An Algorithm For Factoring Integers

Yingpu Deng and Yanbin Pan

Key Laboratory of Mathematics Mechanization,
Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing 100190, People's Republic of China
E-mail addresses: {dengyp, panyanbin}@amss.ac.cn

Abstract We propose an algorithm for factoring a composite number. The method seems new.

Keywords Integer factorization, Binomial coefficient, Combinatorial sum, RSA modulus

1. Introduction

Integer factorization and primality testing are two well-known computational problems, and the later had been proven to be an 'easy' problem by Agrawal, Kayal and Saxena [1] in 2004. However, integer factorization is much more hard, there are several modern methods such as continued fraction method, class group method, elliptic curve method, quadratic sieve and number field sieve, etc. For the details of these methods, see [3, 4, 5] and the references therein. The best known method on integer factorization problem is the general number field sieve [5] and its running time is sub-exponential.

Since the invention of the general number field sieve in 1993, there is no substantial progress on this problem. There is no new method to appear for a long time. In this paper, we propose an alternate algorithm for factoring a composite number. Our method seems new to the best of our knowledge and maybe this is a new idea on integer factorization problem, although the asymptotic computational complexity of our method is unknown. We suspect that the tight asymptotic complexity of our method is hard to obtain, and it may be related with some deep unknown mathematical theory.

We implemented our method on a PC using Shoup's NTL library version 5.4.1 [7]. Unfortunately, we have to say that the practical effect of our method is not good on a single PC, it is worse than many known algorithms on this problem. However, we believe that our method yields interesting phenomena, and it is worth for further study.

The paper is organized as follows. We give the necessary mathematical knowledge of our method and we describe the basic form of our algorithm in Section 2. We give some possible variants of our method in Section 3. We consider factorization of RSA moduli using our method in Section 4. We give some partial experimental results about our method in Section 5. Finally, some open problems and a short conclusion are given.

2. Basic Principle

2.1. Using general a with $\gcd(n, a) = 1$

Let $n \in \mathbb{Z}, n > 1$ be a composite. We want to find a non-trivial divisor of n, i.e. a divisor $d \mid n$ with $1 < d < n$. The following is a key observation.

Proposition 2.1. There exists an integer j such that $1 < j < n - 1$ and $1 < \gcd(n, \binom{n}{j}) < n$.

Proof. We distinguish two cases:

Case (i): n has a square divisor.

Then n has a prime divisor p such that $p^k \parallel n$ with $k > 1$. Since

$$\binom{n}{p} = \frac{n(n-1)(n-2)\cdots(n-p+1)}{p!}$$

and $p \nmid (n - i)$ for $1 \leqslant i \leqslant p - 1$, we have $p^{k-1} \parallel \binom{n}{p}$. Hence $1 < \gcd(n, \binom{n}{p}) < n$.

Case (ii): n is square-free.

Then n has two prime divisors p and q with $p < q$. Obviously $q \mid \binom{n}{p}$. Hence $1 < \gcd(n, \binom{n}{p}) < n$. □

By Proposition 2.1, to obtain a non-trivial divisor of n, a natural way is to expanding the polynomial $(X + 1)^n$, then computing the gcd's of the coefficients with n. However, this will take exponential time. We can do similarly as in Agrawal, Kayal and Saxena [1], and this leads to the following definition.

Definition 2.2. Let $a \in \mathbb{Z}$ be an integer with $\gcd(n, a) = 1$. Let r be a positive integer. Suppose

$$(X + a)^n \equiv \sum_{i=0}^{r-1} a_i X^i \mod(X^r - 1, n)$$

with $a_i \in \mathbb{Z}$ and $0 \leqslant a_i \leqslant n - 1$ for $0 \leqslant i \leqslant r - 1$. Here X is an indeterminate over $\mathbb{Z}$. If there is an i such that $0 \leqslant i \leqslant r - 1$ and $\gcd(n, a_i)$ is a non-trivial divisor of n, then we call that r is a factorization-friendly number of n with respect to a.

Proposition 2.3. $n - 1$ is a factorization-friendly number of n with respect to arbitrary a with $\gcd(n, a) = 1$.

Proof. Since

$$(X + a)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} X^i,$$

so

$$(X + a)^n \equiv (na + a^n) + (1 + na^{n-1})X + \sum_{i=2}^{n-2} \binom{n}{i} a^{n-i} X^i \mod(X^{n-1} - 1, n).$$

Now the result follows from Proposition 2.1. □

Definition 2.4. The least factorization-friendly number of n with respect to a is called the factorization number of n with respect to a and is denoted by $\mathrm{FAC}(n, a)$. So by Proposition 2.3, we have $\mathrm{FAC}(n, a) \leqslant n - 1$.

We have done numerous experiments, these experiments show a remarkable fact that the FAC(n, a), even FAC$(n, 1)$, are surprisingly small relative to n. They grow very slowly with n.

Question: How small FAC(n, a) can be for some fixed a (e.g. for $a = 1$ or $a = -1$)? How small Min$_a$FAC(n, a) can be, where a is taken from some specific set?

