·通信与信息工程·

基于游程统计的自同步扰码多项式阶数估计

黄芝平，周 靖，苏绍璟，刘纯武，吕喜在

(国防科学技术大学机电工程与自动化学院 长沙 410073)

【摘要】为了获取数字通信中未知线路的自同步加扰信息，提出了一种自同步扰码多项式的阶数估计方法。通过对自同步加扰序列的游程特性进行深入研究，发现了自同步加扰序列的游程统计结果与自同步扰码多项式阶数的对应关系。利用这一关系，通过对接收到的自同步加扰序列进行不同长度的游程统计，根据统计结果的极值变化可以估计出未知线路自同步扰码多项式的阶数。实验结果验证了理论分析的正确性和自同步扰码多项式阶数估计方法的有效性。

关 键 词 数字通信; 阶数估计; 游程统计; 自同步扰码多项式

中图分类号 TN911 文献标志码 A doi:10.3969/j.issn.1001-0548.2013.04.002

Order Estimation of Self-Synchronizing Scrambling Polynomial Based on Run Statistic

HUANG Zhi-ping, ZHOU Jing, SU Shao-jing, LIU Chun-wu, and LÜ Xi-zai

(College of Mechatronics Engineering and Automation, National University of Defense Technology Changsha 410073)

Abstract In order to get the self-synchronizing scrambling information of the unknown lines in digital communication, a method of order estimation of self-synchronizing scrambling polynomial is presented. Through deeply studying the run property of self-synchronizing scrambled sequences, the corresponding relationship between the result of run statistic and the order of the self-synchronizing scrambling polynomial is gained. With this relationship, after calculating the run statistic with different length using the received scrambled sequence, the self-synchronizing scrambling polynomial order of the unknown line will be estimated through the distribution of the run extremums. The results of experiment verify the correctness of the theory analysis and the validity of the whole method.

Key words digital communication; order estimation; run statistic; self-synchronizing scrambling polynomial

在数字通信领域中，为了改善信号的传输特性以及对信息进行加密保护，在信号的发送端会对信源进行加扰处理，在接收端再进行相应的解扰处理[1-2]。扰码技术通常包括帧同步扰码和自同步扰码。近年来，针对未知线路的帧同步扰码多项式及其初态的盲恢复研究较为广泛而深入[3-4]，自同步扰码多项式的盲检测却鲜有文献涉及，这和由自同步扰码技术本身的特点决定的抗攻击性是密不可分的[5]。国外关于自同步扰码的文献主要集中于自同步流密码的设计与实现[6-7]，而在盲识别方面尚未见到公开的资料。文献[8]和文献[9]分别从不同的角度研究了自同步加扰序列所具有的性质，具有一定的参考价值。

在实际的数字通信系统中，信源普遍具有0、1不平衡性[10-11]，即信息序列中0、1 bit的频次并不是各占1/2。基于这一前提，文献[12]提出了利用解方程组还原自同步扰码多项式的方法，并在工程实现时采用快速Walsh变换来减小运算量。但该方法需要预先知道通信线路自同步扰码多项式的阶数，这对于绝大部分未知线路来讲是比较困难的。

游程通常只是被用来描述伪随机序列的特性，很少用于密码分析或信道编码盲识别领域。文献[13]首次将游程统计的方法用于识别同步流密码的攻击，文献[14]提出了利用信道序列的游程特点粗略估计自扰多项式阶数的范围，但也仅是工程方法的介绍，并没有给出理论依据。本文从理论上推导自同

收稿日期：2011－11－03；修回日期：2012－05－21
基金项目：部级项目
作者简介：黄芝平(1965－)，男，教授，博士生导师，主要从事数字化测试、信息侦测方面的研究.

