Factoring RSA moduli with weak prime factors

Abderrahmane Nitaj[1⋆] and Tajjeeddine Rachidi[2]

[1] Laboratoire de Mathématiques Nicolas Oresme
Université de Caen Basse Normandie, France
abderrahmane.nitaj@unicaen.fr
[2] School of Science and Engineering
Alakhawayn University in Ifrane, Morocco
T.Rachidi@aui.ma

Abstract. In this paper, we study the problem of factoring an RSA modulus $N = pq$ in polynomial time, when p is a weak prime, that is, p can be expressed as $ap = u_0 + M_1u_1 + \ldots + M_ku_k$ for some k integers $M_1, \ldots, M_k$ and $k+2$ suitably small parameters a, $u_0, \ldots u_k$. We further compute a lower bound for the set of weak moduli, that is, moduli made of at least one weak prime, in the interval $[2^{2n}, 2^{2(n+1)}]$ and show that this number is much larger than the set of RSA prime factors satisfying Coppersmith's conditions, effectively extending the likelihood for factoring RSA moduli. We also prolong our findings to moduli composed of two weak primes.

Keywords: RSA, Cryptanalysis, Factorization, LLL algorithm, Weak primes

1 Introduction

The RSA cryptosystem, invented in 1978 by Rivest, Shamir and Adleman [17] is undoubtedly one of the most popular public key cryptosystems. In the standard RSA [17], the modulus $N = pq$ is the product of two large primes of the same bit-size. The public exponent e is an integer such that $1 \leq e < \phi(N)$ and $\gcd(e, \phi(N)) = 1$ where $\phi(N) = (p-1)(q-1)$ is the Euler totient function. The corresponding private exponent is the integer d such that $ed \equiv 1 \pmod{\phi(N)}$. In RSA, the encryption, decryption, signature generation, and signature verification require substantial CPU cycles because the time to perform these operations is proportional to the number of bits in public or secret exponents [17]. To reduce CPU time necessary for encryption and signature verification, one may be tempted to use a small public exponent e. This situation has been proven to be insecure against some small public exponent attacks (see [8] and [9]). To reduce the decryption and signature generation time, one may also be tempted to use a small private exponent d. Unfortunately, RSA is also vulnerable to various powerful short secret exponent attacks such as, the attack of Wiener [20],

⋆ Partially supported by the French SIMPATIC (SIM and PAiring Theory for Information and Communications security).

and the attack of Boneh and Durfee [4] (see also [3]). An alternate way for increasing the performance of encryption, decryption, signature generation, and signature verification, without reverting to small exponents, is to use the multi-prime variant of RSA. The multi-prime RSA is a generalization of the standard RSA cryptosystem in which the modulus is in the form $N = p_1 p_2 \cdots p_k$ where $k \geq 3$ and the p_i's are distinct prime numbers. Combined with the Chinese Remainder Theorem, a multi-prime RSA is much more efficient than the standard RSA (see [5]).

In Section 4.1.2 of the X9.31-1998 standard for public key cryptography [1], some recommendations are presented regarding the generation of the prime factors of an RSA modulus. For example, it is recommended that the modulus should have $1024 + 256x$ bits for $x \geq 0$. This requirement deters some factorization attacks, such as the Number Field Sieve (NFS) [12] and the Elliptic Curve Method (ECM) [11]. Another recommendation is that the prime difference $|p-q|$ should be large, and $\frac{p}{q}$ should not be near the ratio of two small integers. These requirements guard against Fermat factoring algorithm [19], as well as Coppersmith's factoring attack on RSA [6] when one knows half of the bits of p. For example, if $N = pq$ and p, q are of the same bit-size with $|p - q| < N^{1/4}$, then $\left|p - \left[\sqrt{N}\right]\right| < N^{1/4}$ (see [16]) where $\left[\sqrt{N}\right]$ is the nearest integer to $\sqrt{N}$, which means that half of the bits of p are those of $[\sqrt{N}]$ which leads to the factorization of N (see [6] and [19]). Observe that the factorization attack of Coppersmith applies provided that one knows half of the bits of p, that is p is in one of the forms

$$p = \begin{cases} M_1 + u_0 & \text{with known } M_1 \text{ and unknown } u_0 \leq N^{\frac{1}{4}}, \\ M_1 u_1 + M_0 & \text{with known } (M_1, M_0) \text{ and unknown } u_1 \leq N^{\frac{1}{4}}. \end{cases}$$

Such primes are called Coppersmith's weak primes. In the case of $p = M_1 u_1 + M_0$ with known M_1 and M_0, the Euclidean division of q by M_1 is in the form $q = M_1 v_1 + v_0$. Hence $N = pq = (M_1 u_1 + M_0)(M_1 v_1 + v_0)$ which gives $M_0 v_0 \equiv N \pmod{M_1}$. Hence, since $gcd(M_0, M_1) = 1$, then $v_0 \equiv N M_0^{-1} \pmod{M_1}$. This means that when p is in the form $p = M_1 u_1 + M_0$ with known M and u_0, then q is necessarily in the form $q = M_1 v_1 + v_0$ with known v_0. Coppersmith's attack is therefore applicable only when small enough parameters M_0 and v_0 can be found such that $p = M_1 u_1 + M_0$ and $q = M_1 v_1 + v_0$. This reduces the applicability of the attack to the set of moduli such that p and q are of the form defined above.

