)$$

In our attack, the equation is $ex - \phi(N)y = (p+q-1)z$ with $|y-z|x < \frac{\sqrt{2}}{4} \sqrt{N}$ where $p+q-1$ is P_0 -smooth. To apply the attack of Blömer-May to the equation $ex - \phi(N)y = (p+q-1)z$, the parameters x , y or z should satisfy the conditions ((1)) with $x' = x$, $y' = y$ and $z' = (p+q)z$. We discuss here the possibility of applying this attack.

- First, x should satisfy $1 \leq x \leq \frac{1}{3} N^{\frac{1}{4}}$ while our condition is

$$|y-z|x < \frac{\sqrt{2}}{6} \sqrt{N} \text{ which can be satisfied even for } x > \frac{1}{3} N^{\frac{1}{4}}.$$

- Second, for $1 \leq x \leq \frac{1}{3}N^{\frac{1}{4}}$, the condition $|z'| = (p+q-1)|z| < N^{\frac{3}{4}}ex$ should be satisfied. Using Lemma (1), this implies that z should satisfy

$$|z| < \frac{N^{\frac{3}{4}}ex}{p+q-1} \leq \frac{N^{\frac{3}{4}}e \cdot \frac{1}{3}N^{\frac{1}{4}}}{2\sqrt{N}} = \frac{e}{6N},$$

which is impossible if $e < N$. In contrast, we notice that our attack can be applied in this situation.

The following numerical example shows that our attack can factor the RSA modulus, while the method of Blömer-May can not, even when $1 \leq x \leq \frac{1}{3}N^{\frac{1}{4}}$.

Consider the RSA modulus N and the public exponent e :

$$\begin{aligned} N &= 1109588733951066161621444541807594330953054059731781682203833 \\ &\quad 144561998894299860092108429, \\ e &= 5279260403171254865054708811525772310380190134076606252599984 \\ &\quad 60242050783295418708150391. \end{aligned}$$

One of the convergents of $\frac{e}{N}$ is

$$\frac{y}{x} = \frac{4141907516815061251110241}{8705412438008399695743224}.$$

We take

$$x = 8705412438008399695743224, \quad y = 4141907516815061251110241,$$

from which we get

$$\begin{aligned} M = |ex - Ny| &= 4976027368320236249605032824108992876193 \\ &\quad 1415223327678306879195. \end{aligned}$$

By applying ECM to M with the bound $P_0 = 10^6$, we get the list of P_0 -smooth factors, namely $M = 5 \cdot 37^4 \cdot 109 \cdot 1249^3 \cdot 9851^2 \cdot 78721^4 \cdot M'$, where the prime factors are larger than P_0 . Using the divisor $g = 5 \cdot 37^4 \cdot 1249^3 \cdot 9851^2 \cdot 78721^4$, we get

$$p+q-1 = 68043596921586139175071580389567209356324045,$$

which combined with $pq = N$ give

$$p = 40942350213056433546764732109616278918346813,$$

$$q = 27101246708529705628306848279950930437977233.$$

We note that $d \equiv e^{-1} \pmod{\phi(N)} > N^{0.991}$ which means that the method of Wiener, as well as the method of Boneh-Durfee can not be applied for this example. On the other hand, as required in the method of Blömer-May the convergent $\frac{y}{x}$ should satisfy $1 \leq x \leq \frac{1}{3}N^{\frac{1}{4}}$, and $|ex - \phi(N)y| < N^{-\frac{3}{4}}ex$. This is not the case. Moreover, the inequality $|ex' - \phi(N)y'| < N^{-\frac{3}{4}}ex'$ is never satisfied with any convergent $\frac{y'}{x'}$ of $\frac{e}{N}$. This shows that the method of Blömer-May can not be applied to find the factorization of the modulus N in this example.

7. Conclusion

In this work, we presented an attack on the RSA cryptosystem with a modulus $N = pq$ by studying a specific arithmetic relation verified by the public exponent e , the Euler totient function $\phi(N)$, and the sum $p + q - 1$. The attack is efficient and works when the following conditions are satisfied : the sum $p + q - 1$ is composed by sufficiently small prime factors, the public exponent e is such that $ex - \phi(N) = (p + q - 1)z$, and the parameters x , $|y - z|$ are suitably small. The technique used in the attack combines the continued fraction algorithm and the elliptic curve factorization method (ECM). Moreover, we showed that such RSA moduli as well as the exponents that are vulnerable to our attack represent non-negligible parts of all possible moduli and exponents in an interval.

