对大整数 $n = pq$ 分解的一个有效的搜索算法

董庆宽,傅晓彤,肖国镇
(西安电子科技大学综合业务网国家重点实验室,陕西西安 710071)

摘 要: 本文通过构造一个简单的基于调差思想的搜索算法和一个快速的开方算法对满足一定条件的大整数 $n = pq$(p、q 为大素数)进行快速分解.从而指出基于因子分解的密码体制中存在着相当多的弱密钥,而且很难避免选取这些弱密钥.这对于我们分析基于因子分解的公钥体制的安全性是很有意义的.

关键词: 因子分解; 密码学; RSA
中图分类号: TN918.1 **文献标识码**: A **文章编号**: 0372-2112 (2001) 10-1436-03

An Effective Searching Algorithm for Factoring Large Integer n = pq

DONG Qing-kuan, FU Xiao-tong, XIAO Guo-zhen
(*National Key Lab of Integrated Service Networks, Xidian University, Xi'an, Shanxi 710071, China*)

Abstract: An effective searching algorithm based on a new idea, difference-adjusted, and a fast algorithm for calculating square root are presented to factorize large integer $n = pq$ (p、q are large primes) in some conditions. Consequently, a great deal of weakness existing in the cryptosystem based on the factorization are found and it is very difficult to avoid selecting them in practice. It can be well used in the analysis of the secure in the public key cryptosystem based on the factorization.

Key words: factoring; cryptography; RSA

1 引言

在过去的二十多年中,从事数论与密码学的研究人员对整数因子分解问题,特别是大整数 $n = pq$(p、q 为大素数)的分解,从不同的角度进行了分析,取得了很大的进展[1].通常有两种思路:一种是与具体体制参数结合进行攻击,如 RSA 体制中的一些参数(e, d),这里 $ed \equiv 1 \bmod \phi(n)$,如果 e, d 选择不当,n 可被快速分解[1,2];另一种是直接对 $n = pq$ 进行分解已有很多著名算法对一定条件下的大整数 n 可快速分解[2,3],不再赘述.

设 $n = pq$,其中 p, q 为大素数,则在基于因子分解的公钥密码体制中参数 p, q 必须满足如下条件[2]:(Ⅰ) p, q 必须为强素数.(Ⅱ) $p-1$ 与 $q-1$ 的最大公因子要小.(Ⅲ) p、q 要足够大.(Ⅳ) p、q 之差要足够大。这里条件(Ⅳ)是远远不够的,在新提出的搜索算法中将指出,即使 p 与 q 的差值很大,如果选取不当也可被快速分解.

本文通过构造一个简单的基于调差思想的搜索算法,以调差因子 k 为变量进行搜索,找出更多的不安全的(p, q)对,这是先前 RSA 等基于因子分解的密码体制参数选择条件所不能及的,这些由搜索所产生的大量的弱密钥使得基于该难题的密码体制的安全性受到质疑.国内外的研究中虽对此两参数的差值有所提及,但均没有给予足够重视.论文的第 2 节具体叙述了搜索算法;第 3 节给出了算法的有效性分析和时间复杂度分析;第 4 节给出三个扩展方法以加大搜索成功的可能性;文章的最后给予总结.

2 算法描述

2.1 调差思想

设 $n = pq$(p, q 为大素数),令 $k = st$ 称为调差因子,s、t 为正整数,则如下恒等式成立:

$(sp + tq)^2 = (sp - tq)^2 + 4stpq$,即

$$(sp + tq)^2 = (sp - tq)^2 + 4kn \tag{1}$$

令正整数 m 满足 $$m^2 < 4kn < (m+1)^2 \tag{2}$$

现在做如下减法运算, $$(m+1)^2 - 4kn \tag{3}$$

假设调差值 $abs(sp - tq)$ 满足条件:

$$(sp - tq)^2 < (m+2)^2 - 4kn \tag{4}$$

那么由式(1)和(2)得 $m^2 < (sp + tq)^2 < (m+2)^2$,即 $(sp + tq)^2 = (m+1)^2$.因此在条件(4)下,式(3) $= (sp - tq)^2$ 是一个平方数.可见调整 s、t 的值,条件(4)成立时,式(3)即是一个整数的平方,设为 x^2.由于 $k = st$,所以调整 s、t 相当于调整调差因子 k.实际运算中将以 k 为变量对 $(m+1)^2 - 4kn$ 进行搜索,若式(3)是一个整数的平方,即 $(m+1)^2 - 4kn = x^2$,由欧几里德算法,n 得以分解.