In this paper, we consider the generalization of Coppersmith's attack by considering a more satisfiable decomposition of any of the multipliers of p or q, i.e., ap or aq not just p or q, effectively leading to an increased set of moduli that can be factored. We describe two new attacks on RSA with a modulus $N = pq$. The first attack applies in the situation that, for given positive integers $M_1, \ldots, M_k$, one of the prime factors, p say, satisfies a linear equation $ap = u_0 + M_1 u_1 + \ldots + M_k u_k$ with suitably small integers a and $u_0, \ldots, u_k$. We call such prime factors *weak primes* for the integers $M_1, \ldots, M_k$. The second attack applies when both factors p and q are weak for the integers $M_1, \ldots, M_k$. We note

that, for $k = 1$, the weak primes are such that $ap = u_0 + M_1 u_1$. This includes the class of Coppersmith's weak primes. For both attacks, we give an estimation of the RSA moduli $N = pq$ with a prime factor $p \in \left[2^n, 2^{n+1}\right]$ which is weak for the integers $M, M^2, \ldots, M^k$ where $M = \left\lceil 2^{\frac{n}{2k}} \right\rceil$. We show that the number of moduli with a weak prime factor is much larger than the number of moduli with a Coppersmith's weak prime factor.

The rest of the paper is organized as follows. In Section 2, we give some basic concepts on integer factorization and lattice reduction as well as an overview of Coppersmith's method. In Section 3, we present an attack on an RSA modulus $N = pq$ with one weak prime factor. In Section 4, we present the second attack an RSA modulus $N = pq$ with two weak prime factors. We conclude the paper in Section 5.

2 Preliminaries

In this section we give the definitions and results that we need to perform our attacks. These preliminaries include basic concepts on integer factorization and lattice reduction techniques.

2.1 Integer factorization: the state of the art

Currently, the most powerful algorithm for factorizing large integers is the Number Field Sieve (NFS) [12]. The heuristic expected time $T_{NFS}(N)$ of the NFS depends on the bitsize of the integer N to be factored:

$$T_{NFS}(N) = \exp\left((1.92 + o(1))(\log N)^{1/3}(\log\log N)^{2/3}\right).$$

If the integer N has small factors, the Elliptic Curve Method (ECM) [11] for factoring is substantially faster than the NFS. It can compute a non-trivial factor p of a composite integer N in an expected runtime T_{ECM}:

$$T_{ECM}(p) = \exp\left(\left(\sqrt{2} + o(1)\right)(\log p)^{1/2}(\log\log p)^{1/2}\right),$$

which is sub-exponential in the bitsize of the factor p. The largest factor found so far with the ECM is a 83 decimal digits (275 bits) prime factor of the special number $7^{337} + 1$ (see [18]).

2.2 Lattice reduction

Let m and n be positive integers with $m \leq n$. Let $u_1, \ldots, u_m \in \mathbb{R}^n$ be m linearly independent vectors. The lattice $\mathcal{L}$ spanned by $u_1, \ldots, u_m$ is the set

$$\mathcal{L} = \left\{ \sum_{i=1}^{m} a_i u_i \mid a_i \in \mathbb{Z} \right\}.$$

The set $\{u_1, \ldots, u_m\}$ is called a lattice basis for $\mathcal{L}$. The dimension (or rank) of the lattice $\mathcal{L}$ is $\dim(\mathcal{L}) = m$, and $\mathcal{L}$ is called full rank if $m = n$. It is often useful to represent the lattice $\mathcal{L}$ by the $m \times n$ matrix M whose rows are the coefficients of the vectors $u_1, \ldots, u_m$. The determinant (or volume) of $\mathcal{L}$ is defined as $\det(\mathcal{L}) = \sqrt{M \cdot M^t}$. When $\mathcal{L}$ is full rank, the determinant reduces to $\det(\mathcal{L}) = |\det(M)|$. The Euclidean norm of a vector $v = \sum_{i=1}^{m} a_i u_i \in \mathcal{L}$ is defined as $\|v\| = \sqrt{\sum_{i=1}^{m} a_i^2}$. As a lattice has infinitely many bases, some bases are better than others, and a very important task is to find a basis with small vectors $\{b_1, \ldots, b_m\}$ called the reduced basis. This task is very hard in general, however, the LLL algorithm proposed by Lenstra, Lenstra, and Lovász [13] finds a basis of a lattice with relatively small vectors in polynimial time. The following theorem determines the sizes of the reduced basis vectors obtained with LLL (see [15] for more details).

Theorem 1. *Let $\mathcal{L}$ be a lattice spanned by a basis $\{u_1, \ldots, u_m\}$. The LLL algorithm applied to $\mathcal{L}$ outputs a reduced basis $\{b_1, \ldots, b_m\}$ with*

$$\|b_1\| \leq \|b_2\| \leq \ldots \leq \|b_i\| \leq 2^{\frac{m(m-1)}{4(m-i+1)}} \det(L)^{\frac{1}{m+i-1}}, \text{ for } i = 1, 2, \ldots, m.$$

The existence of a short nonzero vector in a lattice is guaranteed by a result of Minkowski stating that every m-dimensional lattice $\mathcal{L}$ contains a non-zero vector v with $\|v\| \leq \sqrt{m} \det(L)^{\frac{1}{m}}$. On the other hand, the Gaussian Heuristic asserts that the norm γ_1 of the shortest vector of a random lattice satisfies

$$\gamma_1 \approx \sqrt{\frac{\dim(\mathcal{L})}{2\pi e}} \det(\mathcal{L})^{\frac{1}{\dim(\mathcal{L})}}.$$

Hereafter, we will use this result as an estimation for the expected minimum norm of a non-zero vector in a lattice.

