大整数分解算法综述

杨江帅
（中国电子信息产业集团有限公司第六研究所，北京 100083）

摘　要： RSA 是目前最有影响力和最常用的公钥加密算法，它广泛应用于各个领域，能够抵抗到目前为止已知的绝大多数密码攻击。RSA 密码算法的安全性基于大整数分解的困难性，因此对大整数分解问题的深入研究具有重要的理论意义和应用价值。主要概括大整数分解的研究现状，回顾当前常用的大整数分解算法，分别详细介绍它们的基本原理以及在应用方面的优缺点，最后分析展望大整数分解的研究趋势。

关键词： RSA；安全性；大整数分解
中图分类号： TN918　　**文献标识码：** A　　DOI：10.19358/j.issn.2096-5133.2018.11.004
引用格式： 杨江帅. 大整数分解算法综述[J]. 信息技术与网络安全，2018，37(11)：12-15.

Survey of large integer factorization algorithm

Yang Jiangshuai
(The 6th Research Institute of China Electronics Corporation, Beijing 100083, China)

Abstract: RSA is currently the most influential and most commonly used public key encryption algorithm. It is widely used in various fields and is resistant to most of the known attacks. The security of RSA cryptographic algorithm is based on the difficulty of large integer factorization. Therefore, the in-depth study of large integer factorization problem has important theoretical significance and application value. The article mainly summarizes the research status of large integer factorization, reviews the commonly used large integer decomposition algorithms, introduces their basic principles and their advantages and disadvantages in application, and finally analyzes the research trend of large integer factorization.

Key words: RSA; security; integer factorization

0　引言

大整数分解是数论中的一个基本问题，从其诞生到现在已有数百年历史，真正引起数学家、计算机科学家以及密码学家的极大关注是近几十年的事情，它不仅是公钥加密算法 RSA 的最直接的攻击手段，也是 RSA 安全性分析最关键的切入点，因而整数分解问题的任何一点进展都将引起密码学界的关注。大数分解问题既未被证明是多项式时间可解的 P 问题，也未被证明是 NP 完备问题，另外，大整数分解问题的研究直接影响到数论及通信领域中其他一些问题的解决，因而对其研究具有极其重要的理论意义和应用价值。

1　RSA 密码算法简述

(1) 选取大素数 p 和 q，计算模数 $n = pq$ 和欧拉函数 $\varphi(n) = (p-1)(q-1)$；

(2) 选取一个和 $\varphi(n)$ 互素的整数 e 作为公钥，求出整数 d 作为私钥，其中 d 满足 $ed \equiv 1 (\mathrm{mod}\varphi(n))$；

(3) 将明文 m 加密：$c = m^e (\mathrm{mod}n)$，其中 m 满足 $0 < m < n$；

(4) 将密文 m 加密：$m = c^d (\mathrm{mod}n)$，其中 c 满足 $0 < c < n$。

由上述过程可知，如果能将算法中的模数 n 进行分解，得到 p 和 q，那么很容易将明文 m 恢复出来，因此，有效的大整数分解算法能够破解 RSA 密码算法。

2　大整数分解算法介绍

2.1　试除法

试除法是分解整数 n 最简单直接的方法。先从最小的素数 2 开始去试除，看其是否能够整除，若能整除，然后依次再用其他的素数重复上面的试除过程，直到最后的商是素数，这样能把给定的整数 n 分解。假如只用 2 到 $\sqrt{n}$ 的所有素数去试除，根据素数计数定理，只需尝试 $O(\sqrt{n}/\log n)$ 个素数，就可以找到 n 的素因子。所以当分解的数没有大素因子时，试除法适合分解，而当分解的数较大且没有小素因子时，试除法基本上无能为力。