Definition 2.5. Let m and r be two positive integers. Let i and a be two integers. We define

$$\begin{bmatrix} m \\ i \end{bmatrix}_r^a = \sum_{\substack{0 \leqslant k \leqslant m \\ k \equiv i(\bmod r)}} \binom{m}{k} a^{m-k}.$$

So, in Definition 2.2, we have $a_i \equiv \begin{bmatrix} n \\ i \end{bmatrix}_r^a$ (mod n) for $0 \leqslant i \leqslant r-1$. We denote

$$\begin{bmatrix} m \\ i \end{bmatrix}_r = \begin{bmatrix} m \\ i \end{bmatrix}_r^1 = \sum_{\substack{0 \leqslant k \leqslant m \\ k \equiv i(\bmod r)}} \binom{m}{k}.$$

Lemma 2.6. Let $\zeta \in \mathbb{C}$ be a primitive $r-$th root of unity. Then, for $0 \leqslant i \leqslant r-1$, we have

$$\begin{bmatrix} m \\ i \end{bmatrix}_r^a = \frac{1}{r} \sum_{j=0}^{r-1} (\zeta^j + a)^m (\zeta^j)^{-i}.$$

Proof. Since $(X + a)^m = \sum_{i=0}^m \binom{m}{i} a^{m-i} X^i$, we have

$$\sum_{j=0}^{r-1} (\zeta^j + a)^m (\zeta^j)^{-i} = \sum_{j=0}^{r-1} \left[\sum_{s=0}^{m} \binom{m}{s} \zeta^{js} a^{m-s} \right] (\zeta^j)^{-i}$$

$$= \sum_{s=0}^{m} \binom{m}{s} a^{m-s} \sum_{j=0}^{r-1} \zeta^{(s-i)j} = r \cdot \sum_{\substack{0 \leqslant s \leqslant m \\ s \equiv i(\bmod r)}} \binom{m}{s} a^{m-s}.$$

□

2.2. Using $a = 1$

In this sub-section, we suppose $a = 1$. Let

$$(X + 1)^n \equiv \sum_{i=0}^{r-1} a_i X^i \mod (X^r - 1, n),$$

where $a_i \equiv \begin{bmatrix} n \\ i \end{bmatrix}_r$ (mod n) for $0 \leqslant i \leqslant r-1$.

For $r = 1$, we have $a_0 \equiv 2^n(\bmod n)$. For $r = 2$, we have $a_0 = a_1 \equiv 2^{n-1}(\bmod n)$. Therefore, if n is even and n is not a power of 2, we have $\mathrm{FAC}(n, 1) = 1$; if n is odd or n is a power of 2, we have $\mathrm{FAC}(n, 1) \geqslant 3$.

Proposition 2.7. We have

$$\left[\begin{array}{c} n \\ 0 \end{array}\right]_3 = \frac{1}{3}\left(2^n + 2\cos\left(\frac{n\pi}{3}\right)\right),$$

$$\left[\begin{array}{c} n \\ 1 \end{array}\right]_3 = \frac{1}{3}\left(2^n + 2\cos\left(\frac{(n-2)\pi}{3}\right)\right),$$

$$\left[\begin{array}{c} n \\ 2 \end{array}\right]_3 = \frac{1}{3}\left(2^n + 2\cos\left(\frac{(n+2)\pi}{3}\right)\right).$$

Proof. Set $i = \sqrt{-1}$. By Lemma 2.6, let $r = 3$, and so $\zeta = e^{i\frac{2\pi}{3}} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i, \zeta^2 = -\frac{1}{2} - \frac{\sqrt{3}}{2}i, 1+\zeta = \frac{1}{2} + \frac{\sqrt{3}}{2}i = e^{i\frac{\pi}{3}}, 1+\zeta^2 = \frac{1}{2} - \frac{\sqrt{3}}{2}i = e^{-i\frac{\pi}{3}}, (1+\zeta)^n = e^{i\frac{n\pi}{3}}, (1+\zeta^2)^n = e^{-i\frac{n\pi}{3}}$, thus we have

$$\left[\begin{array}{c} n \\ 0 \end{array}\right]_3 = \frac{1}{3}\sum_{j=0}^{2}(1+\zeta^j)^n = \frac{1}{3}\left(2^n + e^{i\frac{n\pi}{3}} + e^{-i\frac{n\pi}{3}}\right) = \frac{1}{3}\left(2^n + 2\cos\left(\frac{n\pi}{3}\right)\right),$$

$$\left[\begin{array}{c} n \\ 1 \end{array}\right]_3 = \frac{1}{3}\sum_{j=0}^{2}(1+\zeta^j)^n\zeta^{-j} = \frac{1}{3}\left(2^n + 2\cos\left(\frac{(n-2)\pi}{3}\right)\right),$$

$$\left[\begin{array}{c} n \\ 2 \end{array}\right]_3 = \frac{1}{3}\sum_{j=0}^{2}(1+\zeta^j)^n\zeta^{-2j} = \frac{1}{3}\left(2^n + 2\cos\left(\frac{(n+2)\pi}{3}\right)\right).$$

□

Proposition 2.8. If n is even and n is not a power of 2, then $\mathrm{FAC}(n, 1) = 1$; if n is a power of 2, then $\mathrm{FAC}(n, 1) = 3$.

Proof. Suppose n is a power of 2, i.e. $n = 2^m, m > 1$. If m is even, then $n \equiv 1(\bmod 3)$, and $\frac{n+2}{3}$ is even, so $a_2 \equiv \left[\begin{array}{c} n \\ 2 \end{array}\right]_3 \equiv \frac{1}{3}(2^n + 2)(\bmod n)$, thus $\gcd(n, a_2) = 2$. If m is odd, similarly, we have $\gcd(n, a_1) = 2$. □

Of course, we do not need to factorize an even composite number, Proposition 2.8 just illustrates a fact that the $\mathrm{FAC}(n, 1)$ are very small.