2.3 Coppersmith's Method

In 1996, Coppersmith [6] presented two techniques based on LLL to find small integer roots of univariate modular polynomials or of bivariate integer polynomials. Coppersmith showed how to apply his technique to factorize an RSA modulus $N = pq$ with $q < p < 2q$ when half of the least or the most significant bits of p is known.

Theorem 2. *Let $N = pq$ be an RSA modulus with $q < p < 2q$. Let M_0 and M_1 be two positif integers. If $p = M_1 + u_0$ with $u_0 < N^{\frac{1}{4}}$ or if $p = M_1 u_1 + M_0$ with $u_1 < N^{\frac{1}{4}}$, then N can be factored in time polynomial in $\log N$.*

Coppersmith's technique extends to polynomials in more variables, but the method becomes heuristic. The problem of finding small roots of linear modular polynomials $f(x_1, \ldots, x_n) = a_1 x_1 + a_2 x_2 + \ + a_n x_n + a_{n+1} \pmod{p}$ for some unknown p that divides the known modulus N has been studied using Coppersmith's technique by Herrmann and May [10]. The following result, due to Lu, Zhang and Lin [14] gives a sufficient condition under which modular roots can be found efficiently.

Theorem 3 (Lu, Zhang, Lin). *Let N be a composite integer with a divisor p^u such that $p \geq N^\beta$. Let $f(x_1, \ldots, x_n) \in \mathbb{Z}[x_1, \ldots, x_n]$ be a homogenous linear polynomial. Then one can find all the solutions $(y_1, \ldots, y_n)$ of the equation $f(x_1, \ldots, x_n) = 0 \mod p^v$, $v \leq u$ with $\gcd(y_1, \ldots, y_n) = 1$ and $|y_1| < N^{\delta_1}, \ldots, |y_n| < N^{\delta_n}$ if*

$$\sum_{i=1}^{n} \delta_i \leq \frac{u}{v}\left(1 - \left(1 - \frac{u}{v}\beta\right)^{\frac{n}{n-1}} - n\left(1 - \sqrt[n-1]{1 - \frac{u}{v}\beta}\right)\left(1 - \frac{u}{v}\beta\right)\right).$$

The time complexity of the algorithm for finding such sulution $(y_1, \ldots, y_n)$ is polynomial in $\log N$.

3 The Attack with One Weak Prime Factor

3.1 The Attack

In this section, we present an attack to factor an RSA modulus $N = pq$ when p satisfies a linear equation in the form $ap = u_0 + M_1u_1 + \ldots M_ku_k$ for a suitably small positive integer a and suitably small integers $u_0, u_1, \ldots, u_k$ where $M_1, \ldots, M_k$ are given positive integers. Such prime factor p is called a weak prime for the integers $M_1, \ldots, M_k$.

Theorem 4. *Let $N = pq$ be an RSA modulus such that $p > N^\beta$ and $M_1, \ldots, M_k$ be k positive integers with $M_1 < M_2 < \ldots < M_k$. Suppose that there exists a positive integer a, and $k+1$ integers u_i, $i = 0, \ldots, k$ such that $ap = u_0 + M_1u_1 + \ldots + M_ku_k$ with $\max(u_i) < N^\delta$ and*

$$\delta < \frac{1}{k+1}\left(1 - (1-\beta)^{\frac{k+1}{k}} - (k+1)\left(1 - \sqrt[k]{1-\beta}\right)(1-\beta)\right).$$

Then one can factor N in polynomial time.

Proof. Let $M_1, \ldots, M_k$ be k positive integers such that $M_1 < M_2 < \ldots < M_k$. Suppose that $ap = u_0 + M_1u_1 + \ldots + M_ku_k$, that is $(u_0, \ldots, u_k)$ is a solution of the modular polynomial equation

$$x_0 + M_1x_1 + \ldots + M_kx_k = 0 \pmod{p}. \tag{1}$$

Suppose that $|u_i| < N^\delta$ for $i = 0, \ldots, k$. Using $n = k+1$, $u = 1$ and $v = 1$ in Theorem 3, means that the equation (1) can be solved in polynomial time, i.e., finding $(u_0, \ldots, u_k)$ if

$$(k+1)\delta < \left(1 - (1-\beta)^{\frac{k+1}{k}} - (k+1)\left(1 - \sqrt[k]{1-\beta}\right)(1-\beta)\right),$$

which gives the bound

$$\delta < \frac{1}{k+1}\left(1 - (1-\beta)^{\frac{k+1}{k}} - (k+1)\left(1 - \sqrt[k]{1-\beta}\right)(1-\beta)\right).$$

This terminates the proof. □

Remark 1. For a balanced RSA modulus, the prime factors p and q are of the same bit size. Then $p > N^\beta$ with $\beta = \frac{1}{2}$. Hence, the condition on δ becomes

$$\delta < \frac{1}{k+1}\left(1 - \left(\frac{1}{2}\right)^{\frac{k+1}{k}}\right) - \frac{1}{2}\left(1 - \left(\frac{1}{2}\right)^{\frac{1}{k}}\right). \tag{2}$$

In Table 1, we give the bound for δ for given β and k.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
$\beta=0.5$	0.125	0.069	0.047	0.036	0.029	0.024	0.021	0.018	0.016	0.015
$\beta=0.6$	0.180	0.101	0.071	0.054	0.044	0.037	0.032	0.028	0.025	0.022
$\beta=0.7$	0.245	0.142	0.100	0.077	0.063	0.053	0.046	0.046	0.036	0.032

Table 1. Upper bounds for δ by Theorem 4.