2.2　蒙特卡洛方法

蒙特卡洛方法[1]是一种基于随机数序列的分解整数的方法，是由 PALLARD J M 于 1975 年提出的，该方

法的基本思想如下：假设 n 是给定的待分解的整数，p 是 n 的一个素因子。$f(x)$ 是整数环 Z 上的一元多项式。首先选择随机整数 $0 < a_0 < n$，然后考虑如下序列 $\{a_0, a_1, \cdots a_i, \cdots\}$，$a_{i+1} = f(a_i) \pmod n$，如果有 $i \neq j$ 使得 $a_i \equiv a_j \pmod p$，则有 $1 < \gcd(a_i - a_j, n) < n$，$a_i \neq a_j \pmod n$，这样找到 n 的一个非平因子。

蒙特卡洛方法的实质就是在一条拟随机序列中找到两个模 p 相等的整数，这个问题与著名的生日问题有关，也就是随机选择 k 个人，其中有两个人是同一天生日的概率大于 $1/2$，求 k 至少是多少，答案是 $k = 23$，与人们的直觉不一样。同样地，为了得到两个整数模 p 同余的概率大于 $1/2$，则至少需要的整数数目 k 应满足：

$$1 - \frac{(p-1)(p-2)\cdots(p-k+1)}{p^{k-1}} > \frac{1}{2}$$

由数学中相关知识得到大约 $k = 1.2\sqrt{p}$。而多项式 $f(x)$ 和初始值 x_0 一般选择 $f(x) = x^2 + 1$ 或者 $f(x) = x^2 - 1$，$x_0 = 2$。蒙特卡洛方法适合分解具有较小素因子的整数。基于蒙特卡洛方法，BRENT P R 等人[2]成功地分解了第 8 个费马数。

2.3 $p-1$ 方法

$p-1$ 方法[3]是 POLLARD J M 在 1974 年提出的分解整数的方法。假设当 n 有一个素因子 p 且 $p-1$ 的素因子较小时，使用这种方法是比较有效的。$p-1$ 方法的分解原理来源于数论中的费马小定理。设 n 是待分解的整数，p 是 n 的一个素因子，根据费马小定理，如果 $p-1 \mid Q$，则对于与 n 互素的整数 a，有 $p \mid a^Q - 1$，因此 $d = \gcd(a^Q - 1, n) > 1$，如果 $d < n$，那么就得到了 n 的一个非平凡因子。可以通过以下方法得到 $p-1$ 的倍数 Q：给定一个上界 B，将小于 B 的所有素数乘起来得到 Q。通过 $p-1$ 方法，BRENT P R 找到了梅森数 $M_{977} = 2^{977} - 1$ 的一个 32 位的因子。

2.4 $p+1$ 方法

$p+1$ 方法[4]是 WILLAMS H C 于 1982 年提出的，$p+1$ 方法与 $p-1$ 类似。Willams 的 $p+1$ 方法是通过对 POLLARD J M 的 $p-1$ 方法做过详细研究后并使用卢卡斯序列的等价于幂运算的特性得到的。$p+1$ 方法适合分解这一类合数 n：包含一些素因子 p，且使得 $p+1$ 是由比较多的小素数相乘而得出的。

2.5 费马分解法

费马分解法[5]是借助平方差公式去分解整数，若 n 是两个整数的平方差，例如 $n = a^2 - b^2$，则可以得到 $n = (a+b)(a-b)$ 是整数 n 的一个分解。而要将 n 表示成平法差，可令 $b = 1, 2, 3\cdots$，然后去考察 $n + b^2$，看其是否为一完全平方数，如果对某一个整数 b，$n + b^2$ 为一个完全平方数，那么相应地，就得到了 n 的一个分解。当且仅当 n 是两个差不多对等规模因子乘积的时候，这种方法就很有效。因为只有当 $a+b$ 和 $a-b$ 这两个整数差不多对等规模，$b = \frac{(a+b)-(a-b)}{2}$ 才会比较小，尝试试除的次数也就相应的比较少。实际上，不需要对所有的整数 b 都去计算 $n + b^2$，因为如果 $n + b^2$ 是一个完全平方数，那么对所有整除 b 的奇素数 p，有勒让德符号 $\left(\frac{n}{p}\right) = 0$ 或者 1，所以只需要考察那些不含使得 $\left(\frac{n}{p}\right) = -1$ 的奇素因子 p 的整数 b 即可。由此可知，费马分解法适合分解当 n 含有几乎相等的两个因子的一类整数。

