doi:10.3969/j.issn.1000-436x.2013.01.016

自同步扰码的盲识别方法

廖红舒，袁叶，甘露
（电子科技大学 电子工程学院，四川 成都 611731）

摘 要：在非合作通信领域中，侦察方在信道解码后需要对扰码编码参数进行盲估计以实现解扰，进而恢复出原始信息。在信源不平衡的条件下，针对自同步扰码的盲识别问题，以比特状态统计概率分布与均匀分布之间的修正平方欧几里德距离作为比特状态不平衡性的衡量准则，提出了一种自同步扰码生成多项式的盲识别方法，仿真结果验证了理论分析的正确性和所提算法的实际有效性。
关键词：自同步扰码；盲识别；比特状态不平衡；修正平方欧几里德距离
中图分类号：TN929.11 文献标识码：B 文章编号：1000-436X(2013)01-0136-08

Novel blind recognition method for self-synchronized scrambler

LIAO Hong-shu, YUAN Ye, GAN Lu
(School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: In non-cooperative communication systems, an adversary has to blindly estimate the coding parameters of scrambler from intercepted sequence. On the condition biased information source, a new method was proposed which can blindly recover the generating polynomial of self-synchronized scrambler. This method was based on the corrected squared Euclidean distance which was used to measure the distance between the actual stat ical probability and the uniform distribution. The simulation results verify the theory analysis and the validity of the algorithm.
Key words: self-synchronized scrambler; blind recognition; biased characteristic of the state of bits; corrected squared Euclidean distance

1 引言

在实际数字通信过程中，系统往往受到待传送信息序列统计特性的影响。由于传输的信息不一定是随机的，可能出现连续的“0”或连续的“1”，这不仅破坏了系统设计的前提，对系统性能造成不利影响，还导致接收端无法正确获得定时信息。因此需要对信源编码后的数据进行随机化处理以改善其传输特性，这种处理即为扰码。扰码不仅能提高信息比特的定时含量，还使得信号频谱弥散而保持稳恒，进而改善帧同步和自适应时域均衡等子系统的性能。对非合作通信领域而言，正确有效地分析识别出扰码的编码参数如生成多项式和移位寄存器初态并完成解扰，对后续信息分析处理有着重要的理论意义和实用价值，解扰效果的好坏直接影响到信源解码的性能，进而影响信息的正确获取。

扰码分为同步扰码和自同步扰码，现阶段对扰码盲识别的研究主要是生成多项式的确定和同步扰码初态的重构。近年来，对于同步扰码生成多项式及其初态的盲识别研究相对较为广泛[1~6]，而自同步扰码系统由于需要将扰码序列作为线性移位寄存器（LFSR）的反馈输入，其生成多项式的盲识别分析更为复杂，鲜有文献涉及。在实际数字通信系统中，由于语言统计特性和采用的信息编码方案等原因，信源输出序列普遍具有 0、1 不平衡性[1]，即信息序列中比特 0 和比特 1 出现的概率并不是各占 1/2，而加扰对信息序列进行随机化处理，使得加扰之后的序列中的比特 0 与比特 1 出现的概率各

收稿日期：2012-01-11；修回日期：2012-08-07
基金项目：国家自然科学基金资助项目（11176005）
Foundation Item: The National Natural Science Foundation of China (1117 5)

为 1/2，这正是目前公开可见的所有扰码盲识别方法可行的前提条件。

具体来看，自同步扰码盲识别方法可分作 2 类，一是统计的方法，即利用加扰前后数据序列的内在统计特性进行分析识别。文献[6]选取可能的多项式作为测试多项式对扰码序列解扰，用解扰序列构造一个随机变量，Cluzeau M 指出该随机变量在测试多项式与扰码序列的生成多项式相等和不相等这 2 种情况下，分别服从 2 个不同的高斯分布，并由此给出了相应的判决门限，进而根据该随机变量与判决门限的相对大小对测试多项式进行判定，完成生成多项式的识别，算法的门限设置复杂且需要预先知道信源的准确 0、1 比例，这在盲识别中是很困难的。文献[7]则将流密码的分别征服攻击算法引入扰码的盲识别中，用可能的多项式对扰码序列解扰再加扰，根据新产生的密文序列与原始扰码序列的相关度来判定自同步扰码的生成多项式，算法对扰码生成多项式阶数和项数的限制很大，并且计算复杂门限设置困难。而文献[8]侧重于扰码级数的估计，利用对自同步扰码重码统计特性的分析实现扰码级数的盲估计，预先估计扰码级数在很大程度上减少了抽头位置确定部分的搜索量，然而利用重码特性分析扰码级数这一部分引入的计算量和搜索量依然极为庞大。第二类方法是代数方法，即将扰码识别问题归结为有限域上某个低次多元代数方程组的求解问题，文献[9]就是基于 Walsh-Hadamard 变换解含错方程组来还原生成多项式，但是该方法的使用需要已知扰码级数，且计算量随着扰码级数的增大呈指数增长。

本文首先介绍了自同步扰码的相关原理和识别模型，然后在信源不平衡的条件下，对符合抽头位置的比特状态不平衡性进行深入的理论分析，以比特状态统计概率分布与均匀分布之间的修正平方欧几里德距离作为不平衡性的衡量准则，提出了一种自同步扰码生成多项式的盲识别算法，算法不需要预先知道扰码级数和信源不平衡度即可有效地完成生成多项式的盲识别。最后，对算法的性能进行了计算机仿真验证和对比分析。

