对迭代型混沌密码的逆推压缩攻击

张 斌,金晨辉
(解放军信息工程大学电子技术学院,河南郑州 450004)

摘 要: 本文发现了迭代型混沌密码的一个新信息泄漏规律,即对每个时刻 i,由乱数序列求出的混沌映射在该时刻的可能输入(及可能密钥参数)全体都可用若干个区间的并集简单描述,且对多对一混沌映射而言,每个区间内都有等效解的概率很大,并随着时刻 i 的减小,区间的长度急剧缩短.基于该信息泄漏规律,本文提出了攻击迭代型混沌密码的一个新方法,即逆推压缩攻击方法.在一定的条件下,该攻击方法的成功率为 1,且计算复杂性、存储复杂性和数据复杂性都是密钥规模的线性函数.本文对廖旎焕等人在 2006 年发表的混沌扩频序列密码算法在密钥规模为 64 比特时做了 100 例攻击实验,每例实验平均仅需 0.11 秒就可求出等效密钥,从而验证了逆推压缩攻击方法的有效性和正确性.与现有的攻击混沌密码的一般方法相比,本文提出的方法是首个复杂性为密钥长度线性量级的攻击方法.

关键词: 混沌密码; 密码分析; 逆推压缩攻击; 等效密钥
中图分类号: TN918.1 **文献标识码:** A **文章编号:** 0372-2112 (2010) 01-0129-06

Inversion and Compression Attacks to Iterative Chaotic Ciphers

ZHANG Bin, JIN Chen-hui
(*Electronic Technology Institute, Information Engineering University, Zhengzhou, Henan* 450004, *China*)

Abstract: In this paper, we find a new law of information leaking out. We find that for each clock i, all possible inputs (and the parameter-formed key) of a chaotic map at the i-th clock which can produce the key stream forms an union of intervals, and the probability that each interval includes an equivalent solution is very high for many-to-one chaotic maps, and the length of intervals are decreased exponentially as the clock i decrease. Based on the new law of information leaking out, an inversion and compression attack to iterative chaotic ciphers is proposed for the first time. Under some conditions, the success rate of the attack is 1 and the computational complexity, the memory complexity and the data complexity are linear on the length of key. 100 experiments to the chaotic spreading sequences algorithm, presented by Liao Ni-huan et al in 2006, were done for 64 bits keys at a 2.5GHz of Pentium 4 PC. Each experiment just costs about 0.11 seconds in average, which implies the correctness and validity of our attack algorithm. Comparison with the known attack to chaotic ciphers, the inversion and compression attack is the first general attack that the complexity is linear on the length of key.

Key words: chaotic cipher; cryptanalysis; inversion and compression attack; equivalent key

1 引言

混沌序列具有非周期性、随机性、对初始条件的敏感依赖性等特性,因此利用混沌序列设计扩频序列、伪随机序列及加密算法的研究受到越来越多地关注.但是,混沌密码算法的分析理论还不成熟,对混沌密码的攻击方法也不多.目前对混沌密码的攻击方法主要有多分辨率攻击方法[1]、分割攻击方法[2,3]以及特定条件下的分割攻击方法[4,5].多分辨率攻击方法利用在密钥的最低 i 位全为 0 时,混沌映射输出的最低 j 位全为 0 的概率 $p_{i,j}$ 略大所产生的信息泄漏,借助对 $\{p_{i,j}\}_{j\geqslant 1}$ 的统计结果,从而由 $\{p_{i,j}\}_{j\geqslant 1}$ 与 i 之间的关系推断出密钥最低全 0 位的个数.该方法平均将某些混沌密码的密钥熵降低 2 比特,但平均所需的已知明文量和计算复杂性一般都很大;分割攻击方法利用迭代型混沌密码的参数和初态的低位比特对前若干输出信号的影响不大这个信息泄漏,并利用吻合度[3]的分布规律对该信息泄漏进行定量刻化,进而采取先穷举密钥的高比特块,再穷举密钥的低比特块的方法实施分割攻击,从而利用不长的乱数序列大大降低对迭代型混沌密码攻击的计算复杂性.然而,由于该分割攻击方法的计算复杂性本质上仍是指数级的,且吻合度的增长速度一般都低于密钥规模的增

收稿日期:2007-02-09;修回日期:2009-10-27
基金项目:河南省杰出青年科学基金(No.0312001800)

长速度,因而当密钥规模较大时,该攻击方法一般只能降低密钥熵,而难以在可行的时间内求出密钥.

