Estimating the Prime-Factors of an RSA Modulus and an Extension of the Wiener Attack

Hung-Min Sun, Mu-En Wu, and Yao-Hsin Chen

Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan 30013
hmsun@cs.nthu.edu.tw, {mn,saint_chen}@is.cs.nthu.edu.tw

Abstract. In the RSA system, balanced modulus N denotes a product of two large prime numbers p and q, where $q < p < 2q$. Since Integer-Factorization is difficult, p and q are simply estimated as $\sqrt{N}$. In the Wiener attack, $2\sqrt{N}$ is adopted to be the estimation of $p+q$ in order to raise the security boundary of private-exponent d. This work proposes a novel approach, called EPF, to determine the appropriate prime-factors of N. The estimated values are called "EPFs of N", and are denoted as p_E and q_E. Thus p_E and q_E can be adopted to estimate $p+q$ more accurately than by simply adopting $2\sqrt{N}$. In addition, we show that the Verheul and Tilborg's extension of the Wiener attack can be considered to be brute-guessing for the MSBs of $p+q$. Comparing with their work, EPF can extend the Wiener attack to reduce the cost of exhaustive-searching for $2r+8$ bits down to $2r-10$ bits, where r depends on N and the private key d. The security boundary of private-exponent d can be raised 9 bits again over Verheul and Tilborg's result.

Keywords: RSA, continued fraction, the Wiener attack, exhaustive-searching, most significant bit.

1 Introduction

RSA [7] has been conventionally adopted cryptosystem since 1978. The advantage of using RSA is that its security is based on the difficulty of Integer-Factorization. A 1024-bit RSA modulus N, which is a product of two 512-bit prime numbers, (*i.e.*, $N = pq$), is adopted to make the factoring infeasible. However, this system is inefficient for digital signature signing and verifying. Many practical issues have been considered in implementing issues, such as reducing the verifying and signing time [3], [8], [9]. In the real world, powerful computers such as servers are frequently employed to execute the verifying task. Lightweight devices with weak computational power, *e.g.*, smart card, wireless sensors or IC card, are employed to execute the signing task. Therefore, most research is focused on reducing the signing time rather than verifying time.

Since the complexity of signing depends on the bit-length of private-exponent, the most popular method to reduce the signing time is to apply a small private-exponent d. To achieve this purpose, a small private-exponent d is first chosen

J. Katz and M. Yung (Eds.): ACNS 2007, LNCS 4521, pp. 116–128, 2007.

in the RSA key generation algorithm, and the corresponding public-exponent e satisfying $ed \equiv 1 \ (mod \phi(N))$ is then calculated. This RSA variant is called RSA-Small-d. However, RSA-Small-d also causes security problems [1], [4], [5], [11], [12]. Indeed, instances of RSA with $d < N^{1/4}$, can be efficiently broken by Wiener's continued fraction attack, which is called the Wiener attack [11]. Boneh and Durfee's lattice-based attack [2], which was proposed in 1998, indicates that the instance of RSA with $d < N^{0.292}$ should be considered as unsafe system. Although their attack is heuristic, it can work very well.

Verheul and Tilborg [10] proposed a technique to extend the Wiener attack in 1997. Their technique costs an exhaustive-searching for $2r + 8$ bits, where $r = \log_2 d - \log_2 N^{1/4}$ to raise r bits over the security boundary of the Wiener attack. Assume that brute-searching for 56 bits is feasible in terms of current computational ability. Solving r for the equation: $2r + 8 = 56$ yields $r = 24$. Therefore, the boundary of the Wiener attack can be raised 24 bits by Verheul and Tilborg's extension.

This work indicates that Verheul and Tilborg's extension can be considered as brute-guessing for the most significant bits (MSBs) of $p + q$, thus providing a motivation to study how to find out the MSBs of $p + q$ as many as possible. Consequently, this work develops an approach to estimate the appropriate prime-factors of N. Assume that the estimated prime-factors are termed p_E and q_E respectively. These terms p_E and q_E are called the "EPFs of N", where EPF is short for "Estimated Prime-Factor". Using EPF, $p + q$ can be estimated more accurately than simply adopting $2\sqrt{N}$ as the estimated value.

Given a 1024-bit RSA modulus N, which is a product of two 512-bit prime numbers p and q, the values $p_E + q_E$ and $p + q$ generally match 10 to 12 MSBs. Therefore, if EPF is adopted to extend the Wiener attack, then the cost of exhaustive-searching for $2r + 8$ bits is reduced to that of exhaustive-searching for $2r - 10$ bits. Consequently, the security boundary of private-exponent d can be raised 9 bits again over that of Verheul and Tilborg's extension.