2　自同步扰码原理

自同步扰码在 L 级线性移位寄存器 LFSR 的反馈逻辑输出与第一级移位寄存器之间引入异或逻辑，将信息序列与反馈逻辑的输出进行模二求和后作为 LFSR 的反馈输入。如图 1 所示，信道序列$\{y_k\}$的产生不仅与 LFSR 的输出序列$\{s_k\}$有关，还受到输入信息序列$\{x_k\}$的影响，加扰过程可以表示为

$$y_k = x_k \oplus \bigoplus_{i=1}^{L} c_i y_{k-i} \tag{1}$$

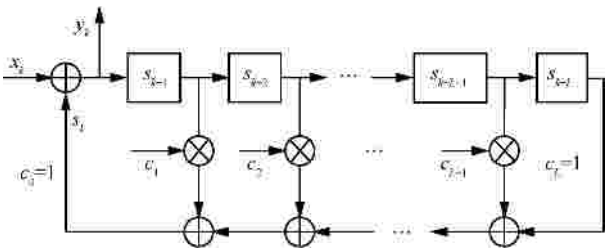

图 1　自同步扰码器

其中，$c_i \in \mathrm{GF}(2)$, $i=0,1,\cdots,L$ 是 LFSR 的反馈系数，GF(2)是{0,1}在模 2 加法和模 2 乘法运算法则下的二元域。L 级线性移位寄存器的反馈多项式可以表示为

$$f(x) = c_0 + c_1 x + c_2 x^2 + \mathrm{L} + c_L x^L,\ c_0, c_L = 1 \tag{2}$$

$f(x)$亦即自同步扰码的生成多项式。由式（1）可得

$$x_k = \bigoplus_{i=0}^{L} c_i y_{k-i} \tag{3}$$

这正是图 2 所示的自同步扰码解扰过程。

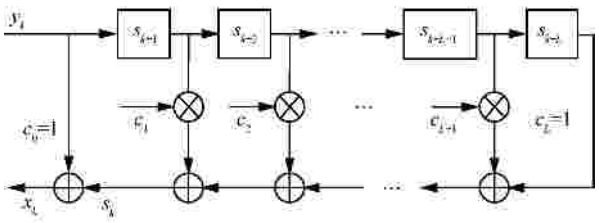

图 2　自同步解扰器

自同步扰码的主要优点是不需要建立同步，在传输开始时或是出现一个错误数据后，扰码器和解扰器可能并不处于相同的状态，然而，由于解扰器不是完整的封闭回路，在无错误的传输 L bit 后，解扰器的 LFSR 可以与加扰器的 LFSR 处于相同的状态，这样同步就建立了，也就是说，加解扰器只要使用相同的 LFSR 反馈多项式，而不需要相同的 LFSR 初态。因此，自同步扰码的盲识别问题，就是根据截获的信道序列$\{y_k\}$识别出线性移位寄存器的反馈多项式 $f(x)$。

值得指出的是，在实际通信中，扰码使用的 LFSR 的级数绝大多数分布在 3~100 之间，且生成

多项式多为稀疏多项式，其项数通常不超过 5 项，同步扰码的生成多项式多为 3 项式或 5 项式，而自同步扰码由于误码扩散率与生成多项式的项数成正比[8]，一般都使用 2 项式或 3 项式，比如 ITU V.34 中使用的就是 3 项式 $1+x^{18}+x^{23}$。因此有限的多项式项数和 LFSR 的级数是目前自同步扰码多项式盲识别最根本的前提。

3 扰码多项式的估计

3.1 估计原理

对于图 1 所示自同步扰码，其生成多项式可以表示为

$$f(x)=c_0+c_1x+c_2x^2+\cdots+c_Lx^L=\sum_{j=1}^{r_v}x^{i_j},\quad 0=i_1<i_2<\cdots<i_{r_v}=L \tag{4}$$

其中，L 是 LFSR 的反馈逻辑输出级数，r_v 是生成多项式 $f(x)$的项数。自同步扰码器输入的信息序列为 $\{x_k\}$，序列中 0 所占的比例为$1/2+e$ ，e 为信源的 0、1 不平衡度，扰码器输出序列为$\{y_k\}$。定义 GF(2)上的 r 项多项式 $g(x)=\sum_{j=1}^{r}x^{i_j},\ 0=i_1<i_2<\cdots<i_r$ ，按照 $g(x)$ 对扰码序列抽取并进行模二求和得

$$z_k=\bigoplus_{j=1}^{r}y_{k-i_j},\ k\geq i_r \tag{5}$$

当 $g(x)=f(x)$时，比较式(5)和式(3)可知，得到 z_k 的过程实际上就是自同步扰码的解扰过程。

$$z_k=\bigoplus_{j=1}^{r}y_{k-i_j}=\bigoplus_{j=1}^{r_v}y_{k-i_j}=x_k \tag{6}$$

自同步扰码生成多项式系数为 1 的项对应于式(5)中 $y_{k-i_1}y_{k-i_2}\cdots y_{k-i_r},0=i_1<i_2<\cdots<i_r$ 比特位置。以生成多项式为 $1+x+x^{15}$ 的自同步扰码为例，扰码器输入 x_k 的值为 0 或 1 时，$y_ky_{k-1}y_{k-15}$ 比特可能的状态如表 1 前 2 列所示。

