基于混沌时间序列分析的信息注入技术研究

李肖坚，夏春和，吕艳丽
（北京航空航天大学计算机学院　　北京　100083）

〔摘要〕 初始序列号猜测是在信息对抗中谋求信息优势的一项关键技术。目前常用的初始序列号猜测算法是 Michal 算法。在对 Michal 算法进行分析的基础上，结合混沌时间序列分析方法和线性回归法，提出了一种新的初始序列号猜测算法并加以验证；结果表明，序列号猜测对信息注入是有效的，可以减少注入的不确定性，目前操作系统的 ISN 生成算法仍然存在安全风险。

〔关键词〕 初始序列号，混沌时间序列，线性回归，信息熵　　〔中图分类号〕 TP393
〔文献标识码〕 A　　〔文章编号〕 1000-7857(2005)05-0025-03

RESEARCH ON INFORMATION INJECTION TECHNOLOGY BASED ON CHAOTIC TIME SERIES

LI Xiao-jian, XIA Chun-he, LV Yan-li
(School of Computer, Beihang University, Beijing 100083, China)

Abstract: Initial Sequence Number (ISN) prediction is a key technology of striving for information superiority. On the base of Michal arithmetic, and combined with chaotic time series and linear regression, a new ISN prediction method is presented and validated by formal proof .The result indicates that present Operation Systems' TCP ISN generation methods still have serious security risk.

Key Words: Initial Sequence Number, chaotic time series, linear regression, information linear regression
CLC Number: TP393　　**Document Code**: A　　**Article ID**: 1000-7857(2005)05-0025-03

1 引言

以信息技术为核心的高科技在军事领域的迅猛发展，武器装备信息化水平的不断提高，同时也深刻地改变了战争理论，出现了一种全新的战争形势——信息战。“信息战”是指在军事行动的准备和实施中，为夺取和保持对对方的信息优势，按照统一的意图和计划所采取的信息保障、信息对抗和信息防护的综合措施。信息战中，信息赋予了全新涵义，成为重要战争资源。信息注入是一种重要的信息对抗手段，即给目标主机注入虚假信息使对方能够接受并认为是正确的。

信息注入的一种方法就是伪造对方认为正确的数据包，在面向连接的 TCP 协议中，一个正确的 TCP 包要包含正确的源地址、目的地址、源端口、目的端口和序列号。对于攻击者而言，获得源地址、目的地址、源端口和目的端口是容易的，攻击的主要困难是获得目前连接的初始序列号(ISN)。为了确保 TCP 连接的完整性，每一个单向连接数据流都要分配一个独一无二的序列号。攻击如果要用假地址建立连接，或者把恶意数据插入数据流中来破坏 TCP 连接的完整性，必须首先估计 ISN[1]。

ISN 猜测的一般方法是，攻击者与被攻击主机的一个端口建立起正常的连接。通常，这个过程被重复若干次，并将目标主机最后所发送的 ISN 存储起来。攻击时根据所存储的 ISN 和 RTT 计算特定时间的目标主机使用的 ISN。

1985 年，Bob Morris 第一次确定了 TCP 协议潜在的威胁，其中包括序列号猜测[2]。

1989 年，Steve Bellovin 发现“morris”攻击可以用来模仿服务器来攻击客户端，并且提出了加强序列号生成器的方法[3]。

1995 年，CERT 协调中心发布了 CA-1995-01，第一次报道了因特网上这种攻击的大规模使用[4]。

1995 年，操作系统逐渐开始增加 TCP 协议栈 ISN 生成的随机性。

1995 年，Laurent Joncheray 进一步描述了黑客劫持一次 TCP 连接的过程。如果确切知道当前序列号，通过在网络路径上部署嗅探器和生成器，可以实现 TCP 连接重定向[5]。

