序列密码的分析与设计

关杰 丁林 张凯 著

科学出版社

序列密码的分析与设计

关杰 丁林 张凯 著

科学出版社

北京

内 容 简 介

本书介绍序列密码的设计理念、设计原理和分析技术，以21世纪为划分点，本书将序列密码加密模型分成传统模型和新型模型，介绍各个模型的代表算法，包括蓝牙系统 E_0 加密算法、GSM 手机 A5 加密算法，NESSIE 计划候选算法 SNOW 系列算法、我国设计的新一代无线移动通信系统标准 ZUC 算法，以及 eSTREAM 计划胜出算法 Trivium、Grain、MICKEY、Salsa20 等，CAESAR 竞赛中认证加密算法 MORUS 等典型序列密码算法的设计特点，同时介绍针对这些序列密码算法的典型分析方法及最新的攻击结果，以期为密码设计和分析者提供参考和借鉴。

本书可作为密码学和信息安全专业高年级本科生和研究生的教材，也可作为从事相关专业的教学、科研和工程技术人员的参考书。

图书在版编目(CIP)数据

序列密码的分析与设计 / 关杰，丁林，张凯著. —北京：科学出版社，2019.11

ISBN 978-7-03-062969-2

Ⅰ. ①序… Ⅱ. ①关… ②丁… ③张… Ⅲ. ①密码学 Ⅳ. ①TN918.1

中国版本图书馆CIP数据核字(2019)第242859号

责任编辑：张艳芬 赵微微 / 责任校对：杨聪敏

责任印制：吴兆东 / 封面设计：蓝 正

科学出版社出版

北京东黄城根北街16号

邮政编码：100717

http://www.sciencep.com

北京中石油彩色印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2019年11月第 一 版 开本：720×1000 1/16

2019年11月第一次印刷 印张：16 3/4

字数：325 000

定价：108.00元

(如有印装质量问题，我社负责调换)

前　言

序列密码是密码学的一个重要分支，它在密码学中处于基础地位，其设计理念和分析思想对密码学各大分支的发展发挥着重要作用。序列密码的研究历经三起三落，却始终是密码研究者关注的一个方向。自从 2004 年 eSTREAM 计划启动以来，序列密码重新成为密码领域的研究热点。

与分组密码算法相比，序列密码算法的设计呈现出多样性、迥异性、个性化等特点。以 21 世纪为划分点，本书将序列密码加密模型分成传统模型和新型模型：传统模型主要包括前馈模型、钟控模型等；新型模型主要包括基于非线性移位寄存器型、表代替型、类分组型、SNOW 型等序列密码模型。

本书将某些结构相似、设计原理相同的一系列典型序列密码算法抽象成模型，从序列密码设计模型出发，介绍典型算法的设计特点和研究现状，并给出针对这些序列密码算法的典型分析方法和最新的攻击结果。

本书撰写分工如下：第 1 章～第 3 章由关杰撰写，第 4 章、第 6 章和第 8 章由丁林撰写，第 5 章和第 7 章由张凯撰写。关杰对全书进行了统稿。

本书凝结了作者及其科研团队的研究成果，在此，对作者指导的所有硕士研究生和博士研究生一并表示感谢。本书的撰写工作得到了战略支援部队信息工程大学密码工程学院密码理论与技术实验室全体师生的积极配合，特别是李俊志博士、施泰荣博士、周琮伟博士、刘帅博士及李昂硕士、黄俊君硕士、马宿东硕士等给予了全力配合，在此一并对他们表示衷心的感谢！

本书的出版得到了战略支援部队信息工程大学教材建设基金的资助。此外，本书部分成果来自国家自然科学基金项目(61572156、61202491)的资助。在此一并表示感谢。