虽然加扰对 0、1 比例统计不平衡的信息序列 $\{x_k\}$进行了随机化处理，但是从式（6）可以看出，扰码序列$\{y_k\}$在符合扰码生成多项式抽头位置处比特状态的统计概率与扰码器输入序列$\{x_k\}$中的 0 和 1 出现的统计概率有关。针对本例而言，若$\{x_k\}$的 0 和 1 出现的统计概率不相等，$y_ky_{k-1}y_{k-15}$ 比特的 8 种状态出现的统计概率也一定不相等（如表 1 所示），即 $y_ky_{k-1}y_{k-15}$ 比特的状态具有不平衡性。

表 1 $y_ky_{k-1}y_{k-15}$ 比特状态统计

$y_ky_{k-1}y_{k-15}$	x_k	概率
000	0	0.187 5
001	1	0.062 5
010	1	0.062 4
011	0	0.187 5
100	1	0.062 5
101	0	0.187 5
110	0	0.187 6
111	1	0.062 5

由式（6）可知，$\bigoplus_{j=1}^{r}y_{k-i_j},0=i_1<\cdots<i_r=L$ 的取值与 x_k 有关，且

$$P\left\{\bigoplus_{j=1}^{r}y_{k-i_j}=0,\ 0=i_1<\cdots<i_r=L\right\}=P\{x_k=0\}=\frac{1}{2}+e \tag{7}$$

又因为使得 $\bigoplus_{j=1}^{r}y_{k-i_j}=0,\ 0=i_1<\cdots<i_r=L$ 的 2^{r_v-1} 个状态等概率出现，则这 2^{r_v-1} 个状态的概率均为

$$p_0=\frac{\left(\frac{1}{2}+e\right)}{2^{r_v-1}} \tag{8}$$

同理，使$\bigoplus_{j=1}^{r}y_{k-i_j}=1,\ 0=i_1<\cdots<i_r=L$ 的 2^{r_v-1} 个状态的概率为

$$p_1=\frac{\left(\frac{1}{2}-e\right)}{2^{r_v-1}} \tag{9}$$

当 $g(x)?f(x)$时，式（5）所示的处理过程不再是解扰过程，扰码序列$\{y_k\}$中的 0 和 1 等概率出现，因此，在错误抽取时每种比特状态出现的概率相等，即

$$p_w=\left(\frac{1}{2}\right)^{r_w} \tag{10}$$

其中，r_w 表征比特抽取时所参照的多项式集合 Q 非生成多项式的其他多项式的项数，即 $g(x)?f(x)$时，$g(x)$项数表示为 r_w。

综上所述，基于比特流抽头位置的状态不平衡性，对于截获的扰码序列，按照式（5）选定 r 个相对固定的比特位置，抽取大量数据，作 r 比特的状态统计，比特状态统计最不平衡的那组比特位置对应的多项式即为待识别的生成多项式。以生成多项

式为 $1+x+x^{15}$ 的自同步扰码为例，加扰器输入数据流中 0 的比例为 $1/2+e, e=0.25$，取长度为 10^4 bit 的数据，根据可能的多项式 $\sum_{j=1}^{r} x^{i_j}, 0=i_1<i_2<\cdots<i_r$，统计 $y_{k-i_1}y_{k-i_2}\cdots y_{k-i_r}(0=i_1<i_2<\cdots<i_r)$ 比特的各种状态出现的概率。当按照自同步扰码的生成多项式进行比特状态统计时，相应的状态概率分布如表 1 所示，可以看出此时 $y_k y_{k-1} y_{k-15}$ 比特的状态呈现明显的统计不平衡；而当按照非生成多项式的其他多项式进行统计时，比特状态却是统计平衡的（表 2 是按照多项式 $1+x^2+x^3+x^5+x^{15}$ 进行比特状态统计时，$y_k y_{k-2} y_{k-3}\ y_{k-5} y_{k-15}$ 比特的状态统计概率分布情况）。由此反推认为该自同步扰码器的正确生成多项式应该是 $f(x)=1+x+x^{15}$。

表 2　$y_k y_{k-2} y_{k-3}\ y_{k-5} y_{k-15}$ 比特状态统计

比特状态	出现概率	比特状态	出现概率	比特状态	出现概率
00000	0.031 2	01011	0.031 3	10110	0.031 3
00001	0.031 2	01100	0.031 2	10111	0.031 2
00010	0.031 2	01101	0.031 3	11000	0.031 2
00011	0.031 2	01110	0.031 2	11001	0.031 3
00100	0.031 2	01111	0.031 3	11010	0.031 2
00101	0.031 3	10000	0.031 2	11011	0.031 2
00110	0.031 3	10001	0.031 3	11100	0.031 3
00111	0.031 3	10010	0.031 3	11101	0.031 3
01000	0.031 3	10011	0.031 3	11110	0.031 2
01001	0.031 2	10100	0.031 3	11111	0.031 3
01010	0.031 2	10101	0.031 2		