Proposition 2.9. Let m, r be two positive integers, and let i be an integer. Then we have

$$\left[\begin{array}{c} m \\ i \end{array}\right]_r = \left[\begin{array}{c} m \\ m-i \end{array}\right]_r.$$

Proof. We have

$$\left[\begin{array}{c} m \\ m-i \end{array}\right]_r = \sum_{\substack{0 \leqslant k \leqslant m \\ k \equiv m-i(\bmod r)}} \binom{m}{k}$$

$$= \sum_{\substack{0 \leqslant m-k \leqslant m \\ m-k \equiv i(\bmod r)}} \binom{m}{m-k} = \left[\begin{array}{c} m \\ i \end{array} \right]_r .$$

□

2.3. Using $a = -1$

In this sub-section, we suppose $a = -1$. We have

$$(X-1)^n \equiv \sum_{i=0}^{r-1} a_i X^i \mod(X^r - 1, n),$$

where

$$a_i \equiv \left[\begin{array}{c} n \\ i \end{array} \right]_r^{-1} \equiv \sum_{\substack{0 \leqslant k \leqslant n \\ k \equiv i(\bmod r)}} \binom{n}{k} (-1)^{n-k} (\bmod n)$$

for $0 \leqslant i \leqslant r-1$.

Obviously, for $r = 1$, we have $a_0 = 0$, so $\mathrm{FAC}(n, -1) \geqslant 2$ for all n. For $r = 2$, we have $a_0 \equiv (-1)^n 2^{n-1} (\bmod n)$ and $a_1 \equiv (-1)^{n+1} 2^{n-1} (\bmod n)$. Hence, if n is even and n is not a power of 2, we have $\mathrm{FAC}(n, -1) = 2$; if n is a power of 2 or n is odd, we have $\mathrm{FAC}(n, -1) \geqslant 3$.

Proposition 2.10. We have

$$\left[\begin{array}{c} n \\ 0 \end{array} \right]_3^{-1} = 3^{\frac{n}{2}-1} \cdot 2 \cos\left(\frac{5n\pi}{6}\right),$$

$$\left[\begin{array}{c} n \\ 1 \end{array} \right]_3^{-1} = 3^{\frac{n}{2}-1} \cdot 2 \cos\left(\frac{(5n-4)\pi}{6}\right),$$

$$\left[\begin{array}{c} n \\ 2 \end{array} \right]_3^{-1} = 3^{\frac{n}{2}-1} \cdot 2 \cos\left(\frac{(5n-8)\pi}{6}\right).$$

Proof. Set $i = \sqrt{-1}$. By Lemma 2.6, let $r = 3$, and so $\zeta = e^{i\frac{2\pi}{3}} = -\frac{1}{2} + \frac{\sqrt{3}}{2} i, \zeta^2 = -\frac{1}{2} - \frac{\sqrt{3}}{2} i, \zeta - 1 = -\frac{3}{2} + \frac{\sqrt{3}}{2} i = \sqrt{3} e^{i\frac{5\pi}{6}}, \zeta^2 - 1 = -\frac{3}{2} - \frac{\sqrt{3}}{2} i = \sqrt{3} e^{-i\frac{5\pi}{6}}, (\zeta - 1)^n = \sqrt{3^n} e^{i\frac{5n\pi}{6}}, (\zeta^2 - 1)^n = \sqrt{3^n} e^{-i\frac{5n\pi}{6}}$, thus we have

$$\left[\begin{array}{c} n \\ 0 \end{array} \right]_3^{-1} = \frac{1}{3}((\zeta - 1)^n + (\zeta^2 - 1)^n) = 3^{\frac{n}{2}-1} \cdot 2 \cos\left(\frac{5n\pi}{6}\right),$$

$$\left[\begin{array}{c} n \\ 1 \end{array} \right]_3^{-1} = \frac{1}{3}((\zeta - 1)^n \zeta^{-1} + (\zeta^2 - 1)^n \zeta^{-2}) = 3^{\frac{n}{2}-1} \cdot 2 \cos\left(\frac{(5n-4)\pi}{6}\right),$$

$$\left[\begin{array}{c} n \\ 2 \end{array}\right]_3^{-1} = \frac{1}{3}((\zeta-1)^n\zeta^{-2} + (\zeta^2-1)^n\zeta^{-4}) = 3^{\frac{n}{2}-1} \cdot 2\cos\left(\frac{(5n-8)\pi}{6}\right).$$

□

Proposition 2.11. If n is even and n is not a power of 2, then $\mathrm{FAC}(n,-1) = 2$; if n is a power of 2, then $\mathrm{FAC}(n,-1) = 3$.

Proof. Suppose n is a power of 2, i.e. $n = 2^m, m > 1$. Then $\frac{5n-4}{6} = \frac{5\cdot 2^{m-1}-2}{3}$ and $\frac{5n-8}{6} = \frac{5\cdot 2^{m-1}-4}{3}$, so $\frac{5n-4}{6}$ or $\frac{5n-8}{6}$ must be even, thus $\left[\begin{array}{c} n \\ 1 \end{array}\right]_3^{-1}$ or $\left[\begin{array}{c} n \\ 2 \end{array}\right]_3^{-1}$ is $2 \cdot 3^{2^{m-1}-1}$. Hence we have $\gcd(n, a_1) = 2$ or $\gcd(n, a_2) = 2$. □

Proposition 2.12. Let m, r be two positive integers, and let i be an integer. Then we have

$$\left[\begin{array}{c} m \\ i \end{array}\right]_r^{-1} = (-1)^m \cdot \left[\begin{array}{c} m \\ m-i \end{array}\right]_r^{-1}.$$

Proof. We have

$$\left[\begin{array}{c} m \\ m-i \end{array}\right]_r^{-1} = \sum_{\substack{0 \leqslant k \leqslant m \\ k \equiv m-i (\bmod r)}} \binom{m}{k}(-1)^{m-k}$$

$$= \sum_{\substack{0 \leqslant m-k \leqslant m \\ m-k \equiv i (\bmod r)}} \binom{m}{m-k}(-1)^k \cdot (-1)^{m-2k} = (-1)^m \cdot \left[\begin{array}{c} m \\ i \end{array}\right]_r^{-1}.$$

□

2.4. An algorithm for factoring integers

Now suppose n is an odd composite number. The following algorithm will find a non-trivial divisor of n. Set $\log n = \log_2 n$.