The remainder of this paper is organized as follows: Section 2 briefly reviews some basic results used in the paper, including continued fraction, the Wiener attack, and Verheul & Tilborg's extension. Section 3 then proposes the approach of EPF and shows the experiment results. Next, Section 4 gives another look on Verheul and Tilborg's extension and applies EPF to improve its performance. Conclusions are finally drawn in Section 5, along with recommendations for future work.

1.1 Our Contribution

The contributions of this work are listed in the following:

(1) A novel approach, called EPF, is provided to evaluate the appropriate prime-factors of N.
(2) Verheul and Tilborg's extension of the Wiener attack is considered as brute-guessing for the MSBs of $p + q$.
(3) Combine the results of (1) and (2), the exhaustive-searching for the extension of the Wiener attack can be reduced from $2r + 8$ bits to $2r - 10$ bits.

2 Preliminary

Some background knowledge is reviewed in this section, including continued fractions, the Wiener attack, and the Verheul and Tilborg's extension.

2.1 Continued Fractions

First we give the definition of continued fractions and some related theorems. The details can be referenced in [6].

Definition 1. *For any poistive real number ξ_0, define $a_i = \lfloor \xi_i \rfloor$, $\xi_{i+1} = 1/(\xi_i - a_i)$ for $i = 0, 1, 2, ..., n$, until ξ_n is an integer. Then ξ_0 can be expanded into the following form:*

$$\xi_0 = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + ... + \cfrac{1}{a_n}}} \tag{1}$$

(1) is called the continued fraction expression of ξ_0. For simplicity, we write (1) to be $\xi_0 = (a_0, a_1, a_2, ..., a_n)$. Besides, $(a_0, a_1, ..., a_i)$ is denoted as the i'th convergent of the continued fraction expansion of ξ_0.

Theorem 2. *If ξ_0 is a rational number, then the process of calculating continued fraction expression would be finished in some finite index n. Otherwise, if ξ_0 is an irrational number, the process would not stop and n is approaching to infinite.*

Theorem 3. *For any positive real number ξ_0, suppose $\frac{h_n}{k_n}$ is the i'th convergent of the continued fraction expression of ξ_0. Define $h_{-2} = 0$, $h_{-1} = 1$; $k_{-2} = 1$, $k_{-1} = 0$, then $h_i = a_i h_{i-1} + h_{i-2}$ and $k_i = a_i k_{i-1} + k_{i-2}$ for $i \geqq 0$.*

Theorem 4. *The convergents $\frac{h_n}{k_n}$ are successively close to ξ_0, that is*

$$\left| \xi_0 - \frac{h_n}{k_n} \right| < \left| \xi_0 - \frac{h_{n+1}}{k_{n+1}} \right| .$$

Furthermore, if ξ_0 is an irrational number, then $\lim_{n \to \infty} \frac{h_i}{k_i} = \xi_0$.

Theorem 5. *Let ξ_0 denote any real number. If there is a rational number $\frac{a}{b}$ with $1 \leq b$ satisfying*

$$\left| \xi_0 - \frac{a}{b} \right| < \frac{1}{2b^2} ,$$

then $\frac{a}{b}$ is one of the convergents of the continued fraction expression of ξ_0.

2.2 The Wiener Attack

Wiener [11] first applied the technique of continued fraction to attack RSA-Small-d. He observed that RSA equation $ed = k\varphi(N) + 1$ can be rewritten as the following form:

$$\left|\frac{e}{\varphi(N)} - \frac{k}{d}\right| = \left|\frac{1}{d\varphi(N)}\right|. \tag{2}$$

Replacing $\frac{e}{\varphi(N)}$ in (2) by $\frac{e}{N}$ yields

$$\left|\frac{e}{N} - \frac{k}{d}\right| < \frac{1}{2d^2}. \tag{3}$$

According to Theorem 5, if (3) is hold, then $\frac{k}{d}$ equals one of the convergents of the continued fraction expression of $\frac{e}{N}$. Since $\gcd(k, d) = 1$, the values of d and k can be extract out actually. Since $N^{1/2} \approx p \approx q$ and $d \approx k$, the left side of (3) reduces to

$$\left|\frac{e}{N} - \frac{k}{d}\right| = \frac{Nk - ed}{Nd} = \frac{k\,(p + q - 1) - 1}{Nd} \approx \frac{2}{N^{1/2}}. \tag{4}$$

In order to apply Theorem 5 again, we have to set

$$\frac{2}{N^{1/2}} < \frac{1}{2d^2}\ ,$$

which leads to:

$$d < \tfrac{1}{2}N^{1/4}. \tag{5}$$

After ignoring the small constatnt $\frac{1}{2}$ in (5), the Wiener attack shows that RSA is insecure when the private-exponent d is smaller than $N^{1/4}$. For instance of 1024-bit RSA modulus, d should be chosen larger than 256 bits.