2.2 开方算法

该算法即是求满足 $m^2 \le N < (m+1)^2$ 的正整数 m.从 m

收稿日期:2000-06-02;修回日期:2001-04-26
基金项目:国家自然科学基金(No. 60073051);基于数学的密码理论研究(No. G1999035804)

的高比特位开始依次确定 m 的各比特位的值. 以$|x|$表示 x 的比特长度,设$|m| = t$,$|N| = L$,易知,当 L 为偶数时有 $t = L/2$,反之有 $t = (L+1)/2$. 记 $m = m_{t-1}2^{i-1} + m_{i-2}2^{i-2} + \cdots + m_1 2^1 + m_0$,下面给出求 m 的一个归纳性描述. (Ⅰ)对于第一个有效比特位 m_{t-1},显然有 $m_{t-1} = 1$. (Ⅱ)现假设前 $i-1$ 个比特位的值均已求出,记 $s_{i-1} = m_{t-1}2^{t-1} + m_{t-2}2^{t-2} + \cdots + m_{t-i+1}2^{t-i+1}$,我们来求第 i 比特位的值 m_{t-i}. 设 $m_{t-i} = 1$,则有 $s_i = s_{i-1} + m_{t-i}2^{t-i} = s_{i-1} + 2^{i-1}$. 然后计算 s_i^2 并与 N 比较大小,如果 $s_i^2 \le N$ 则假设成立,取 $m_{t-i} = 1$,否则,取 $m_{t-i} = 0$,相应的 $s_i = s_{i-1}$. 由此可确定 m 的全部比特位的值($m = s_t$). 显然,在计算第 i 比特位时,只有前 i 比特为有效位,后面均为 0. 基于此特性及以上归纳性描述,下面给出一个有效的递归处理,使计算量得到极大的简化.

令 $u_i = s_i/2^{t-i} = 2^{t-1} + m_{t-2}s^{i-2} + \cdots + m_{t-i+1}2^1 + m_{t-i}$,则 $u_t = s_t = m$. $s_1 = 2^{t-1}$,对应的有 $u_1 = 1$,$u_1^2 = 1$. 假设 u_{i-1}及 u_{i-1}^2已求出,现在求 u_i 及 u_i^2(相应的即求 m_{t-i}). 设 $m_{t-i} = 1$,则 $u_i = 2u_{i-1} + 1$,$u_i^2 = (u_{i-1}^2 + u_{i-1})2^2 + 1$. 易知,$s_i$ 的最低有效位为第 i 位,所以 s_i^2 的最低有效位后面的 $2t-2i$ 个比特位的值均为 0. 这样在比较 s_i^2 与 N 的大小时,只需将 u_i^2 左移 $2t-2i$ 位后比较 u_i^2 与 N 的大小. 如果大于 N,则取 $u_i = 2u_{i-1}$,$u_i^2 = 2^2 u_{i-1}^2$(即取 $m_{t-i} = 0$),否则按假设取值. 直到求出 u_t 及 u_t^2. 在最后一次比较时,如果 $u_t^2 = N$,则说明 N 是平方数,$m = u_t$ 是平方根,并置 Flag = 1,否则置 Flag = 0,此时 $m = u_t$ 即为所求. 下面我们给出计算流程:

置 Flag = 0; $u = 1$; usquare = 1;// usquare 表示 u^2

从 $i = 2$ 到 t 做($t = L/2$)

{ y = add(u, usquare); y = add($y \ll 2$, 1); $u = u \ll 1$;// 计算 $2^2(u + u^2) + 1$,u 乘 2

如果 compare($y \ll (2t-2i)$, N) ≤ 0(表示前者小于或等于后者), u = add(u, 1), usquare = y;

否则 usquare = usquare $\ll 2$;// u^2 乘 4}// 循环结束

$m = u$; msquare = y;// msquare 表示 m^2

如果 compare(y, N) = 0(表示相等)则置 Flag = 1;返回 m、msquare 和 Flag 的值.

为分析方便把移位看作加法(移位比加法简单). 因加 1 运算均是在左移位后进行,只相当于最低比特位取反,可忽略不计. 所以最坏情况下全过程需 $4(t-1)$ 次加法.

2.3 搜索算法的流程

由前述调差思想有下面流程:

置 $N1 = 4n$; $N = 0$;

从 $k = 1$ 到 Kc 做 // Kc 为某一个限,它的取值留在第 3 节说明

{ N = add(N, N1);// 根据递推式 4kn = 4(k - 1)n + 4n 使每次乘 k 运算变为一次简单的加法 unin = sqrt(N);// 开方; unin 为结构体,包含三个域值 unin. m、unin. msquare 及 unin. Flag

m2 = unin. m ≪ 1; m2 = add(m2, 1); y = add(unin. msquare, m2);// 计算 $m^2 + 2m + 1$

y = sub(y, N);// 计算 $m^2 + 2m + 1 - N = (m+1)^2 - N$,对 y 开方

如果 T. Flag = 1 则此时 N = 4kn 为平方数,跳出循环;

}// 注:如 3.2 节分析,其中的加 1 运算可忽略不计.