2001 年，Tim Newsham 提出了在已知 ISN 的值和已经发送数据包的大小的情况下，可以实现会话劫持和会话中断；如果不知道 ISN 的值，也可以猜测一个合理的范围，然后发送序列号在这一范围内的包，直到猜中为止[6]。

目前大多数系统都采用 PRNG(Pseudo Random Number Generators)来生成随机 ISN。Tim Newsham 在《The Problem with Random Increments》一文中证明了目前操作系统 TCP 协议栈的 ISN 生成算法所产生的 ISN 仍然是可猜测的，该算法不足以保护 TCP 连接[6]。本文在分析 Michal 算法[7]的基础上，结合混沌时间序列和线性回归法，提出改进后的 ISN

收稿日期：2005-04-08
基金项目：航空基金(03F51060)、总装备部“十五”预研资助项目(418010703)
作者简介：李肖坚，北京市海淀区学院路 37 号北京航空航天大学计算机学院，副教授，主要研究方向为虚拟现实与可视化、网络安全技术研究；E-mail: xiaojian@vrlab.buaa.edu.cn

猜测算法，该算法在猜测速度和运行精度方面较 Michal 算法有很大提高，提高了猜测算法的实用性。

2 Michal 算法分析

重复、准确地执行指令是计算机的特点，同一算法在相同的条件下执行总会得到同样的结果，因此计算机不可能产生完全的随机数，即便产生完全随机的 ISN 值，将导致 ISN 空间值的互相重复，所以计算机生成的 ISN 只是一个“伪随机数”。目前主要使用的 ISN 猜测算法是 Michal 算法。本小节将对 Michal 算法进行描述和分析，并提出该算法可以进一步提高的地方。

2.1 Micahl 算法描述

Michal 序列号猜测方法[7]主要是采用延迟坐标重构法来猜测目标操作系统 TCP 协议栈 ISN。其基本步骤如下[7]。

第一步，采样目标主机 TCP 协议栈的 ISN 值，记为 seq[t]，其中 t 取 0,1, …….,49 999。

第二步，对所采样的 50 000 个 ISN 序列 seq[t] 按照方程式

$$y[t]=\text{seq}[t-1]-\text{seq}[t-2]$$

来构建空间模型 A。A 呈现的空间模型的 3D 特性越强，它的可分析性就越好。

$$\begin{aligned} x[t]&=\text{seq}[t]-\text{seq}[t-1]\\ y[t]&=\text{seq}[t-1]-\text{seq}[t-2]\\ z[t]&=\text{seq}[t-2]-\text{seq}[t-3] \end{aligned} \quad (1)$$

第三步，在远程主机上探测目标主机 TCP 协议栈的 3 个序列号，记为 seq[t−1],seq[t−2],seq[t−3]，并根据这 3 个值，按照下列公式求空间的一条直线 L：

$$\begin{aligned} y&=\text{seq}[t-1]-\text{seq}[t-2]\\ z&=\text{seq}[t-2]-\text{seq}[t-3] \end{aligned} \quad (2)$$

第四步，选取 L 与 A 交点之间的所有点。如果 L 与 A 没有交点，则选取离 L 最近的 A 的部分点。

第五步，把在 A 附近（半径<=R1 的空间范围内）点集 M 找出来，由这些点的相应的 x 值根据一定的算法推导相应的 seq[t]（推导算法参见文献[7]）。

在实际环境中对 Michal 算法进行了试验，目标主机操作系统是 Windows XP(SP1)。在 R1 取 10，R2 取 251，事先采样 50 000 个目标主机的 ISN。

为了更清楚地了解 Michal 算法可准确猜测的各位数的情况及概率，在下面的试验中，我们只选择选取 L 与 A 的交点，如果 L 与 A 没有交点，我们则选取离 L 最近的 A 的点。目标主机操作系统是 Windows XP(SP1)。试验结果如下表所示。

表 1 Windows 2000 操作系统 ISN 猜测结果统计

操作系统	正确猜测位数	出现次数	出现频率
Windows 2000	3	17	0.17
Windows 2000	4	58	0.58
Windows 2000	5	25	0.25

