大整数因子分解新算法及对 RSA 密码制的解密*

王 泽 辉
（中山大学 科学计算与计算机应用系，广东 广州 510275）

摘 要：对一大类大整数的因子分解构造算法 WZH，可在 $O(L(\ln m)^2)+O(\ln m)^3$（$L<m$）时间内求其标准分解式 $m=p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_s^{\alpha_s}$，并求出 p_i-1 的部分素因子，从而揭示 m 的内部结构；对于 RSA 密码体制，可在 $O(L\ln m)+O((\sqrt{p}-\sqrt{q})^2/L\cdot(\ln m)^2)+O((\ln m)^3)$ 时间内完整作因子分解求出私匙 p、q、α，算法对 RSA 适合全部正整数，特殊情况下时间复杂性降为 $O((\ln m)^k)$，k 为正整数。

关键词：大整数标准分解；RSA 密码体制；密匙完全求解；多项式时间
中图分类号：O157.4；TN918.4 文献标识码：A 文章编号：0529-6579（2003）05-0015-04

目前不少信息系统（尤其在网络环境下）的加密与解密，多采用公匙密码体制，公匙密码体制最著名是 RSA 密码体制及其变种，以及 RLGamal 密码体制及其变种（如椭圆曲线密码体制）这两大体系，其中 RSA 体制的安全性是基于大整数因子分解的困难性，即由两个大素数 p、q 相乘作为公匙，却难于在多项式时间内分解出 p、q，本文针对该问题进行求解，先考虑更广泛一类问题：任意大整数 m 的因子分解。

1 对一般大整数分解的结果

设 $m=p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_s^{\alpha_s}$ 为大整数 m 的标准分解式，待求，记 $\delta_N(a)$ 为 a 对模 N 之阶（即满足 $a^L\equiv 1(\mathrm{mod}\ N)$ 之最小正整数 L_0），$\gcd(a,b)$ 表示 a，b 之最大公约数，$L=[l_1,l_2,\cdots\cdots,l_s]$ 为 $l_1,l_2,\cdots\cdots,l_s$ 之最小公倍数。

预解算法 WZH0（m，G，L），筛选一组正整数 g_i，$\gcd(g_i,m)=1$，$i=1,\cdots\tau$，τ 为小正整数（如 $\gcd(g_i,m)\neq 1$ 已求出 m 真因子）。在 τ 台计算机上并行计算，对 g_i 求 $\delta_m(g_i)$，将第一个出现的解记为 $L=\delta_m(g)$，g_i 记为 g，其他中间结果暂存并转算法 WZH1，WZH0 之时间复杂性为 $\tau O(\ln m)=O(L\ln m)$（$L<m$），特殊时总时间降为 $O(\ln m)^2$）。

大整数 m 对于特选的 g 可存在偏小的阶 L，明显的例子是：$N=a^k-1$，$\delta_N(a)=k=O(\ln N)$，$N=a^{(k-1)n}+\cdots+a^{2n}+a^n+1$，$\delta_N(a)=kn=O(\ln N)$。降低 L 值的算法涉及复杂的估计，限于篇幅另文述之。

主算法 WZH1（m，g，L，h，a，b，C，ER），求 h 使 $L\nmid h!$，$L\mid(h+1)!$，间接得到 L 所有素因子幂，算法伪代码为

```
f0←1, T←0
for i←1 to L
    f1←f0 *i (mod L)
    if (f1+1) ≡0 (mod (i+1)) and L ≡0 (mod (i+1))
    then {T←T+1, C(T)←i+1}
    if f1=0 then exit for
    f0←f1
next i
    h ← i−1, b ← g^f0 (mod m)−1,
        a ← gcd(b, m)                    (1)
```

ER 是退出标志，ER=0 表示算法正常退出，a 存 m 某个真因子，ER=1 表示非正常退出，$a=1$。WZH1 中计算 a 之时间复杂性为 $O((\ln m)^3)$，b 计算耗时 $O(\ln f_0\cdot(\ln m)^2)$ 上界为 $O(\ln m)^3$），循环执行时间为 $O(h(\ln m)^2)=O(L(\ln m)^2)$，WZH1 总时间为 $O(L(\ln m)^2)+O((\ln m)^3)$。WZH0 + WZH1 总时间为 $O(L(\ln m)^2)+O((\ln m)^3)$。

有关算法时间复杂性的更详细证明另文述之。

定理 1 算法 WZH1 得到

$$f_0=h!(\mathrm{mod}\ L)\neq 0,\quad \text{且 } L\nmid h!,\ L\mid(h+1)! \tag{2}$$

