计算机科学2002Vol.29№.8

基于合数标识法的数字签名

Digital Signature Based on Marking Method of Compsite Number

吴李瀚　陈四清　吴中福
（重庆大学计算机学院　重庆400044）

Abstract This paper explains how to marking out the composite number on the natural number axis, and finds out high prime numbers which are applied in RSA Digital Signature method.

Keywords Prime number, Marking of composite number, Digital signature

一、引言

数字签名是用公钥能顺利地解密一个数据，则该数据一定是私钥加密的，它解决了密码学中的认证和非否认问题。认证允许一个人在电子世界中确认数据和身份；非否认则防止一个人否认其在电子世界中的行为。在电子世界，数字签名提供了任何其他方式难以实现的安全能力。自1985年，ELGAmal第一次在有限域上基于离散对数问题设计了ELGAmal数字签名方案，之后，又有很多好的方案涌出，如今最流行的用于数字签名的公钥算法还是RSA，它容易理解且易于实现，是第一个较完善的公开密钥算法。

基于寻找大素数的数字签名是一类纯数学基础理论与计算机技术相结合的一个零身份认证的数字签名技术。一些成熟的数字签名技术如RSA、DSA、Rabin、Williams、ELGAmal等都归于此类。欧洲委员会的信息社会技术IST，在欧洲21世纪密码候选标准中，规划出资33亿欧元支持的NESSIE(数字签名，完整性和加密欧洲方案)一项工程中的七个方案(ACE Sign；ECDSA；ESIGN；FLASH；QUARTZ；RSA—PSS；SFLASH)[4]中，基于寻找大素数方法仍占很大比重。

RSA的安全是基于大数分解的难度，因此要增加破译的时间开销，必须寻找出更大更新的大素数。因为并非所有的素数都会对你选取的公钥e起作用，在找到两个合适的素数之前必须舍弃一些素数，那么这两个素数必须从足够大的集合中进行选取。但现在实际使用的产生大数素数的方法仍不能满足人们的期望，所以考察新的产生大数素数方法是非常重要的。文中将引用一种至今尚未在数字签名算法中采用的寻找大素数方法，即合数标识法[1]，该方法具有简易、快速、准确的特点。它能充分利用计算机强大的储存、计算和查询的功能。它在标识[0,N]的合数的同时，也完成了对[0,N]内所有素数的标识，从而可以快速准确地确定该区间的全部素数，并建立一份规范的素数表，便于查找和调用。

二、数字签

数字签名通过单向散列(HASH)函数对要传送的消息M进行处理得到消息摘要(Message Digest)H(M)，然后使用消息发送者的私有密钥D对摘要加密得到签名DH(M)，再将消息和加密的摘要一起发送给接收者。散列函必须是有如下性质：

(1)H能用于任何大小的数据分组；

(2)H能产生等长输出H(M)；

(3)对任何给定的M，H(M)要相对易于计算，能快速得出H(M)；

(4)对任何给定的h，寻找x使得H(x)=h在计算上是不可行的，这就是单向性质；

(5)抗冲突，分弱抗冲突和强抗冲突。任何给定分组M_1，寻找$M_1 \neq M_2$，使得$H(M_1)=H(M_2)$在计算上是不可行的称为弱抗冲突。任意两个M_1，M_2使得$H(M_1)=H(M_2)$在计算上也是不可行的称为强抗冲突；

为此散列函数的目的是为消息产生一个“数字指纹”(Digital Fingerprint)，用它来判断该文档有无被修改过。通过数字签名的发散性就可知道消息是否被篡改过。