如果 $k \le Kc$,则说明搜索成功,取 $x = T.m$, $m = unin.m$, 如 2.1 节所述,求出.

3 算法分析

3.1 有效性分析

在 2.1 节已经给出,只要找出某个 $k(k = st \le K_c)$,使式(4)成立就不难分解 n. 现在从(p, q)对的选取进行分析. 在实际密码体制中所选的(p, q)对如能满足式(4)且 $k = st \le K_c$ 那么就称这一(p, q)对为不安全对,即弱密钥. 对某一个固定的素数 q,试给出素数 p 的一个集合,使相应的(p, q)对在前述算法下是不安全对.

由式(4)可得 $abs(sp - tq) \le \sqrt{(m+2)^2 - 4kn}$

于是有 $t/s\,q - 1/s\,\sqrt{(m+2)^2 - 4kn} < p < t/s\,q + 1/s\,\sqrt{(m+2)^2 - 4kn}$ (5)

固定 s, t 和 q,且 $k = st \le K_c$,改变 p 值,将所有满足式(5)的 p 值归入一个集合并记为 $U_{(s,t,q)}$,称为关于 q 的不安全的 p 集合. 改变 s, t 的取值可得到不同的不安全 p 集合. 注意这里 $gcd(s, t) = 1$,这是由算法的执行特性决定的. 假如 $gcd(s, t) = d > 1$ 且式(4)成立,令 $s = s_0 d$, $t = t_0 d$,可以证明 $k_0 = s_0 t_0$ $(< k)$ 时式(4)亦成立,搜索到 k_0 时算法已终止.

综上对于某固定素数 q,构成不安全(p, q)对的全部素数 p 的集合描述如下:

$U_q = \bigcup_{st \le K_c,\, gcd(s,t)=1} U_{(s,t,q)} = \{$素数 $p \mid t/s\,q - 1/s \cdot \sqrt{(m+2)^2 - 4kn} < p < t/s\,q + 1/s\,\sqrt{(m+2)^2 - 4kn}$, $k = st \le K_c$, $gcd(s, t) = 1\}$. 下面对集合 $U_{(s,t,q)}$ 的大小作进一步说明. 易知$(m+2)^2 - 4kn > 2m + 3$,保守的估计以 $2m+3$ 代$(m+2)^2 - 4kn$. m 随 p 的增减而增减,增加 p 值至不满足式(5),取此时 m 记为 M,减少 p 值至不满足式(5),取此时 m 记为 M_{min},那么至少在区间$(t/s\,q - 1/s \cdot \sqrt{2M_{min} + 3}$, $t/s\,q + 1/s \cdot \sqrt{2M + 3})$上的素数 p 属于集合 $U_{(s,t,q)}$,区间长度 $Len = 1/s \cdot (\sqrt{2M_{min} + 3} + \sqrt{2M + 3})$. 由素数定理 $\pi(n) \ge \ln 2(n/\ln n)$[4] 得素数密度 $\rho = \pi(n)/n = \ln 2/\ln n$. 由此可得在区间$(t/s(q - \sqrt{2M_{min} + 3})$, $(t/s(q + \sqrt{2M + 3}))$上素数个数大于 $\rho \times Len$. q 很大时该值是很大的,也就是说有相当多的不安全(p, q)对.

现在举例说明. 取 $n = pq$ 为 512 比特,则 p, q 约为 256 比特,$M = \sqrt{4stn} > 257$ 比特(不计 s, t 的贡献),那么 $\sqrt{2M+3}$大于 129 比特,取最小值 2^{128}. 为分析方便取 $Len = 1/s \cdot (\sqrt{2M+3}) > 1/s 2^{128}$,$\rho = \ln 2/\ln q \approx 2^{-8}$. 这时 $\rho \times Len = 1/s \cdot 2^{120}$. 当 $s = t = 1$ 时即为第 2 节中条件(Ⅳ),相应的不安全集合 $U_{(1,1,q)}$的阶大于 $\rho \times Len > 2^{120}$. 当 $s = 1$, $t = 2$ 时$|U_{(1,2,q)}| > \rho \times Len \times 2^{120}$,此时 p 与 $2q$ 的差值小,而 p, q 的比特长不相等,差值很大,这说明仅凭 p, q 的差值大,是不能保证安全

的,只要调差的结果小,n 就会被快速分解. 相应的当 $s=2$ 时有 $|U_{(2,1,q)}|>Len\rho>1/2\ 2^{120}=2^{119}$,$|U_{(2,3,q)}|>Len\rho>1/2\ 2^{120}=2^{119}$ 等等,其 p,q 的差值都很大,却可被快速分解. 由上可见在最保守的估计下,对于某一固定素数 q 的不安全对的个数也达到了相当大的数量级(上例中至少是大于 2^{120}),所以说弱密钥是相当多的. 在实际体制中不可能穷尽 s,t 的值去进行验证. 因此该搜索算法有很大的有效性,它对因子分解体制构成了一定威胁.

