基于流密码代数攻击的研究*)

张 莉 刘海波 白恩健 肖国镇

(西安电子科技大学综合业务网国家重点实验室 西安710071)

摘 要 概述了流密码代数攻击的提出和发展,介绍了它的基本思想。简要描述了代数攻击的一般算法和可攻击的流密码类型,针对带记忆和不带记忆非线性组合流密码的代数攻击,阐述了 Courtois 等人的具体工作,并给出了两个新的选择非线性布尔函数的标准。笔者提出了密码非线性部分未知情况下的一般代数攻击方法,最后总结了代数攻击的贡献和不足之处。

关键词 流密码,相关攻击,代数攻击,多元方程,XL 算法,布尔函数

A Study of the Algebraic Attacks on Stream Ciphers

ZHANG Li LIU Hai-Bo BAI En-Jian XIAO Guo-Zhen

(National Key Lab. of Integrated Service Networks, Xidian Univ., Xi'an 710071)

Abstract In this paper, the basic contents of the algebraic attacks on stream ciphers with linear feedback are introduced. A simple description is given to the general arithmetic of algebraic attacks and the type of stream ciphers that can be attacked, The rule discovered by Courtois at algebraic attacks on stream ciphers with (and without) memory is presented, and two new criterion on selecting good Boolean function are presented. A general algebraic attacks on ciphers which the nonlinear combiner unknown is put forward by writer, The contribution and shortage of algebraic attacks are summarized.

Keywords Stream ciphers, Correlation attack, Algebraic attacks, Multivariate equations, XL algorithm, Boolean function

1 引言

我们研究的是基于线性反馈移位寄存器(LFSR)的流密码,它包括若干 LFSR 和一个非线性组成器。由于这种密码体制具有简洁性的特点,而且适当选择 LFSR 和非线性组合函数可使密钥流序列具有良好的统计特性和高的线性复杂度,因而它在现实生活中被广泛使用,其安全性能也一直受到很大的关注。为了抵抗许多已知对流密码的有效攻击(例如快速相关攻击[10],条件相关攻击[11]和插入攻击[12]等),Golic 在文[8]中给出了流密码体制设计需要遵循的一系列准则。而且为了抵抗不同类型的相关攻击,许多密码专家致力于构造既有高线性复杂度又有高阶相关免疫的布尔函数。

相关攻击的基本思想是通过识别发生器的输出和它的内部块某一块之间的相关性。然后,通过观察输出序列,获得关于其内部输出的一些信息。用这些信息和其它的相关性,搜索其它内部输出的相关性,直到整个发生器被破译。其本质是看作在某概率下解多元线性方程组的问题。最近相关攻击的应用范围扩展到了高阶相关攻击,在文[3]中,作者就利用 Toyocrypt 密码体制内部的相关特性,获得了状态变量的非线性低阶多元方程组,代数攻击的思想初具端倪。另外 XL 算法是解决低阶非线性多元方程组的有效算法,它的提出[13]使代数攻击开始蓬勃发展。

代数攻击的主要思想是,建立起初始密钥和输出密钥流比特之间的代数方程,然后运用线性化手段(或者 XL 算法)来解方程获得秘密的初始密钥,因此实际上是已知明文攻击。代数攻击首先应用在分组密码和公开密钥密码系统[2,7]。它首次应用于流密码是在 Courtois 对 Toyocrypt 的分析中[3]。由于此方法随后扩展到对 LILI-128的分析,从而引起了极大的注意。在文[5,6]中,Courtois 和 Meier 对不带记忆的线性反馈流密码的代数攻击做了研究,并且总结了快速攻击方法。早些时候,带记忆部分的流密码被认为可以更好的抵抗此类攻击,但是不久文[1,4]的出现,表明即使是这样的情况,流密码仍有可能受到攻击。

2 流密码代数攻击的基本模型和一般攻击方法

2.1 可被攻击的流密码基本模型

我们仅考虑同步流密码,即每一个状态独立于明文由前一个状态生成。或者考虑规则步进流密码,即流密码在已知方式下步进。然而,这个条件有时候可以放松,参看文[4～6]中描述的对 LILI-128的攻击。

为简单起见,我们限制在二进制流密码上讨论,这里状态和密钥流都是比特序列,且一次生成一比特。我们也限制这样的情况:计算下一状态的"连接函数"在 $GF(2)$ 上是线性的。例如一个 LFSR 对应的连接多项式,或者若干个 LFSR 的并列组合。我们设 L 为"连接函数",假设 L 是公开的,只有状态是秘密的。也假设由线性部分状态计算输出比特的非线性函数 F 是公开的,且与密码的初始密钥无关。