3.2　不平衡性衡量准则

对于比特状态不平衡性的判断，需要引入评价标准。根据 3.1 节的分析，算法选取最不平衡的一组状态位置，这就意味着选取距离均匀分布最远的那一组状态概率分布对应的多项式作为正确的生成多项式。由于每组状态的概率是独立的，而离散概率分布可以看作是一个向量，也就是笛卡尔坐标系或欧几里德空间中的一个点，因此可以用几何距离来度量离散概率分布之间的距离[10]。这里考虑最常用的平方欧几里德距离。

$$d_{\mathrm{sqe}}=\sum_{i=0}^{n-1}\left(P_i-Q_i\right)^2 \tag{11}$$

其中，$P_i(i=0,\cdots,n-1)$和 $Q_i(i=0,\cdots,n-1)$表示 2 个不同的离散概率分布，式(11)要求 $P_i(i=0,\cdots,n-1)$和 $Q_i(i=0,\cdots,n-1)$必须等长，而且具有对等的值，即相同的 i。

对于截获的扰码序列，假设统计得到的全部比特状态数为 M（一旦扰码序列的长度确定，则 M 恒定），一共有 2^r 种比特状态，各种比特状态出现的数目表示为 M_i，i 是二进制比特状态对应的十进制表示，$i\in\{0,1,\cdots,2^r-1\}$。则由式（11）可得到平方欧几里德距离作为比特状态不平衡性的衡量准则。

$$d_{\mathrm{sqe}}=\sum_{i=0}^{2^r-1}\left(\frac{M_i}{M}-\frac{1}{2^r}\right)^2 \tag{12}$$

d_{sqe} 越大，比特状态概率分布与均匀分布的距离越远，也就是说，比特状态统计越不平衡。

理论上来看，当比特抽取参照的多项式正好是扰码生成多项式时，衡量不平衡性的指标值应该只与信源不平衡度 e 有关，即对于确定的信源不平衡度，该指标值恒定；而当抽取参照的多项式取自多项式集合 Q 中非生成多项式的其他多项式时，根据前面的理论分析，此时各比特状态都是统计平衡的，相应的不平衡性指标值应该相等。即当以比特状态概率分布与均匀分布的平方欧几里德距离 d_{sqe} 作为衡量准则时，若抽取参照的多项式取自多项式集合 Q 中非生成多项式的其他多项式，则对应的 d_{sqe} 应该相等且为 0，这显然只在数据长度 $N\to\infty$ 时才成立。由于受到数据长度的限制，实际比特状态统计概率分布与均匀分布之间的距离 d_{sqe} 是一个随机变量，因此，以比特抽取所参照的多项式取自多项式集合 Q 中非生成多项式的其他多项式时的情况为例，考察式（12）作为比特状态不平衡性的衡量准则性能。

对于截获的一段长度为 N 的扰码序列，按照 3.1 节所述的方法对扰码序列进行比特状态统计。以项数为 r_w 的多项式 $g_m(x)$（$g_m(x)\in Q$ 且 $g_m(x)?f(x)$）作为比特抽取的参照多项式，计算统计所得的实际比特状态概率分布与均匀分布之间的修正欧几里德距离 d_{sqew}。根据前面的分析，作为衡量不平衡性的准则，简单直观地看，当数据长度确定时，d_{sqew} 的期望应该是一个定值，并与多项式 $g_m(x),m=1,2,\cdots$ 无关。对 d_{sqew} 求取期望

$$\begin{aligned}\mathrm{E}\left[d_{\mathrm{sqew}}\right]&=\sum_{i=0}^{2^{r_w}-1}\mathrm{E}\left[\frac{M_i}{M}-\frac{1}{2^{r_w}}\right]^2\\&=\sum_{i=0}^{2^{r_w}-1}\left\{\mathrm{E}\left[\left(\frac{M_i}{M}\right)^2\right]-\frac{1}{2^{r_w-1}}\mathrm{E}\left[\frac{M_i}{M}\right]+\frac{1}{2^{2r_w}}\right\}\end{aligned} \tag{13}$$

仍然假定此时统计得到的全部比特状态数为 M（对于给定长度 N，M 恒定），这 M 个状态分别表示为 $X_1, X_2, \cdots, X_M$。$X_j, j=1,2,\mathrm{L},M$ 是一个随机事件，为了统计状态 i 出现的次数，定义一个随机序列 $Y_{i1}, Y_{i2}, \mathrm{L}, Y_{iM}\ (i \in \{0,1,\mathrm{L},2^{r_w}-1\})$。对于 $i, i \in \{0,1,\mathrm{L},2^{r_w}-1\}$，如果状态 X_j 的十进制表示正好等于 i 这一事件出现，把该事件定义成 $Y_{ij}=1$，$\mathrm{P}\{Y_{ij}=1\}=\dfrac{1}{2^{r_w}}$，而状态 X_j 的十进制表示不等于 i 的概率则为 $\mathrm{P}\{Y_{ij}=0\}=\dfrac{(2^{r_w}-1)}{2^{r_w}}$，这样就将随机事件与随机数联系起来，且有

$$M_i=\sum_{j=1}^{M} Y_{ij} \tag{14}$$

由上式可得

$$\mathrm{E}\left[M_i\right]=\mathrm{E}\left[\sum_{j=1}^{M} Y_{ij}\right]=\sum_{j=1}^{M}\mathrm{E}\left[Y_{ij}\right]=\frac{M}{2^{r_w}} \tag{15}$$

$$\begin{aligned}\mathrm{E}\left[M_i^{\,2}\right]&=\mathrm{E}\left[\left(\sum_{j=1}^{M} Y_{ij}\right)^2\right]=\sum_{j=1}^{M}\sum_{n=1}^{M}\mathrm{E}\left[Y_{ij}Y_{in}\right]\\&=\frac{M}{2^{r_w}}+\frac{M^2-M}{2^{2r_w}}\end{aligned} \tag{16}$$