Algorithm A:

```
Input:  An odd composite $n$.
Output:  A non-trivial divisor of $n$.
1.  For $r = 3$ to $n-1$ do
2.      Compute $\sum_{i=0}^{r-1} a_i X^i = (X+1)^n \mod (X^r - 1, n)$
3.          For $i = 0$ to $r-1$ do
4.              Compute $d_i := \gcd(n, a_i)$
5.              If $1 < d_i < n$, then output it and halt
```

The correctness of the algorithm is obvious. Now we analyze computational complexity of Algorithm A. First, Algorithm A will terminate when r attains to $r = \mathrm{FAC}(n,1)$. For a fixed r, Step 2 will take $O\tilde{}(r\log^2 n)$ time ([9] Corollary 8.27, p.233). In Step 4, one gcd computation will take $O\tilde{}(\log^2 n)$ time, so computing r gcd's will take $O\tilde{}(r\log^2 n)$ time.

Hence, for a fixed r, Algorithm A takes $O\tilde{}(r \log^2 n)$ time. The total time complexity of Algorithm A is $O\tilde{}(\mathrm{FAC}(n,1)^2 \log^2 n)$.

Of course, using Proposition 2.9, one needs only to compute $\frac{r}{2}$ gcd's in Step 4 of Algorithm A.

3. Some Variants

3.1. Randomized algorithms

We can use some randomized numbers to modify Algorithm A. For example, we can first select randomly a number a with $\gcd(n, a) = 1$, then we perform the computation $\sum_{i=0}^{r-1} a_i X^i = (X + a)^n \mod (X^r - 1, n)$ for all r in some range and then we compute r gcd's just as in Algorithm A. The formal description is as follows.

Algorithm B:

```
Input:   An odd composite $n$.
Output:   A non-trivial divisor of $n$.
1.  Selectly $a$ randomly with $1 \leqslant a \leqslant n-1$
2.  If $1 < \gcd(n,a) < n$, then output it and halt
3.  For $r = 1$ to $n-1$ do
4.          Compute $\sum_{i=0}^{r-1} a_i X^i = (X+a)^n \mod (X^r - 1, n)$
5.                  For $i = 0$ to $r-1$ do
6.                          Compute $d_i := \gcd(n, a_i)$
7.                          If $1 < d_i < n$, then output it and halt
```

Algorithm B will terminate when r attains to $r = \mathrm{FAC}(n,a)$, so the time complexity of Algorithm B will depend on the size of $\mathrm{FAC}(n,a)$. Our experiments show that $\mathrm{FAC}(n,a)$ is much less than $\mathrm{FAC}(n,1)$ for some a.

We can also select some r randomly with r not too big when we expand the polynomial $(X+a)^n \mod (X^r - 1, n)$. If we fail to find a non-trivial divisor of n, we can try another r.

Furthermore, we can select a polynomial $f(X)$ randomly in $\mathbb{Z}[X]$, then we perform the computation $f(X)^n \mod (X^r - 1, n)$. Our experiments show that this works well also.

3.2. Deterministic bounded algorithm

We use two bounds A and R, where $A \in \mathbb{Z}, 0 < A \leqslant \frac{n-1}{2}$ and $R \in \mathbb{Z}, 0 < R \leqslant n-1$.

Algorithm C:

```
Input:   An odd composite $n$.
Output:   A non-trivial divisor of $n$.
1.  For $a = 1$ to $A$ do
2.  If $1 < \gcd(n,a) < n$, then output it and halt
3.  For $r = 1$ to $R$ do
4.          Compute $\sum_{i=0}^{r-1} a_i X^i = (X+a)^n \mod (X^r - 1, n)$
5.                  For $i = 0$ to $r-1$ do
6.                          Compute $d_i := \gcd(n, a_i)$
7.                          If $1 < d_i < n$, then output it and halt
8.          Compute $\sum_{i=0}^{r-1} a_i X^i = (X-a)^n \mod (X^r - 1, n)$
```

```
9.          For $i = 0$ to $r - 1$ do
10.             Compute $d_i := \gcd(n, a_i)$
11.             If $1 < d_i < n$, then output it and halt
```

Of course, if the algorithm fails to find a non-trivial divisor of n, we can increase the bound A or R.

We can also run Algorithm C for $a \in A$ and $r \in R$, where $A \subseteq \{1, \ldots, n-1\}, R \subseteq \{1, \ldots, n-1\}$ are two specific subsets.

3.3. Using polynomials of several variables

We can also use polynomials of several variables $f(X, Y, \ldots, Z) \in \mathbb{Z}[X, Y, \ldots, Z]$, then we perform the computation $f(X, Y, \ldots, Z)^n \mod (X^{r_x} - 1, Y^{r_y} - 1, \ldots, Z^{r_z} - 1, n)$, where $r_x, r_y, \ldots, r_z$ are positive integers, not necessarily the same, and then we compute the gcd's of the coefficients and n to find a non-trivial divisor of n. Our experiments show that this works well also.