本文从新的角度出发,考查了迭代型混沌密码信息泄漏的规律和利用方法.我们发现了混沌密码的一个新的信息泄漏规律:对于每个时刻 i,由乱数序列求出的混沌映射在每个时刻的可能输入(和可能密钥参数)全体都可用若干个区间的并集简单地描述,且对多对一混沌映射(如 Logistic 映射、Tent 映射、Chebyshev 映射、立方映射、逐段线性映射和逐段非线性映射[6]等)而言,每个区间都有等效解的概率很大,且随着时刻 i 的减小,区间的长度急剧缩短.基于上述信息泄漏规律,我们提出了对混沌密码的一种新的攻击方法,即逆推压缩攻击方法.该方法首先由第 t 时刻和第 $t-1$ 时刻的乱数,求出第 $t-1$ 时刻混沌映射的可能输入全体形成的所有区间,然后再求出这些区间关于混沌映射的原像集合中能产生第 $t-2$ 时刻乱数的点形成的所有区间,依此类推,最后求出混沌映射的初态或等效的初态,完成对混沌密码密钥的求解.在一定的条件下,该攻击方法的成功率为 1,且计算复杂性、存储复杂性和数据复杂性都是密钥规模的线性函数,因而可快速完成对任一规模密钥的破解.本文还对廖旎焕等人提出的混沌扩频序列密码算法[7]在密钥长度为 64 比特时做了 100 例攻击实验,每例实验平均仅需 0.11s 就可求出等效密钥,从而验证了逆推压缩攻击方法的有效性和正确性.与现有攻击混沌密码的一般方法[1~5]相比,本文提出的方法是首个复杂性为密钥规模的线性量级的攻击方法.

2 对迭代型混沌密码的分析

2.1 迭代型混沌密码的基本模型

迭代型混沌密码最基本的模型包括两个密码变换,一个是混沌映射 $f(x)$,另一个是量化函数 $g(x)$.混沌密码算法一般分为三步:

(1)产生混沌状态序列:以 $x_0\in[-1,1]$ 为密钥,对于 $i\geqslant 1$,由 $x_i=f(x_{i-1})$ 产生混沌状态序列 $\{x_i\}_{i=1}^{\infty}$;

(2)利用量化函数产生乱数序列:对于 $i\geqslant 1$,由 $b_i=g(x_i)$ 产生乱数序列 $\{b_i\}_{i=1}^{\infty}$;

(3)利用乱数序列 $\{b_i\}_{i=1}^{\infty}$ 对明文序列 $\{p_i\}_{i=1}^{\infty}$ 加密产生密文序列 $\{c_i\}_{i=1}^{\infty}$.加密方法一般采用 $c_i=p_i\oplus b_i$.

为简单起见,本文以下都针对上述模型进行分析,不再一一具体指出,但所得的结论原则上对将混沌映射的参数和初态同时作为密钥的模型仍然适用,具体的分析工作我们留待以后进行.

备注:数字化混沌密码必须以有限精度方式实现.有限精度实现的含义有以下两方面的内容:

(1) 将无限精度小数 $x\in[-1,1]$ 转换为 n 精度小数的方法.这类方法有很多,其中最简单的方式是将 $x=\sum_{i=0}^{\infty}x_i2^{-i}$ 用 $x^{(n)}=\lfloor 2^n x\rfloor/2^n=\sum_{i=0}^{n-1}x_i2^{-i}$ 代替,这里 $x_i\in\{0,1\}$,$i\geqslant 1$ 或者 $x_i\in\{0,-1\}$,$i\geqslant 1$.为简单起见,本文以下均按该方式分析,对其它转换方式也可类似地分析.

(2)函数 $f(x)$ 的有限精度实现方式.主要方法有两种.第一种是首先将输入 x 用 $x^{(n)}$ 代替,并将每个指令的运算结果都转化为 n 精度小数,然后将之作为下步运算的输入;第二种是首先将输入 x 用 $x^{(n)}$ 代替,然后按实数运算计算出 $f(x^{(n)})$,最后将 $f(x^{(n)})$ 的 n 精度小数 $[f(x^{(n)})]^{(n)}$ 作为函数 $f(x)$ 在有限精度实现时的输出.为简单起见,本文以下均假设函数 $f(x)$ 以第二种方式完成有限精度实现,但所用的分析方法原则上也适用于第一种实现方式.

本文提出的对迭代型混沌密码的逆推压缩攻击方法本质上是穷举攻击方法,但由于混沌映射存在逆像压缩性这种新的信息泄漏规律导致在一定条件下逆推压缩攻击方法的计算复杂性、存储复杂性均降为密钥规模的线性函数,使得该方法对迭代型混沌密码的破译很有效.