Remark 2. We note that Coppersmith's weak primes correspond to moduli $N = pq$ with $q < p < 2q$ where one of the prime factors is of the form $p = M_1 + u_0$ or $p = M_1 u_1 + M_0$ with $u_0, u_1 < N^{0.25}$ as mentioned in Theorem 2. This a special case of the equation of Theorem 4. Indeed, we can solve the equations $p = M_1 + u_0$ and $p = M_1 u_1 + M_0$ when $|u_0|, |u_1| < N^{\frac{1}{4}}$. Alternatively, Coppersmith's weak primes correspond to the cell $(k, 2\beta) = (1, 0.25)$ in Table 1.

3.2 Numerical Examples

Example 1. Let

$$N = 100097528863121099880227782275505778370812151920051298647846851857440468018795774211860316385574268129624076883575119637091 41,$$

be a 412-bit RSA modulus with $N = pq$ where $q < p < 2q$. Then p and q are balanced and $p \approx N^{\frac{1}{2}} \approx 2^{206}$. Hence for $\beta = 0.5$, we have $p > N^\beta$. Suppose that p satisfies an equation of the form $ap = u_0 + Mu_1 + M^2 u_2$. Typically, $M^2 \approx N^{\frac{1}{2}}$, that is $M \approx N^{\frac{1}{4}}$. So let $M = 2^{100}$. For $\beta = 0.5$ and $k = 2$, Table (1) gives the bound $\delta < 0.069$. Assume therefore that the parameters u_i satisfy $|u_i| < N^{0.069} \approx 2^{28}$ for $i = 0, 1, 2$. By applying Theorem 4 we should find u_0, u_1 and u_2 as long as $u_0, u_1, u_2 < 2^{28}$. We apply the method of Lu et al. [14] with $m = 4$ and $t = 1$. This gives a 35-dimensional lattice. Applying the LLL algorithm [13], we find a reduced basis with multivariate polynomials $f_i(x_1, x_2, x_3) \in \mathbb{Z}[x_1, x_2, x_3]$, $i = 1, \ldots, 3$. Applying the Gröbner basis technique for solving a system of polynomial equations, we get $u_0 = 9005$, $u_1 = 7123$,

$u_2 = 3915$. Using these values, we can compute $ap = u_0 + Mu_1 + M^2u_2$ from which we deduce $p = \gcd(u_0 + Mu_1 + M^2u_2, N)$, that is

$$p = 123356126338704841740132972382836883609800988209539117002682143.$$

Finally, we can compute $q = \frac{N}{p}$, that is

$$q = 811451622502140724659803961925628218026979706614326237650389 87.$$

Note here that there is no linear decomposition of p in the form $p = M_1 + u_0$ nor $p = M_1u_1 + M_0$ with $u_0, u_1 < N^{0.25}$ that makes p vulnerable to the attack of Coppersmith. This shows that the modulus N is vulnerable to our attack, while it is not vulnerable to Coppersmith's attack. Finally, the overall recorded execution time for our attack using an off-the-shelf computer was 17 seconds.

Example 2. In [2], Bernstein et al. discovered many prime factors with special forms. Many of these primes were found by computing the greatest common divisor of a collection of RSA moduli. Others were found by applying Coppersmith's technique. We show below that our attack can find some primes among the list of Bernstein et al. One of these primes is

$$\begin{aligned} p =& 0xc000000000000000000000000000000000000000000000000000000000000000 \\ & 000000000000000000000000000000000000000000000000000000000000002f9, \\ =& 100558559474569478246805187486543845956095243654442950333292671082 \\ & 791323022555160232601405723625177570767523893639864538140315412 10 \\ & 8959927459825236754563833. \end{aligned}$$

Using $M = 2^{510}$, we get $p = 3M + 761 = Mu_1 + u_0$ where $u_1 = 3$ and $u_0 = 761$. We have $u_1, u_0 < N^{\delta}$ with $\delta \approx 0.007$ which is less than the bound 0.125 in Table 1 for a 1024 bit-size RSA modulus N with $\beta = 0.5$, and $k = 1$. This implies that the conditions for Theorem 4 are satisfied and our method finds p when used in any RSA modulus.

Example 3. Now, consider this other example from the list of Bernstein et al. [2]

$$\begin{aligned} p =& 0xc000b800000000000000000000000000000000000000000000000000000000 \\ & 000006800000000000000000000000000000000000000000000000000000000251 \\ =& 100560029943006619091785857474102967729151903474112071240937611 5 \\ & 252074921606554559888603722177799493811165931923242874631881248 7 \\ & 609513837263772711701709393 \end{aligned}$$

Then p has the form $p = 3145774M^7 + 27262976M^3 + 593 = M^7u_7 + M^3u_3 + u_0$ where $M = 2^{70}$. The coefficients u_7, u_3 and u_0 satisfy $u_7, u_3, u_0 < N^{\delta}$ with $\delta \approx 0.016$ while the bound of Theorem 4 is 0.021 (see Table 1 for $k = 7$ and $\beta = 0.5$). Again, this shows that our method will find the factorization of any RSA modulus that is a multiple of p.