至此，式（13）可化简为

$$\begin{aligned}\mathrm{E}\left[d_{\mathrm{sqew}}\right]&=\sum_{i=0}^{2^{r_w}-1}\left\{\mathrm{E}\left[\left(\frac{M_i}{M}\right)^2\right]-\frac{1}{2^{r_w-1}}\mathrm{E}\left[\frac{M_i}{M}\right]+\frac{1}{2^{2r_w}}\right\}\\&=\sum_{i=0}^{2^{r_w}-1}\left\{\frac{\dfrac{M}{2^{r_w}}+\dfrac{M^2-M}{2^{2r_w}}}{M^2}-\frac{1}{2^{r_w-1}}\frac{1}{2^{r_w}}+\frac{1}{2^{2r_w}}\right\}\\&=\frac{2^{r_w}-1}{2^{r_w}M}\end{aligned} \tag{17}$$

从上述分析可知，对于给定长度的扰码序列，作 r 比特的状态统计，当比特抽取所参照的多项式 $g_m(x)\in Q$ 且 $g_m(x)?f(x)$ 时，$\mathrm{E}[d_{\mathrm{sqe}}]$ 并不是常数而与 $g_m(x), m=1,2,\cdots$ 的项数 r_w 有关，即直接以平方欧几里德距离作为衡量准则并不满足前面分析得出的一致不平衡性的要求，因此需要对平方欧几里德距离进行如下修正。

$$d_{\mathrm{sqe}}=\frac{2^r}{2^r-1}\sum_{i=0}^{2^r-1}\left(\frac{M_i}{M}-\frac{1}{2^r}\right)^2 \tag{18}$$

显而易见，当参照多项式 $g_m(x)\in Q$ 且 $g_m(x)$? $f(x)$ 对给定长度的扰码序列进行抽取时，实际概率分布与均匀分布之间的修正平方欧几里德距离 $\%_{\mathrm{sqew}}$ 的期望为

$$\mathrm{E}\left[\%_{\mathrm{sqew}}\right]=\frac{1}{M} \tag{19}$$

即对于给定长度的扰码序列，$\mathrm{E}\left[\%_{\mathrm{sqew}}\right]$ 是常数，满足一致不平衡性的要求。此外，当截获的扰码序列足够长（即 $N\to\infty$）时，$\%_{\mathrm{sqew}}$ 满足

$$\%_{\mathrm{sqew}}=\frac{2^{r_w}}{2^{r_w}-1}\sum_{i=0}^{2^{r_w}-1}\left[\left(\frac{1}{2}\right)^{r_w}-\frac{1}{2^{r_w}}\right]^2=0 \tag{20}$$

相应地，$\mathrm{E}\left[\%_{\mathrm{sqew}}\right]=0$。

考虑比特抽取所参照的多项式正好是扰码生成多项式即 $g(x)=f(x)$ 的情况，若扰码序列足够长，则相应的修正平方欧几里德距离为

$$\begin{aligned}\%_{\mathrm{sqev}}&=\frac{2^{r_v}}{2^{r_v}-1}\left[\sum_{\substack{i=0\\ \mathrm{ieven}}}^{2^{r_v}-1}\left(\frac{\dfrac{1}{2}+e}{2^{r_v-1}}-\frac{1}{2^{r_v}}\right)^2+\sum_{\substack{i=0\\ \mathrm{iodd}}}^{2^{r_v}-1}\left(\frac{\dfrac{1}{2}-e}{2^{r_v-1}}-\frac{1}{2^{r_v}}\right)^2\right]\\&=\frac{4e^2}{2^{r_v}-1}\end{aligned} \tag{21}$$

要正确地完成自同步扰码生成多项式的识别就要求

$$\%_{\mathrm{sqev}}>\%_{\mathrm{sqew}}\Rightarrow\frac{4e^2}{2^{r_v}-1}>0 \tag{22}$$

综上所述，对于给定长度的扰码序列，按照式（5）选定 r 个相对固定的比特位置，作 r 比特状态统计，得到相应的概率分布，并按照式（18）计算该概率分布与均匀分布之间的修正平方欧几里德距离，与均匀分布之间距离最大的那组比特状态对应的多项式即为待识别的生成多项式。

4 仿真实验及性能比较

4.1 仿真实验

在第 2 节中已经指出，扰码使用的 LFSR 级数绝大多数分布在 3~100 之间，且自同步扰码多使用 2 项式或 3 项式。为给出较为实际有效的正确识别率，仿真通过遍历的方式采用 3 组多项式作为自同步扰码的生成多项式：第一组，由 3~100 阶的全部 98 个 2 项式 $f(x)=1+x^L$（L=3,4,…,100）组成；第二