2.2 迭代型混沌密码的逆像压缩性和逆推压缩攻击

混沌映射 f 通常是连续函数或具有有限个间断点的间断连续函数,这个性质除了导致输入低位的变化对输出高位的变化影响不大[2,3]外,还将导致值域中一个区间 I 的逆像 $f^{-1}(I)=\{x:f(x)\in I\}$ 仍是一个区间或是有限个区间的并,因而只需计算出这些区间的端点,就可利用它们描述逆像的取值范围,而不需要穷举和存储这些区间中的每个点,从而可以大大减少求解 f 的全部可能输入时的计算量和存储量.

此外,由于混沌映射通常为多对一的映射且具有扩大输入的差异的作用,因而在单调区间上的两个数 x,y 一般均使 $|f(x)-f(y)|>|x-y|$ 成立.例如 Logistic 映射、Tent 映射、Chebyshev 映射、逐段线性映射和逐段非线性映射等混沌映射都具有该特性.该特性将导致 $f^{-1}(I)$ 中各区间的长度均小于区间 I 的长度,且各区间在函数 f 下的像通常也是区间 I.这就是混沌映射的逆像压缩性.

量化函数 $g(x)$ 通常设计为直接抽取当前混沌状态的若干比特,易证在第 i 时刻的乱数 b_i 已知的条件下,能产生该乱数的当前状态的所有可能值也构成一个或若干个区间.因此,能够使混沌映射的输出落入指定区间 I 且能够量化为指定乱数的所有可能输入也构成一个区间或若干个区间的并.同时,乱数提供的信息

将导致输入的所有可能区间的总长度一定小于区间 I 的长度.上述特性将导致对混沌密码的一种更加有效的攻击方法,即逆推压缩攻击.逆推压缩攻击是一种已知明文攻击方法,其基本思路如下:

设已知混沌密码的前 N 个乱数 $b_1, b_2, \cdots, b_N$.我们首先求出并存储满足 $gf(x) = b_N$ 和 $g(x) = b_{N-1}$的所有可能点 x 构成的全部区间 $I_1^{(N-1)}, I_2^{(N-1)}, \cdots, I_{j_{N-1}}^{(N-1)}$的端点;接着求出并存储满足 $f(x) \in \bigcup_j I_j^{(N-1)}$ 和 $g(x) = b_{N-2}$的所有可能点 x 构成的全部区间的端点;如此继续下去,对 $i = N-2, N-3, \cdots, 2$,依次求出并存储满足 $f(x) \in \bigcup_j I_j^{(i)}$ 和 $g(x) = b_{i-1}$的所有可能点 x 构成的那些区间 $I_1^{(i-1)}, I_2^{(i-1)}, \cdots, I_{j_{N-1}}^{(i-1)}$的端点.最后求出并存储满足 $f(x) \in \bigcup_j I_j^{(1)}$ 的所有可能点 x 构成的那些区间 $\bigcup_j I_j^{(0)}$ 作为密钥的候选值.由于区间 $\bigcup_j I_j^{(i)}$ 的总长度随着 i 的减小而急剧缩短,因而上述算法只需很少步骤就可求出可能的密钥.由于求逆时仅需计算出每个逆像区间的端点,而不需穷举和存储区间中的所有点,因此上述算法的运算速度很快.该逆推压缩攻击方法本质上是逆推攻击方法,但由于算法只需存储各区间的端点,因而大大减少了逆推压缩攻击所需要的存储量.如果再采用回溯法设计具体的攻击算法,则攻击算法的存储量就会更加减少.下面研究制约逆推压缩攻击方法性能的几个指标.

定义 1(逆推压缩比)　设 $I_1^{(t)}, I_2^{(t)}, \cdots, I_{j_t}^{(t)}$ 和 $I_1^{(t-1)}, I_2^{(t-1)}, \cdots, I_{j_{t-1}}^{(t-1)}$都是不交的区间,且

$$\{x : f(x) \in \bigcup_{j=1}^{j_t} I_j^{(t)} \text{ 且 } g(x) = b_{t-1}\} = \bigcup_{j=1}^{j_{t-1}} I_j^{(t-1)}$$

记 L_t 是区间 $I_1^{(t)}, I_2^{(t)}, \cdots, I_{j_t}^{(t)}$的总长度,则称 $\rho_t = L_{t-1}/L_t$ 为混沌密码的逆推压缩比,并称 $\xi = j_{t-1}/j_t$ 为区间个数扩张比.