步加扰序列的游程统计特性，并提出一种可以精确检测自同步扰码多项式阶数的游程统计方法。

1 自同步扰码问题描述

自同步加扰与解扰的实现均以线性移位寄存器为基础，自同步加扰器由线性反馈移位寄存器构成，而自同步解扰器由线性前馈移位寄存器实现，如图1所示。

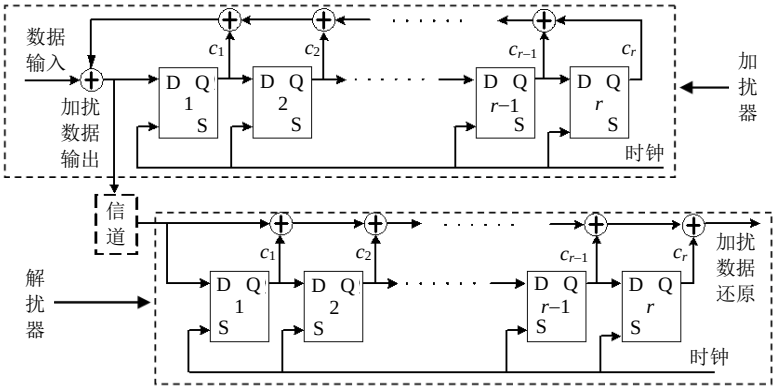

图1 自同步加/解扰器结构图

将式(1)带入式(2)，即可还原出加扰前的原始序列为：

$$\hat{d}_j=[d_j\oplus c_1z_{j-1}\oplus c_2z_{j-2}\oplus\cdots\oplus c_rz_{j-r}]\oplus c_1z_{j-1}\oplus c_2z_{j-2}\oplus\cdots\oplus c_rz_{j-r}=d_j \tag{3}$$

加扰器中的反馈移位寄存器在二元域里可以表示为：

$$f(x)=1+c_1x+c_2x^2+\cdots+c_rx^r \tag{4}$$

式中，$c_i(i=1,2,\cdots,r)$ 与图1所示的反馈移位寄存器抽头系数相对应，取值为0或1。为了取得对信源数据良好的扰乱或加密效果，通常 $f(x)$ 为 r 阶本原多项式。于是有 $c_r=1$，此外，因为本原多项式一定是奇数项，所以 $C=\sum c_i$ 为偶数。

若输入信息序列中各比特间彼此独立，则经过自同步扰码后的输出比特序列彼此独立，且0、1 bit出现的概率近似相等，这也是自同步扰码器的基本特性之一。

2 自同步加扰序列的游程统计特性

定义 1 在二元序列中形如“$100\cdots001$”和“$011\cdots110$”的子序列称为序列的0游程和1游程，0游程中“0”的个数或1游程中“1”的个数称为游程的长度。

定理 1 设独立同分布信源中“1”的概率为 p_1，则根据式(1)加扰后的信道序列中长度为 i 的1游程(记为 $x_i=(z_0,z_1,\cdots,z_{i+1})$)出现的概率满足：

$$P(x_i)=\begin{cases}(1/2)^{i+2} & i\leqslant r-2\\ p_1(1/2)^r & i=r-1\\ (1/2)^r(1-p_1)^2p_1^{i-r} & i\geqslant r\end{cases} \tag{5}$$

证明：根据1游程的定义，x_i 满足 $z_0=z_{i+1}=0$，$z_1=z_2=\cdots=z_i=1$。

1) 当 $i\leqslant r-2$ 时：

$$P(x_i)=P(z_0=0)P(z_{i+1}=0)\prod_{j=1}^{i}P(z_j=1)=(1/2)^{i+2}$$

2) 当 $i=r-1$ 时：

由式(1)可得 $z_{i+1}=d_{i+1}\oplus c_1z_i\oplus c_2z_{i-1}\oplus\cdots\oplus c_rz_{i+1-r}=d_{i+1}\oplus(c_1\cdot1)\oplus(c_2\cdot1)\oplus\cdots\oplus(c_r\cdot0)$。因为 $C=\sum_{i=1}^{r}c_i$ 为偶数，且 $c_r=1$，所以 $C'=\sum_{i=1}^{r-1}c_i$ 为奇数，$(c_1\cdot1)\oplus(c_2\cdot1)\oplus\cdots\oplus(c_r\cdot0)=1$，$z_{i+1}=d_{i+1}\oplus1=\overline{d}_{i+1}$。所以 $P(x_i)=P(z_0=0)\cdot P(z_{i+1}=0)\cdot\prod_{j=1}^{i}P(z_j=1)=(1/2)^{i+1}\cdot P(z_{i+1}=0)=(1/2)^r\cdot P(d_{i+1}=1)=p_1\cdot(1/2)^r$。