组，由 3~100 阶的全部 4 949 个 3 项式 $f(x)=1+x^k+x^L$（L=3,4,⋯,100，k=1,2,⋯,L-1）组成；第三组，由 3~100 阶的 5 项式组成，由于 3~100 阶的 5 项式的总数过于庞大，因此在各个阶数对应的 5 项式中分别随机抽取 10 个或 11 个多项式，若相应阶数的 5 项式不足 10 个（比如 3~6 阶）则该阶数对应的 5 项式全部抽取，最终组成一共 1 000 个 5 项式。

在实际通信中，e 通常为 0.05 或 0.1[1]，因此在用每一组多项式中的每一个式子作为自同步扰码生成多项式时，按照信源不平衡度 e 以 0.002 的步长从 0~0.1 变化，分别随机生成长度为 $N=10^4$bit 的序列作为扰码器的输入信息序列，使用本文的算法对相应的扰码序列进行识别；每个多项式进行 500 次蒙特卡洛仿真，得到仿真结果如图 3 所示。

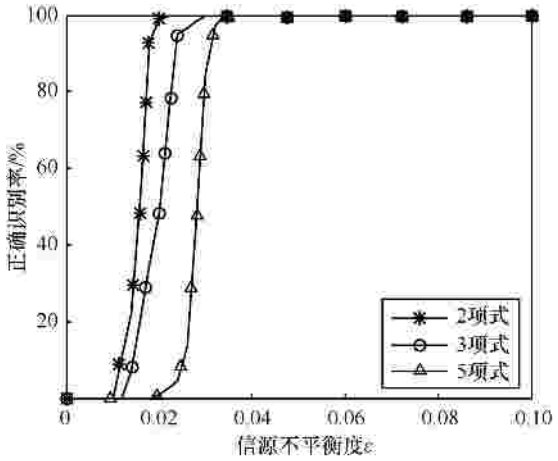

图 3　不同项数的多项式作为扰码生成多项式时 e 对识别率的影响

图 3 的仿真结果表明，在扰码生成多项式项数 r_v 恒定、扰码序列长度 N 一定的情况下，随着 e 的增大，正确识别率增大；由式（21）可知，e 越小，正确抽取时的状态概率分布与均匀分布的距离 $\hat{d}_{\mathrm{sqev}}$ 越小，使得正确抽取时的 $\hat{d}_{\mathrm{sqev}}$ 与错误抽取时的 $\hat{d}_{\mathrm{sqew}}$ 越接近而无法区分，进而导致误判增加。从图 3 还可以看到，自同步扰码所使用的生成多项式项数 r_v 越少，正确识别率随着 e 的增大而增大得越快；因为 r_v 越小，式（21）中 e^2 的系数越大，相应地 $\hat{d}_{\mathrm{sqev}}$ 增长更快，$\hat{d}_{\mathrm{sqev}}$ 与错误抽取时的 $\hat{d}_{\mathrm{sqew}}$ 之间的差异越明显，越容易识别。

上述仿真参数覆盖了 3~100 阶的全部 2 项式、3 项式以及各个阶数下随机抽取的一部分 5 项式，e 从 0~0.1 变化，完全符合实际通信系统参数，这样大量的典型实验使得本文给出的识别概率比较可靠。从图 3 中可以看出，对于 3~100 阶的 2 项式和 3 项式，当信源不平衡度为 0.025 时，本文算法的正确识别率已经达到 95%以上；即使是对于 3~100 阶的 5 项式，本文算法在信源不平衡度为 0.03 时的正确识别率也达到 95%以上。

4.2　性能比较

为了更好地展示本文所提算法的性能，将本文所提的方法与文献[6]和文献[9]的方法从实际系统参数的计算量和仿真实验这 2 个方面进行比较。

假设扰码序列的长度为 N，本文算法的计算量主要集中在多项式的遍历和比特状态的统计这 2 部分。自同步扰码一般都选用 3~100 阶的 2 项式或 3 项式（r=2,3）作为生成多项式，在自同步扰码阶数 L 已知时，相应的 2 项式即为 $f(x)=1+x^L$，而 3 项式一共有 L-1 个，即 $f(x)=1+x^k+x^L$，k=1,2,⋯,L-1，因此待测试的多项式一共 L 个；对于每一个多项式需要遍历一次扰码序列以统计比特状态，而比特状态间隔为 L，引入的计算量为 N-L。故而本文算法的计算量为 $o((N-L)L)$。文献[6]的算法计算量集中在解扰和多项式的遍历 2 部分，对于每个测试多项式，需要对 N-L 个比特进行解扰，Cluzeau M 指出该算法多项式遍历部分的计算量为 $o(1/e^2)$，故文献[6]的计算量为 $o((N-L)/e^2)$。对扰码级数 L 已知的情况，在 $e=0.1$ 的时候，文献[6]的计算量与本文算法的计算量相当，e 越小，文献[6]的计算量越大；当扰码级数 L 未知时，本文算法对 3~100 阶的全部 2、3 项式（一共 5 047 个多项式）进行搜索遍历，这样的搜索量也并不大。