设 ρ_t 都近似为 ρ,当混沌映射 f 在$[-1, 1]$上定义时,利用 N 个乱数进行逆推攻击所得的 $I_1^{(t)}, I_2^{(t)}, \cdots, I_{j_t}^{(t)}$的总长度近似为 $2\rho^{N-t+1}$.因此,如果混沌密码采用 n 精度小数实现,则当 $2\rho^{N-t+1} < 2^{-n}$时,区间 $I_1^{(t)}, I_2^{(t)}, \cdots, I_{j_t}^{(t)}$中含有的 n 精度数的数学期望为 $2^{n+1}\rho^{N-t+1} < 1$,因而此时所保留的基本上都是正确密钥或其等效密钥,故逆推压缩比 ρ 反映了逆推压缩攻击的压缩速度,因而是逆推压缩攻击的一个重要指标.由此可见,将区间 $I_1^{(t)}, I_2^{(t)}, \cdots, I_{j_t}^{(t)}$压缩成单点集所需要的步骤 $-(n+1)/\log_2\rho$ 是密钥规模 $n+1$ 的线性函数.此时,再借助乱数 $b_1, b_2, \cdots, b_t$,就可利用逆推压缩攻击方法,最终求出正确密钥.

另一方面,由于逆推压缩攻击方法求出的 $\bigcup_j I_j^{(t)}$ 就是能产生乱数 $b_t, b_{t+1}, \cdots, b_N$ 的所有可能的 x_t 构成的集合,因而与利用 $b_t, b_{t+1}, \cdots, b_N$ 穷举攻击 x_t 所得的结果相同,故逆推压缩攻击已经利用了所有可利用的信息,与穷举攻击的差异仅在于二者的实现方式、计算复杂性和存储复杂性不同,但二者的数据复杂性相同.因此,当乱数是 r 比特数时,每个乱数将提供 r 比特的信息,故攻击密钥规模为 $n+1$ 的混沌密码算法平均需要 $(n+1)/r$ 个乱数.换句话说,当直接沿用上述思路设计逆推攻击算法时,逆推压缩攻击平均只需$(n+1)/r$ 步.这同时也说明混沌密码的逆推压缩比近似为 $1/2^r$.实验结果也证实了这个推测.

区间个数的扩张比是制约逆推攻击算法性能的一个重要指标.当求出的小区间个数 j_t 的增长速度不快时,逆推攻击方法存储复杂性的增长速度不会影响到算法的实现.但一般而言,如果小区间个数 j_t 增长很快,就会造成需要存储的区间端点个数增长过快,造成最终存储不下的问题.为解决这个问题,在攻击算法的具体设计时,我们可以使用回溯法,从而通过以时间换空间的方法,降低攻击算法所需的存储量.但这种方法同时将会增加攻击算法的计算复杂性.

下面首先给出等效密钥的概念.

定义 2　设 k_1 和 k_2 是某序列密码算法的两个密钥,如果它们产生的乱数序列相同,则称 k_1 和 k_2 为等效密钥.

接着,我们基于回溯法,给出对混沌密码的逆推压缩攻击算法.

基于回溯法的逆推压缩攻击算法:

输入:乱数序列 $b_1, b_2, \cdots, b_N$;

输出:混沌密码的密钥或等效密钥.

逆推压缩攻击算法的流程:

Step1　初始化 $t = N$;

Step2　求出使$\{x : g(x) = b_t\} = \bigcup_{j=1}^{j_t} I_j^{(t)}$ 的不交区间 $I_1^{(t)}, I_2^{(t)}, \cdots, I_{j_t}^{(t)}$,执行 $M[t][0] \leftarrow j_t$ 并将第 j 个区间 $I_j^{(t)}$ 的端点赋予 $M[t][j]$,执行 $m[t] \leftarrow 1$;

Step3　如果 $m[t] > j_t$,则当 $t = N$ 时,终止算法,当 $t < N$ 时,执行 $t++$ 和 $m[t]++$ 后返回执行 Step3;如果 $m[t] \leqslant j_t$,则分两种情况:

(1)设$\{x : f(x) \in I_{m[t]}^{(t)}$且 $g(x) = b_{t-1}\} = \Phi$,则执行 $m[t]++$ 后返回执行 Step3;

(2)设$\{x : f(x) \in I_{m[t]}^{(t)}$且 $g(x) = b_{t-1}\} = \bigcup_{j=1}^{j_{t-1}} I_j^{(t-1)} \neq \Phi$ 是不交区间 $I_1^{(t-1)}, I_2^{(t-1)}, \cdots, I_{j_{t-1}}^{(t-1)}$的并,则当 $t = 2$ 时,执行 Step4,当 $t > 2$ 时,执行 $t \leftarrow t-1$ 后执行 $M[t][0] \leftarrow j_t$ 并将第 j 个区间的端点赋予 $M[t][j]$;执行 $m[t] \leftarrow 1$;执行 Step3.