2.3 Verheul and Tilborg's Extension of the Wiener Attack

While considering the private-exponent d which is slightly larger than $N^{1/4}$, the Wiener attack would be failed. Hence, in order to avoid this situation, Verheul and Tilborg [10] propose a technique to raise the security boundary of $N^{1/4}$ with exhaustive-searching for $2r + 8$ bits, where $r = \log_2 d - \log_2 N^{1/4}$. They consider the following identity:

$$\frac{k}{d} = \frac{p_{j+1}U + (U\Delta + V)\,p_j}{q_{j+1}U + (U\Delta + V)\,q_j}, \tag{6}$$

where $\frac{p_i}{q_i}$ is the i'th convergent of the continued fraction of $\frac{e}{N}$. "U" and "V" are unknown numbers with upper bound: $\log_2 U \le r + 4$ and $\log_2 V \le r + 4$ respectively. The item "Δ" is a small number, *e.g.*, 1 or 2., thus we omit its uncertainty. Consequently, the uncertainty of $\frac{k}{d}$ in (6) is about $2r + 8$ bits, which

means we need to do an exhaustive-searching for about $2r + 8$ bits to extract out the correct value of $\frac{k}{d}$.

Assume that brute-guessing a number with quantity 2^{56} is feasible in terms of current computational ability. Solving r for $2r + 8 = 56$ yields the boundary of Verheul and Tilborg's result. That is, Verheul and Tilborg's extension can further extend the security boundary of d up to 24 bits over Wiener's result. Thus, the instance of RSA with $d < N^{1/4}2^{24}$ can be totally broken by the technique of continued fraction. In this paper, we show Verheul and Tilborg's extension can be regarded as brute-guessing the MSBs of $p+q$. Furthermore, we reduce the cost of original exhaustive-searching for $2r + 8$ bits to $2r - 10$ bits, where $r = \log_2 d - \log_2 N^{1/4}$.

3 The Proposed Approach (EPF) to Estimate the Prime-Factors of $N = pq$

In this section, a novel approach, called EPF, to estimate the prime-factors of N $(= pq)$ is proposed. The point is to find out two numbers, p_E and q_E, by imitating the properties of p and q as similar as possible. The properties includes the bit-length, the most significant bits, and the product of p_E and q_E. Also, we name p_E and q_E "EPFs of N", where EPFs is short for "Estimated Prime-Factors".

First, we focus on how to estimate $p + q$. Note that the hardness of finding $p + q$ is the same as the hardness of finding p and q due to the formula:

$$(p - q)^2 = (p + q)^2 - 4N. \tag{7}$$

Thus solving p and q is obvious by computing $p - q$ with the formula (7).

3.1 How to Estimate $p + q$?

Suppose $N = pq$, where p and q are two large prime-numbers with the same bit-length. Without loss of generality, we assume that $q < p < 2q$. Define D_p to be the difference of $\sqrt{N}$ and p. Similarly, define D_q to be the difference of q and $\sqrt{N}$. That is,

$$p = \sqrt{N} + D_p \text{ and } q = \sqrt{N} - D_q \tag{8}$$

Applying (8) to $N = pq$ we have

$$N = pq = (\sqrt{N} + D_p)(\sqrt{N} - D_q) = N + \sqrt{N}(D_p - D_q) - D_pD_q \tag{9}$$

After simplifying (9) yields

$$D_pD_q = \sqrt{N}(D_p - D_q) \tag{10}$$

Dividing by $\sqrt{N}D_pD_q$ in both sides of (10) leads to

$$\frac{1}{\sqrt{N}} = \frac{D_p - D_q}{D_pD_q}. \tag{11}$$

To estimate the appropriate quantities of D_p and D_q, we compute the i'th convergent of $1/\sqrt{N}$, denoted h_i/k_i, in the continued fraction expression. Hence, according to Theorem 4, $\{h_i/k_i\}_i$ is a rational sequence such that

$$\frac{h_i}{k_i} \to \frac{1}{\sqrt{N}} = \frac{D_p - D_q}{D_p D_q}, \text{ as } i \to \infty. \tag{12}$$