3) 当 $i\geqslant r$ 时：

由式(1)可得 $z_j=d_j\oplus(c_1\cdot1)\oplus(c_2\cdot1)\oplus\cdots\oplus(c_r\cdot1)=d_j$，$j>r+1$。$z_{r+1}=d_{r+1}\oplus(c_1\cdot1)\oplus(c_2\cdot1)\oplus\cdots\oplus(c_r\cdot0)=d_{r+1}\oplus1=\overline{d}_{r+1}$。所以，$P(x_i)=P(z_0=0)\times P(z_{i+1}=0)\prod_{j=1}^{r-1}P(z_j=1)\prod_{j=r}^{i}P(z_j=1)=(1/2)P(d_{i+1}=0)\times$

$(1/2)^{r-1}P(d_{r+1}=0)\prod_{j=r+1}^{i}P(d_j=1)=(1/2)^r(1-p_1)^2p_1^{i-r}$。

定理 1 得证。

定理 2　设独立同分布信源中“0”的概率为 p_0，则根据式(2)加扰后的信道序列中长度为 i 的0游程(记为 $y_i=(z_0,z_1,\cdots,z_{i+1})$)出现的概率满足：

$$P(y_i)=\begin{cases}(1/2)^{i+2} & i\leqslant r-2\\ p_0(1/2)^r & i=r-1\\ (1/2)^r(1-p_0)^2p_0^{i-r} & i\geqslant r\end{cases}\tag{6}$$

定理2的证明与定理1完全类似。

3　基于游程统计的自同步扰码多项式阶数估计方法

长为 l ($l\gg r$)的自同步加扰序列中包含 $l-i+1$ 个长为 i 的子序列。设该序列中长为 i 的1游程个数和0游程个数分别为 $K(x_i)$ 和 $K(y_i)$。则有：

$$K(x_i)=(l-i+1)P(x_i)\tag{7}$$

$$K(y_i)=(l-i+1)P(y_i)\tag{8}$$

由于 $l\gg r$，进而有：

$$\eta_1(i)=\frac{K(x_{i+1})}{K(x_i)}=\frac{(l-i)P(x_{i+1})}{(l-i+1)P(x_i)}\approx\frac{P(x_{i+1})}{P(x_i)}\tag{9}$$

$$\eta_0(i)=\frac{K(y_{i+1})}{K(y_i)}=\frac{(l-i)P(y_{i+1})}{(l-i+1)P(y_i)}\approx\frac{P(y_{i+1})}{P(y_i)}\tag{10}$$

将式(5)带入式(9)，式(6)带入式(10)分别得：

$$\eta_1(i)\approx\begin{cases}1/2 & i\leqslant r-3\\ p_1 & i=r-2\\ (1-p_1)^2/p_1 & i=r-1\\ p_1 & i\geqslant r\end{cases}\tag{11}$$

$$\eta_0(i)\approx\begin{cases}1/2 & i\leqslant r-3\\ p_0 & i=r-2\\ (1-p_0)^2/p_0 & i=r-1\\ p_0 & i\geqslant r\end{cases}\tag{12}$$

因为 $p_1+p_0=1$，令 $\eta(i)=\eta_1(i)-\eta_0(i)$，则有：

$$\eta(i)\approx\begin{cases}0 & i\leqslant r-3\\ p_1-p_0 & i=r-2\\ (p_0^3-p_1^3)/p_1p_0 & i=r-1\\ p_1-p_0 & i\geqslant r\end{cases}\tag{13}$$