Step4 计算满足 $gf(x)\in\bigcup_{j=1}^{j_1} I_j^{(1)}$ 的所有可能点 x，检验其中是否存在等效密钥.若存在等效密钥,则当算法只需求出一个等效密钥时,算法终止;当算法需要求出全部等效密钥时,执行 $m[t]++$ 后返回执行 Step3.若不存在等效密钥,执行 $t\leftarrow N$ 和 $m[N]++$ 后返回执行 Step3.

在上述算法中,$m[t]$用来记录当前处理的第 t 层小区间 $I_j^{(t)}$ 的序号 j,以便在小区间 $I_j^{(t)}$ 未被否定时,可以利用乱数继续对 $I_j^{(t)}$ 进行检验,直到 $I_j^{(t)}$ 被否定或者找到一个等效密钥为止,但在小区间 $I_j^{(t)}$ 被否定后,可以利用回溯法查出需要检验的小区间 $I_{j+1}^{(t)}$的序号,直接检验下个需要检验的区间.

对一些特殊的混沌映射,如 Logistic 映射、Tent 映射、Chebyshev 映射、逐段线性映射和逐段非线性映射等混沌映射,由于它们都是多对一的,因而导致混沌密码具有大量的等效密钥.如果只要求攻击算法求出一个等效密钥,则攻击算法的运行速度可以达到密钥规模的线性量级.为此我们引入下面的概念:

定义 3 如果 $f(x_1)=f(x_2)=y$,则称 x_1 和 x_2 为 $f(x)=y$ 的等效解,并称集合 $f^{-1}(y)=\{x:f(x)=y\}$中元素个数为等效解数.

定义 4(等效区间个数膨胀比) 设 I 是$[-1,1]$中一个区间,$f^{-1}(I)=\bigcup_{j=1}^{j_I} I_j$ 是 j_I 个不交区间的并,又设等效解数平均为 m,则称 j_I/m 为等效区间个数膨胀比.

等效区间个数膨胀比 j_I/m 反映了求出一个等效解平均需要搜索的小区间个数.下面证明对于大多数常用混沌映射,等效区间个数膨胀比 j_I/m 都近似为 1.

定理 1 如果混沌映射 f 为偶函数,则 f 至少有 x 和 $-x$ 是等效解.

易证,Logistic 映射 $f(x)=1-2x^2$,Chebyshev 映射 $f(x)=\cos(4\cos^{-1}x)$均为偶函数.

定理 2 (1)对于 Logistic 映射

$$f(x)=4x(1-x),x\in[0,1],$$

x 和 $1-x$ 是$f(x)=y$ 的等效解;

(2)对于 Tent 映射

$$f_p(x)=\begin{cases}x/p,x\in[0,p];\\(1-x)/(1-p),x\in(p,1].\end{cases}$$

py 和 $1-(1-p)y$ 是$f(x)=y$ 的等效解;

(3)对于 Chebyshev 映射

$$f(x)=\cos(4\cos^{-1}x),x\in[-1,1],$$

下述 4 点是 $f(x)=y$ 的等效解

$$\cos(\frac{1}{4}\cos^{-1}y),-\cos(\frac{1}{4}\cos^{-1}y),$$

$$\sin(\frac{1}{4}\cos^{-1}y),-\sin(\frac{1}{4}\cos^{-1}y);$$

(4)对于逐段线性映射

$$f_p(x)=\begin{cases}x/p,&x\in[0,p);\\(x-p)/(0.5-p),&x\in[p,0.5];\\f_p(1-x),&x\in[0.5,1).\end{cases}$$

py,$1-py$,$p+(0.5-p)y$ 和 $1-p-(0.5-p)y$ 是$f(x)=y$ 的等效解;

(5)对于逐段非线性映射

$$f(x)=\begin{cases}1-\sqrt{1-2x},x\in[0,0.5);\\\sqrt{2x-1},&x\in[0.5,1].\end{cases}$$

$y-\frac{1}{2}y^2$ 和$\frac{1}{2}(1+y^2)$是 $f(x)=y$ 的等效解.

(6)对于上述 5 个混沌映射,等效区间个数膨胀比都是 1.

证明 (1)至(5)的结论显然成立.(6)上述 5 个混沌映射均是多对一的连续映射且在其单调区间上均存在一个等效解,因此等效解数与该混沌映射单调区间的个数相等.对于区间 I,可在混沌映射的每个单调区间上分别求得 I 的一个逆像小区间,各逆像小区间的并就是区间 I 的逆像集,即逆像小区间个数等于单调区间的个数.因此上述 5 个混沌映射的等效区间个数膨胀比均是 1.