Since $\sqrt{N}$ must be an irrational number, or we can factor N immediately, the three values $1/\sqrt{N}$, $D_p - D_q$ and $D_p D_q$ in (12) are irrational numbers as well. Due to the reason of inconvenience for operations on irrational numbers, we consider the rational number $\frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lfloor D_p D_q \rfloor}$, which is close to $\frac{D_p - D_q}{D_p D_q}$. The integer parts of $D_p - D_q$ and $D_p D_q$ are almost the same as the integer parts of $\lceil D_p \rceil - \lfloor D_q \rfloor$ and $\lfloor D_p D_q \rfloor$ respectively. Setting $p = \left\lfloor \sqrt{N} \right\rfloor + \lceil D_p \rceil$ and $q = \left\lfloor \sqrt{N} \right\rfloor - \lfloor D_q \rfloor$, we have

$$p + q = 2 \left\lfloor \sqrt{N} \right\rfloor + \lceil D_p \rceil - \lfloor D_q \rfloor . \tag{13}$$

According to (13), we know that the information of $\lceil D_p \rceil - \lfloor D_q \rfloor$ is still uesful for us to estimate $p + q$.

Next, Theorem 6 shows that $\frac{h_n}{k_n}$ and $\frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lfloor D_p D_q \rfloor}$ are quite near. This implies that adopting $\frac{h_n}{k_n}$ to be the estimation of $\frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lfloor D_p D_q \rfloor}$ is reasonable. Moreover, we give the upper bound of the difference between $\frac{h_n}{k_n}$ and $\frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lfloor D_p D_q \rfloor}$.

Theorem 6. *If $k_n < D_p D_q$, we have*

$$\left| \frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lfloor D_p D_q \rfloor} - \frac{h_n}{k_n} \right| < \frac{3}{\lfloor D_p D_q \rfloor}. \tag{14}$$

Proof. Since $k_n < D_p D_q$ and $k_n < k_n k_{n+1}$, we have

$$\frac{1}{k_n k_{n+1}} < \frac{1}{k_n} < \frac{1}{D_p D_q} < \frac{1}{\lfloor D_p D_q \rfloor}.$$

Now we prove (14) by triangle inequality:

$$\left| \frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lfloor D_p D_q \rfloor} - \frac{h_n}{k_n} \right| = \left| \frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lfloor D_p D_q \rfloor} - \frac{D_p - D_q}{D_p D_q} + \frac{D_p - D_q}{D_p D_q} - \frac{h_n}{k_n} \right|$$

$$\leqq \left| \frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lfloor D_p D_q \rfloor} - \frac{D_p - D_q}{D_p D_q} \right| + \left| \frac{D_p - D_q}{D_p D_q} - \frac{h_n}{k_n} \right| < \frac{2}{\lfloor D_p D_q \rfloor} + \frac{1}{k_n k_{n+1}} < \frac{3}{\lfloor D_p D_q \rfloor}.$$

Done. ■

Since $\lfloor D_p D_q \rfloor$ is much larger than 3, $\frac{3}{\lfloor D_p D_q \rfloor}$ is close to 0. Consequently, the value of $\frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lfloor D_p D_q \rfloor}$ is almost the same as the vaule of $\frac{h_n}{k_n}$. Also, if the bit-length of h_n is equal to or slightly smaller than the bit-length of $\lceil D_p \rceil - \lfloor D_q \rfloor$, h_n may be considered as the estimation of $\lceil D_p \rceil - \lfloor D_q \rfloor$ reasonably. Hence, to select the suitable index n, we apply the following rule:

$$h_n < \lceil D_p \rceil - \lfloor D_q \rfloor < h_{n+1}$$

Notice that h_n is smaller than h_{n+1} due to Theorem 3. An apparent question is that how to choose the right value of n without the information of $\lceil D_p \rceil - \lfloor D_q \rfloor$. To solve this problem, in fact, we choose several index as candidates according to the statistical result. The experiment shows the average of n is 299, with the standard deviation 12. Therefore, while searching the right value of n for each modulus N, it may increase a little complexity. Here we simply estimate the appropriate quantity which is slightly smaller than the vaule of $\lceil D_p \rceil - \lfloor D_q \rfloor$. Thus we choose h_n as estimated value rather than h_{n+1}. However, it has no theory to justify the difference of bit-lengths of h_n and $\lceil D_p \rceil - \lfloor D_q \rfloor$. Thus we show that the bit-length of h_n is actually slightly smaller than the bit-length of $\lceil D_p \rceil - \lfloor D_q \rfloor$ by implementing experiments. Table 1 gives the results for 1024-bit and 2048-bit RSA modulus respectively. We take 100 instances for each case and compute the average bit-length and its standard deviation. According to our experiments, for 1024-bit RSA modulus, h_n is about 502 bits with standard deviation 2.01. As for 2048-bit RSA modulus, h_n is about 1011 bits with standard deviation 4.42.