3.3 The Number of Single Weak Primes in an Interval

In this section, we consider two positive integers n and M and present a study of the weak primes with M, that is the primes $p \in \left[2^n, 2^{n+1}\right]$ such that there exists a positive integer a that gives the decomposition

$$ap = \sum_{i=0}^{k} M^i u_i$$

where $|u_i| < N^\delta$ and δ satisfies Theorem 4. We show that the number of the RSA moduli N in the interval $[2^{2n}, 2^{2(n+1)}]$ with a weak prime factor $p \in \left[2^n, 2^{n+1}\right]$ is polynomial in 2^n. That is, this number is lower bounded by 2^η where $\eta > \frac{1}{2}$. We call such a class weak RSA Moduli in the interval $[2^{2n}, 2^{2(n+1)}]$.

Theorem 5. *Let n be a positive integer. For $k \geq 1$, define $M = \left\lceil 2^{\frac{n}{k}} \right\rceil$. Let $\mathcal{N}$ be the set of the weak RSA moduli $N \in \left[2^{2n}, 2^{2(n+1)}\right]$ such that $N = pq$, p and q are of the same bitsize, $p > q$, and $p = \left\lfloor \frac{\sum_{i=0}^{k} M^i u_i}{a} \right\rfloor + b \in \left[2^n, 2^{n+1}\right]$ for some small integers b, $a < N^\delta$ and $|u_i| < N^\delta$ for $i = 0, \ldots, k$ with*

$$\delta = \frac{1}{k+1}\left(1 - \left(\frac{1}{2}\right)^{\frac{k+1}{k}}\right) - \frac{1}{2}\left(1 - \left(\frac{1}{2}\right)^{\frac{1}{k}}\right).$$

Then the cardinality of $\mathcal{N}$ satisfies $\#\mathcal{N} \geq 2^\eta$ where

$$\eta = (1 + 2(k+1)\delta)n + \log_2\left(\frac{(n-1)}{n(n+1)\log(2)}\right).$$

Proof. Let N be an RSA moduli. Suppose that $N \in \left[2^{2n}, 2^{2(n+1)}\right]$ with $N = pq$ where p and q are of the same bitsize. Since $p \approx N^{\frac{1}{2}}$, then $p \in \left[2^n, 2^{n+1}\right]$. Suppose further that for some positive integer a, we have $ap = \sum_{i=0}^{k} M^i u_i$. Then

$$M^k = \frac{ap - \sum_{i=0}^{k-1} M^i u_i}{u_k} \approx \frac{a}{u_k} p,$$

which implies $M \approx p^{\frac{1}{k}} \approx N^{\frac{1}{2k}}$. Now, define

$$M = \left\lceil N^{\frac{1}{2k}} \right\rceil = \left\lceil 2^{\frac{n}{k}} \right\rceil,$$

where $\lceil x \rceil$ is the integer greater or equal to x. This yields $2^n \leq M^k \leq 2^{n+1}$. Consider the set

$$\mathcal{P} = \left\{ p = \left\lfloor \frac{\sum_{i=0}^{k} M^i u_i}{a} \right\rfloor + b,\ p \text{ is prime},\ \mid p \in \left[2^n, 2^{n+1}\right],\ a < N^\delta,\ |u_i| < N^\delta \right\},$$

where δ satisfies (2). Here b is as small as possible so that $\left\lfloor \frac{\sum_{i=0}^{k} M^i u_i}{a} \right\rfloor + b$ is prime. Also, since M^k is the leading term, then observe that

$$\frac{\sum_{i=0}^{k} M^i u_i}{a} - M^k = \frac{u_k - a}{a} M^k + \frac{\sum_{i=1}^{k} M^i u_i}{a}.$$

To ensure $p \in \left[2^n, 2^{n+1}\right]$, we consider only the situation where $u_k \geq a$. Hence, using the bounds $a < N^\delta$ and $|u_i| < N^\delta$ for $i = 0, \ldots, k-1$, we get a lower bound for the number of possibilities for a and for u_i, which themselves set a lower bound for the cardinality of $\mathcal{P}$ as follows:

$$\#\mathcal{P} \geq \left\lfloor N^\delta \right\rfloor \left\lfloor N^\delta \right\rfloor^k \approx N^{(k+1)\delta} \approx 2^{2(k+1)n\delta}. \tag{3}$$

On the other hand, the prime number theorem asserts that the number $\pi(x)$ of the primes less than x is

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

Hence, the number of primes in the interval $\left[2^n, 2^{n+1}\right]$ is approximately

$$\pi\left(2^{n+1}\right) - \pi\left(2^n\right) \approx \frac{2^{n+1}}{\log\left(2^{n+1}\right)} - \frac{2^n}{\log\left(2^n\right)} = \frac{(n-1)2^n}{n(n+1)\log(2)}. \tag{4}$$

It follows that the number of RSA moduli $N = pq \in \left[2^{2n}, 2^{2(n+1)}\right]$ with a weak factor $p \in \mathcal{P}$ and $q \in \left[2^n, 2^{n+1}\right]$ is at least $\#(\mathcal{N}) \geq \#\mathcal{P} \times \left(\pi\left(2^{n+1}\right) - \pi\left(2^n\right)\right)$. Using 3 and 4, we get

$$\begin{aligned}
\#(\mathcal{N}) &\geq 2^{2(k+1)n\delta} \times \frac{(n-1)2^n}{n(n+1)\log(2)} \\
&= \frac{(n-1)}{n(n+1)\log(2)} \times 2^{(1+2(k+1)\delta)n} \\
&= 2^\eta,
\end{aligned}$$

where

$$\eta = (1 + 2(k+1)\delta)n + \log_2\left(\frac{(n-1)}{n(n+1)\log(2)}\right).$$

This terminates the proof. □

Table 2 presents a list of values of the bound η in terms of k and n. In Table 2,