2.2 存在问题分析

Michal 算法具有较高的“准确猜测位数”。但算法依然存在需要改进的地方。

首先，采样点数过多。Michal 算法中，需要采集 50 000 个目标操作系统 TCP 协议栈的 ISN。在实际的信息对抗应用中，以 1 个/秒的采样速率，采样 50 000 个序列号则需要 50 000 秒，即 13.89 小时。13.89 小时足以错过信息对抗的最佳时机或者让对方有所察觉而无法获得信息优势。其次，Michal 算法只是借鉴混沌计算中的相空间重构方法，可以尝试已经有广泛应用各领域进行预测的混沌时间序列。且混沌时间序列猜测结果对采样点数依赖关系不明显。此外，多种方法组合使用进行猜测也可取得更好的效果。如以等时间间隔连续采样 Windows 2000 操作系统 TCP 协议栈的 80 个 ISN，结果如图 1 所示。从图 1 中可以看出，该操作系统 TCP 协议栈的 ISN 在短时间内基本呈线性增加。因此线性回归法也是一种有效的猜测方法。

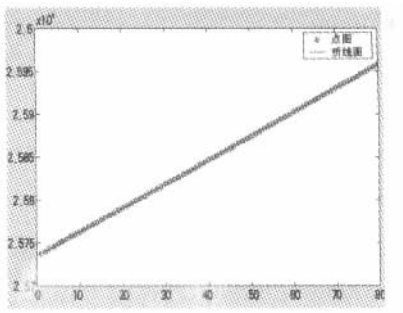

图 1 Windows 2000 操作系统 TCP 协议栈 ISN 采样

3 ISNPLM 算法及其验证

本小节在第 2 小节中对 Michal 算法的分析基础上，提出结合混沌时间序列和线性回归法的 ISN 猜测算法，记为 ISNPLM。并以 Windows XP 和 Windows 2000 为例在实际环境中验证该算法的有效性。

3.1 ISNPLM 算法描述

混沌时间序列预测[8]是 20 世纪 80 年代末发展起来的一种非线性预测新方法，它是混沌理论的一个重要应用领域。在时间序列的分析中，决定序列的客观因素很多，而且相互作用的动力学方程往往是非线性的。目前广泛采用的是延迟坐标状态空间重构法对序列动力学因素的分析。该方法的基本思想是：系统中的任一分量的演化都是由与之相互作用着的其它分量所决定的[9]。相空间重构的具体描述是，给定的混沌时间序列

$$\{x_i,\ i=1,\ 2,\ 3,\ \cdots,\ \text{n}\} \quad (3)$$

设 $m<n$，将时间序列(3)中的 n 个数据延拓成 $n-(m-1)$个 m 维空间的矢量：

$$Y(j)=(x_j,\ x_{j+1},\ x_{j+2},\ \cdots,\ x_{j+m-1}) \quad (4)$$

其中 j=1, 2, …, n−(m−1), m 称为维数。m 维矢量在重构相空间的轨迹将重现混沌运动的某些特性和混沌运动的规律。

混沌时间序列的预测方法是，通过找出预测点的若干邻近相关点，用这些相关点与其后续时间序列的函数关系，近似代替预测点后与其后续时间序列的函数关系，以估计轨迹下一点的走向，最后从预测出的轨迹点的坐标中分离出所需的预测值。为此，可采用局域线性近似方法[10]。

ISNPLM 算法主要思想是，先对采样值采用局域线性近似方法进行一次线性回归，再对回归值和样本值的差值进行混沌时间序列猜测，最后求出猜测值。改进后的算法对采样点数没有具体要求，具体描述如下。