因为 $\eta(r-1)\approx(p_0^3-p_1^3)/p_1p_0=-(p_1-p_0)\times(p_1^2+p_1p_0+p_0^2)/p_1p_0$，所以，当 $p_1-p_0>0$ 时，$\eta(i)$ 将在 $i=r-1$ 处取得极小值，在 $i=r-2$ 和 $i\geqslant r$ 等处取得极大值；当 $p_1-p_0<0$ 时，$\eta(i)$ 将在 $i=r-1$ 处取得极大值，在 $i=r-2$ 和 $i\geqslant r$ 等处取得极小值；当 $p_1-p_0=0$，即信源平衡时，$\eta(i)$ 值均约等于0。表1给出不同 p_1 情况下 $\eta(i)$ 的取值。

从表1可以看出，$|p_1-0.5|$ 越大，$\eta(i)$ 的极大值与极小值之差也越大。也就是说，信源不平衡性越强，自同步加扰后序列的游程统计所体现的极值也越明显。而极值出现的位置与自扰码多项式的阶数密切相关，这一点对自扰多项式的盲识别具有非常重要的理论指导意义。

表 1　不同 p_1 情况下 $\eta(i)$ 的取值

$\eta(i)$	$i\leqslant r-3$	$i=r-2$	$i=r-1$	$i\geqslant r$
$p_1=0.3$	0	–0.4	1.505	–0.4
$p_1=0.4$	0	–0.2	0.633	–0.2
$p_1=0.5$	0	0	0	0
$p_1=0.6$	0	0.2	–0.633	0.2
$p_1=0.7$	0	0.4	–1.505	0.4

为了利用游程统计法测出自扰多项式的阶数 r，游程长度 i 至少要取 $r+1$，同时要求 $K(x_i)\geqslant1$ 且 $K(y_i)\geqslant1$。根据式(7)和式(8)，加扰序列截取的长度 l 应满足：

$$\begin{cases}(l-r-2)\left(\frac{1}{2}\right)^r(1-p_0)^2p_0\geqslant1\\ (l-r-2)\left(\frac{1}{2}\right)^r(1-p_1)^2p_1\geqslant1\end{cases}\tag{14}$$

考虑到通常情况下 $0.3\leqslant p_0\leqslant0.7$，$0.3\leqslant p_1\leqslant0.7$，式(14)可近似为：

$$(l-r-2)\geqslant15.87\times2^r\tag{15}$$

再结合 $l\gg r$，式(14)可近似为：

$$l>2^{r+4}\tag{16}$$

统计长度为 r 的游程 $K(x_i)$ 和 $K(y_i)$，需要进行的比较次数 $N_{K(x_i)}$ 和 $N_{K(y_i)}$ 满足：

$$N_{K(x_i)}=N_{K(y_i)}=(l-r-2)(r+2)\tag{17}$$

而整个算法还需要计算多个 $K(x_i)$ 和 $K(y_i)$，i 在 r 前后取值。因此基于游程统计的自扰多项式阶数估计算法的计算复杂度约为 $o(2\times(r+2)\times l)\geqslant o((r+2)\times2^{r+5})$。

4　实验验证

实验1　随机生成不同 p_1 情况下的信源序列，长度均为40 000 bit，加扰多项式为 $f(x)=1+x^3+x^7$。对加扰序列进行1游程和0游程统计，结果如表2、表3所示。

从实验 1 结果可以看到，$|\hat{\eta}(i)|$ 最大的位置均出现在 $i=6$ 处，准确反映了真实的自扰多项式 $f(x)=$

$1+x^3+x^7$ 的阶数信息，即 $r=i+1=7$，识别成功。

实验2 随机生成不同 p_1 情况下的信源序列，长度均为160 000 bit，加扰多项式为 $f(x)=1+x^4+x^9$。对加扰序列进行1游程和0游程统计，结果如表4、表5所示。

从实验2结果可以看到，$|\hat{\eta}(i)|$ 最大的位置均出现在 $i=8$ 处，准确地反映了真实的自扰多项式 $f(x)=1+x^4+x^9$ 的阶数信息，即 $r=i+1=9$，识别成功。