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$n = \frac{1}{2}\log_2(N) = 512$	759	715	698	689	684	680	677
$n = \frac{1}{2}\log_2(N) = 1024$	1526	1438	1404	1386	1375	1368	1362
$n = \frac{1}{2}\log_2(N) = 2048$	3061	2885	2818	2782	2759	2744	2733

Table 2. Lower bounds for η under Theorem 5.

we see that in the situation $(\beta, k) = (0.5, 1)$, the number $\#(\mathcal{N})$ of 1024-bits RSA moduli $N = pq \in \left[2^{1024}, 2^{1026}\right]$ with a weak factor p is at least $\#(\mathcal{N}) \geq 2^{759}$. This is much larger than the number of RSA moduli with a weak Coppersmith prime factor in the same interval, which is actually $N^{0.25} \approx 2^{256}$. This remark is also valid for 2048-bits and 4096-bits RSA moduli.

4 The Attack with Two Weak Prime factors

4.1 The Attack

In this section, we present an attack on RSA with a modulus $N = pq$ when both the prime factors p and q are weak primes.

Theorem 6. *Let $N = pq$ be an RSA modulus and M be a positive integer. Let $k \geq 1$. Suppose that there exist integers a, b, u_i and v_i, $i = 1, \ldots, k$ such that $ap = \sum_{i=0}^{k} M^i u_i$ and $bq = \sum_{i=0}^{k} M^i v_i$ with $|u_i|, |v_i| < N^\delta$ and*

$$\delta < \frac{1}{2k+1} + \frac{\log\left(2k^3\right)}{2(2k+1)\log(N)} + \frac{\log(2k+1) - \log(2\pi e)}{4\log(N)} - \frac{\log\left(4k^3\right)}{4\log(N)}.$$

Then one can factor N in polynomial time.

Proof. Suppose that $ap = \sum_{i=0}^{k} M^i u_i$ and $bq = \sum_{i=0}^{k} M^i v_i$. Then multiplying ap and bq, we get

$$abN = \sum_{i=0}^{2k} M^i w_i, \quad \text{with} \quad w_i = \sum_{j=0}^{i} u_j v_{i-j}.$$

This can be transformed into the equation

$$M^{2k} x_{2k} + M^{2k-1} x_{2k-1} + \ldots + M x_1 - yN = -x_0, \tag{5}$$

with the solution $(x_{2k}, x_{2k-1}, \ldots, x_1, y, x_0) = (w_{2k}, w_{2k}, \ldots, w_1, ab, u_0 v_0)$. For $i = 0, \ldots, k$, suppose that $|u_i|, |v_i| < N^\delta$. Since for $i = 0, \ldots, 2k$, the maximal number of terms in w_i is k, we get

$$|x_i| = |w_i| \leq k \max_j(|u_j|) \cdot \max_j(|v_j|) < kN^{2\delta}. \tag{6}$$

Let C be a constant to be fixed later. Consider the lattice $\mathcal{L}$ generated by the row vectors of the matrix

$$M(\mathcal{L}) = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & CM^{2k} \\ 0 & 1 & \ldots & 0 & 0 & CM^{2k-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & CM \\ 0 & 0 & 0 & \ldots & 0 & -CN \end{bmatrix}. \tag{7}$$

The dimension of the lattice $\mathcal{L}$ is $\dim(\mathcal{L}) = 2k+1$ and its determinant is $\det(\mathcal{L}) = CN$. According to the Gaussian Heuristic, the length of the shortest non-zero vector of the lattice $\mathcal{L}$ is approximately $\sigma(\mathcal{L})$ with

$$\sigma(\mathcal{L}) \approx \sqrt{\frac{\dim(\mathcal{L})}{2\pi e}} \det(\mathcal{L})^{\frac{1}{\dim(\mathcal{L})}} = \sqrt{\frac{2k+1}{2\pi e}} (CN)^{\frac{1}{2k+1}}.$$

Consider the vector $v = (x_{2k}, x_{2k-1}, \ldots, x_1, -Cx_0)$. Then, using (5), we get

$$(x_{2k}, x_{2k-1}, \ldots, x_1, -Cx_0) = (x_{2k}, x_{k-1}, \ldots, x_1, y) \cdot M(\mathcal{L}).$$

This means that $v \in \mathcal{L}$. Consequently, if C satisfies $\|v\| \leq \sigma(\mathcal{L})$, then, by the Gaussian Heuristic, v is the shortest vector of L. Using the bound (6), the length of the vector v satisfies

$$\|v\|^2 = C^2 x_0^2 + \sum_{i=1}^{2k} x_i^2 \leq \left(C^2 + \sum_{i=1}^{2k} k^2 \right) N^{4\delta} = \left(C^2 + 2k^3 \right) N^{4\delta}.$$

Let C be a positive integer satisfying $C \leq \sqrt{2k^3}$. Then the norm of the vector v satisfies $\|v\|^2 < 4k^3 N^{4\delta}$. Hence, using the Gaussian approximation $\sigma(\mathcal{L})$, the inequality $\|v\| \leq \sigma(\mathcal{L})$ is satisfied if

$$2k^{\frac{3}{2}} N^{2\delta} \leq \sqrt{\frac{2k+1}{2\pi e}} \left(2^{\frac{1}{2}} k^{\frac{3}{2}} N \right)^{\frac{1}{2k+1}}.$$