数字签名由签名和检验两部分组成，签名和检验过程如图1所示。

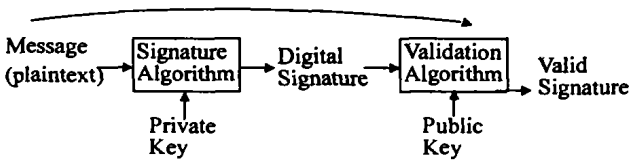


图1

用RSA算法来进行数字签名是算法作用于消息摘要上形成的。如Alice要发送一份有自己签名的消息M出来，先通过HASH函数进行摘要得到H(M)，再用私钥D_A加密得到签名的$S_A(M)$，有$S_A(M)=D_A(H(M))$。接收方收到消息M及签名$S_A(M)$后，再计算H(M)，然后使用Alice的公钥E_A解密签名。如消息完整，就有$H(M)=E_A(S_A(M))=E_A(D_A(H(M)))=H(M)$[3]。

RSA是迄今为止理论上最为成熟完善的一种公钥密码体制，它的安全基于大数分解的难度，体制构造是基于Euler定理。

先选取两个大素数p和q为100到200位的十进制数。为了获得最大程度的安全性，两数的长度一样。

计算：$n=pq$

由Euler定理得知：$\Phi(n)=(p-1)(q-1)$

然后随机选取加密密钥e，使得e和$\Phi(n)$互素。用Euclid扩展算法计算解密密钥d。

$ed\equiv 1 \pmod{(p-1)(q-1)}$

则　$d\equiv e^{-1} \pmod{(p-1)(q-1)}$

吴李瀚　硕士生，主要研究方向为网络安全。陈四清　副教授。吴中福　博士导师。

而(d,n)=1,e 和 n 是公开密钥,d 是私人密钥。两个素数 p 和 q 不再需要,它们应被舍去,但绝不可泄露

加密消息 m 时,首先将它分成 n 个小的数据组(采用二进制数,选取小于 n 的2的最大次幂),也就是说,p 和 q 为100位的素数,那么 n 将有200位,每个消息分组 m_i 应小于200位长(如加密固定的消息分组,左边可用0来填充)。加密后的密文 C,将由相同长度的分组 C_i 组成。

加密:$C_i \equiv m_i^e \pmod n$

解密:$m_i = c_i^d \pmod n$

由于

$$c_i^d = (m_i^e)^d = m_i^{ed} = m_i^{k\phi(n)+1} = m_i * m_i^{k(p-1)(q-1)}$$
$$= m_i * l = m_i \quad \text{全部} \pmod n$$

这个公式能恢复出明文。现小结如下:

公钥 n: n=pq p,q 两素数,必须保密。e:e 和 Φ(n)互素

私钥 d:$d \equiv e^{-1} \pmod{(p-1)(q-1)}$

加密 $c \equiv m^e \pmod n$

解密 $m \equiv c^d \pmod n$

从一个公开密钥和密文中恢复出明文的难度等价于分解 n 为两个大素数之积。所以大部分关于 RSA 密码分析的讨论都集中在对 n 进行因子分解上。给定 n 确定 Φ(n)就等价于对 n 进行因子分解。给定 e 和 n 时使用目前已知的算法求出 d 似乎在时间开销上至少和因子分解问题一样大。因此我们可以把因子分解的性能作为一个评价 RSA 安全性的基准。而为了避免选择容易分解的数值 n,算法发明人建议对 p,q 施加以下限制:

(1)p 和 q 的长度应该只差几个数字;

(2)(p−1)和(q−1)都应该包含大的素因子;

(3)(p−1,q−1)应该很小。

综上所述,RSA 的关键是寻找一对能较好地满足安全需要的大素数。公开密钥的算法需要大素数,实际使用的网络出于安全考虑也需要许多这样的大素数。

三、寻找大素数的合数标识

1. 若干准备

我们知道,自然数包含素数,合数,还有1。

关于任一自然数能够唯一分解为若干素数积的形式,有著名的算术基本定理:任何一个大于1的自然数 a,都有:

$a = p_1^{\alpha_1} \cdots p_i^{\alpha_i} \cdots p_n^{\alpha_n}$

其中,p_i 是一组各不相同的素数,α_i 是正整数,$n \geqslant 1$。

由此定理可知,合数 a 是基于素因子 P_i 的幂指数形式所表示,在合数标识法中称为标识。下文要用到标识矩阵$\{p_i\}$。

$\{p_i\}$为 $p_i * (1\ 2\ 3\ 4\ 5\ 6 \cdots)$

表示的是一个无穷序列:

$p_i * 1 \quad p_i * 2 \quad p_i * 3 \quad p_i * 4 \quad p_i * 5 \quad p_i * 6$

其中,p_i 为素数。为了解决寻找大素数问题,还要引用如下矩阵。

W_1:
2(3 4 5 6 7 ……)
3(3 4 5 6 7 ……)
5(5 6 7 ……)
7(7 ……)
……
p_i(p_i p_{i+1})

2. 求任意多个素数的数学模

这里主要是构造一把“标识尺”。抽象地设想,有一把和标识矩阵$\{p_i\}$相应的刚性标识尺$\{p_i\}$。其一端为原点 O,另一端向右无限远处伸延。该尺以0为起点,上有与被标识的对象一自然数列轴相同的刻度与标数。自然数列轴上依序出现的标数排列起来,就是自然数序列。除此以外,在尺上还用记号标出矩阵$\{p_i\}$的所有元素,即由$\{p_i\}$表示的合数。这只需把尺上那些与矩阵$\{p_i\}$元素相同的自然数及其相应的点打上记号就可以了。显然对于某矩阵组中的每一个矩阵$\{p_i\}$,都相应有,也只有这样一把尺。这把尺叫$\{p_i\}$的合数标识尺或简称为标识尺。当说到用某矩阵组中矩阵$\{p_i\}$标识,就是指用相应的$\{p_i\}$标识尺去对自然数列轴进行合数标识,并有以下关于进行合数标识的操作规程:

当把$\{p_i\}$标识尺的原点 O、尺身与自然数列轴的原点0、轴线相重合,则得两类点的重合:一是$\{p_i\}$标识尺上的自然数和其刻度,与自然数列轴上的自然数和其刻度一一重合;二是$\{p_i\}$标识尺上标有记号的合数点和自然数列轴上相应的那部分自然数点相重合。凡与尺上合数点相重合的自然数列轴上的点,就被标识为合数点;该点相应的标数,就被标识为合数。凡与尺上合数点不相重合的自然数列轴上的点,就被标识为相对$\{p_i\}$的可能素数点;该点相应的自然数,就被标识为可能素数。凡与所有应参加标识的标识尺的合数点不相重合的点,被标识为素数点,该点相应的标数,就被最后标识为素数。

现给出求自然数列轴上任意素数,直至所有素数的方法和步骤。

(1)最初的几个素数,是根据素数定义计算出来的。令它们依序是2,3,5,7,11,…,p_{i-1},p_i,以它们为首因子的相应矩阵是{2},{3},{5},{7},…,$\{p_{i-1}\}$,$\{p_i\}$,对(0,N),其中 $p_i = \sqrt{N}$,以 W_1中小于等于$\sqrt{N}$的所有素数为首因子的标识矩阵$\{p_1\}$,$\{p_2\}$,…,$\{p_i\}$作合数标识就最后定出(0,N)内的全部合数和素数,于是我们就得到一组新的大于 p_i 的素数,它们是 $p_{i+1}, \cdots, p_k$。

(2)接着,$[p_i, p_i^2)$中出现的素数 $p_{i+1}, p_{i+2}, \cdots, p_{k-1}, p_k$ 相应的矩阵$\{p_{i+1}\}$,$\{p_{i+2}\}$,…,$\{p_{k-1}\}$,$\{p_k\}$,同原先那组矩阵一起对$[p_i^2, p_k^2)$进行标识,就得到该区间所有的合数和素数,其中新素数是 $p_{k+1}, \cdots, p_{l-1}, p_l$,且 $p_{k+1} < \cdots < p_{l-1} < p_l$,其相应的标识矩阵是$\{p_{k+1}\}, \cdots, \{p_{l-1}\}, \{p_l\}$如图2。

(3)然后,矩阵$\{p_{k+1}\}, \cdots, \{p_{l-1}\}, \{p_l\}$和前二组矩阵一起,又去标识新的区间$[p_k^2, p_l^2]$,并最后确定其内所有的新素数和合数。