定理 2 的结论说明,如果只要求攻击算法求出一个等效密钥,对于利用定理 2 中的混沌映射构造的密码算法,由于等效区间个数膨胀比均为 1,即每个非空区间均包含一个能产生相应乱数的等效解,因而由每个区间最终都能求出混沌密码的一个等效解.因此,基于回溯法的逆推压缩攻击算法在实际的执行过程中,基本上没有回溯的过程,故求出一个等效密钥的攻击算法的计算复杂性就是其数据复杂性,因而是密钥规模的线性函数.

定理 3 设 $f(x)$是定理 2 中一个混沌映射且其平均等效解数是 m,基于该混沌映射设计的迭代型混沌密码的输出乱数 b_i 为 r 比特,设 ξ 为该混沌密码算法的区间个数扩张比,T 为由乱数 b_i 求出的混沌状态 x_i 构成的小区间平均个数,则基于回溯法的逆推压缩攻击算法的成功率为 1,数据复杂性近似为 $N=(n+1)/(r+\log_2 m)$,存储复杂性约为 $2T+1+(2\xi+1)(n+1)/(r+\log_2 m)$,如果仅需求得混沌密码算法的一个等效密钥,则计算复杂性约为 $2T+2T\xi(n+1)/[m(r+\log_2 m)]$.

证明 基于回溯法的逆推压缩攻击算法本质上是穷举攻击算法,只是实现方式上不同,因此最终一定能够得到正确密钥或等效密钥,成功率为 1.

当混沌密码算法以 n 精度实现时,由于混沌映射在$[-1,1]$上取值,故该算法的密钥规模为 $n+1$ 比特.

由于逆推压缩攻击方法本质上是穷举攻击,当利用乱数 b_i 进行攻击时,逆像区间总长度平均缩小为原像区间总长度的 $1/2^r$,即混沌状态 x_{i-1}的变化量平均仅为 x_i 变化量的 $1/2^r$,又因混沌映射 $f(x)$的平均等效解数为 m,故混沌状态 x_{i-1}的每个等效解的变化量平均仅为 x_i 变化量的 $1/(m2^r)$.因此,在利用乱数提供的信息的同时,由于混沌映射自身存在等效解,逆推压缩攻击方法每逆推一次所提供的等效密钥的信息量为 $r+\log_2 m$.故由信息论知,基于回溯法的逆推压缩攻击算法的数据复杂性为 $N=(n+1)/(r+\log_2 m)$.

对于 $t=N$,需存储由乱数 b_N 决定的混沌状态 x_N 可能取值的 T 个区间及 j_N 的值.由于混沌密码算法的区间个数扩张比为 ξ,故对于 $1\leqslant t\leqslant N-1$,有 $j_t\approx\xi$,需存储 j_t 的值及 j_t 个区间的端点值,故存储复杂性近似为

$$2T+1+(2\xi+1)N$$
$$=2T+1+(2\xi+1)(n+1)/(r+\log_2 m).$$

由于 Step1 平均产生 x_N 的 T 个区间,故 Step1 的计算复杂性为 $2T$.由于集合$\{f^{-1}(x_{N+1})\}$中平均有 m 个元素,故当 Step1 所产生的某个区间与集合$\{f^{-1}(x_{N+1})\}$的交集非空时,由定理 2 知 $f(x)$的等效区间膨胀比为 1,即利用该区间进行逆推攻击所得的每个区间均包含等效解,故攻击算法仅需依次搜索诸 x_t 的第 1 个区间就可找到等效密钥而终止,此时对于 $1\leqslant t\leqslant N-1$,仅需计算出 $j_t\approx\xi$ 个区间的端点值,故计算复杂性平均为 $2\xi(n+1)/(r+\log_2 m)$;当 Step1 所产生的某个区间与集合$\{f^{-1}(x_{N+1})\}$不交时,如果利用该区间进行逆推攻击,则攻击算法在依次搜索诸 x_t 的第 1 个区间后时,该区间一定被排除,此时攻击算法直接否定由 Step1 所产生的该区间进而对 Step1 所产生的下个区间重新开始逆推压缩,计算复杂性同样近似为 $2\xi(n+1)/(r+\log_2 m)$.在 T 个区间中只需平均搜索 T/m 个区间就可搜索到 1 个包含集合$\{f^{-1}(x_{N+1})\}$的某元素的区间,故 Step2 和 Step3 的计算复杂性近似为

$$(T/m)\times 2\xi(n+1)/(r+\log_2 m)$$
$$=2T\xi(n+1)/[m(r+\log_2 m)].$$

因此攻击算法的计算复杂性为

$$2T+2T\xi(n+1)/[m(r+\log_2 m)]$$

定理 3 说明基于回溯法的逆推压缩攻击算法的数据复杂性和存储复杂性均为密钥规模的线性量级,当混沌密码算法的等效区间个数膨胀比为 1 且仅需求出一个等效密钥时,该攻击算法的计算复杂性也为密钥规模的线性量级,此时攻击算法非常有效.