4. Factoring RSA modulus

Now we suppose $n = pq$ be a RSA modulus, where $p < q$ are two distinct odd primes. In this section we provide some upper bounds for $\mathrm{FAC}(n, 1)$ with RSA modulus n. We have to say that these bounds are rather rough, see the following Section 5.

Lemma 4.1. Let $m > 1$ be a positive integer and let k be an integer with $0 < k < m$ and $\gcd(m, k) = 1$. Then we have $m \mid \binom{m}{k}$.

Proof. Since

$$\binom{m}{k} = \frac{m!}{k!(m-k)!} = \frac{m}{k} \cdot \frac{(m-1)!}{(k-1)!(m-k)!} = \frac{m}{k} \cdot \binom{m-1}{k-1}$$

is an integer and $\binom{m-1}{k-1}$ is also an integer, we have $k \mid m\binom{m-1}{k-1}$. As $\gcd(m, k) = 1$, hence $k \mid \binom{m-1}{k-1}$, and so $m \mid \binom{m}{k}$. □

Now for $0 \leqslant k \leqslant n$, if $\gcd(n, k) = 1$, i.e. $p \nmid k$ and $q \nmid k$, by Lemma 4.1, we have $n \mid \binom{n}{k}$. So we need only to consider $k = 0$ or $k = n$ or $k = pi (0 < i < q)$ or $k = qj (0 < j < p)$.

Lucas' Theorem. (See [2] p.28) Let p be a prime, and let $a = a_0 + a_1 p + \cdots + a_k p^k, b = b_0 + b_1 p + \cdots + b_k p^k$, where $0 \leqslant a_i, b_i < p$ for $i = 0, 1, \ldots, k$. Then

$$\binom{a}{b} \equiv \prod_{i=0}^{k} \binom{a_i}{b_i} \pmod{p}.$$

Since $q \nmid pi$ for $0 < i < q$, by Lucas' Theorem, we have $q \mid \binom{n}{pi}$. Similarly, we have $p \mid \binom{n}{qj}$ for $0 < j < p$. By Lucas' Theorem, we have $\binom{n}{qj} \equiv \binom{p}{j} (\mathrm{mod}\ q)$ for $0 < j < p$. Since $p < q$, then $q \nmid \binom{p}{j}$, so $q \nmid \binom{n}{qj}$ for $0 < j < p$.

Proposition 4.2. p is a factorization-friendly number of a RSA modulus $n = pq$ with respect to 1, so $\mathrm{FAC}(n, 1) \leqslant p < \sqrt{n}$.

Proof. Write $q = ap + k$ with $a > 0$ and $0 < k < p$. Set $I = \{k + ps \mid s \geqslant 0\}$. Obviously, $0, n$ and $pi(0 < i < q)$ are all not in I and q is in I. If qj is in I for some j with $0 < j < p$, i.e. $qj = k + ps$ for some $s \geqslant 0$. Then $k \equiv qj \equiv kj \pmod{p}$, thus $j \equiv 1 \pmod{p}$, so $j = 1$. Hence

$$\left[\begin{array}{c} n \\ k \end{array}\right]_p = \sum_{\substack{0 \leqslant t \leqslant n \\ t \equiv k(\bmod p)}} \binom{n}{t} \equiv \binom{n}{q} \pmod{n}.$$

By above analysis, we have $p \mid \binom{n}{q}$ and $q \nmid \binom{n}{q}$, hence

$$\gcd\left(n, \left[\begin{array}{c} n \\ k \end{array}\right]_p\right) = p.$$

Therefore p is a factorization-friendly number of a RSA modulus $n = pq$ with respect to 1. □

The upper bound for $\mathrm{FAC}(n, 1)$ in Proposition 4.2 is rather rough, in fact, our experiments show that $\mathrm{FAC}(n, 1)$ is much less than p, see the following Section 5. We have $\mathrm{FAC}(3q, 1) = 3$ for an odd prime $q > 3$ from Proposition 4.2.

Proposition 4.3. Let $n = pq$ be a RSA modulus with $p < q < 2p$. Write $q = p + k$ with $0 < k < p$. Let c be a positive integer. Suppose $k \leqslant p - 3c - 1$. Then $p - c$ is a factorization-friendly number of n with respect to 1, so $\mathrm{FAC}(n, 1) \leqslant p - c$.

Proof. Since k is even, we have $p \geqslant 3c + 3$, so $c + 1 < p - c$ and $k + c < p - c$. Since $(p - c) + c = p$ is a prime, we have $\gcd(p - c, c) = 1$. Similarly, $\gcd(p - c, k + c) = 1$. So the sets $\{-ci(\bmod (p - c)) \mid i = 1, 2, \ldots, c + 1\}$ and $\{(k + c)j(\bmod (p - c)) \mid j = 1, 2, \ldots, c\}$ have $c + 1$ elements and c elements, respectively. So we can choose an element a from the first set such that a is not in the second set. Let $a \equiv -ci_1(\bmod (p - c))$ with $1 \leqslant i_1 \leqslant c + 1$ and $0 < a < p - c$.