Table 1. The Bit-lengths of Estimated and Real Values

Modulus N	h_n	$\lceil D_p \rceil - \lfloor D_q \rfloor$	h_{n+1}
1024 bits	502 bits	503 bits	505 bits
Standard deviation	2.01	1.43	2.10
2048 bits	1011 bits	1012 bits	1013 bits
Standard deviation	4.42	4.19	4.40

3.2 Estimated Prime-Factors of N (EPFs of N)

Here we show how to estimate the prime-factors of N, where $N = pq$. The estimated prime-factors are denoted as p_E and q_E, which are called "EPFs of N". Since $\lfloor D_p D_q \rfloor \approx \lceil D_p \rceil \cdot \lfloor D_q \rfloor$, k_n can be regarded as the estimation of $\lceil D_p \rceil \cdot \lfloor D_q \rfloor$. In other words, we have three fractions,

$$\frac{h_n}{k_n} \approx \frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lfloor D_p D_q \rfloor} \approx \frac{\lceil D_p \rceil - \lfloor D_q \rfloor}{\lceil D_p \rceil \cdot \lfloor D_q \rfloor} ,$$

which are all close to each other. Besides, the bit-lengths of their numerator and denominator are almost the same. Hence, h_n and k_n can be regarded as the estimations of $\lceil D_p \rceil - \lfloor D_q \rfloor$ and $\lceil D_p \rceil \cdot \lfloor D_q \rfloor$, that is,

$$h_n \approx \lceil D_p \rceil - \lfloor D_q \rfloor \text{ and } k_n \approx \lceil D_p \rceil \cdot \lfloor D_q \rfloor$$

Computing $\lceil D_p \rceil + \lfloor D_q \rfloor$ by the formula:

$$\begin{aligned} (\lceil D_p \rceil + \lfloor D_q \rfloor)^2 &= (\lceil D_p \rceil - \lfloor D_q \rfloor)^2 + 4 \lceil D_p \rceil \cdot \lfloor D_q \rfloor \\ &= h_n^2 + 4k_n \end{aligned} \tag{15}$$

Solving $\lceil D_p \rceil$ and $\lfloor D_q \rfloor$ from (15) we get

$$\lceil D_p \rceil = \left\lceil \frac{\sqrt{h_n^2+4k_n}+h_n}{2} \right\rceil \text{ and } \lfloor D_q \rfloor = \left\lfloor \frac{\sqrt{h_n^2+4k_n}-h_n}{2} \right\rfloor$$

Therefore, the EPFs of N are set to

$$p_E = \left\lfloor \sqrt{N} \right\rfloor + \left\lceil \frac{\sqrt{h_n^2+4k_n}+h_n}{2} \right\rceil \text{ and } q_E = \left\lfloor \sqrt{N} \right\rfloor - \left\lfloor \frac{\sqrt{h_n^2+4k_n}-h_n}{2} \right\rfloor.$$

3.3 The Accuracy of EPFs of N

Here we show the accuracy of EPFs of N by experiments. The statistical data in Table 2 shows the difference of EPFs and practical prime-factors, *i,e,*, $p - p_E$, $q_E - q$. We also compute the average bit-length of $N - N_E$, where $N_E = p_E q_E$. The data comes from the average of 100 samples for each case. Note that the p_E and q_E are the same 7 MSBs with p and q respectively for 1024-bit RSA modulus. Also, for the case of 2048-bit RSA modulus, p_E and q_E are the same 9 MSBs with p and q respectively.

Table 2. The Accuracy of EPFs

the bit-length of N	1024 bits	2048
the average bit-length of $p - p_E$	505 bits	1015 bits
standard deviation of $p - p_E$	1.52	2.57
the average bit-length of $q_E - q$	505 bits	1015 bits
standard deviation of $q_E - q$	1.49	2.57
the average bit-length of $N - N_E$	510 bits	1022 bits
standard deviation	1.56	2.06

4 Another Look on Verheul and Tilborg's Extension and Its Improvement

We show that Verheul and Tilborg's extension can be regarded as brute-guessing for the MSBs of $p + q$ in this section. By applying EPF to improve the Wiener attack, the new result raises the security boundary of d again. In the remainder of the paper, we suppose the estimation of $p + q$ is $2A$, *i.e.*, $A \approx \frac{p+q}{2}$. Under such assumption $\phi(N) = (N + 1) - (p + q)$ is estimated as $(N + 1) - 2A$.

