有限自动机理论的公钥加密算法的研究与改进

阎 浩，严筱永，沈维艳
（金陵科技学院，江苏 南京 210001）

摘 要：该文讨论的是基于有限自动机理论的公钥加密体制。该体制的工作原理是利用有限自动机的可逆性来完成加密、解密以及数字签名等功能。文中对该密码体制进行分析讨论并提出了构造可逆自动机的一种改进方法，提高了可逆自动机的复杂度，进一步增强了算法的安全性，同时也保持了密钥短，速度快等优点。

关键词：有限自动机；公钥加密体制；可逆性

Research and Improve of an Asymmetric Cryptosystem Algorithm Based on the Theory of the Finite Automaton

YAN Hao, YAN Xiao-yong ,SHEN Wei-yan

Abstract: In this paper we discuss one public key cryptosystem which based on the theory of the finite automaton. This cryptosystem works by using the invertibility theory of finite automata to complete the encrypting, decrypting and signature function. The paper will discuss and analysis the cryptosystem, improve on the construction of the invertibility finite automat, to advance its complexity, so security will be boosted, at the same time its retains the advantages of the faster speed, relatively short public key etc.

Key words:finite automata; public key cryptosystem; invertibility

1 引言

信息安全问题是当前信息社会中存在的一个棘手但又不得不面对的问题。密码学技术的发展和应用对于解决信息安全问题有着不可估量的作用，大量的学者和研究人员投入到了密码学的研究工作中，提出了各种加密体制，加密算法，其中公钥加密体制是密码学发展中的一个里程碑，开创了现代密码学的一个新的领域[2]。

公钥加密具有两个密钥，一个是对外公开的加密密钥，另一个是保密的解密密钥。其工作原理为：任一用户A，想与用户B秘密通信，首先A用B的公开密钥对欲传送的明文加密得到密文，传送给用户B，当用户B接收到发送来的密文后，用自己的解密密钥对密文进行解密得到A发送的明文。这种加密算法具有安全性高，难以破译等优点[3]。

本文所讨论的加密算法也属于公钥加密体制，它利用有限自动机可逆性的原理，实现加、解密过程以及数字签名功能[4]。公开密钥和私有密钥都是该可逆有限自动机的某些状态和某些设定的已知参数，工作过程就是由这些参数引起自动机状态的变化并输出一定的结果。这种加密体制具有安全性高，密钥短，速度快等特点。

2 有限自动机的理论

2.1 有限自动机的概念

有限自动机(Finite Automata)理论是自动机理论的一个的分支。从本质上讲，有限自动机是从一些具体的动态离散数字系统中抽象出来的数学模型，一般情况下，一个有限自动机M可以表示成一个五元组 $M=<X,Y,S,\delta,\lambda>$，其中X，Y和S是3个非空有限集，分别称为M的输入字母表，输出字母表和状态字母表；δ 是迪卡儿积集S×X到S的单值映射，称为M的下一状态函数，λ 是S×X到Y的单值映射，称为M的输出函数[1]。若在M的初始状态s(0)下输入序列 $x(0)$，$x(1),\cdots$，产生的输出序列 $y(0)$，$y(1),\cdots$，则有：

$$y(i)=\lambda(s(i),x(i))$$
$$s(i+1)=\delta(s(i),x(i)),\ i=0,1,2,\cdots$$

另外，对于集合A，用 A^N 表示A上所有有限长序列的集合（包括空序列），A^a 表示A上所有无限长序列的集合。按下面的方法可以将 δ 和 λ 扩充到 $S\times X^N$ 和 $(S\times(X^N\cup X^a))$ 上：

$$\delta(s,\varepsilon)=s,\delta(s,\alpha x)=\delta(\delta(s,\alpha),x)$$
$$\lambda(s,\varepsilon)=\varepsilon,\lambda(s,\beta x)=\lambda(s,x)\lambda(\delta(s,x),\beta)$$

其中

$$s\in S,x\in X,\alpha\in X^n,\beta\in(X^N\cup X^a)$$。

如果 Φ 是 $Y^k\times X^{h+1}$ 到Y的单值映射，而有限自动机 $M=<X,Y,S,\delta,\lambda>$ 由下式定义：

$$y(i)=\Phi(y(i-1),\cdots,y(i-k),x(i),x(i-1),\cdots,x(i-h))$$
$$i=0,1,2,\cdots$$

则M称为一个(h,k)阶存贮有限自动机，记作 $M_\Phi=<X,Y,Y^k\times X^h,\delta,\lambda>$。此时如果k=0，则 M_2 称为一个h阶存贮有限自动机。