限于作者的水平，本书难免存在不足之处，恳请读者批评指正。

作　者

2019 年 3 月 6 日

目　录

第1章 概 述

序列密码，也称流密码，最初主要应用于军事、政治等要害部门，目前世界上绝大多数国家和地区的军事、政府、外交领域的保密通信仍采用序列密码。随着互联网和无线通信应用的日益广泛，序列密码已广泛应用于商业、个人的信息加密，并且因其自身独特的特点和优势，具有广阔的应用前景。

序列密码和分组密码是对称密码体制的两大重要分支。两者在设计理念、算法结构、应用场景等方面既有很大区别又有紧密联系。两者最重要的区别体现在“记忆性”上。分组密码通常是按固定规模将明文分组，对每组均使用一个固定的加密变换来进行运算，是“无记忆”的；序列密码是由少量的真随机数按照固定规则生成密钥序列，密钥序列再和明文分组(一个明文的独立符号单位)结合生成密文，因此其加密变换是随时间变化的，具有时序性，是“有记忆”的。

欧洲两个密码征集计划 NESSIE(New European Schemes for Signatures, Integnity, and Encryption)[1]和 ECRYPT(European Network of Excellence for Cryptology)[2]极大地促进了现代序列密码的研究。许多经过精心设计和公开分析的序列密码算法与同级别的分组密码算法相比，占用的资源更少，速度更快。多数密码学家认为，序列密码在资源极端受限的硬件领域和需要极高加解密速度的领域两个方面具有较大优势。未来序列密码研究主要围绕新型序列密码模型、新型序列密码分析方法、序列密码的初始化过程及基于序列密码的认证加密算法的设计与分析等方面展开，这些是密码工作者研究的热点和重点。

1.1 序列密码的发展历史

序列密码的历史比较悠久，可以追溯到古典密码的多表驱动。例如，维吉尼亚密码就是一种序列密码。在第二次世界大战期间，德国的 Enigma 密码和日本的“紫密”密码是典型的机械式序列密码，因其在战争和外交中的突出作用而受到重视。

1.1.1 维吉尼亚密码

维吉尼亚密码的密钥空间、明文空间和密文空间均为英文字母的序号集合 $Z_{26}=\{0,1,\cdots,25\}$，加密变换为对英文字母的加密变换：

$$c=(m+k)\bmod 26$$

使用长度为 d 的密钥 $k=(k_1,k_2,\cdots,k_d)$，加密时对每个明文 $m=(m_1,m_2,\cdots,m_n)$ 进行加密变换得到密文 $c=(c_1,c_2,\cdots,c_n)$，这里

$$c_i=(m_i+k_i)\bmod 26,\quad i=1,2,\cdots,n$$

当 $i\geqslant d+1$ 时，将密钥 k 按周期重复使用即可。

将维吉尼亚密码中密钥的周期 d 扩大为无限，即是“一次一密”密码体制的雏形。

1.1.2 “一次一密”密码体制

设明文序列 $(m_i)_{i=1}^n$ 是二元明文序列，$(k_i)_{i=1}^n$ 是二元密钥序列，$(c_i)_{i=1}^n$ 是二元密文序列，且 $\forall i\geqslant 1$，都有 $c_i=m_i\oplus k_i$，则当且仅当 $k_1,k_2,\cdots,k_n$ 相互独立且都在密钥空间 K 上服从均匀分布时，该密码体制称为“一次一密”密码体制。

当明文、密文不是二元序列时，只要保证密钥相互独立且在密钥空间 K 上服从均匀分布，将异或运算“$\oplus$”替换为拉丁方变换，则“一次一密”密码体制即可扩展为下述“一次一密”密码体制，结构如图 1.1.1 所示。