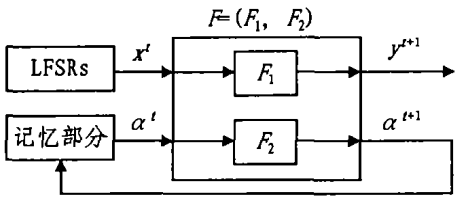
这样一个流密码的密码分析问题可以如下描述。设 $k=(k_0,\cdots,k_{n-1})$ 是一个 n-比特初始密钥,$s=k$ 是密码线性部分的初始状态,经每一步运动,计算 $s\leftarrow L(s)$ 得到新的状态。假设 F 使用 n 比特中的 k 比特生成下一状态,称 $x_0^{(t)},\cdots,x_{k-1}^{(t)}$ 为 $s_0^{(t)},\cdots,s_{n-1}^{(t)}$ 的固定子集,t 表示时刻。F 中有 l 比特内部记忆状

*)基金项目:国家自然科学基金项目(60073051);973项目(No:G1999035804)。张 莉 硕士研究生,研究方向:信息安全和密码学。刘海波 硕士研究生,研究方向:无线通信网和安全。白恩健 博士研究生,研究方向:信息安全和密码学。肖国镇 教授,博士生导师,长期从事密码学,编码学的研究。

态(无记忆状态是 $l=0$),表示为 $a_0^{(t)},\cdots,a_{l-1}^{(t)}$,其初始状态未知。也假设 F 有 m 比特的输出(通常情况下 $m=1$),输出状态表示为 $y_0^{(t)},\cdots,y_{m-1}^{(t)}$。因此,非线性组成部分 F 可表示为 $F=(F_1,F_2):GF(2)^{k+l}\to GF(2)^{m+l}$,给出当前记忆部分状态和输入,就可以计算下一个记忆状态和输出。

$$F:\begin{cases}(y_0^{(t+1)},\cdots,y_{m-1}^{(t+1)})=F_1(x_0^{(t)},\cdots,x_{k-1}^{(t)},a_0^{(t)},\cdots,a_{l-1}^{(t)})\\(a_0^{(t+1)},\cdots,a_{l-1}^{(t+1)})=F_2(x_0^{(t)},\cdots,x_{k-1}^{(t)},a_0^{(t)}\cdots,a_{l-1}^{(t)})\end{cases}$$

我们也可以用图形来直观的表示:



2.2 一般代数攻击的步骤

流密码的一般代数攻击如下:

① 通过某些方法(因不同的密码体制而异),发现状态比特和一些连续 M 比特输出之间的低阶为 d 的多元关系 Q,例如:$Q(s_0,s_1,\cdots,s_{n-1},y^{(0)},\cdots,y^{(M-1)})=0$。

② 由于密码构造的递归性,相同的方程可用于所有连续的 M 比特输出状态:

$$Q([L^t(k)]_0,[L^t(k)]_1,\cdots,[L^t(k)]_{n-1},y^t,\cdots,y^{(t+M-1)})=0$$

③ 观察密码的输出,用已知值代替 $y^{(t)},\cdots,y^{(t+M-1)}$。另外由于 L 的线性性,这些方程的阶均为 d。这样,只要有足够多的密钥流比特,我们就能获得一个超定义方程组(即大量以 k_i 为变量的阶为 d 的多元方程)。

④ 应用2000年欧密会上提出的 XL 算法来解低阶多元方程组。

⑤ 如果我们有充足数量的密钥流,XL 算法就可用所谓的线性化方法来代替,那样计算会更简单。n 个变量 k_i 可组成 $T\approx\binom{n}{d}$ 个阶不大于 d 的单项式(假设 $d\leqslant n/2$),将每一个单项式看作一个新的变量 V_j,则给出大约 $\binom{n}{d}+M$ 比特密钥流和 $R\geqslant\binom{n}{d}$ 个基于连续 M 比特输出的方程,我们就得到了 $R\geqslant T$ 个,带有 $T=\binom{n}{d}$ 个变量 V_j 的线性方程组,可用高斯消元法解决。

⑥ 理论上,线性化方法需要 T^ω 次运算,其中 T 为方程量的大小,$\omega\leqslant 2.376$,参看文[8]。然而我们所知道的最快的算法是 Strassen 的算法,需要 $7\cdot T^{\log_2 7}$ 次运算。