Set $I = \{a + (p - c)s \mid s \geqslant 0\}$. Obviously, $0 \notin I$. Since $n = pq \equiv c(k + c)(\bmod (p - c))$ and by the choice of a, we have $n \notin I$. For $pi(0 < i < q)$, if $pi \in I$, i.e. $pi \equiv a(\bmod (p - c))$, i.e. $ci \equiv -ci_1(\bmod (p - c))$, we have $i \equiv -i_1(\bmod (p - c))$. Hence $i = p - c - i_1 + (p - c)s, s \geqslant 0$. Since $p - c - i_1 + p - c = 2p - 2c - i_1 \geqslant 2p - 3c - 1 \geqslant p + k = q$, we have $i = p - c - i_1 := i_0$. For $qj(0 < j < p)$, if $qj \in I$, i.e. $qj \equiv a(\bmod (p - c))$, i.e. $(k + c)j \equiv a(\bmod (p - c))$, this equation has a unique solution j_0 with $0 \leqslant j_0 < p - c$ and $j \equiv j_0(\bmod (p - c))$. By the choice of a, we have $j_0 \geqslant c + 1$. Because $j_0 + p - c > p$, so $j = j_0$. Thus, we have

$$\left[\begin{array}{c} n \\ a \end{array}\right]_{p-c} \equiv \binom{n}{pi_0} + \binom{n}{qj_0}(\bmod n).$$

Obviously,

$$q \nmid \left[\begin{array}{c} n \\ a \end{array}\right]_{p-c}.$$

Since $q - i_0 + 1 \leqslant p$, we have $p \mid \binom{n}{pi_0}$, thus

$$p \mid \left[\begin{array}{c} n \\ a \end{array}\right]_{p-c}.$$

Hence

$$\gcd\left(n, \left[\begin{array}{c} n \\ a \end{array}\right]_{p-c}\right) = p.$$

Therefore $p - c$ is a factorization-friendly number of n with respect to 1. □

Corollary 4.4. Let $n = pq$ be a RSA modulus with $p < q < 2p$. Write $q = p+k$ with $0 < k < p$. Suppose $k < \varepsilon p, 0 < \varepsilon < 1$. Suppose $p \geqslant \frac{3}{1-\varepsilon}$. Then $\mathrm{FAC}(n, 1) \leqslant p - \lfloor \frac{1-\varepsilon}{3} p \rfloor \approx \frac{2+\varepsilon}{3} p$.

Proof. Put $c = \lfloor \frac{1-\varepsilon}{3} p \rfloor$. The result then follows from Proposition 4.3. □

Proposition 4.5. Let $n = pq$ be a RSA modulus with $p < q < 2p$. Write $q = p + k$ with $0 < k < p$. Suppose $n \equiv -1(\bmod 4)$. Suppose $2 < k < \frac{2}{3}p$. Then $r := \frac{p}{2} + \frac{3}{4}k$ is a factorization-friendly number of n with respect to 1, so $\mathrm{FAC}(n, 1) \leqslant \frac{p}{2} + \frac{3}{4}k$.

Proof. Since $n \equiv -1(\bmod 4)$, then p and q must be the case: the one is $\equiv 1(\bmod 4)$ and the other is $\equiv -1(\bmod 4)$. So $k \equiv 2(\bmod 4)$, thus r is a positive integer and $r < p$. Let $a \in \mathbb{Z}$ with $0 \leqslant a < r$ and $a \equiv (r-1)p(\bmod r)$. Since $r > 1$, we have $a > 0$. Set $I = \{a + rs \mid s \geqslant 0\}$. A similar analysis as the proof of Proposition 4.3, we have $0, n \notin I$ and only $p(r-1)$ and $q(r-3)$ are in I. Thus

$$\left[\begin{array}{c} n \\ a \end{array}\right]_{r} \equiv \binom{n}{p(r-1)} + \binom{n}{q(r-3)} (\bmod n).$$

Similarly, we have

$$\gcd\left(n, \left[\begin{array}{c} n \\ a \end{array}\right]_{r}\right) = p.$$

Therefore r is a factorization-friendly number of n with respect to 1. □

Corollary 4.6. Keeping the notations in Proposition 4.5, further suppose $k < \varepsilon p, 0 < \varepsilon \leqslant \frac{2}{3}$, then $\mathrm{FAC}(n, 1) \leqslant (\frac{1}{2} + \frac{3}{4}\varepsilon)p$.

Remark. Comparing Corollaries 4.4 and 4.6, it is easy to see that, when $\varepsilon = \frac{2}{5}$, then $\frac{2+\varepsilon}{3} = \frac{1}{2} + \frac{3}{4}\varepsilon$; when $\varepsilon > \frac{2}{5}$, then $\frac{2+\varepsilon}{3} < \frac{1}{2} + \frac{3}{4}\varepsilon$; when $\varepsilon < \frac{2}{5}$, then $\frac{2+\varepsilon}{3} > \frac{1}{2} + \frac{3}{4}\varepsilon$.

Obviously, all the above bounds for $\mathrm{FAC}(n, 1)$ hold also for $\mathrm{FAC}(n, a)$ with arbitrary a such that $\gcd(n, a) = 1$. We conclude this section by giving the following interesting result. This seems mean that an easily factorized number is also easily factorized by our method.

Proposition 4.7. Let $n = pq$ be a RSA modulus. Suppose $q = p + 2$, i.e. p and q are twin primes. Then $\mathrm{FAC}(n, 1) \leqslant 6$.

