第 36 卷 第 2 期 计 算 机 工 程 2010 年 1 月
Vol.36 No.2 Computer Engineering January 2010
·博士论文· 文章编号：1000—3428(2010)02—0001—03 文献标识码：A 中图分类号：TP309

基于 DNA 计算的 RSA 密码系统攻击方法

杨学庆[1,2]，柳重堪[1,2]
(1. 北京航空航天大学数学、信息与行为教育部重点实验室，北京 100083；2. 北京航空航天大学电子信息工程学院，北京 100083)

摘 要：针对 RSA 公钥密码体制的陷门库特点，提出一种新的 DNA 计算模型：并类计算模型，阐述基于该模型的 RSA 密码系统的攻击方法。该方法采用 DNA 分子编码陷门库与公钥，通过组合、设置、分离、清除等操作筛选出陷门，由电泳确定陷门的值，再用陷门计算私钥的值。该方法所需的时间复杂度为 $O(1bn)^3$，DNA 分子的体积不超过 1 m^3。

关键词：DNA 计算；RSA 公钥密码；并类计算模型

RSA Cryptosystem Attack Method Based on DNA Computing

YANG Xue-qing[1,2], LIU Zhong-kan[1,2]
(1. Key Laboratory of Mathematics, Informatics and Behavioral Semantics of Ministry of Education of China, Beihang University, Beijing 100083;
2. School of Electronic and Information Engineering, Beihang University, Beijing 100083)

【Abstract】 In terms of the feature of trapdoor base of the RSA public-key cryptosystem, parallel kind computing model, this paper proposes a new model of DNA computing. Based on the model it presents a method on attacking RSA public-key cryptosystem. DNA molecule encodes the trapdoor base and public key. The trapdoor is sorted out by combination, separation, set, and clear and private key is computed by trapdoor. The expected bio-steps in the technique is only $O(1bn)^3$ and volume of DNA is no more than 1 m^3.

【Key words】 DNA computing; RSA public-key; parallel kind computing model

1 概述

DNA 计算由 Adleman[1]于 1994 年创建，因其天生极大的并行计算性与海量储存密度，故它一经创立就被尝试用于密码攻击[2-3]，但这些攻击都是针对分组密码进行的，在查阅过的文献中，尚没见过用 DNA 计算攻击公钥密码的报道。RSA 密码体制[4]是第 1 个能同时用于加密和数字签名的公钥密码算法，是沿用至今最过有生命力的一种公钥密码体制，国际上一些标准化组织均已接受 RSA 作为标准。现在个人使用的 RSA 算法通常使用的密钥长度是 1 024 bit。本文从 RSA 公钥密码的陷门库的特点出发，设计了一种具有较佳的数据移动性的 DNA 计算模型，阐述了基于该模型的 RSA 公钥密码的攻击方法。

2 并类计算模型

并类计算模型根据数据在算法运行过程中的变化将数据分为 3 类：(1)数据的初始值不是 0，其值在整个计算过程中会有变化；(2)数据的初始值不是 0，但其值在整个计算过程中都不变；(3)数据的初始值是 0，其值在整个计算过程中也会有变化。例如，用并类计算模型攻击公钥密钥系统，公开的密钥是第 1 类数据，陷门是第 2 类数据，从公钥计算陷门的过程中涉及到的其他数据都是第 3 类数据。

并类计算模型的 DNA 链分为长、短 2 种。长的 DNA 单链称为存储链，短的 DNA 单链称为粘附链。当粘附链与存储链结合时，表示该区域所编码的比特值为 1；否则，如存储链上某个编码数据的区域没与其相对应的粘附链结合，表示该区域所编码的比特值为 0。在并类计算模型中，通过 DNA 存储链两端的 DNA 片段来标志该存储链所编码的数据种类。一条长的 DNA 单链上若干部分与短的 DNA 链结合形成的复杂组合体结构被称为 DNA 复合体，这种 DNA 复合体包括 2 种限制酶的识别位点。