文献[9]在进行自同步扰码多项式的识别时，需要预先已知扰码级数 L，其计算量主要集中在 Hadamard 变换这部分，Hadamard 变换将 2^{L+1} 维的向量与$(2^{L+1},2^{L+1})$维的 Hadamard 矩阵相乘，计算量为 $o(2^{L+1}\cdot 2^{L+1})$，即便使用快速 Walsh-Hadamard 变换，文献[9]也仍然需要 $o(2^{L+1}\cdot(L+1))$的计算量，此外，Hadamard 矩阵的存储还需要附加 2^{2L+2} 的存储空间。综上所述，文献[9]所需要的计算量和存储空间都随着扰码级数 L 的增加而呈指数增长，远大于本文所需要的计算量。

仿真对比实验使用 ITU V.34（自同步扰码多项式 $1+x^{18}+x^{23}$）和 ITU G.709（自同步扰码多项式 $1+x^{43}$）这 2 个实际系统所使用的生成多项式。对于每一个多项式，按照 e =0.05 和 e =0.1 分别随

机生成长度从 0~5 000bit 以 200bit 步长变化的序列作为扰码器的输入信息序列，分别使用本文、文献[6]和文献[9]的算法对相应的自同步扰码序列进行仿真测试。即分别使用这 3 种方法对表 3 的 4 组参数所对应产生的不同长度的自同步扰码序列进行蒙特卡洛仿真实验，蒙特卡洛仿真次数为 500 次。

表 3 4 组仿真参数设置

仿真参数编号	信源不平衡度 e	生成多项式
第 1 组仿真参数	0.05	$f(x)=1+x^{18}+x^{23}$
第 2 组仿真参数	0.1	$f(x)=1+x^{18}+x^{23}$
第 3 组仿真参数	0.05	$f(x)=1+x^{43}$
第 4 组仿真参数	0.1	$f(x)=1+x^{43}$

前面已经指出，文献[9]所需要的计算量和存储空间都随着扰码级数 L 的增加呈指数增长，在利用文献[9]的方法对第 3 组和第 4 组参数进行仿真时的计算量和存储量实在过于庞大，计算机未能完成其仿真测试，因而并未给出文献[9]对第 3 组和第 4 组参数的仿真结果，即图 4(c)和图 4(d)中未画出文献[9]相应的曲线。

图 4 的仿真结果表明，本文算法、文献[6]和文献[9]的算法的正确识别率均随着扰码序列长度的增加而增大。就本文的算法而言，扰码序列越长，统计信息越充分，统计得到的实际比特状态概率分布越接近于理论分布，错误抽取时的 $d_{\%sqew}$ 和正确抽取时的 $d_{\%sqev}$ 之间的区分更明显，从而能获得更高的正确识别率。从图 4 可以看出，对于相同的仿真参数，为得到同样的正确识别率，本文算法所需要的数据量小于另外 2 种算法所需要的数据量。再者，对比图 4(a)和图 4(b)，图 4(c)和图 4(d)可以看到，本文算法对于信源不平衡度 e 的适用范围更广，性能更好。比如生成多项式为 $1+x^{18}+x^{23}$ 的情况，当 $e=0.1$，扰码序列长 2 000bit 时，3 种方法的正确识别率均为 100%；而当 e 变为 0.05，对长为 2 000bit 的扰码序列进行分析识别时，本文算法的正确识别