1）采集目标主机 TCP 协议栈的 ISN，记为 $\{\text{isn}_1, \text{isn}_2, \cdots, \text{isn}_n\}$，进行一次线性回归。对应回归值记为 $\{\text{isn}'_1, \text{isn}'_2, \cdots, \text{isn}'_n\}$，并得出下一个 ISN 的回归记为 isn'_{n+1}。

2）计算采样值 $\{\text{isn}_1, \text{isn}_2, \cdots, \text{isn}_n\}$ 和回归值 $\{\text{isn}'_1, \text{isn}'_2, \cdots, \text{isn}'_n\}$ 之间的对应相差值 $\{\text{isn}_1-\text{isn}_1', \text{isn}_2-\text{isn}_2', \cdots, \text{isn}_2-\text{isn}_2'\}$，记为 $\{x_1, x_2, \cdots, x_n\}$。

3）根据$\{x_1, x_2, \cdots, x_n\}$计算 m 维空间的矢量，记为$\{Y(t_1), Y(t_2), \cdots, Y(t_{n-(m-1)})\}$。

4）根据 $L_{nb}=\min\{\|Y(t_i)-Y(t_j)\|, j\neq i\}$

其中 $\|g\|$ 为欧氏模。$Y(t_{n-(m-1)})$通过计算各点和 $Y(t_{n-(m-1)})$之间的欧氏距离，找出和 $Y(t_{nb-(m-1)})$欧氏距离最小的点 $Y(t_{n-(m-1)})$，即为 $Y(t_{n-(m-1)})$的近似矢量。

5）上述两个近似矢量之间的关系可以表示为以下线性方程组：

$$Y(t_{n-(m-1)})=A+BY(t_{nb-(m-1)}) \quad (5)$$

可以用线性回归方法计算回归系数 A，B。

6）利用关系式 $x_{n+1}=A+Bx_{nb+1}$ 猜测 x_{n+1}。

得出下一个 ISN 的值为：$x_{n+1}+\text{isn}'_{n+1}$。

3.2 实验验证及与 Michal 算法效果对比

本文通过两组实验来验证改进后的 ISN 猜测算法的有效性及猜测精度。实验环境包括两台主机，采样主机 A 和目标主机 B。采样主机的操作系统是 Windows XP，目标主机的操作系统为 Widows XP (SP1)。

在相同的实验环境下，ISNPLM 和 Michal 算法都使用相同的采样点来猜测行改进后的 ISN 猜测算法。运行数据如表 2 所示，图形界面如图 2 至图 7 所示。图中，横坐标表示 ISN 值，纵坐标表示 ISN 值的编号，虚线表示样本值，实线表示猜测值。