$C(1:T)$ 存放 L 之所有 T 个素因子

证明 由程序易知 f_0 表达式及意义，第 i 个循环中 $f_1=i!(\mathrm{mod}\ L)$ 根据威尔逊（Wilson）定理易

* 收稿日期：2003-06-28
作者简介：王泽辉（1963 年生），男，副教授；E-mail：info@gdcts.com

证：$i+1$ 为 L 素因子之充要条件为 $(i+1)\mid(f_1+1)$ 且 $(i+1)\mid L$

因为 $L\mid 1\times 2\times\cdots\times(h+1)$

所以 L 所有素因子在循环 for 中出现并存放 $C[1:T]$ 中，证毕。

求出 l 之素因子后，可用试除法求出因子之幂，存放于 $C(T+1)\cdots C(T+T)$ 中。

定理2 设 g 对 p_i 之阶为 l_i，$i=1,2,\cdots s$，则

$$L=[l_1,l_2,\cdots,l_s]\,p_1^{\alpha_1-1}p_2^{\alpha_2-1}\cdots p_s^{\alpha-1} \tag{3}$$

证明 可用数学归纳法证明：

$g^{l_i}\equiv 1(\bmod p_i)\Rightarrow g^{l_i\cdot p_i^{k-1}}\equiv 1(\bmod p_i^k)$，$k\geqslant 1$

故 g 对 $p_i^{\alpha_i}$ 之阶为 $l_i\cdot p_i^{\alpha-1}$，再用数学归纳法可证 g 对 $m=(p_1^{\alpha_1})(p_2^{\alpha_2})\cdots(p_s^{\alpha_s})$ 之阶 $L=[l_1p_1^{\alpha_1-1},l_2p_2^{\alpha_2-1},\cdots,l_sp_s^{\alpha_s-1}]$，证毕。

推论1 当 $\alpha_i\geqslant 2$，$p_i^{\alpha_i}\mid m$ 时，p_i 必存于数组 $C[1:T]$ 中，用试除法可分离出 m 之重因子 $p_i^{\alpha_i}$，$\alpha_i\geqslant 2$。

注1 m 含有重因子会增大 L 之值，但密码学中 m 少出现重因子，下假定

$$m=p_1p_2\cdots p_s,$$
$$\gcd(p_i,p_j)=1,\ i\neq j,$$
$$\gcd(g,m)=1 \tag{4}$$

引理1 算法WZH1中求得之 b 满足

$$b=(g^{h!}-1)(\bmod m),\text{且}\ 1\leqslant b\leqslant m-2 \tag{5}$$

证明 在（1）中设 $b_0=g^{f_0}\pmod m$，由（2）$f_0\equiv h!\pmod L\neq 0$，故存在 λ_1，使

$$h!=f_0+\lambda_1 L,$$
$$g^{h!}(\bmod m)=g^{f_0+\lambda_1 L}(\bmod m)=$$
$$g^{f_0}(g^L)^{\lambda_1}(\bmod m)=g^{f_0}\times 1(\bmod m)=b_0$$

因为 $b=b_0-1=(g^{h!}-1)(\bmod m)$，易证 $1\leqslant b\leqslant m-2$，证毕。

定理3 记 $l_i=\delta_{p_i}(g)$，当 $l_i\mid h!$ 时，$p_i\mid\gcd(b,m)$，$p_i\mid a$，ER=0。

证明 设

$$b_0=g^{h!}(\bmod m)$$

所以存在 λ_2 使

$$g^{h!}-b_0=\lambda_2 m,\ p_i\mid m\Rightarrow p_i\mid(g^{h!}-b_0),$$
$$b_0\equiv g^{h!}(\bmod p_i),$$

如 $l_i\mid h!$ 设

$$h!=\lambda_3 l_i,$$
$$b_0\equiv g^{\lambda_3 l_i}(\bmod p_i)=$$
$$(g^{l_i})^{\lambda_3}(\bmod p_i)\equiv 1(\bmod p_i),$$
$$p_i\mid(b_0-1)$$

或 $p_i\mid b$，又 $p_i\mid m$，$p_i\mid\gcd(b,m)$，或 $p_i\mid a$，证毕。

注2 定理3表明WZH1可绕开 p_i-1 之大素因子，对于偏小的 $L=[l_1,l_2\cdots l_s]$，l_i 更小，但只要有一个 $l_i\mid h!$ 便可解出 m 之真因子。