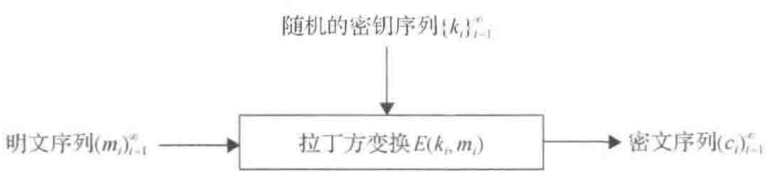


图 1.1.1　“一次一密”密码体制的结构

定义 1.1　设 X、Y、Z 都是具有 n 个点的有限集，$f:X\times Y\to Z$。若对 $\forall x_0\in X$，$y_0\in Y$，以 y 为变量的映射 $f(x_0,y)$ 是 Y 至 Z 的双射，且以 x 为变量的映射 $f(x,y_0)$ 是 X 至 Z 的双射，则称 f 是拉丁方变换。

“一次一密”密码体制是 Mauborgne 和 Vernam 于 1917 年发明的，在研究电报通信的实际工作中，Mauborgne 发现可以使用便笺来记录密钥，每页纸上有排列好的随机字母或数据，两份相同的便笺分发给收方和发方，其他人不能够获得或预测出便笺上的任何信息，由于每页纸上的每个数据只使用一次，又称其为“一次性便笺加密体制”。

1949 年，Shannon 在“保密系统的通信理论”一文中证明了“一次一密”密码体制在唯密文攻击下，即使是具有无限计算资源的攻击者也不能识别出真正的明文[3]。“一次一密”密码体制是迄今为止唯一一个理论上无法攻破的加密体制，是完全保密的密码体制。它在密码学的发展中意义重大，是序列密码设计理念的源头，也促进了量子密码的发展。

1.1.3 序列密码的设计理念

“一次一密”密码体制利用随机的密钥序列对明文序列加密得到密文序列。由于随机的密钥序列必须与明文等长，其生成、分配、存储和使用都存在一定的困难，因此人们设想使用少量的真随机数按固定规则生成“伪随机”的密钥序列代替真正的随机序列，这就是序列密码的设计思想。序列密码中使用的少量真随机数就是序列密码的密钥，也称为种子密钥。序列密码的这种设计理念达到了只需分配和存储少量的真随机数(种子密钥)就可对任意长度的明文加密的目的。因此，序列密码脱胎于“一次一密”密码体制。

序列密码的设计思想是将种子密钥通过密钥流生成器产生的伪随机序列(也称为乱数序列)与明文简单结合生成密文。称与明文结合的元素为密钥流元素(也称为乱数)，称产生密钥流元素的部件为密钥流发生器或乱数发生器。一个序列密码方案是否具有很高的密码强度主要取决于密钥流发生器的设计。序列密码中的密钥序列是由少量真随机数按固定规则生成的，因而不可能是真正随机的。因此，如何刻画密钥序列的伪随机性，如何保证密钥序列的伪随机性，使其不会造成加密算法在实际中被破解，是序列密码设计中需要解决的问题。序列密码的安全性基础即是无法有效地将伪随机序列(乱数序列)与随机序列区分开。“区分问题”也可以拓展到其他类型的密码体制，如分组密码、杂凑函数等，如何将密码算法输出和(伪)随机序列有效区分，是密码分析的基础问题。

序列密码中涉及的随机概念贯穿于密码学各分支中，奠定了序列密码在密码学中的基础地位。序列密码的安全性问题，即区分问题也是密码设计者需要考虑的一个首要问题。可以说，序列密码在密码学中处于基础地位，发挥着重要作用。

1.1.4 序列密码的传统模型

以21世纪为划分点，本书将序列密码加密算法分为传统算法和新型算法。称21世纪前的序列密码算法为传统算法，21世纪后的序列密码算法为新型算法。

20世纪初，线性反馈移位寄存器(linear feedback shift register，LFSR)的产生使得序列密码拥有了很好的驱动部件，但是到了20世纪60年代，B-M算法的提出使得序列密码的设计者不得不将非线性改造增加到设计中去。因此，20世纪80年代，序列密码的设计主要采用LFSR和非线性改造相结合的方式，包括前馈模型和钟控模型等，主要以蓝牙(bluetooth)技术中用于数据加密的 E_0 算法[4]、用于手机通信中蜂窝语音和数字加密的A5系列算法[5]等典型算法为代表。