3 有限自动机加密原理

3.1 FAPKC 简介

1985 年，陶仁骥和陈世华利用有限自动机可逆性理论提出了有限自动机公钥密码体制(FAPKC)。同时，也提出了一个具体的有限自动机公钥密码体制 FAPKC0，在 1986 年，陶仁骥和陈世华又提出了 FAPKC 的另外两种变形 FAPKC1 和 FAPKC2，随后又相继提出了 FAPKC3 和 FAPKC4，构成了一个有限自动机加密的算法系列[7]。这几个算法共同点是公钥中都有一个组合自动机 $C^*(M_1,M_0)$，但是每个算法又有不同点，主要体现在其中的两个有限自动机 M_1 和 M_0 上：(1)FAPKC0：M_1 是延迟 0 步弱可逆的 r 阶输入存储非线性有限自动机，M_0 是延迟 τ 步弱可逆的 (τ,τ) 阶存储线性有限自动机。(2)FAPKC1：M_1 是延迟 0 步弱可逆的 r 阶输入存储非线性有限自动机，M_0 是延迟 τ 步弱逆的 r 阶输入存储线性有限自动机。(3)FAPKC2：M_1 是延迟 τ 步弱可逆 r 阶输入存储非线性有限自动机，M_0 是延迟 τ 步弱逆的 r 阶输入存储线性有限自动机。(4)FAPKC3：M_1 是延迟 τ 步弱可逆 h_1 阶输入存储有限自动机，M_0 是延迟 τ 步弱可逆的 (h_0,k_0) 阶存储有限自动机。(5)FAPKC4：M_1 是延迟 τ 步弱可逆的 h_1 阶输入存储有限自动机，M'_0 是延迟 τ 步弱可逆的 (h_0,k_0) 阶存储有限自动机。

3.2 FAPKC3 加密原理

在 FAPKC3 公钥加密体制中，公钥中的复合有限自动机是由一个存储有限自动机和输入存储有限自动机组合而成的，其中，M_0 是延迟 τ 步弱可逆的 (h_0,k_0) 阶存储有限自动机，M_1 是延迟 τ 步弱可逆的 h_1 阶输入存储有限自动机。密钥是由两个存储有限自动机和它们的初始状态组成的。每个公钥中的有限自动机和它在密钥中相应的有限自动机是延迟一定步数弱可逆的。密文的长度要比相应的明文长一定的位数，这个多出的位数就是自动机的延迟步数[6]。

设 X 和 Y 分别为有限域 GF(q) 上 l 维和 m 维列向量空间。预先规定共同的 q 和 l，并取 m = l，这样有利于签名机制，也就是说所有的用户利用相同的字母表来互相通信。

一个用户 A，按照如下方式来构造其公钥和密钥：

(1) 构造一个 (h_0,k_0) 阶存储有限自动机 $M_0=<X,Y,S_0,\delta_0,\lambda_0>$ 和一个 (τ_0+h_0,k_0) 阶存储有限自动机 $M'_0=<Y,X,S'_0,\delta'_0,\lambda'_0>$ 满足下列条件：

1) 对于 M_0 的任何状态 $s_0=<y_{-k_0},\cdots y_{-1},x_{-h_0},\cdots x_{-1}>$ 和任意的 X 中的元素 $x_0,x_1,\cdots$，如果有 $y_0y_1\cdots=\lambda_0(s_0,x_0x_1\cdots)$，则有 $\lambda'_0(s'_{t_0},y_{t_0},y_{t_0+1}\cdots)=x_0x_1\cdots$，其中，$s'_{t_0}=<x_{t_0-h_0},\cdots,x_{t_0-1},y_{t_0-k_0},\cdots,y_{t_0-1}>$。

2) 对于集合 $\{\delta'_0(s',y_0\cdots y_{t_0-1})\mid s'\in S'_0,y_0,\cdots,y_{t_0-1}\in Y\}$ 中的任何状态 $s''_0=<x'_{-h_0},\cdots,x'_{-1},y_{-\tau_0-k_0},\cdots,y_{-1}>$ 和自动机 M_0 的状态 $s_0=<y_{-k_0-\tau_0},\cdots,y_{-\tau_0-1},x'_{-h_0},\cdots,x'_{-1}>$，对于 Y 中的任何元素 $y_0,y_1,\cdots$，如果 $x'_0x'_1\cdots=\lambda'_0(s''_0,y_0y_1\cdots)$，则 $\lambda_0(s_0,x'_0x'_1\cdots)=y_{-\tau_0}\cdots y_{-1}y_0y_1\cdots$。