重复上述过程得到结论是每组新标识出来的素数中最大者 p_q,其平方 p_q^2 必大于此素数区间的上界 p_k^2,(p_k^2, p_q^2)被$\{p_q\}$及以前所有素数相应矩阵标识,一组新素数又最后被确定。然后象滚雪球似的,自然数数列轴上被此标识的合数和我们不断所收获的素数,不断地向数轴的右边延伸。在上述延伸过程中将会产生一个问题,在$[p_i^2, p_k^2)$中会有新素数 p_{k+1}存在吗?回答是肯定的。数论专家在研究相邻两个素数 p_{n+1}和 p_n 的差 d_n 分布方面的成就中,目前有关 d_n 最好的结果是霍斯勒(M. N. Huxley)得到的。他指出,对于任何 $\epsilon > 0$,皆存在仅依赖于 ϵ 的常数 $n_0(\epsilon)$,当 $n > n_0$时,在 n 与 $n + n^{\frac{7}{12}+\epsilon}$之间恒存在一个素数。n 为充分大的自然数,故在$[p_i^2, p_k^2)$中必存在至少一个素数。在此数学模型下,我们就得到一张合数标识图(见附图1)。该图直接展示了素数分布图象和素数分布规律,为我们查找和调用提供了方便。

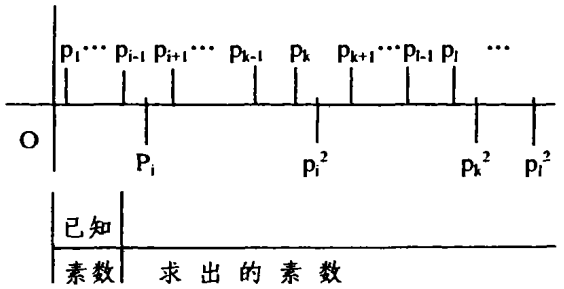


图2

至此，在求素数、合数的问题上，除了众所周知的用定义来识别的原始方法、操作型的 Eratosthenes 筛法外，还有“合数标识法”。很可能还有其余的方法，但估计是未曾公开或鲜为人知的。

结论 可以设想在计算机配合下，完全可以用合数标识法编制一张迄今还没有的[0，N]内自然数分析表。它标记了[0，N]内的合数、素数，以便于随时调用。正因为合数标识法用简单、机械、重复的“打记号”方式代替了繁琐的计算，计算机就可用最简单的程序进行大素数的寻找，使寻找大素数的速度极大地提高。总之，寻找大素数的合数标识法具有简易、快速、准确的特点。因而，它为数字签名中寻找所需的素数提供了一种较好的手段。