2.6 连分数分解法

19 世纪 20 年代，文献[6]拓展了费马费解法，其思想是：不要求存在整数 a 和 b 使得 $a^2 - b^2 = n$，而是要求它是 n 的一个倍数，即满足同余式 $a^2 - b^2 \equiv 0 \pmod n$。1931 年，Lehmer 和 Powers 根据这一思想提出了使用连分式分解整数的方法，Morrison 和 Brillhart 也在 1975 年发布了根据上面方法构造适合计算机运算的算法。连分数分解法过程如下：对于非负整数 m，令 A_m/B_m 为 $\sqrt{kn}$ 的 m 次渐进分数，则有 $A_m^2 - knB_m^2 = (-1)^{m+1}Q_m$，其中 Q_m 可由一个简单计算的递推关系得出。因此，对于每个 m，得到：

$$A_m^2 \equiv (-1)^{m+1}Q_m \pmod n,$$

也就是说 $(-1)^{m+1}Q_m$ 是一系列模 n 的二次剩余。在这些二次剩余中，可以选出一些组成一个集合 S，如果集合 S 中各个元素的乘积又刚好是一个完平方数，那么得到了一个形如 $a^2 \equiv b^2 \pmod n$ 的同余式，这样也就得到了 n 的一个分解。

2.7 二次筛选法

二次筛选分解法是 POMERANCE C[6]在线性筛选法的基础上于 1981 年提出的。二次筛选法用到多项式选取 $\varpi(x) = (\lfloor\sqrt{n}\rfloor + x)^2 - n$，这里 $|x| < n^{\varepsilon}$，ε 为一任意小的数。二次筛选法大体分为选择分解基、筛过程(包含选筛区间)、建立矩阵、寻找线性关系等几个重要步骤，其中最耗时的部分为筛过程，过程的快慢与选取的分解基和筛区间有很大关系。由于建立的是一个很大的稀疏矩阵，这必然占用计算机的很大一部分资源，中间用到的高斯消元需要占用巨大的存储空间。二次筛法有多种优化的方法，使用多个一元二次多项

式代替 $\varpi(x)$，从而使每个多项式的筛选区间远小于使用单个的 $\varpi(x)$，从而提高算法的效率。

2.8 椭圆曲线分解法

椭圆曲线分解法[7]是由荷兰科学家 LENSTRA H W 于 1987 年提出的一种整数分解法，这种方法使用了椭圆曲线的群结构，随机选取一条椭圆曲线 $E: y^2z = x^3 + axz^2 + bz^3$，如果 (x,y,z) 满足该方程，且 $c \neq 0 \bmod p$，则 (cx,cy,cz) 也满足该方程，因此 (x,y,z) 和 (cx,cy,cz) 可以看成是等价的。用 $(x:y:z)$ 表示包含 (x,y,z) 一类等价点的等价类。在这个群 $E_{a,b}$ 中的加法零元 O 是 $(1:0:1)$，此时 $z \equiv 0 \pmod p$，如果在某一步运算中得到了加法零元 $O = (x,y,z)$，那么通过计算 $\gcd(z,n)$ 得到了一个因子，可能将 n 分解。BRENT P R 使用椭圆曲线方法成功分解了第十个和第十一个费马数，Wagstaff 于 2012 年使用此方法分解出了一个 79 位数的因子。

2.9 数域筛选法

数域筛选法[8]是目前最快的整数分解算法，它是由 POLLARD J M 首先提出的。这个算法仍然是以解决 $a^2 \equiv b^2 \pmod n$ 为目的的，首先选取一个合适的整系数多项式 $f(x)$ 和一个有理整数 m，使得 $f(m) \equiv 0 \pmod n$，α 是多项式 $f(x)$ 的一个复根，可以定义一个从 $\mathbf{R} = \mathbf{Z}[\alpha]$ 到整数域的一个环同态：$\varphi: \mathbf{R} \to \mathbf{Z}_n$，$\varphi(\alpha) = m \pmod n$，则有：

$$X^2 = \varphi(\beta)^2 = \varphi(\beta^2) = \varphi\left(\prod_i a_i + b_i\alpha\right) = \prod_i (a_i + b_i m) = Y^2 \pmod n$$