Solving for δ, we get

$$\delta < \frac{1}{2k+1} + \frac{\log\left(2k^3\right)}{2(2k+1)\log(N)} + \frac{\log(2k+1) - \log(2\pi e)}{4\log(N)} - \frac{\log\left(4k^3\right)}{4\log(N)}.$$

If δ satisfies the former bound, then the LLL algorithm, applied to the lattice $\mathcal{L}$ will output the vector $v = (x_{2k}, x_{2k-1}, \ldots, x_1, -Cx_0)$ from which, we deduce

$$w_{2k} = |x_{2k}|,\ w_{2k-1} = |x_{2k-1}|, \ldots, w_1 = |x_1|,\ w_0 = \frac{|-Cx_0|}{C}.$$

Using the coefficients w_i, $i = 1, \ldots, 2k$, we construct the polynomial $P(X) = w_{2k}X^{2k} + w_{2k-1}X^{2k-1} + \ldots + w_1 X + w_0$. Factoring $P(X)$, we get

$$P(X) = \left(\sum_{i=0}^{k} M^i u_i \right) \left(\sum_{i=0}^{k} M^i v_i \right),$$

from which we deduce all the values u_i and v_i for $i = 1, \ldots, k$. Using each u_i and v_i for $i = 1, \ldots, k$, we get $ap = \sum_{i=0}^{k} M^i u_i$ and finally obtain $p = \gcd\left(\sum_{i=0}^{k} M^i u_i, N \right)$ which in turn gives $q = \frac{N}{q}$. This terminates the proof. □

In Table 3, we give the bound for δ for a given k and a given size of the RSA modulus.

4.2 Examples

Example 4. Consider the 234 bits RSA modulus

$N = 181287275221777294353476345871682929689873183168124359321741177743400 29.$

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\log_2(N)=1024$	0.332	0.199	0.141	0.109	0.089
$\log_2(N)=2048$	0.333	0.199	0.142	0.110	0.090

Table 3. Upper bounds for δ with Theorem 6.

Let $M = 2^{50}$. Suppose further that the prime factors p and q are such that $ap = M^2u_2 + Mu_1 + u_0$ and $bq = M^2v_2 + Mv_1 + v_0$, that is $k = 2$ with the notation of Theorem 6. We built the matrix (7) with $C = \sqrt{2k^3} = 4$ and applied the LLL algorithm [13]. We got a new basis, where the last row is:

$(w_4, w_3, w_2, w_1, -Cw_0) =$
$(30223231819936, 68646317659290, 109044283791446, 80821741694637, -162291153390444).$

From this, we form the polynomial $P(X) = w_4X^4+w_3X^3+w_2X^2+w_1X^1+w_0$. which factors as:

$$P(X) = \left(4678994X^2 + 5832048X + 4871673\right)\left(6459344X^2 + 6620037X + 8328307\right).$$

From this, we deduce

$$u_2 = 4678994,\ u_1 = 5832048,\ u_0 = 4871673,$$
$$v_2 = 6459344,\ v_1 = 6620037,\ v_0 = 8328307.$$

Using these values, we compute

$$ap = M^2u_2 + Mu_1 + u_0 = 59313295525642905665289652194515573 69,$$
$$bq = M^2v_2 + Mv_1 + v_0 = 81881912986806196686803624641586187 39.$$

and obtain

$$p = \gcd(ap, N) = 126198501118389160989977983392586327,$$
$$q = \gcd(bq, N) = 143652478924221397696146709897519627.$$

This leads to the factorization of $N = pq$. We note that the first attack described in Section 3 does not succeed to factor N. Indeed, we have $\frac{\log(\max_i(|v_i|))}{\log N} \approx 0.098$ which is larger than the value $\delta = 0.069$ for $k = 2$ and $\beta = 0.5$ in Table 1. Finally, the overall recorded execution time for our attack using an off-the-shelf computer was 12 seconds.

4.3 The Number of Double Weak Primes in an Interval

In this section, we consider two positive integers n and M and present a study of the double weak primes with M, that is the primes $p, q \in \left[2^n, 2^{n+1}\right]$ such that there exists positive integer a and b that give the decompositions:

$$ap = \sum_{i=0}^{k} M^i u_i, \quad bq = \sum_{i=0}^{k} M^i v_i$$

where $|u_i| < N^{\delta}$, $|v_i| < N^{\delta}$ and δ satisfies Theorem 6. We show that the number of the RSA moduli N in the interval $[2^{2n}, 2^{2(n+1)}]$ with a weak prime factors $p, q \in \left[2^n, 2^{n+1}\right]$ is lower bounded by 2^{η_2} where $\eta_2 > \frac{1}{2}$.

Theorem 7. *Let n be a positive integer. For $k \geq 1$, define $M = \left\lceil 2^{\frac{n}{k}} \right\rceil$. Let $\mathcal{N}$ be the set of the weak RSA moduli $N \in \left[2^{2n}, 2^{2(n+1)}\right]$ such that $N = pq$ with $p = \left\lfloor \frac{\sum_{i=0}^{k} M^i u_i}{a} \right\rfloor + u$, $q = \left\lfloor \frac{\sum_{i=0}^{k} M^i v_i}{b} \right\rfloor + v$, $p, q \in \left[2^n, 2^{n+1}\right]$ for some small integers u, v, $a < N^{\delta}$, $b < N^{\delta}$, $|u_i| < N^{\delta}$ and $|v_i| < N^{\delta}$ for $i = 0, \ldots, k$ with*

$$\delta = \frac{1}{k+1}\left(1 - \left(\frac{1}{2}\right)^{\frac{k+1}{k}}\right) - \frac{1}{2}\left(1 - \left(\frac{1}{2}\right)^{\frac{1}{k}}\right).$$

Then the cardinality of $\mathcal{N}$ is at least $\#\mathcal{N} \geq 2^{\eta_2}$ where $\eta_2 = 4(k+1)n\delta$.