1. 前馈模型

前馈模型是序列密码的一个基本模型。LFSR 产生的输出序列由于具有线性制约性，不能直接作为乱数序列使用。因此，前馈模型的基本思想就是利用非线性变换对 LFSR 的输出序列或状态序列进一步加工，达到破坏其线性制约性的目的，并保持 m 序列的周期长和统计特性好的优点。

当初始乱源发生器是 LFSR 时，根据初始乱源发生器是一个 LFSR 还是多个 LFSR，又可将前馈模型分为非线性滤波模型和非线性组合模型。

非线性滤波模型是将 LFSR 的若干抽头的输出经过非线性函数作用而产生乱数。

非线性组合模型是将若干 LFSR 的输出组合起来经过非线性函数作用后产生乱数。

2. 钟控模型

钟控模型的主要思想是利用一条由密钥决定的未知序列(如某 LFSR 的输出序列)来控制另外一条 LFSR 序列的动作方式(动作次数、状态更新方式)等。钟控模型主要有他控、自控和互控等三种模型。

3. 单表驱动模型

到了 20 世纪 90 年代，随着计算机技术的发展，出现了第一个面向软件的序列密码算法 RC4[6]，其软件速度要比当时的数据加密标准(data encryption standard，DES)快很多。它是由 Rivest 于 1987 年设计的密钥长度可变的序列密码加密算法簇。RC4 算法的结构十分简洁，通过表驱动的方式实现随机数的生成，和前面基于 LFSR 的、易于硬件实现的序列密码设计思想完全不同。这种表驱动型序列密码易于软件实现，广泛用于商业密码产品中，包括 Lotus Notes 和 Oracle 等。

1.2 序列密码的研究现状

进入 21 世纪，序列密码研究呈现蓬勃发展的趋势。日本在 2000～2003 年开展了 CRYPTREC(Cryptography Research and Evaluation Committees)密码征集计划，欧洲的两个序列密码征集计划——NESSIE 和 eSTREAM，更是极大地促进了序列密码的研究，逐渐使其成为密码研究领域的热点。

1.2.1 NESSIE 计划

NESSIE 是欧洲的一个征集签名、完整性和加密的密码征集计划，开始于2000 年并在 2004 年结束。NESSIE 计划的主要目标是通过公开透明的征集和评价，建立一整套高效安全的密码标准。NESSIE 共征集到 42 个不同类型的密码算法，有 6 个序列密码算法(LEVIATHAN[7]、LILI-128[8]、BMGL[9]、SOBER-t32[10]、SNOW[11]、SOBER-t16[12])进入了第二阶段的评估；但由于 6 个候选序列密码算法都存在弱点，因而最终都被淘汰。这也反映了序列密码的研究没有分组密码成熟，同时这也引起了众多序列密码学者的关注，并激发出他们的研究兴趣，随后人们对序列密码的研究力度也明显增加。

1.2.2 CRYPTREC 计划

日本政府在 2000 年实施了密码征集计划 CRYPTREC[13]，并参考了 NESSIE 计划的做法，对征集到的密码算法的安全性和效率等方面进行评估。2003 年 5 月，CRYPTREC 计划推荐了 3 个序列密码算法：MUGI[14]、MULTI-S01[15]和 128 比特的 RC4，其中 RC4-128 限定只能应用于 SSL3.0 和 TSL1.0 或 TSL 随后的版本中。与 NESSIE 计划相比，CRYPTREC 计划对序列密码发展的影响较小。

1.2.3 eSTREAM 计划

ECRYPT 是欧洲的三十几个大学和公司在密码学与水印方面进行合作研究的联盟，它于 2004 年 11 月启动了欧洲序列密码征集计划 eSTREAM，该计划的目的就是公开征集序列密码算法并筛选出可以广泛应用的序列密码算法，它要求序列密码必须具有如下特点之一。