并类计算模型由以下 5 种基本的操作组成：

(1)组合：如果将 DNA 看成是字母表{*A*, *G*, *C*, *T*}上的多重集，则组合运算相当于多重集里的并运算。

(2)分离：将一个集合里的字符串根据指定位的值进行分类，把指定位的值为 0 的字符串归为一个集合，指定位的值为 1 的字符串归为另一个集合。

(3)设置：将一个集合里的每个字符串的指定位的值都设置为 1。

(4)清除：将一个集合里的每个字符串的指定位的值归 0。

(5)拼接：拼接运算(图 1)相当于在同类数据中进行的替换运算。

图 1 拼接运算示意图

这 5 种基本运算的实验实现方法如下：

(1)组合：将来自 2 个试管的 DNA 溶液都倒进第 3 个试管里，就能实现组合操作。

(2)分离：记编码该指定位的粘附链为 *S*，先合成粘附链 *S*，接着用分子标记技术进行亲合标记，然后使用亲合层析技术将 *S* 固定在层析柱上，把需要进行分离操作 DNA 溶液上柱，收集流出层析柱里的溶液，置于试管 T_1 中，最后用缓冲液洗脱层析柱上的 DNA 复合体，置于试管 T_0 中。

基金项目：国家自然科学基金资助重点项目(11037705)
作者简介：杨学庆(1972－)，女，博士研究生，主研方向：DNA 计算，密码破译；柳重堪，教授、博士生导师
收稿日期：2009-09-23 **E-mail**：yxqjessica@sohu.com

(3)设置：需进行设置的区域的粘附链为 S，向需要进行设置操作的试管中加入过量的粘附链 S，充分反应后，分离出未使用的粘附链 S。

(4)清除：将需要进行清除操作的试管进行热变性，然后使用亲合层析术将相应的粘附链固定在层析柱上，收集流出层析柱里的溶液，再降低温度使其复性。

(5)拼接：首先向需要进行拼接操作的试管中加入 2 种限制性内切酶：限制性内切酶 AtaII 与限制性内切酶 EcorRI，充分反应后，使用亲和分离技术提取出要替换的 DNA，接着加入待插入的DNA与DNA连接酶，最后提取出所需的DNA，置于新的试管中。

3 RSA 公钥密码体制简述

独立地选取 2 个大素数 p 和 q (为 100 bit~200 bit 十进制数字)，为获得最大程度的安全性，2 个数的长度一样。计算乘积：

$$n = pq$$

然后随机选取加密密钥 e，使 e 和 $(p-1)(q-1)$ 互素。最后用欧几里德扩展算法计算解密密钥 d，以满足：$ed \equiv 1 \bmod (p-1)(q-1)$，则

$$d = e^{-1} \bmod ((p-1)(q-1))$$

其中，e 和 n 是公开密钥；d 是私人密钥。加密消息 m 时，首先将它分成比 n 小的数据分组，即如 n 有 200 位，每个消息分组 m_i 应小于 200 位长。加密后的密文 c，将由相同长度的分组 c_i 组成。加密公式简化为

$$c_i = m_i^e (\bmod n)$$

在解密消息时，取每一个加密后的分组 c_i 并计算：

$$m_i = c_i^d (\bmod n)$$

4 DNA 计算攻击公钥密码系统的算法

4.1 算法概述

RSA 公钥密码中，e 和 n 是公钥，它们的值是已知的。小于 $\sqrt{n}$ 的素数为 RSA 的陷门，称所有小于 $\sqrt{n}$ 的素数的全体为陷门库。p 和 q 是 n 的 2 个素因子，不失一般性，设 $p < q$，则 $p < \sqrt{n}$，由于 p 是 n 的一个因子，因此用陷门库中的陷门除 n 余数为 0 的项即为 p。p 和 q 求出后，根据等式：$d = e^{-1} \bmod ((p-1)(q-1))$ 就能计算出私钥 d 的值。总之，用 DNA 计算攻击 RSA 密码体制的 DNA 算法可以归纳为以下 4 个步骤：

