第五届（2020）全国高校密码数学挑战赛

赛题一

一、赛题名称：序列的 *k*-错线性逼近

二、赛题描述

1.1 基本概念

记 $\mathbf{F}_2$ 表示只含有 0、1 两个元素的有限域。记 $\underline{s} = (s_0, s_1, \ldots, s_{N-1})$表示 $\mathbf{F}_2$ 上长度为 N 的序列，这里 $s_i \in \mathbf{F}_2, 0 \le i \le N-1$。如果序列 $\underline{s}$ 满足线性递归关系：

$$s_i + d_1 \cdot s_{i-1} + \cdots + d_L \cdot s_{i-L} = 0, \quad L \le i \le N-1,$$

其中系数 $d_j \in \mathbf{F}_2, 1 \le j \le L,$ 则称多项式 $x^L + d_1 x^{L-1} + \cdots + d_L \in \mathbf{F}_2[x]$是序列 $\underline{s}$ 的一个特征多项式。次数最小的特征多项式称为序列 $\underline{s}$ 的极小多项式，极小多项式的次数称为序列 $\underline{s}$ 的线性复杂度，记为 $LC(\underline{s})$。特别地，当 $\underline{s}$ 为全 0 序列时，令 $LC(\underline{s}) = 0$。

$\mathbf{F}_2$ 上序列 $\underline{s} = (s_0, s_1, \ldots, s_{N-1})$的汉明重量是指 $s_i = 1, 0 \le i \le N-1$ 的个数，用 $w_H(\underline{s})$表示。$\mathbf{F}_2$ 上两条 N 长序列 $\underline{s}$ 和 $\underline{s}'$ 的汉明距离是指 $s_i \ne s'_i$, $0 \le i \le N-1$ 的个数，用 $d_H(\underline{s}, \underline{s}')$表示，即 $d_H(\underline{s}, \underline{s}') = w_H(\underline{s} + \underline{s}')$。

例如，序列 $\underline{s} = (1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1)$的极小多项式为 $x^4 + x + 1$，线性复杂度 $LC(\underline{s}) = 4$，汉明重量 $w_H(\underline{s}) = 8$。

1.2 问题描述

序列的 *k*-错线性逼近是指：对线性复杂度较高的序列 $\underline{s}$, 寻找一条含有 k 个比特错误、线性复杂度较低的序列来逼近 $\underline{s}$。数学描述如下：

设 $\underline{s} = (s_0, s_1, \ldots, s_{N-1})$是 $\mathbf{F}_2$ 上长度为 N 的序列，如果 $\mathbf{F}_2$ 上 N 长序列 $\underline{s}'$ 满足：

$$LC(\underline{s}') < LC(\underline{s}) \text{且 } d_H(\underline{s}, \underline{s}') = k,$$

则称 $\underline{s}'$是 $\underline{s}$ 的一条 *k*-错线性逼近序列。

密码分析者希望使用尽可能小的 k 值，获得线性复杂度尽可能低的 *k*-错线性逼近序列。请对附件中七条长度为 1000 的 $\mathbf{F}_2$ 上序列，分别寻找它们的 *k*-错线性逼近，使得逼近序列的线性复杂度与 k 值之和尽可能小，每条序列的 k 值不必相同。要求给出逼近序列的极小多项式、逼近序列与原序列不同的错误比特的位置

（序列比特位置从 0 开始编号）。

例如，长度为 18 的序列 $\underline{s}=(1,1,1,0,0,0,1,0,0,1,1,0,1,0,1,1,1,0)$的线性复杂度是 14，改变其最后 1 比特（将 s_{17} 由 0 改为 1），将得到 $\underline{s}$ 的一条线性复杂度为 4 的 1-错线性逼近序列。逼近序列的线性复杂度与 k 值之和为 5，错误比特位置为 17。

1.3 成绩评判

本题共有七条序列，分别标记为：序列 1~序列 7。下面给出第 i 条序列的评分办法，$1 \le i \le 7$：

如果参赛者没有给出结果，或者所给极小多项式无法产生其声称的逼近序列，则该条序列的得分记为 0；

在上述两种情况之外，第 i 条序列的得分采用公式：$(\frac{1000}{k+l}-1.5)\times i$ 计算，这里 l 是参赛者所给逼近序列的线性复杂度，k 是错误比特数，i 是序列的编号。

七条序列的得分之和作为参赛者的最终得分，最终得分高者获胜。

三、赛题背景与研究进展

线性递归序列是许多序列密码和伪随机数发生器的主要部件，利用 Berlekamp-Massey 算法（简称 B-M 算法）容易求取一条序列的极小多项式，并且只需已知连续两倍线性复杂度长的序列片段就可以还原整条线性递归序列。为了抵抗 B-M 算法的分析，伪随机序列必须具有足够大的线性复杂度。上世纪 80 年代末，我国学者肖国镇和丁存生等指出序列仅仅线性复杂度高是不够的，还希望对其改变少量比特后线性复杂度不会大幅下降。同时期，美国学者 M. Stamp 和 C. F. Martin 也独立地提出了衡量序列线性复杂度稳定性的 k-错线性复杂度。对密码分析者而言，希望能够快速求取一条含有少量错误比特、线性复杂度尽可能低的序列来逼近原序列，这就是序列的 k-错线性逼近问题。

求序列的 k-错线性逼近，最直接的方法就是穷尽所有可能的 k 个错误位置再调用 B-M 算法来求线性复杂度，然而当 k 较大时，计算复杂度非常高。通过采用启发式策略结合 B-M 算法，A. Alecu 和 A. Sălăgean 设计了有限长序列的 k-错线性逼近的算法，该算法的计算复杂度仍然是指数级的。对于周期是 2 的方幂形式的周期序列，M. Stamp 和 C. F. Martin 设计了求 k-错线性复杂度的多项式时间

算法，A. G. B. Lauder 和 K. G. Paterson 也设计了求序列错误线性复杂度谱的快速算法。A. Sălăgean 将有限长序列看作是 2 方幂周期序列的一部分，利用 Stamp-Martin 算法设计了求 k-错线性逼近的快速算法。

四、参考文献

[1] J. L. Massey. Shift register synthesis and BCH decoding[J]. IEEE Transactions on Information Theory, 1969, 15(1): 122–127.

[2] M. Stamp, C. F. Martin. An algorithm for the k-error linear complexity of binary sequences with period 2^n[J]. IEEE Transactions on Information Theory, 1993, 39(4): 1398–1401.

[3] A. Sălăgean. On the computation of the linear complexity and the k-error linear complexity of binary sequences with period a power of two[J]. IEEE Transactions on Information Theory, 2005, 51(3): 1145–1150.