(2) 构造一个 h_1 阶输入存储有限自动机 $M_1=<X,Y,S_1,\delta_1,\lambda_1>$ 和一个 (τ_1,h_1) 阶存储有限自动机 $M'_1=<Y,X,S'_1,\delta'_1,\lambda'_1>$ 满足下列条件：

1) 对于 M_1 的任何状态 $s_1=<x_{-h_1},\cdots,x_{-1}>$ 和任意的 X 中的元素 $x_0,x_1,\cdots$，如果有 $x'_0x'_1\cdots=\lambda_1(s_1,x_0x_1\cdots)$，则有 $\lambda'_1(s'_{\tau_1},x'_{\tau_1}x'_{\tau_1+1})=x_0x_1\cdots$，其中 $s'_{\tau_1}=<x_{-h_1},\cdots,x_{-1},x'_{\tau_1-1},\cdots,x'_0>$。

2) 对于 M'_1 的任何状态 $s'_0=<x_{-h_1},\cdots,x_{-1},x'_{-\tau_1},\cdots,x'_{-1}>$ 和任意的 X 中的元素 $x'_0,x'_1,\cdots$，如果有 $x_0x_1\cdots=\lambda'_1(s'_0,x'_0x'_1\cdots)$，则 $\lambda_1(<x_{\tau_1-1},\cdots,x_{\tau_1-h_1}>,x_{\tau_1}x_{\tau_1+1}\cdots)=x'_0x'_1\cdots$

(3) 由 M_0 和 M_1 组合而成的有限自动机为 $M=C^*(M_1,M_0)=<X,Y,S,\delta,\lambda>$

(4) 记 $\tau=\max(\tau_0,\tau_1,h_0)$，任选 $y_{-1,s},\cdots,y_{-\tau-\tau_1,s}\in Y$，任意的 $s0_{-\tau,s}=<x_{-\tau-\tau_1-1,s}',\cdots x_{-\tau-\tau_1-h_0,s}',y_{-\tau-\tau_1-1,s},\cdots,y_{-\tau-\tau_1-\tau_0-k_0,s}>^T\in S_0'$，及任意的 $s1_{\tau,s}=<x_{-\tau-\tau_1-1,s},\cdots x_{-\tau-\tau_1-h_1,s},x_{-\tau-\tau_1-1,s}',\cdots x_{-\tau-2\tau_1,s}'>^T\in S_1'$，记：$x_{-\tau-\tau_1,s}'\cdots x_{-1,s}'=\lambda_0'(s0_{-\tau,s},y_{-\tau-\tau_1,s}\cdots y_{-1,s})$，$s_{0,s}'=\delta_0'(s0_{-\tau,s},y_{-\tau-\tau_1,s}\cdots y_{-1,s})$，$x_{-\tau-\tau_1,s}\cdots x_{-1,s}=\lambda_1'(s1_{-\tau,s},x_{-\tau-\tau_1,s}'\cdots x_{-1,s}')$，$s_{1,s}'=\delta_1'(s1_{-\tau,s},x_{-\tau-\tau_1,s}'\cdots x_{-1,s}')$，令 $s^{out}{}_v=<y_{-1,s},\cdots y_{-k_0,s}>^T$，$s^m{}_v=<x_{-1,s},\cdots x_{-h_0-h_1+\tau_0+\tau_1,s}>^T$ 任选 $s_e=<y_{-k_0,e},\cdots,y_{-1,e},x_{-1,e},\cdots x_{-h_0-h_1,e}>^T\in S$，记：$x_{-h_0,e}'\cdots x_{-1,e}'=\lambda_1'([x_{-h_0-1,e},\cdots x_{-h_0-h_1,e}]^T\in X,x_{-h_0,e}\cdots x_{-1,e})$，令 $s^{out}{}_{1,d}=<x_{-1,e},\cdots x_{-h_1,e}>^T\in X$，$s^{out}{}_{0,d}=<x_{-1,e}',\cdots,x_{-h_0,e}'>$。

(5) 得到公钥：$C^*(M_1,M_0)$，$s^{out}{}_v$，$s^{in}{}_v$，s_e，$\tau_0+\tau_1$

得到密钥：M_0'，M_1'，$s_{0,s}'$，$s_{1,s}'$，$s^{out}{}_{1,d}$，$s^{out}{}_{0,d}$，τ_0，τ_1。