4.1 Improvement of the Wiener Attack

Consider the following question:

Question:
When considering the RSA equation: $ed = k(p-1)(q-1)+1$, where $N = pq$, what range of d would satisfy the following inequalities?

$$|\frac{e}{N+1-2A} - \frac{k}{d}| < \frac{1}{2d^2} < |\frac{e}{N} - \frac{k}{d}| \tag{16}$$

The meaning of (16) is shown as follows: In the right side of (16), the inequality means the range of d that the Wiener attack would fail. Instead, in the left side of (16), the inequality means the the Wiener attack can work successfully. The difference between left inequality and right inequality of (16) is whether applying new estimation of $\phi(N)$, *i.e.*, $N+1-2A$, to replace N. Now we simplify (16) in the following:

Consider the right side of (16), since

$$\left|\frac{e}{N} - \frac{k}{d}\right| = \frac{ed - Nk}{Nd} = \frac{k(p+q-1)-1}{Nd} \ ,$$

the right side inequality of (16) is equivalent to $\frac{1}{2d^2} < \frac{k(p+q-1)-1}{Nd}$, that is

$$N < 2dk[(p+q)-1] - 2d. \tag{17}$$

Similarly, in the left side of (16), since

$$\left|\frac{e}{N+1-2A} - \frac{k}{d}\right| = \left|\frac{ed - k(N+1-2A)}{(N+1-2A)d}\right| = \frac{k[(p+q)-2A]-1}{(N+1-2A)d} \ ,$$

the left side inequality of (16) is equivalent to $\frac{k[(p+q)-2A]-1}{(N+1-2A)d} < \frac{1}{2d^2}$, that is

$$2dk[(p+q)-2A] - 2d < N+1-2A. \tag{18}$$

In order to combine (17)and (18), we rearrange (18) in the following form:

$$2dk[(p+q)-1] - 2d < N + (2dk-1)(2A-1) \ , \tag{19}$$

which is the same format of (17). Consequently, after combining (17) and (19), (16) is equivalent to

$$N < 2dk[(p+q)-1] - 2d < N + (2dk-1)(2A-1). \tag{20}$$

Note that if $A = \frac{p+q}{2}$, the right side of (20) changes to

$$2dk\,(p+q-1) - 2d < N + (2dk-1)\,((p+q)-1) = 2dk\,(p+q-1) + \varphi(N) \ ,$$

which is always hold for any size of private-exponent d.
Solving d in the right inequality of (20) we get its upper bound:

$$d < \frac{N+1-2A}{2k\,(p+q-2A)-2}. \tag{21}$$

According to (21), the private-exponent d should not be chosen smaller than $\frac{N+1-2A}{2k(p+q-2A)-2}$ or RSA system can be totally broken immediately. In addition, the closer the distance between $2A$ and $p+q$ is, the larger upper bound the insecure private-exponent is. Therefore, to raise the security boundary of d, we should try to find the estimated value of $p+q$ as appropriate as possible. This conclusion also implies that the complexity of extending the Wiener attack can be considered as the complexity of finding $p+q$.

4.2 Applying EPF to the Proposed Extension of the Wiener Attack

Now we analyze how much complexity could be reduced when we apply EPF to the proposed extension of the Wiener attack. Define a variable "Λ" to denote the difference of $\frac{p+q}{2}$ and A, *i.e.*, $\Lambda = \frac{p+q}{2} - A$. Note that A is the esitmated value of $\frac{p+q}{2}$, thus Λ is represented the uncertainty part of $\frac{p+q}{2}$. Replacing A by $\frac{p+q}{2} - \Lambda$ into the right inequality of (20) yields

$$\begin{aligned} & 2dk(p+q-1) - 2d \\ & < N + (2dk-1)\,(2(\tfrac{p+q}{2} - \Lambda) - 1) \\ & = 2dk(p+q-1) + \varphi(N) - 2\Lambda(2dk-1) \end{aligned} \tag{22}$$

Eliminating $2dk(p+q-1)$ in both sides of (22), we have

$$2\Lambda(2dk-1) - 2d < \varphi(N). \tag{23}$$

According to (22), we have the following conclusion: The parameters Λ, k, and d in (23) play the main role to determine whether the Wiener attack can work or not. Since d and k are pre-determined parameters in the key-generation of RSA, the only variable we could control is the parameter Λ which represents the uncertainty part of $\frac{p+q}{2}$. This implies the more accuracy A is estimated, the smaller quantity of Λ will be. Therefore, to raise the security boundary of private-exponent d, we should focus our effort on finding out the MSBs of $p+q$ as many as possible. In the following, Table 3 gives the experiment results about how many MSBs of $p+q$ that EPF can be found out.