(1) 软件实现时，序列密码算法具有高吞吐率。

(2) 硬件实现时，序列密码算法仅需要有限资源(如有限的存储空间、与非门数量和功耗等)。

eSTREAM 计划的征集引起了各国学者的广泛关注，截至 2005 年 4 月共征集到 34 个序列密码算法。随后，ECRYPT 每年举行一次学术会议，主要对各个候选序列密码算法进行深入的安全性和效率评估。2008 年 5 月公布的最终评选结果[16]是：eSTREAM 计划候选胜出算法有 7 个，其中 3 个是面向硬件实现的算法，即 Grain v1、MICKEY v2 和 Trivium；有 4 个是面向软件实现的算法，即 HC-128、Rabbit、Salsa20/12 和 SOSEMANUK。

事实上，一个精心设计、经过公开分析的序列密码算法要比同级别的分组密码算法在软件实现上快 3～5 倍，或者需要的硬件资源仅为分组密码的 1/3，因此序列密码算法的应用前景十分广阔，这也正是 eSTREAM 计划得到广泛重视的一

个根本原因。

1.2.4　序列密码的应用场合

与分组密码相比，序列密码在一些特定的应用场合发挥重要的作用。只要能够充分发挥自身优势，序列密码将具有很好的应用价值和发展前景，而不会被分组密码取代。

1. 资源受限的环境

在手机、无线通信、智能卡等一些有资源限制要求，体积小、运算速度高等应用环境中需要使用序列密码进行加密。例如，大家熟知的欧洲蜂窝式移动电话系统加密标准 A5/1 算法，蓝牙网络规范中用于数据加密的 E_0 算法等，均采用了易于硬件实现的 LFSR，是轻量级密码算法。

2. 对数据格式有特殊要求的环境

数据库加密或主机和终端通信加密时，当设备没有存储区或数据缓冲区有限时，数据处理必须是每次一个符号，这时需要使用序列密码。例如，RC4 算法每个时刻输出的乱数是 8 比特，可以一次加密一个字节，其生成乱数的软件速率高、随机性较好，因此广泛应用于 Oracle、Lotus Notes 等软件中。

3. 信道不好的一些特殊应用环境

在无线信号传输等密文信号容易丢失或出错的应用环境，可以利用自同步序列密码有限步的错误传播特性进行加密，分组密码的密码反馈模式就是一种自同步序列密码；对于信号不容易丢失但容易出错的环境，如果明文的冗余度大，明文出现错误不影响效果，可采用序列密码进行数据加密，如在卫星通信中的图像加密、语音加密等环境中，利用分组密码的输出反馈模式就可以设计出这类序列密码。

4. 高度机密的低带宽通信环境

在军事和外交保密通信这些高度机密的特殊条件下，仍然可采用“一次一密”的完全保密(无条件安全)体制进行加密。

1.2.5　序列密码的分类

文献[17]描述了序列密码的基本模型，如图 1.2.1 所示。下面按照乱数发生器的生成方式对序列密码进行分类。

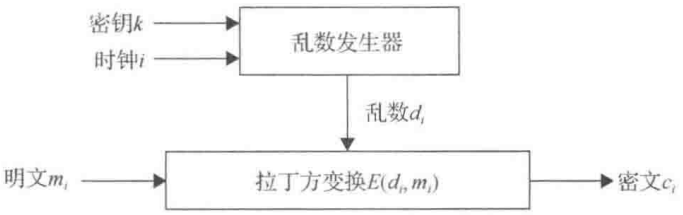


图 1.2.1　序列密码的基本模型

1. 按照初始化过程的结构分类

乱数发生器主要由初始化过程和密钥流生成过程组成。序列密码初始化过程的作用是使得初始密钥 K 和初始值 IV 之间达到充分的混乱和扩散效果，序列密码密钥流生成的作用是在初始化过程的基础上对初始密钥 K 和初始值 IV 做进一步变换，产生对明文加密所使用的乱数流。本书以序列密码初始化过程与密钥流生成过程之间差异程度的大小为标准，将初始化过程的设计分为结构相同型序列密码、结构相似型序列密码和结构迥异型序列密码三类[18-26]。