备注　由于数字化混沌算法都利用有限精度实现,因而在攻击算法的实现时,各小区间的端点可能会出现计算上的误差,该误差就可能导致算法程序中所记录的小区间并不包含真正的等效解.为避免这一情形的出现,在具体的实现时,我们可将小区间适当扩大.由于这种扩大只是增大了逆推压缩比而不增加区间个数扩张比和等效区间个数膨胀比,因而只是稍微增加了攻击算法的计算复杂性而其存储复杂性不会有明显的增加.实验也证实了这个猜测.

3　逆推压缩攻击算法的攻击实例

3.1　混沌扩频序列算法介绍

文献[7]选取 Logistic 映射、立方映射和 Chebyshev 映射构建了一种广义混沌映射:

$$f(x)=a_0+a_1x+a_2x^2+a_3x^3+a_4\cos(k\cos^{-1}x)\qquad(1)$$

当 $a_0=1,a_2=-2,a_1=a_3=a_4=0$ 时,式(1)对应于 Logistic 映射;当 $a_1=3,a_3=-4,a_0=a_2=a_4=0$ 时,式(1)对应于立方映射;当 $a_0=a_1=a_2=a_3=0,a_4=1,k=4$ 时,式(1)对应于 Chebyshev 映射.构建广义混沌映射的目的是在统一的结构模式下,通过一定的参数切换方式实现多种混沌映射,并且参数的切换原则可根据需要任意确定.下面以文献[7]中分析其所产生乱数序列的性质时选用的参数切换原则为例,对该密码算法进行描述和分析.

该混沌扩频序列算法[7]包括三步:

(1)产生混沌序列.以 $x_0^{(n)}\in[-1,1]$为密钥,对于 $i\geqslant 1$,由 $x_i^{(n)}=f(x_{i-1}^{(n)})$产生混沌序列$\{x_i^{(n)}\}_{i=1}^{\infty}$,其中每迭代 300 次切换一次参数,对应的映射依次为 Logistic 映射,立方映射,Chebyshev 映射.

(2)采样混沌序列.对混沌序列$\{x_i^{(n)}\}_{i=1}^{\infty}$进行 d 采样得到新序列$\{x_i^{(n)\prime}\}_{i=1}^{\infty}$,即 $x_i^{(n)\prime}=x_{d\times i}^{(n)}$.

(3)量化采样序列产生乱数序列.对每次参数切换后的单一映射序列分别进行处理,量化函数用门限函数

$$b_i=\begin{cases}0, & \text{if}\quad x_i^{(n)\prime}<h;\\ 1, & \text{if}\quad x_i^{(n)\prime}\geqslant h.\end{cases}$$

式中的 h 为对应各段实值序列的平均值,产生乱数序列$\{b_i\}_{i=1}^{\infty}$.

文献[7]对 $d=3$ 时所产生的乱数序列进行了随机性检验,得出了其随机性很好的结论.

3.2　混沌扩频序列算法分析

由混沌序列的随机特性可知,混沌实值序列平均值 h 的期望 $E(h)=0$,因此其采样序列平均值 h 的期望 $E(h)=0$,故在加密算法的量化函数中可认为 $h=0$.

由于乱数序列是对混沌序列的采样序列量化后得到的,因此对其进行逆推压缩攻击时可将混沌映射的 d 次复合视为新的混沌映射.下面仅对 $d=3,h=0$,实现精度 $n=64$ 比特的加密算法进行分析.

加密算法首先由 Logistic 映射的 3 次复合产生乱数序列,Logistic 映射的 3 次复合构成的映射的等效解数 $m=8$,每个乱数的比特数 $r=1$,区间个数的扩张比 $\xi=4$,等效区间个数膨胀比为 1,故逆推压缩攻击算法的数据复杂性 $N=65/4\approx17<100$,因此在对加密算法[7]攻击时仅需利用由 Logistic 映射产生的乱数序列即可.由量化函数知,由乱数 b_N 可确定混沌状态 $x_N^{(n)\prime}$ 的正负号,即 $T=4.5$,由定理 3 知,利用基于回溯法的逆推压缩攻击算法对混沌扩频序列算法进行攻击时计算复杂性近似为 $2^{6.36}$、存储复杂性近似为 $2^{7.29}$.