这样求得同余式 X,Y 即为所需。为了得到满足 $\prod_i (a_i + b_i\alpha) = \beta^2$ 和 $\prod_i (a_i + b_i m) = Y^2$ 的整数对 (a,b)，需要将 $\mathbf{Z}[\alpha]$ 上的平滑和 $\mathbf{Z}$ 的平滑概念结合起来，寻找整数对 (a,b) 使得 $a + b\alpha$ 在 $\mathbf{Z}[\alpha]$ 的代数因子库上平滑，同时 $a + bm$ 在 $\mathbf{Z}$ 的有理因子库上平滑，当找到足够多的在两个因子库上都平滑的整数对之后，在 $\mathbf{Z}[\alpha]$ 和 $\mathbf{Z}$ 上同时产生一个平方值。数域筛选法包括多项式的生成、数域的生成、分解基的生成、筛过程、分解基矩阵的生成、同态映射、求根等步骤，其中分解基矩阵的生成最耗时。一般情况下努力做到所选的代数数域是唯一分解整环，但大多数时候并不是，那么这个时候需要用到理想数，因为任何代数数都可以分解为素理想。但除了一般形式的模数分解之外，还有类似于 $n = r^e \pm s$ 这样特殊的数需要分解，通常采用特殊数域筛选法。从历史分解记录来看，随着计算机计算能力的扩大，差不多每 10 年会多分解掉 100 比特的数。从实际情况看包括数域筛选法之前的算法都存在复杂性高、算法流程复杂、耗资巨大等实质性问题，较好的结果仅仅是对数域筛选法做出一些改进，没有在算法或者架构上做出跨越式的进展。

3 大整数分解算法比较分析

根据素数定理，试除法的时间复杂度最快可达到 $O(n^{1/2}/\log n)$；POLLARD J M 的蒙特卡洛方法的时间复杂度为 $O(n^{1/4})$；连分式分解法等其他基于同余的分解法的启发式时间复杂度为 $O(e^{C\sqrt{\log n \log\log n}})$，其中 C 为一较小的常数，而且需要想当大的存储空间；椭圆曲线法的启发式时间复杂度为 $e^{(2+o(1))\sqrt{\log p \log\log p}}$，其中 p 为 n 的最小素因子；二次筛选法的启发式时间复杂度为 $O(e^{(1+o(1))\sqrt{\log n \log\log n}})$，在实现过程中会产生大规模的 01 稀疏矩阵，需要大量内存，普通计算机无法满足；数域筛选法的时间复杂度为 $e^{O((64/9+o(1))^{1/3}(\log n)^{1/3}(\log\log n)^{2/3})}$。目前存在的已知算法中，虽然对于一般数的分解还很困难，但对于具有某些特殊形式的数的分解有时还是相对容易的，比如，试除法、$p \pm 1$ 方法、蒙特卡洛方法对于那些具有小素因子的整数分解速度就很快，而 Fermat 方法适用于分解那些是两个相近因子乘积的整数，数域筛选法对形如 $r^e \pm s$ 的整数（其中 r、s 较小，e 可能很大）分解起来比一般整数要快。

4 结论

到目前为止，对于大整数分解问题，大部分研究工作都是对现有算法的改进，因此要找到新的算法，必须应用新的理论知识。文献[9]第一次将整数分解与二项式系数的和联系在一起，虽然还没有给出算法的复杂度分析，但却给出了研究的新思路和新方向。另一方面，一些基于量子的整数分解算法充分利用了量子计算机强大的并行计算能力，使得大整数分解存在多项式时间算法，但现有的量子计算机尚不能实现有实际意义的量子算法。从总体上看，未来大数分解研究方向会在并行的前提下，尽可能提高系统效率，有可能是未来新的发展方向。

参考文献

[1] POLLARD J M. A monte carlor method for factorization [J]. Bit Numerical Mathematics, 1975, 15(3): 331-334.