Proof. Since $\mathrm{FAC}(15, 1) = 3$, we may assume $p \geqslant 5$. It is easy to see that, there is a positive integer k such that $p = 6k - 1$ and $q = 6k + 1$. An easy analysis shows that

$$\left[\begin{array}{c} n \\ 2 \end{array}\right]_{6} \equiv \sum_{\substack{0 < i < q \\ i \equiv 4(\bmod 6)}} \binom{n}{pi} + \sum_{\substack{0 < j < p \\ j \equiv 2(\bmod 6)}} \binom{n}{qj} (\bmod n).$$

It is easy to see that

$$p \mid \left[\begin{array}{c} n \\ 2 \end{array}\right]_6 .$$

By a result of Sun [8], we have

$$\left[\begin{array}{c} p \\ 2 \end{array}\right]_6 = \frac{1}{6}(1 + 2^p - 3^{\frac{p+1}{2}}), \quad \text{for even } k,$$

$$\left[\begin{array}{c} p \\ 2 \end{array}\right]_6 = \frac{1}{6}(1 + 2^p + 3^{\frac{p+1}{2}}), \quad \text{for odd } k.$$

Since $3^{\frac{p+1}{2}} = 3^{\frac{q-1}{2}} \equiv \left(\frac{3}{q}\right) (\bmod q)$, in both cases, we have

$$\left[\begin{array}{c} p \\ 2 \end{array}\right]_6 \equiv \frac{1}{6} \cdot 2^p (\bmod q).$$

Hence

$$q \nmid \left[\begin{array}{c} n \\ 2 \end{array}\right]_6 .$$

Therefore

$$\gcd\left(n, \left[\begin{array}{c} n \\ 2 \end{array}\right]_6\right) = p.$$

□

5. Experimental Results

5.1. Some values of FAC$(n, 1)$ and FAC(n, a) for RSA moduli n

We have done numerous experiments, these experiments show a remarkable fact that the FAC(n, a), even FAC$(n, 1)$, are surprisingly small relative to n. They grow very slowly with n. We list the partial values of FAC$(n, 1)$ for RSA moduli $n = pq$ such that p and q have three digits and some partial values of FAC$(n = pq, a)$ with 5 or 6 digits of p and q (see Tables 1 and 2).

5.2. Comparison of FAC$(n, 1)$ and FAC(n, a) for RSA moduli n

For a fixed n, different choice of a will in general give distinct FAC(n, a). Usually, for some a's, FAC(n, a) will remarkably be less than FAC$(n, 1)$. This indicates, when we choose such a, we can reduce the time complexity of Algorithm B. We list some such examples, see Table 3.

5.3. Some experiments for polynomials of several variables or of high degrees

We can use some polynomial of degree two or three instead of linear polynomial. For example, we use polynomials $f(X) = X^2 + X + 1$ and $f(X) = X^3 + X^2 + X + 1$, then compute $f(X)^n \mod (X^r - 1, n)$, and compute the gcd's of the coefficients and n to find a non-trivial divisor of n. We also denote the smallest r by $\mathrm{FAC}_S(n)$ and $\mathrm{FAC}_C(n)$ respectively and list some such examples, see Table 4.

We can also use polynomials of several variables instead of polynomials of one variable. The algorithms work well also. For example, we use polynomial $f(X, Y) = X + Y + 1$, compute $f(X, Y)^n \mod (X^r - 1, Y^r - 1, n)$, and compute the gcd's of the coefficients and n to find a non-trivial divisor of n. We denote the smallest r by $\mathrm{FAC}_B(n)$. We also compute $f(X, Y)^n \mod (X^{r_1} - 1, Y^{r_2} - 1, n)$ and denote the first sequence (r_1, r_2) we get by $\mathrm{FAC}_B(n, r_1, r_2)$. In addition, we use polynomial $f(X, Y, Z) = X + Y + Z + 1$, compute $f(X, Y, Z)^n \mod (X^r - 1, Y^r - 1, Z^r - 1, n)$, and denote the smallest r by $\mathrm{FAC}_T(n)$. Similarly, we compute $f(X, Y, Z)^n \mod (X^{r_1} - 1, Y^{r_2} - 1, Z^{r_3} - 1, n)$ and denote the first sequence (r_1, r_2, r_3) by $\mathrm{FAC}_T(n, r_1, r_2, r_3)$. We list some such examples, see Table 5.

5.4. The practical effect of our method

The practical effect of Algorithm A is not good on a single PC. For example, it takes about six hours to factorize a number with 15 digits, however, it takes about one hour and 30 minutes to factorize the same number when using Algorithm B. However, our method has two advantages, one is its simplicity, and the other is its parallelism. It is easily adapted to run simultaneously on many computers, e.g. on Internet. We do not perform such experiments.

6. Open Problems and Conclusion

Of course, one open problem is to obtain better theoretic estimate for $\mathrm{FAC}(n, 1)$, even for RSA modulus n, than the estimate given in Proposition 4.2. Another open problem is to give explicit bounds for A and R in Algorithm C which guarantees Algorithm C always find a non-trivial divisor of n. Since the computational complexity of our method depends directly on the size of $\mathrm{FAC}(n, 1)$ or of $\mathrm{FAC}(n, a)$ for some specific a's, so the most interesting thing is to obtain asymptotic tight upper bounds for these numbers.

Integer factorization is a very important computational problem, and it is the foundation stone of the famous RSA cryptosystem [6]. Since the invention of the general number field sieve in 1993, there is no substantial progress on this problem. There is no new method to appear for a long time. Our method seems new to the best of our knowledge and maybe this is a new idea on integer factorization problem.

References

[1] M. Agrawal, N. Kayal, N. Saxena: Primes is in P. *Ann. of Math.* (2) **160** (2004), no. 2, 781–793.

[2] P. J. Cameron: Combinatorics: topics, techniques, algorithms. Cambridge University Press, Cambridge, 1994.

[3] H. Cohen: A course in computational algebraic number theory. Graduate Texts in Mathematics, vol. **138**. Springer-Verlag, Berlin, 1993.