3 非线性组合生成器的代数攻击

3.1 不带记忆非线性组合生成器的代数攻击

这里讨论的是 $l=0,m=1$ 的情况,这时满足 $f(L^t(k_0,\cdots,k_{n-1}))=y^{(t)}$。Courtois 和 Meier 在文[6]中首先证明,虽然 f 函数未必低阶,但通过乘一个适当选择的低阶多元多项式,可以可观地降低方程的阶,其方法类似于 Courtois 和 Pieprzyk 对一些分组密码的攻击方法[7]。文中有一个结论:

定理1[6] 对于有 k 个输入的布尔函数 $f:GF(2)^k\to GF(2)$,存在阶不超过$\lceil k/2\rceil$的多元多项式 g,使得 $f\cdot g$ 的阶也不超过$\lceil k/2\rceil$。

由此发展了基于线性反馈流密码的一般代数攻击,成功的用于攻击 Toyocrypt 密码体制(2^{49}*CPU* 时钟内)和 LILI-128密码体制(2^{59}*CPU* 时钟内),其攻击复杂度至多是先前已知普通攻击复杂度的平方根,即普通攻击为 $\binom{n}{k}^\omega$ 次运算,而代数攻击 $\binom{n}{k/2}^\omega$ 为次运算。

针对代数攻击的思想,作者修改了文[3]中对布尔函数 f 的构造标准,指出 f 不仅需要满足平衡,高代数阶,高非线性度和高阶相关免疫,还应避免以下情况:

1. 对多元多项式 f,存在非零的多元多项式 g,使得 $f\cdot g$ 为低阶 d。

2. 或者对多元多项式 f,存在这样的乘积 $f\cdot g$,使得 $f\cdot g$ 能被一个低阶函数以概率为 $1-\varepsilon$ 逼近,其中 ε 足够小。

3.2 带记忆非线性组合生成器的代数攻击

文[1]中,F. Armknecht 和 M. Krause 研究了带记忆的非线性组成器(这里 $l\geqslant 1,m=1$),给出了一个定理的证明:

定理2[1] 对于任意固定的有 k 比特输入位,l 比特记忆位和1比特输出的组成器,必存在一个阶不超过$\lceil k(l+1)/2\rceil$的多元关系。

这个定理概括了第1节的定理1,即当 $l=0$时,存在阶不超过$\lceil k/2\rceil$的多元关系。

作者将证明结果应用到蓝牙系统的 E_0密钥流生成器,它是一个 $k=2,l=2,m=1$的带记忆生成器,使用了连续4比特的密钥流输出,成功得到了一个阶为4的方程组,计算复杂度约$2^{67.58}$,存储复杂度约$2^{46.14}$,如果运用文[5]的改进算法(第3节介绍),其运算操作数可降至2^{49}。但需要指出的是,当 k 和 l 超过4时,此攻击在实际中就不可行了。

根据文[1]的启发和已有的一些证明思想,Courtois 在文[4]中扩展了对带记忆组成器的研究,给出了类似文[1]中的更一般的结论,并将其推广到多比特输出的情况。同时指出当密码体制一次输出的比特数越多,其存在的代数方程的阶就越低,代数攻击的复杂度呈指数级下降。并且为了论证这个结论,对熟知的三个密码体制 E_0密钥流生成器,LILI-128和 Snow 生成器的改进模型进行了分析。沿用第2节关于流密码类型的定义,作者给出了一个重要定理:

定理3[4] 设 F 是一个密码体制中任意固定的非线性组成器,带 k 比特输入位 x_i,l 比特记忆位 a_i 和 m 比特输出位 y_i。设 d 和 M 是两个这样的整数,满足:

$$2^{Mm}\cdot\sum_{i=0}^{d}\binom{Mk}{i}>2^{Mk+l}$$

则考虑 M 比特连续状态$(t,\cdots,t+M-1)$,必存在一个关联这些状态的输入输出比特,阶不超过 d 的有关 $x_j^{(t)}$ 的多元方程 R。

$$R(x_0^{(t)},\cdots,x_{k-1}^{(t)},\cdots,x_0^{(t+M-1)},\cdots,x_{k-1}^{(t+M-1)};y_0^{(t)},\cdots,y_{m-1}^{(t)},\cdots,y_0^{(t+M-1)},\cdots,y_{m-1}^{(t+M-1)})=0$$

显然,定理3概括了先前的两个定理,是一个更一般的结论。当然,这只是在理论上证明多元代数方程的存在,怎样针对具体的密码体制发现这样的关系,还需要经过艰苦的工作。就像对 E_0密钥流生成器的攻击,是作者通过仔细的手工推算化简获得的。