加密和解密过程：用户首先将欲发送的明文 $x_0x_1\cdots x_n$，扩展 $\tau_0+\tau_1$ 位得到：$x_0x_1\cdots x_m\cdots x_{t_0+t_1}$，用公钥中的 $C^*(M_1,M_0)$ 和 s_e 计算密文 $y_0y_1\cdots y_{n+t_0+t_1}=\lambda(s_e,x_0x_1\cdots x_{n+t_0+t_1})$。当接收方收到密文后，利用密钥中的 M_0' 和 $s^{out}{}_{0,d}$ 以及公钥中的 $y_{-k0},\cdots,y_{-1}$ 计算得到 $x_0'x_1'\cdots x_{n+}=\lambda_0'([x_1',\cdots x_{h_0}',y_{0-1},\cdots y_{t_0}],y_{t_0}\cdots y_{n+t_0+t_1})$，然后再利用 M_1' 和 $s^{out}{}_{1,d}$ 计算得到明文 $x_0x_1\cdots x_n=\lambda_1'([x_{-1},\cdots,x_{-h_1},x_{t_0-1}',\cdots x_0'],x_{t_1}'\cdots x_{t_1+n}')$。

签名和验证过程：把信息 $y_0y_1\cdots y_n$ 扩展 $\tau_0+\tau_1$ 位得到：$y_0y_1\cdots y_{n+t_0+t_1}$，用密钥中的 M_0'，M_1'，$s_{0,s}'$，$s_{1,s}'$ 计算得到签名：$x_0x_1\cdots x_{n+t_0+t_1}=\lambda_1'(s_{1,s}',\lambda_0'(s_{0,s}',y_0y_1\cdots y_{n+t_0+t_1}))$，验证时用公钥中的 $C^*(M_1,M_0)$，$s^{out}{}_v$，$s^{in}{}_v$，找出 $s=<y_{-1,s},\cdots,y_{-k_0,s},x_{t_0+t_1-1,s},\cdots,x_{t_0+t_1-h_0-h_1,s}>^T$，验证 $\lambda(s,y_{t_0+t_1}\cdots y_{n+t_0+t_1})$ 是否等于 $y_0y_1\cdots y_n$，来确认有效性。

3.3 FAPKC3 的实现

利用 FAPKC3 公钥加密体制[5]，取 q=2，l=m。M_0 和 M_0^* 是线性有限自动机。M_0 表示为：

$$y(i)=\sum_{j=0}^{k_0}A_j y(i-j)+\sum_{j=0}^{h_0}B_j x^t(i-j)$$

$i=0,1,\cdots$

非线性有限自动机 M_1 表示为：

$$x^t(i)=\sum_{j=0}^{h_1}F_j f(x(i-j),\cdots,x(i-j-\varepsilon)),\quad i=0,1,\cdots,$$

$$\sum_{j=0}^{h_0+h_1}C_j z^j=\left(\sum_{j=0}^{h_0}B_j z^j\right)\left(\sum_{j=0}^{h_1}F_j z^j\right),$$

$$\sum_{j=0}^{h_0+h_1-\tau}C'_j z^j=\left(\sum_{j=0}^{h_0}B_j z^j\right)\left(\sum_{j=0}^{h_1-\tau}F'_j z^j\right)$$

则复合自动机 $C'(M_1,M_0)$ 为：

$$y(i)=\sum_{j=1}^{k_0}A_j y(i-j)+\sum_{j=0}^{h_0+h_1}B_j x(i-j)+\sum_{j=0}^{h_0+h_1-\tau}B'_j f(x(i-j),\cdots,x(i-j-\varepsilon))$$

$i=0,1,\cdots$

其中的参数 h_0，h_1，k_0，τ_0 和 τ_1 由用户自己选取。容易看出构造自动机 M_0 是延迟 τ 步弱可逆的 (h_0,k_0) 阶存储有限自动机，自动机 M_1 是延迟 τ 步弱可逆的 h_1 阶输入存储有限自动机。

4 一种改进方法

在 FAPKC3 体制中，如果复合自动机中的 M_1 具有输入线性可分弱逆性，那么这个加密体制是脆弱的，同时根据穷尽式搜索方式的特点，如果知道要加密明文的最后一部分的话，那么这个体制也是容易被攻击的[8]。根据以上情况，这里我们给出一个经过改进的算法来避免这两种情况的发生。

令有限自动机 M_0，M_1 与 $C'(M_1,M_0)$ 如在 FAPKC3 算法中定义。

假设攻击者知道或者成功猜测到明文的后 h_0+h_1 位 $x(n-h_0-h_1+1),\cdots,x(n)$，他就可以得到自动机 $C'(M_1,M_0)$ 的 $s(n+1)$：$s(n+1)=\langle y(n),\cdots,y(n-k_0+1),x(n),\cdots,x(n-h_0-h_1+1)\rangle$。这样攻击者就可尝试计算得到状态 $s(n)=\langle y(n-1),\cdots,y(n-k_0),x(n-1),\cdots,x(n-h_0-h_1)\rangle$。其中惟一的未知变量 $x(n-h_0-h_1)$ 可以利用公钥和对应密文对应的等式：

$$y(n)=\sum_{j=1}^{k_0}A_j y(n-j)+\sum_{j=0}^{h_0}B_j x(n-j)+\sum_{j=0}^{h_0+h_1-\tau}B'_j f(x(n-j),\cdots,x(n-j-\varepsilon))$$