References

- [1] J. Blömer, A. May (2003) New partial key exposure attacks on RSA. In: Boneh D. (eds) *Advances in Cryptology - CRYPTO 2003*. CRYPTO 2003. Lecture Notes in Computer Science, vol 2729. Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-540-45146-4_2
- [2] J. Blömer, A. May (2004) A generalized Wiener attack on RSA. In: Bao F., Deng R., Zhou J. (eds) *Public Key Cryptography PKC 2004*. PKC 2004. Lecture Notes in Computer Science, vol 2947. Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-540-24632-9_1

- [3] D. Boneh, Twenty years of attacks on the RSA cryptosystem, *Notices Amer. Math. Soc.* 46 (2), 203-213, (1999)
- [4] D. Boneh, G. Durfee (1999) Cryptanalysis of RSA with private key d less than $N^{0.292}$. In: Stern J. (eds) *Advances in Cryptology EUROCRYPT 99*. EUROCRYPT 1999. Lecture Notes in Computer Science, vol 1592. Springer, Berlin, Heidelberg. doi.org/10.1007/3-540-48910-X_1
- [5] D. Boneh, G. Durfee and Y. Frankel, An attack on RSA given a small fraction of the private key bits, *Advances in Cryptology - ASIACRYPT'98* (Beijing), Lecture Notes in Comput. Sci., Springer, Berlin, 1514 (1998), 25-34. doi: 10.1007/3-540-49649-1_3
- [6] E.R Canfield, P. Erdős, C. Pomerance, On a problem of Oppenheim concerning factorisatio numerorum, *Journal of Number Theory* 17, pp. 1-28 (1983)
- [7] D. Coppersmith, Small solutions to polynomial equations, and low exponent RSA vulnerabilities. *J. Cryptology* 10, 233-260 (1997). doi.org/10.1007/s001459900030
- [8] N. Demytko (1994) A new elliptic curve based analogue of RSA. In: Helleseht T. (eds) *Advances in Cryptology EUROCRYPT 93*. EUROCRYPT 1993. Lecture Notes in Computer Science, vol 765. Springer, Berlin, Heidelberg. doi.org/10.1007/3-540-48285-7_4
- [9] M. Ernst, E. Jochemsz, A. May, B. de Weger (2005) Partial key exposure attacks on RSA up to full size exponents. In: Cramer R. (eds) *Advances in Cryptology EUROCRYPT 2005*. EUROCRYPT 2005. Lecture Notes in Computer Science, vol 3494. Springer, Berlin, Heidelberg. doi.org/10.1007/11426639_22
- [10] G. H. Hardy, E. M. Wright, *An Introduction to Theory of Numbers*, 5th Edition, The Clarendon Press Oxford University Press, New York, 1979.
- [11] M. Hinek, *Cryptanalysis of RSA and Its Variants*, Chapman & Hall/CRC, Cryptography and Network Security Series, Boca Raton, (2009)
- [12] K. Koyama, U.M. Maurer, T. Okamoto, S.A. Vanstone (1992) New public-key schemes based on elliptic curves over the ring $\mathbb{Z}_n$. In: Feigenbaum J. (eds) *Advances in Cryptology CRYPTO 91*. CRYPTO 1991. Lecture Notes in Computer Science, vol 576. Springer, Berlin, Heidelberg. doi.org/10.1007/3-540-46766-1_20
- [13] H. W. Lenstra Jr. (1987) Factoring integers with elliptic curves, *Annals of Mathematics*, Vol. 126, No. 2, 1987, pp. 649-673. doi:10.2307/1971363