The statistics data in Table 3 comes from the averages of computing 100 instances for $\frac{p+q}{2}$, A and Λ. Note that $2A$ is the estimation of $p+q$. Thus $2A$ is set to be $2\left\lfloor\sqrt{N}\right\rfloor + h_n$ according to EPF.

Table 3. The Difference between Estimated and Real Values

Modulus N	$\frac{p+q}{2}$	$A = \frac{2\lfloor\sqrt{N}\rfloor + h_n}{2}$	$\Lambda = \frac{p+q}{2} - A$
1024 bits	512 bits	512 bits	500 bits
Standard deviation	0	0	1.89
2048 bits	1024 bits	1024 bits	1009 bits
Standard deviation	0	0	4.03

In Table 3, for the case of 1024-bit N, the average bit-length of Λ is 500 bits with the standard deviation 1.89. This implies that $\frac{p+q}{2}$ and A usually match in 10 MSBs at least, where 10 is computed from $512 - \lceil 500 + 1.89 \rceil$. For the case of 2048-bit RSA modulus, the average bit-length of Λ is 1009 bits with the standard deviation 4.03. Thus, $\frac{p+q}{2}$ and A usually also match in 10 MSBs at least, where 10 is computed from $1024 - \lceil 1009 + 4.03 \rceil$.

4.3 Better Result Compared with Verheul and Tilborg's Extension

We compare our improvement with Verheul and Tilborg's result. Consider the case of 1024-bit RSA modulus N, in order to further reduce the quantity of Λ, we do an exhaustive-searching for finding out the s MSBs of Λ and write $\Lambda = (2^{500-s})\Lambda_1 + \Lambda_2$, where $\Lambda_1 \in \left[2^{s-1}, 2^s\right]$, and $\Lambda_2 \in \left[2^{500-s}, 2^{501-s}\right]$. Suppose that Λ_1 can be totally gotten by exhaustive-searching and Λ_2 is still an unknown part. Under such assumption, $\frac{p+q}{2}$ can be estimated as $A + (2^{500-s})\Lambda_1$ more accurately instead of just estimating as A. In addition, the values of $\frac{p+q}{2}$ and $A + (2^{500-s})\Lambda_1$ usually match $12 + s$ MSBs from tha above result. Thus the uncertainty part of $\frac{p+q}{2}$, Λ_2, remains $500 - s$ bits.

Now we analyze how much improvement after adopting brute-guessing for Λ_1. Applying uncertainty part of $\frac{p+q}{2}$, Λ_2, to (23) yields

$$2\Lambda_2(2dk - 1) - 2d < \varphi(N). \tag{24}$$

In order to satisfy (24), we should compute the bit-length of each side. Denote $|d|$ and $|k|$ to be represented the bit-length of d and k respectively, thus the bit-length of $2\Lambda_2(2dk - 1) - 2d$ is

$$1 + (500 - s) + 1 + |d| + |k|$$

which is mainly determined by $2\Lambda_2(2dk-1)$ in (24). Furthermore, to satisfy (24) we have to set

$$1 + (500 - s) + 1 + |d| + |k| < 1024 \tag{25}$$

where 1024 is the bit-length of $\varphi(N)$ in (24). Since the bit-length of d and k are almost the same with high probability in the key-generation algorithm of RSA-Small-d, we can assume $|d| = |k|$. Suppose that $|d| = |k| = 256 + r$. *i.e.*, the private-exponent d exceeds 256 bits ($N^{1/4}$) more r bits. Applying $256 + r$.to (25) we get

$$1 + (500 - s) + (1 + 2(256 + r) < 1024\ ,$$

which is equivalent to

$$2r - 10 < s. \tag{26}$$

By (26), we have a conclusion which is simlilar to Verheul and Tilborg's result: To extend the Wiener's boundary r bits, we only have to do an exhaustive-searching for about $2r - 10$ bits, where $r = \log_2 d - \log_2 N^{1/4}$. Compared with

Verheul and Tilborg's result [10], which costs an exhaustive-searching for $2r+8$ bits, our result is 18 bits fewer than Verheul and Tilborg's. Thus it is more efficient to applying our method on the extension of the Wiener attack.