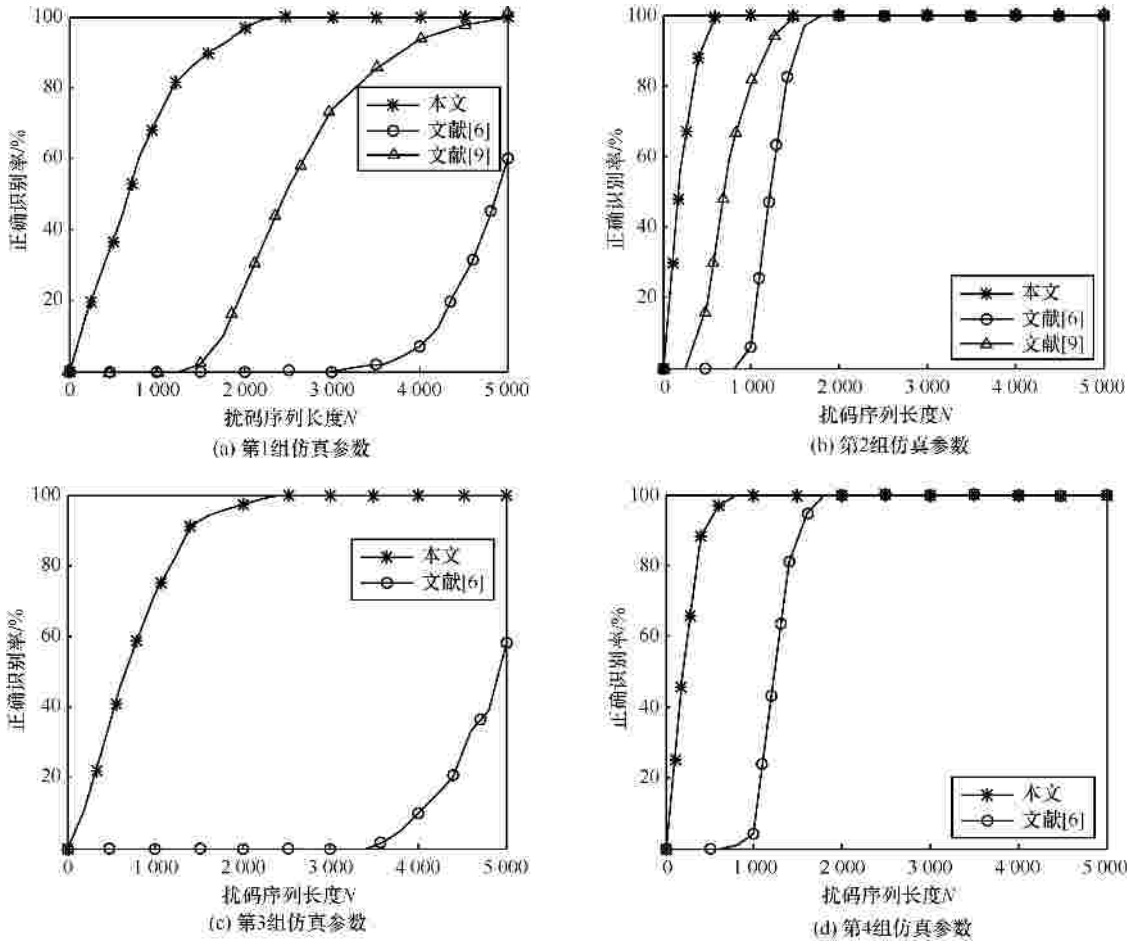

图 4 本文算法、文献[6]和文献[9]的算法对表 3 的 4 组参数的仿真对比

仍在 95%以上，文献[6]却不能够估计出正确的生成多项式，正确识别率为 0，文献[9]的正确算法识别率也降为 20%。综上所述，本文算法具有很好的识别性能和实际应用价值。

5　结束语

本文针对自同步扰码，从扰码序列符合比特流抽头位置上的比特状态不平衡性出发，进行深入分析研究，以比特状态的实际统计概率分布与均匀分布之间的修正平方欧几里得距离作为不平衡性的衡量准则，提出了一种有效的自同步扰码生成多项式盲识别方法。实验结果和对比分析验证了本文理论分析的正确性和所提算法的有效性。本文算法识别性能良好，对先验信息的要求少，且仅需要信息序列具有不平衡性，而并不需要预先知道扰码级数和精确的信源不平衡度，可以更好地应用于实际环境。

参考文献：

[1] CLUZEAU M. Reconnaissance D'un Schéma De Codage[D]. Limousin, France: Université De Limoges, 2006.

[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 W J, HUANG Z P, TANG G L, *et al*. Fast recovery of interfered scrambling code sequence[J]. Acta Armamentarii, 2009, 30(8):1134-1138.

[3] 郝士奇，戚林，王勇. 一种新的伪随机扰码盲识别方法[J]. 电路与系统学报, 2011, 16(4):6-12.
HAO S Q, QI L, WANG Y. A new blind recognition method of pesudo-randomizer code sequence[J]. Journal of Circuits and Systems, 2011, 16(4):6-12.

[4] 罗向阳，沈利，陆佩忠等. 高容错伪随机扰码的快速盲恢复[J]. 信号处理, 2004, 20(6):553-558.
LUO X Y, SHEN L, LU P Z, *et al*. Fast blind restore of LFSR sequences with high error tolerance[J]. Signal Processing, 2004, 20(6): 553-558.

[5] LIU X B, SOO N K, WU X W, *et al*. Reconstructing a linear scrambler with improved detection capability and in the presence of noise[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(1): 208-218.

[6] CLUZEAU M. Reconstruction of a linear scrambler[J]. IEEE Transactions on Computers, 2007, 56(9):1283-1291.

[7] 张永光，王挺，楼才义. 一种自同步扰码生成多项式的盲识别方法[P]. 中国: CN102201912A, 2011.
ZHANG Y G, WANG T, LOU C Y. A Blind Recognition Method of Self-Synchronized Scrambler's Generating Polynomial[P]. China: CN102201912A, 2011.

[8] 吕喜在，苏绍璟，黄芝平. 一种新的自同步扰码多项式盲恢复方法[J]. 兵工学报, 2011, 32(6):680-685.
LV X Z, SU S J, HUANG Z P. A novel blind recovery method of self-synchronizing scrambling polynomial[J]. Acta Armamentarii, 2011, 32(6):680-685.

[9] 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 Z L, LIU Y J. Algorithm research of self-synchronizing scrambler sequence[J]. Information Technology, 2005, (2):30-32.

[10] CHA S H. Comprehensive survey on distance/similarity measures between probability density functions[J]. International Journal of Mathematical Models and Methods in Applied Sciences, 2007, 300-307.

作者简介：

廖红舒（1978-），女，广西贵港人，电子科技大学讲师，主要研究方向为通信信号分析、阵列信号处理。

袁叶（1988-），女，四川乐山人，电子科技大学硕士生，主要研究方向为扰码识别、信道编译码算法等。

甘露（1974-），男，四川成都人，电子科技大学副教授、硕士生导师，主要研究方向为通信信号分析、阵列信号处理。