不妨设 L 之素因子幂为 $q_i^{\beta_i}$ 且已排序，即

$$L=q_1^{\beta_1}q_2^{\beta_2}\cdots q_t^{\beta_t},\ q_1^{\beta_1}<q_2^{\beta_2}\cdots<q_t^{\beta_t}$$

$q_t^{\beta_t}$ 可能存于某个 l_i 中，设为 l_v 即 $q_t^{\beta_t}\mid l_v$，$q_t^{\beta_t+1}\nmid l_v$。

引理2 由 $L\nmid h!$，$L\mid(h+1)!$ 可得① $h+1\geqslant q_k$，$k=1,2\cdots t$，② $h<q_t^{\beta_t}$，③ $h+1$ 至少含 L 一个素因子 q_k，$1\leqslant k\leqslant t$。

证明 反证法易证①、②，下证③，$L\nmid h!$ 表明有 $q_k^{r_k}\mid h!$，$q_k^{r_k+1}\nmid h!$，但 $r_k<\beta_k$，故必须有 $q_k^{\delta_k}\mid(h+1)!$，且 $\delta_k+r_k\geqslant\beta_k$，使 $L\mid h!(h+1)$ 成立，证毕。

定理4 当 $h\geqslant\max\{q_t^{\beta_t-c},q_{t-1}^{\beta_{t-1}}\}$，$1\leqslant c\leqslant\beta_t$ 时，或者（Ⅰ）至少有一个 l_i 与 l_v 无公因子 $q_t^{\beta_t+1-c}$，或者（Ⅱ）所以 l_i（$i\neq v$）与 l_v 有公因子 $q_t^{\beta_t+1-c}$，那么，对（Ⅰ）所有与 l_v 无公因子 $q_t^{\beta_t+1-c}$ 之 l_i 均有 $l_i\mid h!$，从而 $p_i\mid\gcd(b,m)$，ER=0；对于（Ⅱ）ER=1。

证明 只证（Ⅰ），这些 l_i 如含 q_t 之因子幂超过 $q_t^{\beta_t+1-c}$ 则变成（Ⅱ），不超过 l_i 时最大因子幂为 $\max\{q_t^{\beta_t-c},q_{t-1}^{\beta_{t-1}}\}\leqslant h$，即所有 l_i 之因子幂落于 $1,2,\cdots,h$ 中，即 $l_i\mid h!$ 由定理3可证。

定理5 当 $h\geqslant q_{t-1}^{\beta_{t-1}}$ 且 $h<q_t$ 时，由引理2之① 必有 $h+1=q_t$，$h=q_t-1$，这时，或者（Ⅰ）至少有一个 l_i 使 $q_t\nmid l_i$，或者（Ⅱ）所有 $l_1,l_2,\cdots,l_s$ 有公因子 q_t。那么，对（Ⅰ），所有满足 $q_t\nmid l_i$ 之 l_i 有 $l_i\mid h!$ 从而 $p_i\mid\gcd(b,m)$，ER=0；对于（Ⅱ）ER=1。

证明 类似于定理4可证。

定理6 当 $h<q_{t-1}^{\beta_{t-1}}$ 时，由引理2之③，设 $h+1$ 含 L 之素因子为 $q_{k1},q_{k2},\cdots q_{ku}$，$u\geqslant 1$，这时，或者（Ⅰ）至少有一个 l_i 使 $q_{ke}\nmid l_i$，$e=1,2,\cdots u$；或者（Ⅱ）任一 l_i，$\exists e=e(i)$ 使 $q_{ke}\mid l_i$，$e=1,2,\cdots s$。那么，对于（Ⅰ），所有满足 $q_{ke}\nmid l_i$，$e=1,2,\cdots u$ 之 l_i 有 $l_i\mid h!$，从而 $p_i\mid\gcd(b,m)$，ER=0；对于（Ⅱ）可能 ER=0 也可能 ER=1。特别当 $u=1$ 时，即 $h+1$ 仅含 L 一个素因子 q_k 时 $l_1,l_2,\cdots l_s$ 有公因子 q_k。

证明 仅证（Ⅰ），设

$$l_i=q_1^{\beta_{i1}}q_2^{\beta_{i2}}\cdots q_t^{\beta_{ik}},$$
$$\beta_{ik}\leqslant\beta_k,\ k=1,2,\cdots t,\ i=1,2,\cdots s \tag{6}$$

当 $q_{ke}\nmid l_i$，$e=1,2,\cdots u$ 时，如 $l_i\nmid h!$ 即 $\exists q_k,r_k$，使 $q_k^{r_k}\mid h!$，$q_k^{r_k+1}\nmid h!$ 而 $\beta_{ik}>r_k$，这时 $h+1$ 必有因子 q_k，但 $q_k\mid l_i$，与假设矛盾，故必 $l_i\mid h!$，证毕。