作为序列密码算法的重要组成部分，初始化过程的安全性直接影响序列密码算法的安全性，安全高效的序列密码需要以安全高效的初始化过程为基础，因此很有必要对序列密码初始化过程的安全性进行深入研究。

2. 按照乱数生成过程是否独立于明密文分类

根据乱数生成过程是否与明密文有关，可将序列密码分为同步序列密码(synchronous stream cipher，SSC)、自同步序列密码(self-synchronous stream cipher，SSSC)和非同步非自同步序列密码三类。

1) 同步序列密码

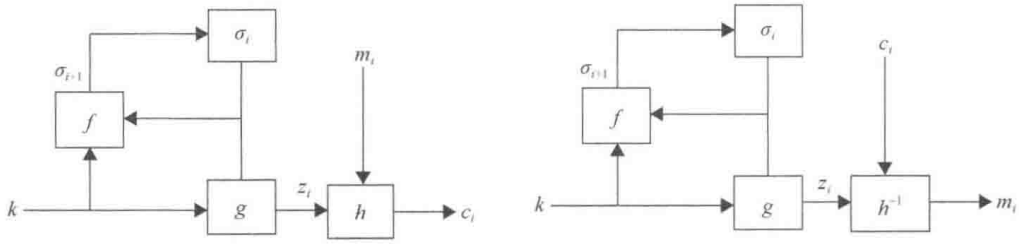
若乱数序列独立于明文、密文，即乱数序列与明文、密文无关，则称此类序列密码算法为同步序列密码[27]。同步序列密码的加密过程可由下式描述：

$$\sigma_{i+1} = f(\sigma_i, k)$$

$$z_i = g(\sigma_i, k)$$

$$c_i = h(z_i, m_i)$$

式中，σ_i 为第 i 时刻的内部状态，i=0 时为初始状态；k 为密钥；f 为状态更新函数；g 为产生密钥流 z_i 的函数；h 为输出函数；m_i 为明文；c_i 为密文。同步序列密码加密与解密过程如图 1.2.2 所示。



(a) 同步序列密码加密图　　(b) 同步序列密码解密图

图 1.2.2　同步序列密码的一般模型

目前大多数序列密码算法均是同步序列密码算法，例如，eSTREAM 计划候选胜出的 7 个序列密码算法均是同步序列密码算法。分组密码的输出反馈模式也是同步序列密码的一个例子。

2) 自同步序列密码

若乱数序列与以前若干个时刻的密文有关，则称其为自同步序列密码[27]。

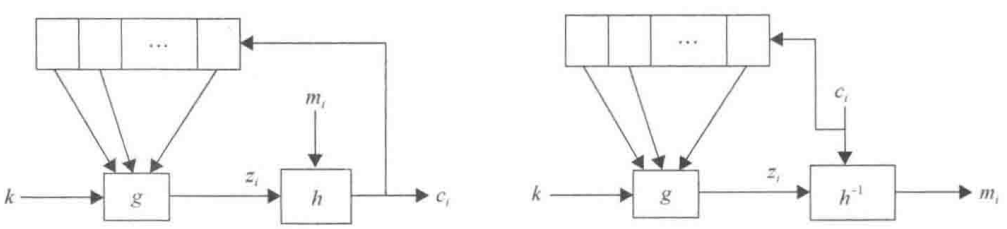
自同步序列密码的加密过程可由下式描述：

$$\sigma_i = f(c_{i-t}, c_{i-t+1}, \cdots, c_{i-1})$$

$$z_i = g(\sigma_i, k)$$

$$c_i = h(z_i, m_i)$$

式中，σ_i 为第 i 时刻的内部状态，可由密钥 k 决定，$\sigma_0 = (c_{-t}, c_{-t+1}, \cdots, c_{-1})$ 为初始状态；f 为状态更新函数；g 为产生密钥流 z_i 的函数；h 为输出函数。自同步序列密码加密与解密过程如图 1.2.3 所示。