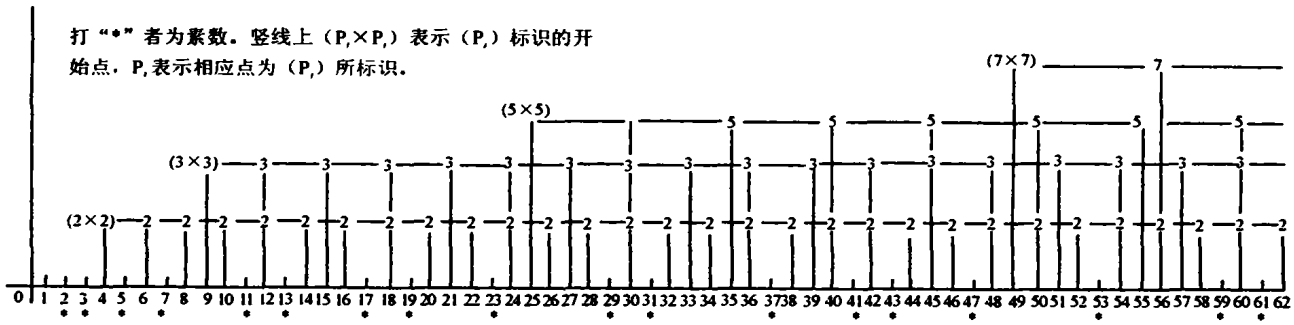
参 考 文 献

1 王长策. 合数标识论. 贵州人民出版社，1999

2 [美]Bruce Schneier，著. 吴世忠，祝世雄，张文政，等译. 应用密码学. 机械工业出版社. 2001

3 徐快，段云所，陈钟. 数字签名与数字证书. 网络安全技术与应用，2001

4 张方国，王育民. 欧洲21世纪密码候选标准. 网络安全技术与应用，2001



附图1 合数标识平面图

（上接第100页）

3）当 $n<L$ 时，根据式(4)更新权矢量；

4）当 $n=L$ 时，根据 $\hat{R}_x(n)$ 及 ε 通过阈值得 $\hat{R}_x(n)$ 的稀疏结构；

5）当 $n\geqslant L$ 时，用共轭梯度法解线性方程组(10)，由(12)式求得下降方向 $\hat{g}(n)$ 并根据式(7)更新权矢量。

整个算法开始时用 *LMS* 算法，从时刻起才采用 *Newton-LMS* 算法。尽管如此，也难以保证 $\hat{R}_x(n)$ 的正定性，为此，在实际中对 $\hat{R}_x(n)$ 还要作对角加载处理，以保证算法的稳定性。

4 计算机仿真

我们在仿真中采用的非最小相位线性信道模型为 $x(n)=0.3482s(n)+0.8704s(n-1)+0.3482s(n-2)$，$s(n)$ 为随机发送的±1信号，信噪比为20*dB*。均衡器长度 N 为20，ε 取0.005，$L=60$。独立运行30次后得到图2所示的文[4]和本文两种算法的学习曲线，从中可以看出，本文算法的收敛性能要好一些。

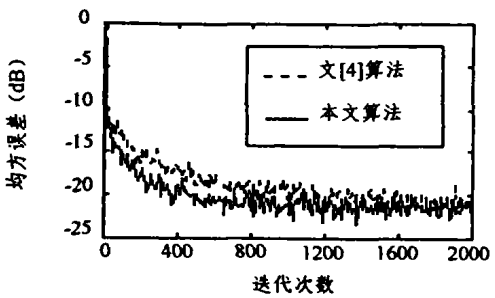


图2 两种均衡算法的学习曲线图

结论 本文主要讨论的是如何利用区间小波变换的特点来提高自适应线性均衡器的收敛速度，以及如何更好地得到稀疏矩阵和减少计算量问题，仿真表明该算法是有效的。同时也可将这一算法运用到判决反馈均衡和盲均衡中去，此时需考虑结合这两种均衡算法本身的特点。另外，对阈值的选取还需作进一步的研究，比如可以把小波分解的次数等因素考虑进去，形成更好的阈值方法，使得相关阵的估计更快更准确一些。

参 考 文 献

1 Lee J C. Performance of transform-domain LMS adaptive digital filters. IEEE Trans on ASSP，1986，34(3)：499～510

2 Erdol N，Basbug F. Wavelet transform based adaptive filters：analysis and new results. IEEE Trans on SP，1996，44(9)：2163～2171

3 Tewfik A H，Kim M. Fast positive definite linear system solvers. IEEE Trans on SP，1994，42(3)：572～585

4 Hosur S，Tewfik A H. Wavelet transform domain adaptive filtering. IEEE Trans on SP，1997，45(3)：617～630

5 Daubechies I. Two recent results on wavelets：Wavelet bases for the interval，and biorthogonal wavelets diagonalizing the derivative operator. Recent Advances in Wavelet Analysis. Academic Press，Inc，1994，237～258

6 Cohen A，Daubechies I，Vial P. Wavelets on the interval and fast transforms. Appl Comput Harmonic Anal，1993，1(1)：54～81

7 张贤达. 信号处理中的线性代数. 北京：科学出版社，1997

8 胡家赣. 线性代数方程组的迭代解法. 北京：科学出版社，1997