上述代数攻击成立还有一个限制是,它要求记忆位比较小,即要求比较小。

3.4 快速代数攻击

到目前为止,我们已经清楚的知道,代数攻击成功与否的关键在于能否建立起初始密钥和输出密钥流之间的代数方程,并且能否有效的解方程。Courtois 在文[5]中总结了基于线性反馈流密码的快速代数攻击,目的是改进算法,降低运算复杂度。显然我们构造的代数方程中只包含两种单项式,一种

是只含初始密钥变量 k_i 的单项式，另一种是同时有初始密钥变量 k_i 和密钥流 y_i 的单项式。假设代数方程的类型是 $K^d \cup K^cY^f$，表示只含 k_i 的单项式最大阶为 d，而同时包含 k_i 和 y_i 的单项式，其 k_i 最大阶为 e，$f \geq 1$。作者利用的是满足 $e<d$ 的方程，证明了对于相同的 n 和 d，更小的 e 可以有一个更快的算法。快速算法的基本内容是给定连续的密钥流比特，利用连续方程间单项式的线性相关性（例如超过 $\binom{n}{d}$ 个的，阶为 d 的单项式组合必线性相关），约去所有阶为 $r(e<r\leqslant d)$ 的单项式，从而得到变量为 k_i 的阶为 e 的方程组，再解方程。此方法有效的降低了运算复杂度，攻击 Toyocrypt 密码体制需 $O(2^{20})$ 次运算，攻击 LILI-128密码体制需 $O(2^{31})$ 次运算，攻击蓝牙系统的 E_0 密钥流生成器需 $O(2^{49})$ 次运算，均是迄今已知的最快攻击。

4 f 函数为未知情况下的一般攻击方法（我们的一点思考）

文[4]中指出，如果一个带有参量(k,l,m)的非线性组成器仅有部分已知（或密钥相关），我们可以假设总共有 l' 比特未知，将这些比特看作是未知的记忆比特，这样我们就得到一个新的等价的非线性组成器$(k,l+l',m)$（其中附加的 l' 记忆比特不用更新，每一步运行都是一样的）。同样的，当 l' 不是太大时，代数攻击有效。

但是，在非线性部分不带记忆又完全未知的情况下，有没有代数攻击的可能呢？这是之前没有提及的需要我们思考的问题，下面我们根据文献的一些启发，提出非线性函数未知情况下的一般攻击方法。

根据定理1的证明，我们可以得到一个非常好的 $f(x)$ 与 $g(x)$ 之间的关系：

$\exists\ b\in\{0,1\}$，使得$\forall\ x\in\{0,1\}^k$，$f(x)=b\Rightarrow g(x)=0$

$f(x)$ 在 t 时刻的值即为 y_t，因此换句话说，一定存在阶不超过$\lceil k/2\rceil$的布尔函数 $g\neq 0$，使得当 $y_t=0$（或者1）时，$g(L^t(k_0,\cdots,k_{n-1}))=0$恒成立。

我们可以做如下运用：

f 函数如果与初始密钥有关，则不可知。在这里我们假设 f 函数虽然不可知，但是确定的。这样的情况是有可能的，例如 f 函数带有一个置换（S-盒），其初始置换与初始密钥作用产生一个全新的，不可知的置换。但是如果 f 函数的输入比特数 k 不大，我们仍可以进行攻击。我们总能找到阶不超过$\lceil k/2\rceil$的布尔函数 $g\neq 0$，使得当 $y_t=0$（或者1）时，$g(L^t(k_0,\cdots,k_{n-1}))=0$恒成立，而不管 $f(x)$ 是什么样的函数。我们只要得到这样的 $g(x)$，就能通过解方程求得初始密钥值。

具体步骤为：

① 假设 $g(x)$ 的阶为 $e(1\leqslant e\leqslant\lceil k/2\rceil)$（从低阶做起）。

② 构造阶不超过 e 的单项式集合$\{1,x_1,x_2,\cdots,x_k,x_1x_2,\cdots\}$，从中任意组合成阶等于 e 的布尔函数 $g(x)$。

③ 我们需要 $\binom{n}{e}$ 比特固定等于1或者0的已知位置的密钥流，解方程 $g(L^t(k_0,\cdots,k_{n-1}))=0$。若方程有解，就可能是我们要的初始密钥，可进行检验；若无解，回到①重复。由于密钥流的随机特性，我们大约需要 $2\times\binom{n}{e}$ 比特密钥流就可以得到所需的方程数。每一个确定的 $g(x)$ 需要 $O(\binom{n}{e}^\omega)$ 次计算。