(a) 自同步序列密码加密图　　(b) 自同步序列密码解密图

图 1.2.3　自同步序列密码的一般模型

在丢失若干密文信号但此后信号不再丢失的条件下，自同步序列密码仍能正确解密此后的密文信号。自同步序列密码可以检测对密文的篡改，提供认证功能。eSTREAM 计划第一轮提交的 34 个候选算法中，只有一个序列密码[28]是自同步序列密码，由于其存在安全问题而在第二轮评选中被淘汰。

3)非同步非自同步序列密码

若乱数序列与明文有关，而与密文无关，则称此类序列密码为非同步非自同步序列密码[16]。eSTREAM 计划进入第二轮的候选算法 Phelix 算法[29]就是明文反馈参与乱数序列生成的，是一类非同步非自同步序列密码，可以检测对明文的篡改，提供认证功能。

3. 按照新型序列密码算法的结构分类

本书对以 eSTREAM 计划候选算法为代表的新型序列密码算法进行了分析研究，发现新型序列密码算法的设计与传统的序列密码算法具有很大不同，结合密码学专家和学者对该项目征集到的 34 个候选算法的分析，根据乱数发生器的设计模型特点，可将新型序列密码算法的设计划分为以下三种类型。

1)基于反馈移位寄存器型序列密码

反馈移位寄存器(feedback shift register，FSR)可分为 LFSR 和非线性反馈移位寄存器(non-linear feedback shift register，NFSR)。

传统序列密码大多基于 GF(2)上的 LFSR 设计，便于硬件实现。新型序列密码算法中，为了考虑软件实现，大多采用有限域 GF(2^m)上的 LFSR。例如 NESSIE 计划候选算法 SNOW 系列[11]，以及 eSTREAM 计划胜出算法 Sosmenuk 都是采用了有限域 GF(2^{32})上的 LFSR 结合有限状态机(finite-state machine，FSM)的设计结构。

新型序列密码算法中还涌现出一批基于 NFSR 设计的密码算法，如 Grain、MICKEY 和 Trivium 等算法。相关分析、代数分析等密码分析方法的发展对基于 LFSR 的序列密码造成了一定的威胁，因此人们把目光放到了 NFSR 的设计。但是，人们对 NFSR 密码学性质的了解不如 LFSR 那样彻底，也缺少合适的数学工具对其进行研究，在采用 NFSR 进行设计时容易出现问题，有时还需结合 LFSR 进行设计以保证其安全性。例如，eSTREAM 计划胜出算法中的 Grain 算法就是采用 LFSR 级联 NFSR 的结构以保证其周期可控。

2)表驱动型序列密码

传统算法中 RC4 算法是典型的单表驱动型序列密码算法，它基于一个代替表，通过交换表中元素位置的方式实现对表的动态更新，其结构简洁，软件实现速率快，但是该算法也存在严重的安全问题，例如前几个时刻的乱数呈现出不随机的特性、使用方式不当可导致该算法被完全破译等。新型序列密码算法的表驱动方式对 RC4 算法的单个表更新方式进行了改进，利用多个状态表的互控更新来构造序列密码，结构较为复杂、安全强度高，例如 eSTREAM 计划胜出算法中的 HC-128 算法，就是采用两个状态表相互交替更新设计而成的。

3)类分组型序列密码

类分组型序列密码是利用分组密码部件或者分组密码思想构造序列密码。该