我们针对以 64 比特精度小数实现的混沌扩频序列算法做了 100 例攻击实验.为保证攻击算法能使各区间包括等效解以及尽量压缩为一点,我们多利用了几个已知乱数,即使用了长度为 $N=27$ 的乱数序列.在主频为 2.5GHz 的 Pentium4PC 机上,100 例攻击实验共用了 11s,即 1 例实验仅需 0.11s 就可求出一个等效密钥.实验结果验证了逆推压缩攻击算法的正确性和有效性.

4 逆推压缩攻击算法和已有的攻击算法比较

文献[1]提出了对混沌密码的多分辨率攻击方法,该方法平均将密钥熵降低 2 比特.但该方法所需的已知明文量和计算复杂性与密钥的分辨率有关.分辨率越小,计算复杂性和数据复杂性越小.由于密钥在随机情况下的分辨率很大,因而该攻击方法平均所需的已知明文量和计算复杂性一般都很大.

对于以混沌初值 x_0 为密钥的迭代型混沌密码算法,分割攻击方法[2~5]的基本思路是将 x_0 由高到低分割成若干块,先穷举 x_0 的高位比特块,利用筛选条件保留所有候选密钥块,再以 x_0 高位比特块的候选值为基础穷举 x_0 的次高位比特块,再利用筛选条件保留所有的候选密钥块,反复迭代这一过程,最终求得密钥 x_0.

由于分割攻击方法必须对分割后的密钥块依次穷举,因此分割攻击方法的计算复杂性通常为指数量级.而逆推压缩攻击方法的计算复杂性在一定条件下为密钥规模的线性量级,因而比分割攻击方法更加有效.

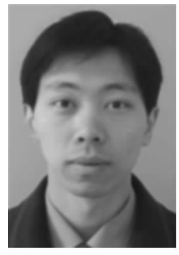
此外,随着迭代次数的增加,迭代型混沌密码初态的低位比特对乱数的影响将越来越大,这就导致分割攻击方法在攻击初态的高位比特时,只能利用前面有限个乱数信号提供的信息,而且分割攻击所能利用的乱数信号长度的增长速度一般低于密钥长度 n 的增长速度.这就导致了分割攻击虽然能够有效地降低密钥熵,但对于较大的 n,分割攻击方法的计算复杂性仍然很大而无法实现对具体密码算法实际上的破译.

逆推压缩攻击方法所需的乱数序列较少,而且在一定条件下计算复杂性、存储复杂性随密钥长度的增大呈线性增长,因此该方法对于攻击特殊的迭代型混沌密码更加有效.

5 结论

本文研究了迭代型混沌密码模型的分析问题,发现了混沌密码的一个新的信息泄漏规律,即对于每个时刻 i,由乱数序列求出的混沌映射在每个时刻的可能输入(和可能密钥参数)全体都可用若干个区间的并集简单地描述,且对多对一混沌映射而言,每个区间都有等效解的概率很大,且随着时刻 i 的减小,区间的长度急剧缩短.基于这个信息泄漏规律提出了逆推压缩攻击方法.在一定的条件下,该攻击方法的成功率为 1,且计算复杂性、存储复杂性和数据复杂性都是密钥长度的线性函数.与已有的攻击混沌密码的方法相比,本文提出的方法是首个复杂性为密钥长度的线性量级的攻击方法.实验结果验证了该攻击方法的有效性和正确性.如何利用逆推压缩攻击的思想对混沌密码的其他模型进行分析,还有待进一步研究.

作者简介:



张 斌 男,1982 年 10 月出生于山西太原,2004 年和 2007 年在解放军信息工程大学电子技术学院获理学学士和军事学硕士学位,现为解放军信息工程大学博士研究生,主要研究方向是密码学.

E-mail:dzjszhangbin@126.com

金晨辉(通信作者) 男,1965 年 3 月出生于河南扶沟县,解放军信息工程大学电子技术学院教授,博士生导师,主要研究方向是密码学和信息安全.

E-mail:jinchenhui@126.com

参考文献:

[1] 李树钧,牟轩沁,纪震,等.一类混沌流密码的分析[J].电子与信息学报,2003,25(4):473 - 479.
Li Shu-jun, Mou Xuan-qin, Ji Zhen, et al. Cryptanalysis of a class of chaotic stream ciphers[J]. Journal of Electronics & Information Technology, 2003, 25(4): 473 - 479. (in Chinese)

[2] 金晨辉.一个基于混沌的分组密码算法的分析[J].中国工程科学,2001,3(6):75 - 80.

(下转第 140 页)