Suppose that the complexity that the current computer can work with is under $O(2^{56})$. This means brute-searching for any number whose bit-length less than 56 is feasible. Verheul and Tilborg's extension can attack successfully on $d < N^{1/4}2^{24}$, where 24 comes from by solving r for $2r+8 = 56$. With our result in (26): Solving r for $2r - 10 = 56$ yields $r = 33$. Hence, the proposed method can attack successfully on $d < N^{1/4}2^{33}$, which is more 9 bits than Verheul and Tilborg's result.

Table 4 shows the comparisons between the original Wiener attack, Verheul and Tilborg's extension (V-T Extension), and our improvement.

Table 4. The comparison between each attack

	Upper Bound of d	Complexity
The Wiener Attack	$d < N^{1/4}$	Polynomial time
V-T Extension	$d < N^{1/4}2^{24}$	exhaustive-searching for $2r + 8$ bits
Our Improvement	$d < N^{1/4}2^{33}$	exhaustive-searching for $2r - 10$ bits

5 Conclusion

This work presents a novel approach, called EPF, to determine the estimated prime-factors of N through the continued fractions. Experiment results shows that the 12 MSBs of $p + q$ can be estimated correctly for the 1024-bit N. This technique reduces the error between the real and estimated $\varphi(N)$, and raises the security boundary of private-exponent d. Besides, We show a result that Verheul and Tilborg's extension of the Wiener attack can be consider as brute-guessing for MSBs of $p+q$. By applying EPF to the proposed result, the security boundary of d can be raised again. Assuming that exhaustive-searching for 56 bits is feasible, Verheul and Tilborg's extension raises 24 bits over the Wiener's boundary. The proposed method raises 9 bits over the Verheul and Tilborg's boundary. Therefore, the instance of RSA with $d < N^{1/4}2^{33}$ can be totally broken by the technique of the continued fractions.

An open problem has been mentioned many times in the past research. Whether exists a better method to evaluate the estimated value of $\varphi(N)$? The boundary of the Wiener attack can be raised again as the accuracy of the estimate of $\varphi(N)$. In the futrue, we will try to design an efficient and accurate method based on the continued fractions and other mathematic materials.

Acknowledgments. This research was supported in part by the National Science Council, Taiwan, under contract NSC95-2221-E-007-030. We are grateful to anonymous reviewers for their valuable comments.

References

[1] D. Boneh, "Twenty Years Attacks on the RSA Cryptosystem", Notices of the American Mathematical Society, Vol. 46, No. 2, pp. 203-213, 1999.

[2] D. Boneh and G. Durfee, "Cryptanalysis of RSA with private key d less than $N^{0.292}$", *IEEE Trans. on Information Theory*, Vol. 46, No. 4, pp. 1339-1349, 2000.

[3] D. Boneh and H. Shacham, "Fast Variants of RSA", CryptoBytes, Vol. 5, No. 1, Springer, 2002.

[4] G. Durfee, P. Q. Nguyen, "Cryptanalysis of the RSA Schemes with Short Secret Exponent form Asiacrypt '99", Proceedings of Cryptology - ASIACRYPT'00, LNCS 1976, Springer-Verlag, pp.1-11, 2000.

[5] H.-S. Hong, H.-K. Lee, H.-S. Lee and H.-J. Lee, "The better bound of private key in RSA with unbalanced primes", *Applied Mathematics and Computation*, Vol. 139, pp. 351-362, 2003.

[6] I. Niven, H. S. Zuckerman, *An Introduction to the Theory of Number*, John Wiley and Sons Inc,1991.

[7] R. Rivest, A. Shamir and L. Aldeman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems", *Communications of the ACM* , Vol. 21, No.2, pp.120-126, 1978.

[8] H.-M. Sun, W.-C. Yang and C.-S. Laih, "On the Design of RSA with Short secret-exponent", Proceedings of Cryptology - ASIACRYPT'99, LNCS 1716, Springer-Verlag, pp.150-164, 1999.

[9] H.-M. Sun and C.-T. Yang, "RSA with Balanced Short Exponents and Its Application to Entity Authentication", Proceeding of Public Key Cryptography 05 - PKC'05, LNCS 3386, Springer-Verlag, pp.199-215, 2005.

[10] E. Verheul and H. van Tilborg, "Cryptanalysis of less short RSA secret-exponents", *Applicable Algebra in Engineering, Communication and Computing*, Vol. 8, Springer-Verlag, pp. 425-435, 1997.

[11] M. J. Wiener, "Cryptanalysis of RSA with short secret-exponents", *IEEE Trans. on Information Theory*, Vol. 36, pp.553-558, 1990.

[12] B. de Weger, "Cryptanalysis of RSA with small prime difference", *Applicable Algebra in Engineering, Communication and Computing*, Vol. 13, pp. 17-28, 2002.