[2] BRENT P R, POLLARD J M. Factorization of the Eigth Fermat Number [J]. Mathematics of Compuutation, 1981, 36(154): 627-630.

[3] POLLARD J M. Theorems on factorizaiton and primality testing [C]. Mathematical Proceedings of the Cambridge Philosophical Society, 1974, 76(3): 521-528.

[4] WILLIAMS H C. A p+1 method of factoring [J]. Mathematics

of Computation，1982，39(159)：225-234.
[5] KOBLITZ N. A course in number theory and cryptograph [M]. Springer-Verlag，1987.
[6] POMERANCE C. The quadratic sieve factoring algorithm [C]. Advance in Cryptology. Lecture Notes in Computer Science，1984，209(4)：169-182.
[7] LENSTRA H W. Factoring integers with elliptic curves [J]. Annals of Mathematics，1987，126(3)：649-673.
[8] LENSTRA A K，LENSTRA H W. The development of the number field sieve [M]. Berlin：Springer，1993.
[9] DENG Y P，PAN Y B. The sum of binomial coefficients and integer factorization [Z]. Integer，2016，16.

(收稿日期：2018-09-30)

作者简介：
杨江帅(1989－)，男，博士，工程师，主要研究方向：信息安全。

(上接第8页)

证明了本文构造的 T 统计量是合理的，从而进行非重叠模板匹配检验后，可得出唯一结论。

参考文献

[1] 张咏. 随机数发生器和随机数性能检验方法研究 [D]. 成都：电子科技大学，2006.
[2] 刘康. 伪随机序列随机性测试系统的设计与实现 [D]. 石家庄：石家庄铁道大学，2016.
[3] 吴若雪，梁笑轩，郑智捷. 改进型 NIST 测试对 ZUC 算法随机序列的可视化检测 [J]. 计算机应用研究，2018，35(1)：253-256.
[4] NIST STS. A statistical test suite for random and pseudorandom number generators for cryptographic applications [Z]. NIST Special Publication 800-22 Revision 1a，2010.
[5] 师国栋，康绯，顾海文. 随机性测试的研究与实现 [J]. 计算机工程，2009，35(20)：145-147，150.
[6] 张永强，李顺波，屈帅，等. NIST 随机性检测方法及应用 [J]. 电脑知识与技术，2014，10(26)：6064-6066.
[7] 冯秀涛. 祖冲之序列密码算法 [J]. 信息安全研究，2016，2(11)：1028-1041.
[8] ORHANOU G，HAJJI S E，BENTALEB Y. Snow 3G stream cipher operation and complexity study [J]. Contemporary Engineering Sciences-Hikari Ltd，2010，3(3)：97-111.
[9] BERBAIN C，BILLET O，CANTEAUT A，et al. Sosemanuk，a fast software-oriented stream cipher [Z]. 2005.
[10] 孙全玲，吕虹，陈万里，戚鹏. m 子序列的密码学性质研究 [J]. 计算机应用研究，2018，35(1)：245-247，256.

(收稿日期：2018-09-30)

作者简介：
王超(1982－)，男，博士，工程师，主要研究方向：信息安全。
温涛(1992－)，男，硕士，主要研究方向：信息安全。
段冉阳(1994－)，男，硕士研究生，主要研究方向：信息安全。

(上接第11页)

[9] BERBAIN C，BILLET O，CANTEAUT A，et al. Sosemanuk，a fast software-oriented stream cipher [Z]. 2005.
[10] DECANNIERE C. Trivium：a stream cipher construction inspired by block cipher design principles [C]. International Conference on Information Security. Springer，Berlin，Heidelberg，2006：171-186.

(收稿日期：2018-09-30)

作者简介：
王超(1982－)，男，博士，工程师，主要研究方向：信息安全。
范国浩(1991－)，男，硕士，主要研究方向：信息安全。
付宝仁(1995－)，男，硕士研究生，主要研究方向：信息安全。