表2 自同步加扰序列的游程统计结果

(p_1=0.6，$f(x)=1+x^3+x^7$)

i	3	4	5
$\hat{K}(x_i)$	1 259	626	319
$\hat{\eta}_1(i)$	0.497	0.510	0.298
$\hat{K}(y_i)$	1 248	611	327
$\hat{\eta}_0(i)$	0.490	0.535	0.615
$\hat{\eta}(i)$	0.007	–0.025	–0.317
i	6	7	8
$\hat{K}(x_i)$	95	136	59
$\hat{\eta}_1(i)$	1.432	0.434	—
$\hat{K}(y_i)$	201	27	22
$\hat{\eta}_0(i)$	0.134	0.815	—
$\hat{\eta}(i)$	1.298	–0.381	—

表3 自同步加扰序列的游程统计结果

(p_1=0.6，$f(x)=1+x^3+x^7$)

i	3	4	5
$\hat{K}(x_i)$	1 229	629	343
$\hat{\eta}_1(i)$	0.512	0.545	0.548
$\hat{K}(y_i)$	1 266	637	280
$\hat{\eta}_0(i)$	0.503	0.440	0.443
$\hat{\eta}(i)$	0.009	0.105	0.105
i	6	7	8
$\hat{K}(x_i)$	188	50	27
$\hat{\eta}_1(i)$	0.266	0.540	—
$\hat{K}(y_i)$	124	116	48
$\hat{\eta}_0(i)$	0.935	0.414	—
$\hat{\eta}(i)$	–0.669	0.126	—

表4 自同步加扰序列的游程统计结果

(p_1=0.3，$f(x)=1+x^4+x^9$)

i	5	6	7
$\hat{K}(x_i)$	1 325	614	303
$\hat{\eta}_1(i)$	0.463	0.494	0.333
$\hat{K}(y_i)$	1326	632	306
$\hat{\eta}_0(i)$	0.477	0.484	0.696
$\hat{\eta}(i)$	–0.013	0.009	–0.363
i	8	9	10
$\hat{K}(x_i)$	101	141	46
$\hat{\eta}_1(i)$	1.396	0.326	—
$\hat{K}(y_i)$	213	18	16
$\hat{\eta}_0(i)$	0.085	0.889	—
$\hat{\eta}(i)$	1.311	–0.563	—

表5 自同步加扰序列的游程统计结果

(p_1=0.6，$f(x)=1+x^4+x^9$)

i	5	6	7
$\hat{K}(x_i)$	1 348	610	300
$\hat{\eta}_1(i)$	0.453	0.492	0.647
$\hat{K}(y_i)$	1277	657	320
$\hat{\eta}_0(i)$	0.512	0.487	0.406
$\hat{\eta}(i)$	–0.062	0.005	0.240
i	8	9	10
$\hat{K}(x_i)$	194	47	29
$\hat{\eta}_1(i)$	0.242	0.552	—
$\hat{K}(y_i)$	130	103	50
$\hat{\eta}_0(i)$	0.792	0.485	—
$\hat{\eta}(i)$	–0.550	0.132	—

在上述实验中，游程统计结果均准确地反映了线路的真实自扰多项式的阶数信息。但由于式(9)～式(13)在推导过程中采用了近似的方法，从而使表2～表5的结果与表1中的理想数值之间存在一定的偏差。此外，一方面，由式(14)～式(16)可知，参与游程统计的加扰序列要足够长(与自扰多项式阶数有关)；另一方面，信源不平衡性的明显程度将直接决定游程统计的理论极值的大小。而在实际的工程应用中，由于线路的自扰多项式阶数信息可能完全未知，再加上加扰前序列的信源不平衡性也是未知的，所以自扰多项式的阶数估计并不容易。如果数据不够长或者信源不平衡性不够明显，游程统计结果将出现很大偏差，甚至导致估计方法失效。

因此，在实际应用中，应充分利用有关未知线路业务特点等先验知识，对自扰多项式的阶数范围进行大致的预测，进而决定参与统计的数据长度。为了克服信源不平衡性可能不明显的问题，可以在不同时间段内分别截取线路的加扰数据参与统计，最终选取游程极值分布较为理想的结果来估计自扰多项式阶数。