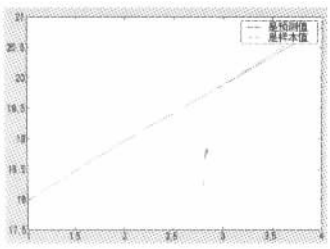

图 2 第一组数据 ISNPLM 猜测结果

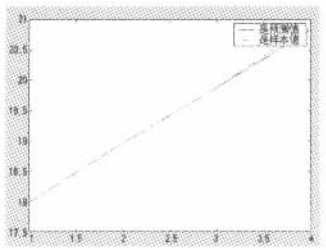

图 3 第一组数据 Michal 猜测结果

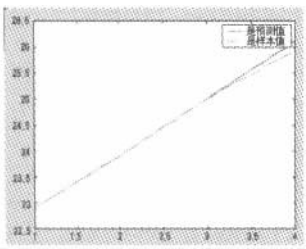

图 4 第二组数据 ISNPLM 算法猜测结果

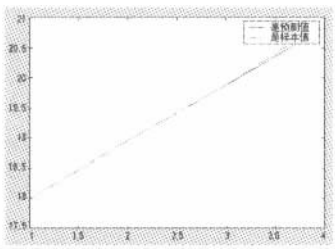

图 5 第二组数据 Michal 算法猜测结果

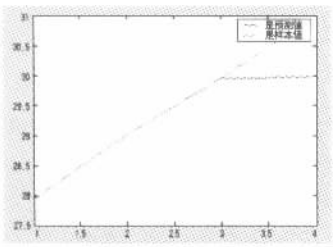

图 6 第三组数据 ISNPLM 算法猜测结果

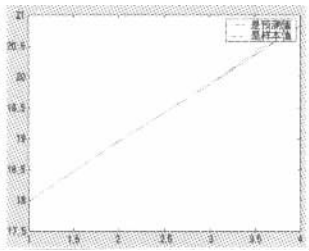

图 7 第三组数据 Michal 算法猜测结果

表 2 Windows XP 操作系统 ISN 猜测结果

采样个数	样本值	ISNPLM 算法猜测值	Michal 算法猜测值
20	2.579338111000000e+009	2.579310763334445e+009	2.579227011843828e+009
25	2.580724063000000e+009	2.580773629674447e+009	2.58919688506306e+009
30	2.582118249000000e+009	2.580773629674447e+009	2.580816589707170e+009

重复实验 100 次，实验结果统计如表 3 所示。

表 3 Windows 操作系统 ISN 猜测结果统计

操作系统	正确猜测位数	出现次数	出现频率
Windows XP	3	5	0.05
Windows XP	4	11	0.11
Windows XP	5	72	0.72
Windows XP	6	12	0.12

从表 2Windows XP 操作系统 ISN 猜测结果表 3 Windows 操作系统 ISN 猜测结果统计可以看出，针对 Windows 2000 操作系统采用混沌时间序列分析方法，能以高 5 位以上的概率猜测到 ISN。

改进后的 ISN 算法在猜测速度和猜测精度方面较 Michal 算法具有较强优势。

1）从实验结果来看，改进后的算法比 Michal 算法“准确猜测位数”增加了 1 位。从 Michal 算法的试验结果来看，“准确猜测位数”基本维持在 4 位，而在改进后的试验结果中，“准确猜测位数”可以基本维持在 5 位，在 5 位及以上的概率在 0.8 以上，最高可达 6 位。

2）从实验结果和采样点数目的依赖关系来看，Michal 算法所需的样本点的个数一般都是 50 000 左右。而改进算法对采样点数依赖不强。从改进算法的试验结果来看，采样点的数目多少对试验结果无明显影响，基于 10 个左右的样本值的猜测和基于 1 000 个样本值的猜测“准确猜测位数”基本上都维持在 5 位左右。

4 结论

本文给出的 ISNPLM，使得“准确猜测位数”比 Michal 算法更逼近了 1 位，且有效地缩小了采样点数目，使初始序列号猜测技术更加可行。该算法为进一步在信息战争中获取信息优势奠定了基础。同时，该算法的实验结果也表明，目前操作系统 TCP 协议栈的 ISN 生成算法仍然存在安全风险。

参考文献(References)

[1] RFC 793；http://www.faqs.org/rfcs/rfc793.html [2]http://www.securiteam.com/securitynews/5PP081P4AC.html

[3] Bellovin. Security Problems in the TCP/IP Protocol Suite [J]. *ACM SIGCOMM Computer Communication Review*，1989，19(3): 10~19

[4] CERT Advisory CA-1995-01

[5] http://www.wcmh.com/uworld/archives/95/security/001.txt.html

[6] http://www.guardent.com/comp_news_tcp.html

[7] http://razor.bindview.com/publish/papers/tcpseq.html

[8] Farmer J D，Sidorowich J J.Predicting chaotic time series [J]. *Physica Review Letter*, 1987, 59(8): 845~848

[9] Packard N H，Crutchfied J P，Famer J Detal.Geometry from a time series[J]. *Phys Rev*. 1980，45(9)：712~716

[10] J.P.Eckmann and D.Ruelle. Ergodic theory of chaos and strange attractors[J]. *Review of Modem Physics*，1985,57(3):617~656

(责任编辑 胡春华)