最坏情况下，我们做到 $e=\lceil k/2\rceil$ 时才得到真正的密钥。这时候，我们共组合了大约 $2^{2^{k-1}}$ 种 $g(x)$，计算量等于 $\sum_{i=1}^{k/2}(2^{\sum_{i=0}^{i-1}\binom{k}{i}}\cdot(2^{\binom{k}{i}}-1)\cdot(\binom{n}{e}^\omega))$。

实际上，如果 $f(x)$ 可约，其最小因子 $r(x)$ 的阶为 d，则一定有 $d\leqslant\lfloor k/2\rfloor$，我们可有 $(f(x)+1)\cdot r(x)=0$，$r(x)$ 即为要求的 $g(x)$。所以一般情况下，我们从低阶做起，可得到一个阶较低的 $g(x)$。

结束语 代数攻击作为一个全新有效的攻击方法，正在引起人们广泛的关注。它的提出告诉我们，虽然一个密码体制可能满足之前的所有设计标准，但是只要密码非线性部分的输入位只用到线性反馈状态的一小部分子集，就是不安全的，易受攻击的。因此，需要增加新的密码体制设计标准，那就是：不存在关联密钥比特和输出比特之间的低阶多元关系。

代数攻击的提出也推动了新一轮的密码分析，除了上述对 Toyocrypt 密码体制，LILI-128密码体制，E_0 密钥流生成器和 Snow 生成器等的分析之外，韩国的 Dong Hoon Lee 等人将代数攻击的思想用于分析使用 n 个 LFSR 的加和生成器[9]，结合他们自己的研究成果，也找到了利用连续$\lceil \log_2 n\rceil+1$比特密钥流的，阶不超过$2^{\lceil\log_2 n\rceil}$的以初始密钥为变量的代数方程。

此外，我们还提出了对未知布尔函数 f 情况下的一般攻击方法，指出如果 f 所用的状态子集大小为 k，就一定能在阶不超过$\lceil k/2\rceil$的范围内找到一个 $g(x)$，使得 $f(x)g(x)=0$或者$(f(x)+1)g(x)=0$。事实上，$\lceil k/2\rceil$的上界对于实际的 $g(x)$绰绰有余。

当然，代数攻击还只是一个很新的攻击方法，涉猎的只是一小部分流密码类型。当 LFSR 的连接多项式不可知，或者步进方式非常复杂并且与整个密钥相关；或输出序列的生成方式非常复杂（例如缩减序列生成器）时，这里所使用的代数攻击方法就不能奏效了。因此，还有很多有趣的问题值得我们的进一步探讨。

参考文献

1 Armknecht F, Krause M. Algebraic attacks on combiners with memory. Advances in Cryptology-Crypto 2003, LNCS 2729, Springer-Verlag, 2003. 162～175

2 Courtois N. The security of Hidden Field Equation (HFE). CT-RSA 2001, LNCS 2020, Springer-Verlag, 2001. 266～281

3 Courtois N. Higher order correlation attacks. XL algorithm and Cryptanalysis of Toyocrypt. ICISC 2002, LNCS 2587, Springer-Verlag, 2002. 182～199

4 Courtois N. Algebraic attacks on combiners with memory and several outputs. E-print achieve, 2003/125

5 Courtois N. Fast algebraic attack on stream ciphers with linear feedback. Advances in Cryptology-Crypto 2003, LNCS 2729, Springer-Verlag, 2003. 176～194

6 Courtois N, Meier W. Algebraic attack on stream ciphers with linear feedback. Advances in Cryptology-Crypto 2003, LNCS2656, Springer-Verlag, 2003. 345～359

7 Courtois N, Pieprzyk J. Cryptanalysis of block ciphers with overdefined systems of equations. Asiacrypt 2002, LNCS 2501, Springer-Verlag, 2002. 267～287

8 Golic J D. On the Security of Nonlinear Filter Generators. FSE'96, LNCS 1039, Springer, 173～188

9 Lee D H, Kim J, Hong J W, Han J W, Moon D. Algebraic Attack on Summation Generators, E-print achieve, 2003/229

10 Meier W, Staffebach O. Fast correlation attacks on certain stream ciphers. Journal of Cryptology, 1989, 1(3): 159～176

11 Anderson R. Serching for the Optimum Correlation Attack, FSE'94, LNCS 1008, Springer, 137～143

12 Golic J D. On the Security of Nonlinear Filter Generator, FSE'96, LNCS 1039, Springer, 173～188

13 Shamir A, Patarin J, Courtois N, Klimov A. Efficient Algorithms for solving Overdefined Systems of Multivariate Polynomial Equations. Eurocrypt'2000, LNCS 1807, Springer, 392～407