5 结 束 语

本文对自同步扰码序列的游程统计特性进行了理论推导，研究发现信源不平衡条件下游程统计结果能够反映自同步扰码多项式的阶数信息。在此基础上，提出了基于游程统计的自同步加扰多项式的阶数估计方法，并对方法的有效性进行了实验验证，从而为有效解决自同步加扰多项式的全盲检测奠定了基础。

参 考 文 献

[1] ZEPERNICK H J, FILGER A. 伪随机信号处理 :理论与应

用[M]. 甘良才, 译. 北京: 电子工业出版社, 2007.
ZEPERNICK H J, FILGER A. Pseudo random signal processing theory and application[M]. Translated by GAN Liang-cai. Beijng: Publising House of Electronics Industry, 2007.

[2] 伍文君, 黄芝平, 唐贵林, 等. 含错扰码序列的快速恢复[J]. 兵工学报, 2009, 30(8): 1134-1138.
WU Wen-jun, HUANG Zhi-ping, TANG Gui-lin, et al. Fast recovery of interfered scrambling code sequence[J]. Acta Armamentarii, 2009, 30(8): 1134-1138.

[3] 伍文君, 唐贵林, 黄芝平. 一种快速相关攻击算法[J]. 计算机工程, 2009, 35(17): 129-131.
WU Wen-jun, TANG Gui-lin, HUANG Zhi-ping. Fast correlation attack algorithm[J]. Computer Engineering, 2009, 35(17): 129-131.

[4] 臧玉亮, 韩文报, 何开成. 一类同步流密码的差分能量攻击[J]. 信息工程大学学报, 2009, 10(3): 325-328.
ZANG Yu-liang, HAN Wen-bao, HE Kai-cheng. Differential power attack on a kind of synchronized stream ciphers[J]. Journal of Information Engineering University, 2009, 10(3): 325-328.

[5] ZHANG Hai-na, LI Lin, YUN Xiao-wang. Fast correlation attack on stream cipher ABC v3[J]. Science in China Series F: Information Science, 2008, 51(7): 935- 947.

[6] SAVAGE J E. Some simple self-synchronizing digital data scramblers[J]. Bell Syst Tech J, 1967, 42(2): 449-487.

[7] ARAZI B. Self synchronizing digital scramblers[J]. IEEE Trans Commun, 1977, 25: 1505-1507.

[8] FAIR I J, BHARGAVA V K, WANG Q. On the power spectral density of self-synchronizing scrambled sequences [J]. IEEE Transactions on Information Theory, 1998, 44(4): 1687-1692.

[9] HEYS H M. An analysis of the statistical self-synchronization of stream ciphers[C]//Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Canada: IEEE, 2001: 897-904.

[10] 丁存生, 肖国镇. 流密码学及其应用[M]. 北京: 国防工业出版社, 1994.
DING Cun-sheng, XIAO Guo-zheng. Compositions for stream cipher and its application[M]. Beijng: National Defense Industry Press, 1994.

[11]游凌, 朱中梁. Walsh函数在解二元域方程组上的应用[J]. 信号处理, 2000, 16(12): 27-30.
YOU Ling, ZHU Zhong-liang. The application of walsh function in resolving of $F(2)$ equations[J]. Signal Processing, 2000, 16(12): 27-30.

[12] 杨忠立, 刘玉君. 自同步扰乱序列的综合算法研究[J].信息技术, 2005(2): 30-32
YANG Zhong-li, LIU Yu-jun. Algorithm research of self-synchronizing scrambler sequence[J]. Information Technology, 2005(2): 30-32.

[13] 朱华安, 谢端强. 基于m序列统计特性的序列密码攻击[J]. 通信技术, 2003(8): 96-98.
ZHU Hua-an, XIE Duan-qiang. Attacks upon stream cipher based on m-sequence's statistical property[J]. Communications Technology, 2003(8): 96-98.

[14] 朱洪斌. 对伪随机扰码和自同步扰码的盲识别[J]. 科技风, 2010(14): 220-221.
ZHU Hong-bin. Blind recognition of pseudo-random scrambling and self-synchronizing scrambling[J]. Technological Wind, 2010(14): 220-221.

编 辑 漆 蓉