推论 2　当 $h+1$ 仅含 L 一个素因子时，ER=1 的必要条件是

$$P_1-1,\ P_2-1,\ \cdots,\ P_S-1$$

有公因子 R（R 已求出）。

在 $h+1$ 含 L 多于一个素因子时，对算法 WZH0 中暂存结果再运算，求出 k 个 g 及对应之 $L=\delta_m(g)$，调高 k 值，$h+1$ 仅含 L 一个素因子的概率较高。

2　在 RSA 密制解密的应用

将上述结果用于 RSA 密码体制，即对应于 $s=2$ 求 $m=pq$ 分解，为避免私匙攻击，常选 $p-1$、$q-1$ 含大素因子且远大于其它因子，不妨设

$$p-1=u_1^{\alpha_1}\cdots u_{e-2}^{\alpha_{e-2}}\cdot u_{e-1}\cdot u_e,$$

$$q-1=v_1^{\beta_1}\cdots v_{f-2}^{\beta_{f-2}}\cdot v_{f-1}\cdot v_f \tag{7}$$

$$u_e>u_{e-1}>\max\{u_1^{\alpha_1},\ \cdots,\ u_{e-2}^{\alpha_{e-2}}\},$$

$$v_f>v_{f-1}>\max\{v_1^{\beta_1},\ \cdots,\ v_{f-2}^{\beta_{f-2}}\} \tag{8}$$

定理 7

(1) 调用 WZH0、WZH1 求得 $L=\delta_m(g)$，当 $u_e\mid L$，$v_f\mid L$，不妨设 $u_e\leqslant v_e$，则或者(Ⅰ)$u_e<v_f$，那么 $\gcd(b,m)=p$，ER=0；或者(Ⅱ)$u_e=v_f=R$，$p-1$，$q-1$ 有大公因子 R，可调用算法 WZH2 求解。

(2) 当 $p-1$、$q-1$ 无大素因子的公因子时（RSA 常这样处理），特选 g 得到 L 可绕开大素因子 u_e，v_f，不妨设

$$u_{e-1}<v_{f-1},\ u_{e-1}\mid L,\ v_{f-1}\mid L,\ u_e\nmid L,\ v_f\nmid L$$

那么

$$\gcd(b,m)=p,\ \mathrm{ER}=0$$

证明　仅证 (1)，由引理 2 之①，②，$h<v_f$，而 $h+1\geqslant v_f$ 故 $h=v_f-1$，或者 $u_e=v_f$ 或者 $u_e<v_f$，这时 u_e，v_f 均为奇数，$h\geqslant u_e+1>u_e$，由定理 4 可证。

算法 WZH2　由 WZH1 所得 $L\mid\varphi(m)$，$R\mid(p-1)$，$R\mid(q-1)$，$R\mid L$，设 $R^2\nmid L$（如 $R^2\mid L$ 可用试除法除去因子使 $R^2\nmid L$），故