[4] A. K. Lenstra: Integer factoring. Towards a quarter-century of public key cryptography. *Des. Codes Cryptogr.* **19** (2000), no. 2-3, 101–128.

[5] A. K. Lenstra, H. W. Lenstra, Jr.(Eds.): The development of the number field sieve. Lecture Notes in Mathematics, vol. **1554**. Springer-Verlag, Berlin, 1993.

[6] R. L. Rivest, A. Shamir, L. Adleman: A method for obtaining digital signatures and public-key cryptosystems. *Communications of the Association for Computing Machinery*, **21**(1978), no. 2, 120–126.

[7] V. Shoup, NTL: A library for doing number theory. Available at http://www.shoup.net/ntl/

[8] Zhi Hong Sun: The combinatorial sum $\sum_{k=0,\ k\equiv r \pmod{m}}^{n} \binom{n}{k}$ and its applications in number theory (I) (In Chinese). Nanjing Daxue Xuebao Shuxue Bannian Kan **9** (1992), no. 2, 227–240.

[9] J. von zur Gathen, J. Gerhard: Modern computer algebra. Cambridge University Press, Cambridge, 1999.

A Some tables

Table 1 The partial values of FAC$(n = pq, 1)$ with 3 digits of p and q

$n = pq$	FAC$(n,1)$	$n = pq$	FAC$(n,1)$	$n = pq$	FAC$(n,1)$
10403=101*103	5	10807=101*107	8	11009=101*109	13
11413=101*113	17	12827=101*127	21	13231=101*131	8
13837=101*137	22	14039=101*139	12	15049=101*149	21
15251=101*151	17	15857=101*157	18	16463=101*163	9
16867=101*167	15	17473=101*173	21	18079=101*179	12
18281=101*181	22	19291=101*191	20	19493=101*193	9
19897=101*197	13	20099=101*199	12	21311=101*211	9
251659=359*701	38	254531=359*709	17	258121=359*719	5
235247=367*641	20	235981=367*643	9	237449=367*647	12
255067=379*673	33	256583=379*677	51	258857=379*683	36
409763=593*691	25	415693=593*701	7	420437=593*709	15
563903=607*929	24	571187=607*941	43	586969=607*967	60
621787=701*887	52	536713=709*757	59	750187=757*991	72
812909=853*953	17	756731=857*883	33	782549=859*911	61
921551=953*967	17	936799=953*983	34	988027=991*997	9

Table 2 The partial values of FAC$(n = pq, a)$ with 5 or 6 digits of p and q

$n = pq$	a	FAC(n,a)
323910211=16453*19687	224606094	25
401112223=16487*24329	254658360	62
556453211=20333*27367	501040105	23
1359410777=32969*41233	1	80
1695762151=35027*48413	278073557	30
2517939323=44351*56773	1970421086	29
2708129327=50753*53359	565530478	91
19366566407=134753*143719	8354295828	16
22473158221=145577*154373	2247187015	207
27367072697=131701*207797	9716426496	80
43567492823=183059*237997	5778311349	40
45444818857=179687*252911	43931485047	29

Table 3 The partial values of FAC$(n, 1)$ and FAC(n, a)

$n = pq$	FAC$(n, 1)$	a	FAC(n, a)
323910211=16453*19687	266	224606094	25
401112223=16487*24329	266	254658360	62
481118119=18371*26189	260	447652040	16
556453211=20333*27367	39	501040105	23
580839353=20201*28753	209	494398594	17
712415273=25237*28229	113	395527894	47
89441974637=276839*323083	712	49599857930	68
91457375567=300721*304127	584	41380446395	123
154709636971=332933*464687	1161	16603703892	162
408187969489=531911*767399	1025	142966224429	67
702358343579=733003*958193	2467	413854661934	245
1039342803007=1012751*1026257	1771	176505030244	198

Table 4 The partial values when using polynomials of high degree

$n = pq$	FAC$(n, 1)$	FAC$_S(n)$	FAC$_C(n)$
399947791=18049*22159	34	150	40
438232609=20731*21139	234	165	43
466390553=16417*28409	132	212	104
540517409=18089*29881	221	100	289
702477619=23057*30467	118	300	109
850932143=28901*29443	148	197	47
437164210933=602489*725597	737	939	1938
588372701219=564701*1041919	1627	2453	1311
617840467649=618571*998819	1886	509	257
626909223527=642937*975071	2073	2938	1359
674776379579=780721*864299	1251	1442	1183
735474271523=857357*857839	507	966	431

Table 5 The partial values when using polynomials of several variables

$n = pq$	FAC$(n, 1)$	FAC$_B(n)$	FAC$_B(n, r_1, r_2)$		FAC$_T(n)$	FAC$_T(n, r_1, r_2, r_3)$		
			r_1	r_2		r_1	r_2	r_3
360379=557*647	45	13	6	2	11	3	3	2
581897=659*883	15	23	6	5	11	6	5	5
599197=601*997	17	17	8	7	6	5	4	4
1685069=1171*1439	45	19	12	5	13	6	4	3
2399219=1231*1949	32	14	8	3	10	8	6	4
2581903=1483*1741	35	12	9	8	17	5	5	4
3202781=1721*1861	70	17	7	3	10	7	6	2
8381519=2069*4051	79	32	11	10	10	6	3	3
10007717=2953*3389	66	41	8	5	6	5	4	2
11963789=3259*3671	17	15	10	7	20	6	5	5
12374501=3079*4019	27	31	11	7	15	7	5	5
13451593=3347*4019	71	26	12	3	17	6	5	3