- [14] S. Maitra, S. Sarkar (2008) Revisiting Wiener's attack - new weak keys in RSA. In: Wu TC., Lei CL., Rijmen V., Lee DT. (eds) Information Security. ISC 2008. Lecture Notes in Computer Science, vol 5222. Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-540-85886-7_16
- [15] G. d. Meulenaer, F. Gosset, G. M. d. Dormale and J. Quisquater, Integer factorization based on elliptic curve method: towards better exploitation of reconfigurable hardware, 15th Annual IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM 2007), 2007, pp. 197-206, doi: 10.1109/FCCM.2007.12
- [16] M. Mumtaz, L. Ping (2019) Forty years of attacks on the RSA cryptosystem: a brief survey, *Journal of Discrete Mathematical Sciences and Cryptography*, 22:1, 9-29, doi: 10.1080/09720529.2018.1564201.
- [17] A. Nitaj, Another generalization of Wiener's attack on RSA, Progress in Cryptology-AFRICACRYPT 2008, 174-190, Lecture Notes in Comput. Sci., 5023, Springer-Verlag, 2008 doi.org/10.1007/978-3-540-68164-9_12
- [18] A. Nitaj (2013) Diophantine and lattice cryptanalysis of the RSA cryptosystem. In: Yang XS. (eds) Artificial Intelligence, Evolutionary Computing and Metaheuristics. Studies in Computational Intelligence, vol 427. Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-642-29694-9_7
- [19] A. Nitaj, A new attack on the KMOV cryptosystem, Bull. Korean Math. Soc. 2014 Vol. 51, No. 5, 1347-1356, doi.org/10.4134/BKMS.2014.51.5.1347
- [20] A. Nitaj, E. Fouotsa (2019) A new attack on RSA and Demytko's elliptic curve cryptosystem, *Journal of Discrete Mathematical Sciences and Cryptography*, 22:3, 391-409, doi: 10.1080/09720529.2019.1587827
- [21] R. Rivest, A. Shamir, L. Adleman, A Method for obtaining digital signatures and public-key cryptosystems, Communications of the ACM Volume 21 Issue 2 Feb. 1978 pp. 120-126 doi.org/10.1145/359340.359342
- [22] S. Sarkar, S. Maitra and S. Sarkar, RSA cryptanalysis with increased bounds on the secret exponent using less lattice dimension, IACR Cryptology ePrint Archive (2008) <https://eprint.iacr.org/2008/315>
- [23] H.M Sun, M.-E. Wu, R. Steinfeld, J. Guo, H. Wang, (2008) Cryptanalysis of short exponent rsa with primes sharing least significant bits. In:

- Franklin M.K., Hui L.C.K., Wong D.S. (eds) Cryptology and Network Security. CANS 2008. Lecture Notes in Computer Science, vol 5339. Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-540-89641-8_4
- [24] G. Tenenbaum, Introduction to Analytic and Probabilistic Number Theory, Cambridge Studies in Advanced Mathematics 46. Cambridge University Press, Cambridge, 1995.
- [25] B. de Weger, Cryptanalysis of RSA with small prime difference. Applicable Algebra in Engineering, Communication and Computing, 13(1), pp. 17-28, 2002 doi:10.1007/s002000100088
- [26] M.J. Wiener (1990) Cryptanalysis of short RSA secret exponents. In: Quisquater J.J., Vandewalle J. (eds) Advances in Cryptology EUROCRYPT 89. EUROCRYPT 1989. Lecture Notes in Computer Science, vol 434. Springer, Berlin, Heidelberg. doi.org/10.1007/3-540-46885-4_36
- [27] Y.-D. Zhao, W.-F. Qi, Small private-exponent attack on RSA with primes sharing bits, in Proc. Information Security Conference 2007, ISC 2007, ser. Lecture Notes in Computer Science, J. Garay et al., Eds. Heidelberg: Springer, 2007, vol. 4779, pp. 221-229, (2007)
- [28] P. Zimmermann, 50 largest prime factors found by ECM. <https://members.loria.fr/PZimmermann/records/top50.html>
- [29] P. Zimmermann, B. Dodson, 20 years of ECM. In: Hess F., Pauli S., Pohst M. (eds) Algorithmic Number Theory. ANTS 2006. Lecture Notes in Computer Science, vol 4076. Springer, Berlin, Heidelberg (2006) doi: 10.1007/11792086_37

Received April, 2021

Revised July, 2021