Proof. As in the proof of Theorem 5, the number of prime numbers $p \in \left[2^n, 2^{n+1}\right]$ such that $p = \frac{\sum_{i=0}^{k} M^i u_i}{a} + u$ with $|u_i| < 2^{2n\delta}$ is

$$\#\mathcal{P} \geq 2^{2(k+1)n\delta}.$$

Then, the number $\mathcal{N}_2$ of RSA modulus $N \in \left[2^{2n}, 2^{2(n+1)}\right]$ with $N = pq$, where both p and q are weak primes is at least

$$\#\mathcal{N}_2 \geq 2^{4(k+1)n\delta} = 2^{\eta_2},$$

where $\eta_2 = 4(k+1)n\delta$. This terminates the proof. □

In Table 3, we present a list of values of the bound η_2 in terms of k and n.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$n=512$	512	424	390	372	361	353	348
$n=1024$	1024	848	780	744	722	707	696
$n=2048$	2048	1696	1560	1489	1444	1414	1392

Table 4. Lower bounds for η_2 under Theorem 7.

5 Conclusions

In this paper we presented and illustrated two attacks based on factoring RSA moduli with weak primes. We further computed lower bounds for the sets of weak moduli -that is, moduli made of at least one or two weak prime respectively- in the interval $[2^{2n}, 2^{2(n+1)}]$ and showed that these sets are much larger than the set of RSA prime factors satisfying Coppersmith's conditions, which effectively extending the likelihood for factoring RSA moduli.

References

1. ANSI Standard X9.31-1998, Digital Signatures Using Reversible Public Key Cryptography for the Financial Services Industry (rDSA).
2. Bernstein, D.J., Chang, Y.A., Cheng, C.M., Chou, L.P., Heninger, N., Lange, T., van Someren, N.: Factoring RSA keys from certified smart cards: Coppersmith in the wild. In Advances in Cryptology-ASIACRYPT 2013. Springer, 2013, pp. 341–360 (2013)
3. Boneh, D.: Twenty years of attacks on the RSA cryptosystem, Notices Amer. Math. Soc. 46 (2), pp. 203–213, (1999)
4. Boneh, D., Durfee, G.: Cryptanalysis of RSA with private key d less than $N^{0.292}$, Advances in Cryptology-Eurocrypt'99, Lecture Notes in Computer Science Vol. 1592, Springer-Verlag, pp. 1–11 (1999)
5. Compaq Computer Corperation. Cryptography using Compaq multiprime technology in a parallel processing environment, 2002. Availabe online at `ftp://ftp.compaq.com/pub/solutions/CompaqMultiPrimeWP.pdf`
6. Coppersmith, D.: Small solutions to polynomial equations, and low exponent RSA vulnerabilities. Journal of Cryptology, 10(4), pp. 233–260 (1997)
7. Hardy, G.H., Wright, E.M.: An Introduction to the Theory of Numbers. Oxford University Press, London (1975)
8. Hastad, J.: On Using RSA with Low Exponent in a Public Key Network, in Proceedings of CRYPTO'85, Springer-Verlag, pp. 403–408 (1986)
9. Hastad, J.: Solving simultaneous modular equations of low degree, SIAM J. of Computing, Vol. 17, pp. 336–341 (1988)
10. Herrmann, M., May, A.: Solving linear equations modulo divisors: On factoring given any bits. In Advances in Cryptology-ASIACRYPT 2008. Springer, 2008, pp. 406–424 (2008)
11. Lenstra, H.: Factoring integers with elliptic curves, Annals of Mathematics, Vol. 126, pp. 649–673 (1987)
12. Lenstra, A.K., Lenstra, H.W. Jr. (eds.): The Development of the Number Field Sieve, Lecture Notes in Mathematics, vol. 1554, Berlin, Springer-Verlag, (1993)
13. Lenstra, A.K., Lenstra, H.W., Lovász, L.: Factoring polynomials with rational coefficients, Mathematische Annalen, Vol. 261, pp. 513–534, (1982)
14. Y. Lu, Y., Zhang, R., Lin, D.: New Results on Solving Linear Equations Modulo Unknown Divisors and its Applications, Cryptology ePrint Archive, Report 2014/343, 2014 `https://eprint.iacr.org/2014/343`.
15. May, A.: New RSA Vulnerabilities Using Lattice Reduction Methods. PhD thesis, University of Paderborn (2003)
16. Nitaj, A.: Another generalization of Wieners attack on RSA, In: Vaudenay, S. (ed.) Africacrypt 2008. LNCS, vol. 5023, pp. 174–190. Springer, Heidelberg (2008)
17. Rivest, R., Shamir, A., Adleman, L.: A Method for Obtaining digital signatures and public-key cryptosystems, Communications of the ACM, Vol. 21 (2), pp. 120–126 (1978)
18. Zimmermann, P.: 50 largest factors found by ECM, `http://www.loria.fr/~zimmerma/records/top50.html`
19. de Weger, B.: Cryptanalysis of RSA with small prime difference, Applicable Algebra in Engineering, Communication and Computing,Vol. 13(1), pp. 17–28 (2002)
20. Wiener, M.: Cryptanalysis of short RSA secret exponents, IEEE Transactions on Information Theory, Vol. 36, pp. 553–558 (1990)