计算求出并进行猜测，依次循环从而破译密文。其中的 s 表示 $x^t\to x$ 的非线性函数。

由于后向前穷尽式搜索攻击并没有在算法设计时避免这种情况的发生，所以它的复杂度是随机的，最坏的情况是 $x(n-h_0-h_1)$ 可由上式直接求出，那么此时该体制就不存在安全性，就算情况复杂一些，也不能避免破译的发生。

因此具体的修改方案为，把原自动机 M_1：

$$x^t(i)=\sum_{j=0}^{h_1}F_j f(x(i-j),\cdots,x(i-j-\varepsilon)),\quad i=0,1,\cdots$$

修改为：

$$y(i)=\sum_{j=0}^{h_1}F_j x(i-j)+\sum_{j=h_1+1}^{2h_1}F_{2h_1-j}x(i-j)$$

$$+\sum_{j=0}^{h_1-\varepsilon}F_j f(x(i-j),\cdots,x(i-j-\varepsilon))+\sum_{j=h_1}^{2h_1-\varepsilon}F_{2h_1-j}f(x(i-j),\cdots,x(i-j-\varepsilon))$$

经过自动机 M_1 的修改，复合自动机 $C'(M_1,M_0)$ 相应修改为：

$$y(i)=\sum_{j=1}^{k_0}A_j y(i-j)+\sum_{j=0}^{2h_0+h_1}C_j x(i-j)+\sum_{j=0}^{2h_0+h_1-\tau}C'_j f(x(i-j),\cdots,x(i-j-\varepsilon))$$

其中，$C_j=\sum_{\substack{k+l=j\\0\le k\le 2h_0\\0\le l\le 2h_1}}B_k F_l$，$C'_j=\sum_{\substack{k+l=j\\0\le k\le 2h_0\\0\le l\le 2h_1-\varepsilon}}B_k F'_l$。

这里我们不妨做以下设定：当 $h_1\le j\le 2h_1-\varepsilon$ 时

$$B_j=B_{2h_0-j},F_j=F_{2h_1-j},F'_j=F'_{2h_1-\varepsilon-j}$$

当 $h_1-\varepsilon<j<h_1$ 时，$F'_j=0$。

利用上述修改方法，既避免了可分有限自动机的产生，又使得后向前穷尽式搜索或前向后穷尽式搜索都变地更加复杂，从而使得破译也变地更加困难，算法的安全性得到了提高。

5 结束

本文首先介绍了公钥加密和有限自动机的概念，然后引入了利用有限自动机理论进行公钥加密的原理和几个具有代表性的算法，最后通过改进原有算法的可逆自动机的构造方法，增强了自动机复杂程度，从而使算法能更加有效地抵制各种方式的攻击，进一步提高了算法的安全性。

参考文献：

[1] 陶仁骥．自动机引论．第一版．北京：科学出版社，1986.

[2] 张焕国，刘玉珍．密码学引论．第 1 版． 武汉：武汉大学出版社，2003.

[3] 杨义先，孙伟，钮心忻．现代密码新理论．第 1 版．北京：科学出版社，2002.

[4] 陶仁骥．有限自动机的可逆性．第 1 版．北京：科学出版社，1979.

[5] 王浩．关于一类有限自动机的可逆性．密码学进展—CHINACRYPT'96，第四届中国密码学学术会议论文集．1996.

[6]Tao Renji,Chen Shihua.Constructing finite automata with invertibility by transformation method.密码学进展—CHINACRYPT'98，第五届中国密码学学术会议论文集．1998.

[7]Tao Renji,Chen Shihua.A note on the public key cryptosystem FAPKC3．密码学进展—CHINACRYPT'98，第五届中国密码学学术会议论文集．1998：69-77.

[8] 戴大为，吴葵，张焕国．有限自动机公钥密码体制的密码分析．中国科学，A 辑．1995.

作者简介：阎 浩 (1980-)，男，助教，硕士，主要研究方向：密码学，可信计算，网络安全；严筱永 (1977-)，男助教，硕士，主要研究方向：网络信息安全，模式识别；沈维艳 (1982-)，女，助教，硕士，主要研究方向：密码学，信息安全。

收稿日期：2008-02-16