(1)用 DNA 分子表示公钥 n。

(2)用 DNA 分子表示小于等于 $\sqrt{n}$ 的所有素数。

(3)通过并行除法运算，筛选出陷门库的陷门除公钥的余数为 0 的项，确定 p 和 q 的值。

(4)计算私钥 d 的值。

4.2 算法第(1)步的细化表述

合成长度为 12 309 个碱基长的 DNA 链(图 2)，将该 DNA 链放入试管 T_n 内，先加入寡聚核苷酸 GAATTC，再加入限制性内切酶 EcorRI，提取出 DNA 复合体，最后根据 n 的值加入充足的粘附链，分离出未使用的粘附链。

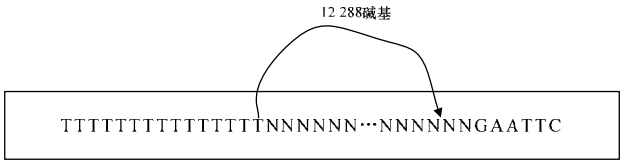

图 2　编码公钥 n 的 DNA 结构示意图(N 为任意碱基)

4.3 算法第(2)步的细化表述

合成 6 144 个碱基长的 DNA(图 3)来编码陷门，将该 DNA 链放入试管 T_p 内，先加入 GAATTC 与 GACGTC2 种寡聚核苷酸，再加入限制性内切酶 EcorRI 与 AtaII，最后提取出 DNA 复合体。将 T_p 的 DNA 分子根据不大于 100 bit(十进制)的素数的数量分成若干份，每一份根据相应的素数值放入充足的粘附链，然后分离出没使用过的粘附链。将所有的试管进行组合，产生试管 T_p，接着组合 T 与 Tp，加入 DNA 连接酶，最后分离出最长的 DNA 复合体。

图 3　编码陷门的 DNA 结构示意图

4.4 算法第(3)步的细化表述

合成 24 616 个碱基长的 DNA 链(图 4)，将该 DNA 链放入试管 T_m 内，加入寡聚核苷酸链 GACGTC。组合 T 与 T_m，加入 DNA 连接酶，提取最长的 DNA 复合体。

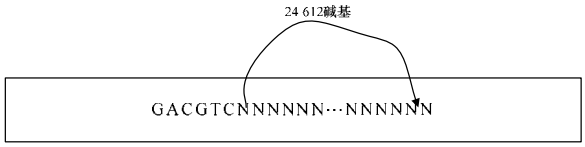

图 4　编码存储中间结果的 DNA 结构示意图

使用 DNA 计算实现并行除法所根据的数学原理如下：

设 α 为被除数，位数为 n，β 为除数，其位数为 $m, n \geq m$。

设 $\alpha = \beta \times \gamma + \varepsilon$，即 α 除以 β 的商为 γ，余数为 ε。

设 $\alpha = a_n \cdots a_1$, $\beta = \beta_m \cdots \beta_1$, $\gamma = \gamma_{n-m} \cdots \gamma_1$。