3.2 时间复杂度分析

下面,首先计算 2.3 节所给流程中一次循环的复杂度. 移位和减法至少不比加法运算量大,所以均看作同比特长加法运算. 流程中需要两次开方运算,现在我们设 $N=4kn$ 的比特长为 L,由 2.2 节分析易知,第一次开方需 $4(t-1)$ 次 L 比特长加法运算(t 为 m 的比特长),第二次开方中被开方数 $y=(m+1)^2-4kn\leq 2m$,因而取 y 的比特长 $l=t+1$,不妨取 l 为偶数,则需 $4(l/2-1)$ 次 l 比特长加法运算,它相当于 $4(l/2-1)/2$ 次 L 比特长加法(这仅当 L 较大时成立,例如是计算机字长的 5 倍以上),此即 $t-1$ 次 L 比特长加法. 所以综上所述共需 $1+4(t-1)+1+1+1+(t-1)=5t-1$ 次 L 比特长加法运算(各项依次为:加、开方、移为、加、减和开方,其中移位和减法作加法计,加 1 运算忽略不计.).

现设 n 的长度为 512 比特,$K_c=2^{21}-1$(此 K_c 即前述的 K_c 限,由计算资源和算法的时间复杂度决定),则 $N=4kn$ 的长度 L 约为 534 比特,对于一般的每秒可做 1000 万次基本加法运算的计算机,仅需约 1.3 小时. 现代计算机的速度远大于此,如果再配合以网络的分布计算,搜索范围将急剧扩大. 由以上分析可见该算法的快速性和有效性.

4 算法的扩展

以 p^2 代 p,q^2 代 q,则 $(m+1)^2-4kn^2$,进行搜索(称为平方调差,其中 $m=\sqrt{4kn^2}$)会增大搜索成功的概率. 因为对某些 (p,q) 对,虽然 $Abs(sp-tq)$ 的调差值很大,但 $Abs(sp^2-tq^2)$ 的值却很小,满足条件式(4). 当然平方调差成功的概率相对要小多了,由于平方因子的存在,比特长加大,计算量也加大了. 另外,在每次循环中,如果不成功可进一步计算 $(m+2)^2-4kn$ 是否是平方数(称为横向搜索),这只需要在 2.3 节的流程中增加一次开方和判断即可. 横向搜索最好不超过 $(m+3)^2-4kn$,否则计算量加大,效率并不高. 如在前述方法中未找到适合的 k,有可能是因为 p,q 相差太大,这时我们可以在 n 上乘以一个因子 r,去搜索 $(m+1)^2-4krn$(其中 $m=\sqrt{4krn}$),这时 p 与 rq(或 rp 与 q)相差很小,有可能搜索成功.

5 结束语

在这篇文章中我们详细论述了调差算法并简要介绍了平方调差、横向搜索和乘因子等加大搜索成功概率的扩展算法,给出了一个快速开方算法. 从有效性分析中可以看出第 1 节中的素数 p,q 的选择条件是远远不够的. 同时也很难证明实际体制中所选 (p,q) 对在上述的搜索算法下是安全的,由此可见该算法对基于因子分解的公钥体制构成了一定的威胁. 因此怎样选取安全的 (p,q) 对是进一步值得研究的问题.

参考文献:

[1] D Boneh. Twenty years of attacks on the RSA cryptosystem [J]. j-NAMS 46 n.2,February 1999:203-213.

[2] 王育民,刘建伟. 通信网的安全——理论与技术 [M]. 西安:西安电子科技大学出版社,1999:197-199.

[3] D R Stinson. Cryptography: Theory and Practice [M]. CRC Press, 1995.

[4] 珂召,孙奇. 数论讲义上册 [M]. 北京:高等教育出版社. 1998:87-88.

[5] 李继红,肖国镇. 一种新的可转换不可否认签名方案的改进 [J]. 西安电子科技大学学报,1999,26(1):22-25.

作者简介:

董庆宽 男. 1973 年 11 月出生于辽宁省海城市. 1998 年 7 月毕业于西安电子科技大学通信工程专业,获工学学士学位,1998 年 8 月开始在西电信息保密研究所攻读密码学硕士学位,并于 2000 年 2 月开始提前攻读密码学博士学位. 主要研究兴趣为密码学,信息安全及信息伪装技术.

傅晓彤 女. 1977 年 3 月出生于陕西省丹凤县. 1999 年 7 月毕业于西安电子科技大学计算数学及其应用软件专业,获理学学士学位. 1999 年 8 月开始在该校信息保密研究所攻读密码学硕士学位. 主要研究兴趣为公钥密码学,网络安全及信息伪装技术.