$$\gcd(R^2,L)=1$$

设

$$n_1=R^2L,$$

$$\varphi(m)=(p-1)(q-1)=pq+1-(p+q)$$

故

$$\varphi(m)=m+1-(p+q)$$

且

$$p+q\equiv(m+1)(\mathrm{mod}R^2),$$

$$p+q\equiv(m+1)(\mathrm{mod}\ L)$$

由中国剩余定理易求出

$$p+q\equiv m_0(\mathrm{mod}n_1),$$

$$n_1=R^2L,$$

$$m+1\geqslant p+q\geqslant 2\sqrt{pq}=2\sqrt{m}$$

设

$$p+q\equiv m_0+\lambda n_1,\left[\frac{m+1}{n_1}\right]=\theta_1,$$

$$[2\sqrt{m}/n_1]=\theta_0$$

则

$$\theta_1\geqslant\lambda\geqslant\theta_0$$

可对 $\lambda=\theta_0,\theta_0+1,\cdots\theta_1$ 逐一试，必找到 $\lambda=\lambda_1$ 使

$$(p+q)^2-4m=(m_0+\lambda n_1)^2-4m$$

为平方数设为 d_0^2，则 $d_0=|p-q|$，又 $p+q=m_0+\lambda_1 n_1$，从而解出 p，q。

算法 WZH2 之时间复杂性上界为 $O(\ln m)^3$）

上述算法可归为“低端攻击方案”，从小往大找偏小的阶 L，而算法 WZH3 可称为“高端攻击方案”，从大往小找阶 L。

算法 WZH3　设

$$\gcd(g,m)=1,\ k=m+1-[2\sqrt{m}]$$

先用 $O((\ln m)^3)$ 时间计算

$$g^k(\mathrm{mod}\ m)\to v,\ g^{-1}(\mathrm{mod}\ m)\to u$$

求出满足 $v\cdot u^f\equiv 1(\mathrm{mod}\ m)$ 之最小 f 记为 f_0，则

$$g^{k-f_0}\equiv 1(\mathrm{mod}\ m)$$

故

$$k-f_0=m+1-[2\sqrt{m}]-f_0\mid\varphi(m)$$

又

$$\varphi(m)=m+1-(p+q)\leqslant$$

$$m+1-2\sqrt{pq}=m+1-2\sqrt{m}\leqslant k$$

故

$$k-f_0=\varphi(m),$$

$$f_0=p+q-[2\sqrt{pq}]\leqslant(\sqrt{p}-\sqrt{q})^2+1$$

解出 $p+q=(m+1)-(k-f_0)$，由 $m=pq$ 用韦达定理易解出 p，q。

算法 WZH3 时间复杂性为

$$O(f_0(\ln m)^2)+O((\ln m)^3),$$

$$f_0\leqslant(\sqrt{p}-\sqrt{q})^2+1$$

如已求出 $L=\delta_m(g)$，则不用搜索 f_0 次，只需对满足 $L\mid k-f_0$ 之 f_0 搜索，时间降为

$$O(f_1(\ln m)^2)+O((\ln m)^3),$$

$$f_1\leqslant(\sqrt{p}-\sqrt{q})^2/L$$

当 f_0 或 f_1 为 $O(\ln m)$ 时，WZH3 为多项式时间算法，时间为 $O(\ln m)^3$），WZH3 适合一切大整数 m。

当 p，q 求出后，私匙 $a=b^{-1}\mathrm{mod}(\varphi(m))$ 易求。

3 结 论

RSA 体制中 $m=pq$，p、q 的选择原则包括①$p-1$、$q-1$ 包含大的素因子；②gcd（$p-1$，$q-1$）应比较小。这样调用算法 WZH0，WZH1 可在 $O(L\cdot(\ln m)^2)+O((\ln m)^3)$ 时间内完成对 RSA 体制密匙完全求解，至少对某一类数存在 $L=O((\ln m)^k)$，即存在多项式时间算法，一般易达到 $L=(m)^{1/s}$，$s>4$ 等。作为纯理论研究，假定对 p，q 不作任何限制，那么调用 WZH0，WZH3 可在

$$O(L\ln m)+O((\sqrt{p}-\sqrt{q})^2/L\cdot(\ln m)^2)+O((\ln m)^3)$$

时间内完全解密，特殊情况下时间降为 $O((\ln m)^k)$，k 为正整数。方法不同于椭圆曲线算法的概率型求解(随机碰撞)，也不同于数域筛法及二次筛法，通过构造 $y^2\equiv x^2 \pmod m$ 求解，求解思路放在寻求模 m 之小整数阶 L，算法显示密匙攻击可绕过大素数因子，通过攻击其“软肋”：偏小素因子入手。

参考文献：

[1] 裴定一，祝跃飞. 算法数论[M]. 北京：科学出版社，2002.

[2] 华罗庚. 数论导引[M]. 北京：科学出版社，1957.

[3] 卢开澄. 计算机密码学[M]. 北京：清华大学出版社，2002.

[4] [美] Stave Burenett, Stephen Paine. 密码工程实践指南[M]. 北京：清华大学出版社，2001.

A Novel Calculation Method for the Factorization of Large Integral Numbers and the Decoding of the RSA Cryptosystem

WANG Ze-hui

(Department of Scientific Computation and Computer Science, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: A novel calculation method, namely WZH, is introduced to factorize a series of large integral numbers. WZH can solve the standard factorization $m=p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_s^{\alpha_s}$ within the time period $O(L(\ln m)^2)+O((\ln m)^3)(L<m)$. It can also calculate some of the prime factors of p_i-1, revealing the interior structure of m. WZH can treat the RSA cryptosystem as a unique factor and solve the cryptogram p, q and α within the time period $O(L\ln m)+O((\sqrt{p}-\sqrt{q})^2/L\cdot(\ln m)^2)+O((\ln m)^3)$. This calculation method can be applied to all positive integers for complete decoding. Under certain condition, the timing complexity is $O((\ln m)^k)$, k is a postive integer.

Key words: standard factorization of lasge integral number; RSA cryptosystem; complete decoding; polynomial timing complexity