记 $X[\gamma] = \{i, \gamma_i = 1\}$，即 $X[\gamma]$ 为数 γ 的值为 1 的那些位的指标的集合。则，

$$\beta \times \gamma = \sum_{i \in X[\gamma]} \beta \times 2^{i-1}$$

$\gamma = 0$

for j=n-m to 1

$\delta = \beta \times 2^{j-1}$

if $a \geq \delta$ then $\gamma_j = 1, \alpha = \alpha - \delta$

else $\gamma_j = 0$

用 DNA 并类计算模型进行并行除法运算的操作方法如下：

For r=1 to 512

记一个空试管为 Ts_{513-r}。

根据 $B_{r+1\ 024}$ 区对 T 进行分离操作，产生 T_1 与 T_0。

组合 T_1 与 Ts_r，产生 Ts_{513-r}，记 T_0 为 T。

Nest r

For b=1to 512

记一个空试管为 T，组合为 T_{S_b} 与 T，产生 T。

For j=1 024 to b

For k=b to 1

根据 $B_{3\ 587-j-k}$ 区对 T 进行分离操作，产生 T_1 与 T_0。

根据 $B_{3\ 587-j-k}$ 区对 T_1 与 T_0 同时进行分离操作，产生 T_{11}, T_{10}, T_{01} 与 T_{00}。

设置 T_{10} 的 $B_{3\ 587-j-k}$，清除 T_{01} 的 $B_{3\ 587-j-k}$ 区。

并行组合 T_{11}, T_{10}, T_{01} 与 T_{00}，产生 T。

Next k

分别记 2 个空试管为 T_h 与 T_x。

For u= 1 to b

根据 B_u 区对 T 进行分离操作，产生 T_1 与 T_0。

根据 $B_{u+2\ 561}$ 区对 T_1 与 T_0 同时进行分离操作，产生 T_{11}, T_{10} , T_{01} 与 T_{00}。

组合 T_{11} 与 T_{00}，产生 T。组合 T_{10} 与 T_h，产生 T_h。组合 T_{01} 与 T_x，产生 T_x。

Next u

组合 T 与 T_h，产生 T。

设置 T 的 $B_{2\ 560-j+b}$ 区，清除 T_h 的 $B_{2\ 560-j+b}$ 区。

For p= 1 024 to 1

根据 B_p 区对 T 进行分离操作，产生 T_1 与 T_0。

根据 B_{2561+p} 区对 T_1 与 T_0 同时进行分离操作，产生 T_{11}, T_{10}, T_{01} 与 T_{00}。

根据 $B_{2\ 561}$ 区对 T_{11}, T_{10}, T_{01} 与 T_{00} 同时进行分离操作，产生 T_{111}, T_{101}, T_{011}, T_{001}, T_{110}, T_{100}, T_{010} 与 T_{000}。

设置 T_{100}, T_{010}, T_{001}, T_{111} 的 $B_{2\ 561+p}$ 区和 T_{010} 的 $B_{2\ 561}$ 区。

清除 T_{110}, T_{101}, T_{011}, T_{000} 的 $B_{2\ 561+p}$ 区和 T_{101} 的 $B_{2\ 561}$ 区。

Next p

For q=1 024 to 1

根据 $B_{2\ 561+p}$ 区对 T 进行分离操作，产生 T_1 与 T_0。

根据 B_q 区对 T_1 与 T_0 同时进行分离操作，产生 T_{11}, T_{10}, T_{01} 与 T_{00}。

设置 T_{10} 的 B_q 区，清除 T_{01} 的 B_q 区。

组合 T_{11}, T_{10}, T_{01} 与 T_{00}，产生 T，清除 T 的 $B_{2\ 561+p}$ 区。

Next q

Next j

记一个空试管为 Ts_b，组合 Ts_b 与 T，产生 Ts_b。

Next b

记一个空试管为 T

For f=1 to 512

组合 T 与 Ts_b，产生 T。

Next f

For i=1 to 1 024

根据 B_i 区对 T 进行分离操作，产生 T_1 与 T_0。

组合 T_0 与 T，产生 T。

Next i

上述操作完成后，p 存储于 $B_{1\ 025}$~$B_{1\ 536}$ 区，q 存储于 $B_{1\ 537}$~$B_{2\ 560}$ 区。用磁珠分离法分离出储存复合体，将储存复合体的粘附链洗提出来，通过走杂交胶判断存在何种粘附链就能确定 p 与 q 的具体值。

4.5 算法第(4)步的细化表述

设 $\varphi(n)=(p-1)(q-1)$。通过求 e 模 $\varphi(n)$ 的逆就能确定私钥 d 的值。其算法如下：

(1) $n_1=\varphi(n)$， $n_2=e$， $b_1=0$， $b_2=1$。

(2) $q=\left\lfloor \dfrac{n_1}{n_2} \right\rfloor$， $r=n_1-qn_2$。

(3)如果 $r\neq 0$，则 $n_2=n_1$， $n_2=r, t=b_2, b_2=b_1-qb_2, b_1=t$，转第(2)步。

(4)如果 $n_2\neq 1$，则 e 模 $\varphi(n)$ 不存在逆元。

(5)如果 $n_2=1$，则 e 模 $\varphi(n)$ 的逆元为 $b_2 \bmod n$，即所求的 d 值为 b_2。

该算法包括除法、减法运算，由于前面已经介绍这 2 种运算的 DNA 计算操作方法，因此不再详述其 DNA 计算实现方法。

计算出私钥 d 的值后，运用解密算法 $x=y^d \bmod n$ 就能将密文解密，求出相应的明文。

5 复杂度分析

5.1 时间复杂度分析

本算法包括并行除法、乘法与减法运算。其中，除法运算包括 3 层循环：第 1 层循环中只包括组合操作；第 2 层循环中包括组合、分离、设置与清除 4 种操作；第 3 层循环中也包括组合、分离、设置与清除 4 种操作，每层循环中每种操作的次数都是 $O(1bn)$ 的倍数，这里 n 为公钥。所以，除法算法的时间复杂度为 $O(1bn)^3$。

乘法运算的时间复杂度为 $O(1bn)^2$，而减法运算的时间复杂度都为 $1bn$，所以，算法的时间复杂度为 $O(1bn)^3$。

5.2 空间复杂度分析

空间复杂度是指该 DNA 算法所需的 DNA 链的数量。从整个算法可以看出小于等于 $\sqrt{n}$ 的所有素数的数量决定了 DNA 链的数量，故需计算素数的个数。

素数定理[5]：设 $x>0$， $\pi(x)$ 为不大于 x 的素数的个数，则

$$\lim_{x\to\infty}\frac{\pi(x)\ln(x)}{x}=1$$

根据该定理，当 x 充分大时，有

$$\pi(x)\approx\frac{x}{\ln x}$$

由此，可估计出长度 100 bit(十进制)的素数大约有

$$\frac{2^{100}}{\ln 2^{100}}-\frac{2^{99}}{\ln 2^{99}}\approx 9\times 10^{27}$$

据估算， 9×10^{27} DNA 的总体积不超过 1 m^3。

6 结束语

由于现有 DNA 计算模型攻击 RSA 公钥密码存在缺陷，因此本文提出了一种新的 DNA 计算模型。因为该模型的同类数据之间可通过拼接运算进行替换，所以数据的可移动性佳，并且可以反复使用。本文还提出了一种基于该模型的 RSA 密码体制的攻击方法，并对方法中的核心技术如陷门库的表示、陷门的筛选等做了详细研究，对方法的时间复杂度、空间复杂度等性能做了评估。由此推出该攻击方法具有较高的可行性。

参考文献

[1] Adleman L. Molecular Computation of Solutions to Combinatorial Problems[J]. Science, 1994, 266(11): 1021-1024.

[2] Adleman L. On Applying Molecular Computation to the Data Encryption Standard[J]. Comp. Biol., 1996, 6(1): 53-63.

[3] Dan B. Breaking DES Using a Molecular Computer[R]. New Jersey, USA: Princeton University, 1995.

[4] Schneirer B. 应用密码学——协议、算法与 C 源程序[M]. 吴世忠, 译. 北京: 机械工业出版社, 2000.

[5] 陈鲁生. 现代密码学[M]. 北京: 科学出版社, 2002.

编辑 索书志