基于单因素方差分析的密码算法统计检验

朱玉倩，王　超，张　艳
（华北计算机系统工程研究所，北京 100083）

摘　要： 在密码学范畴中，随机序列常作为密钥、初始向量或算法参数使用。随机序列的随机性最终决定了整个密码系统的安全性，因此在密码技术中占有重要位置。对于良好的密码算法产生的密文序列，应无法通过统计学方法进行区分。首先对 7 种经典密码算法生成的密文序列进行 NIST 随机性检验，统计失败次数；然后关于密码算法进行单因素方差分析，检验结果在统计学上无显著差异。此统计检验可作为评价密码算法好坏的指标之一。

关键词： 单因素方差分析；密码算法；统计检验；NIST 随机性检测
中图分类号： TP309.7　　**文献标识码：** A　　DOI：10.16157/j.issn.0258-7998.211329

中文引用格式： 朱玉倩，王超，张艳. 基于单因素方差分析的密码算法统计检验[J].电子技术应用，2021，47(9)：43-45，50.
英文引用格式： Zhu Yuqian，Wang Chao，Zhang Yan. Statistical test of cryptographic algorithms based on ANOVA[J]. Application of Electronic Technique，2021，47(9)：43-45，50.

Statistical test of cryptographic algorithms based on ANOVA

Zhu Yuqian，Wang Chao，Zhang Yan
(National Computer System Engineering Research Institute of China，Beijing 100083，China)

Abstract： In the field of cryptography，a random sequence is often used as a key，an initial vector or a time-varying parameter in cryptographic protocol. Actually，the randomness of a random sequence plays a very important role in the cryptography，since it determines the security of the whole system. The ciphertext sequence generated by a good cryptographic algorithm should not be distinguished by statistical methods. In this paper，we count the number of failures by using the national institute of standards and technology(NIST)，which is executed on the ciphertext sequences generated by seven classic cryptographic algorithms. The analysis shows that the results are not statistically significant. Thus，the statistical test used in the paper can be used as one of the indicators to evaluate the quality of cryptographic algorithms.

Key words： one way analysis of variance；cryptographic algorithms；statistical test；national institute of standards and technology (NIST) randomness test

0 引言

在密码学领域，主要使用对称密码算法对信息进行加密，保障信息的机密性。随着密码分析技术的进步和敌手攻击能力的提升，加密密码算法的设计要求不断提高。分析与识别保密系统所采用的密码算法，对于评估信息系统安全性、密码分析和攻击、非法通信监控、恶意代码识别等都有着重要的理论意义[1]。

近几年加密算法生成密文的检验领域成果层出不穷，王瑛[2]等人结合人工智能技术和机器学习方法，研究和设计了网络加密流量检测体系框架和方法。吴杨[3-4]等人通过 NIST 随机性检测标准中的单比特频数检验、块内频数检验和游程检验理论设计统计量，对 OpenSSL 软件库中的 AES、Camellia、DES、3DES 和作者实现的 SM4 分组密码算法生成的 200 组、每组 1 万条密文进行分析，实现上述 5 种分组密码算法的识别。

本文方法具有以下不同点：

(1)支持全部 NIST 随机性检验标准；

(2)统计量为 NIST 检验失败次数；

(3)7 种对称密码算法，涵盖分组密码算法和序列密码算法；

(4)5 种随机方式和 2 种影射方式构造密钥和初始向量得到 10 类密文；

(5)运用单因素方差分析。

1 背景知识

1.1 密码算法

密码算法主要包括序列密码算法、分组密码算法、公钥密码算法以及散列函数。序列密码算法输出的密文序列既与密码算法相关，又与算法加载的密钥和初始向量相关，不同的密钥和初始向量将产生不同的输出；分组密码算法输出的密文序列既与密码算法和工作模式相关，又与算法加载的密钥、明文和初始向量相关。对于给定的分组密码算法、工作模式和固定明文，不同的密

钥和初始向量将产生不同的输出。

本文主要使用序列密码算法和分组密码算法，具体包括 ZUC-128 算法[5]、ZUC-256 算法[6]、Snow 算法[7]、AES 算法[8]、SM4 算法[9]、SOSEMANUK 算法[10]以及 Triviurm 算法[11]，共计 7 种密码算法。

ZUC-128 算法是我国自主研发的序列密码算法，被 3GPP 选为 LTE 加密标准算法。它采用 128 bit 的初始密钥和 128 bit 的初始向量作为输入，输出关于字的密钥流对信息进行加解密。ZUC 执行分为两个阶段：初始化阶段和工作阶段。第一阶段是对密钥和初始向量进行初始化，第二阶段每一个时钟脉冲产生一个 32 bit 的密钥输出。

ZUC-256 算法是 ZUC-128 算法的改进版，包含了初始化阶段、密钥流生成阶段及消息认证码生成阶。初始化阶段采用 256 bit 初始密钥，与 ZUC-128 流密码高度兼容，进一步提高了算法的安全性。

Snow 算法是一种面向字的序列密码算法，其密钥长度为 128 bit，初始向量长度为 128 bit。

SM4 算法是一种分组密码算法，作为国家密码管理局公布的第一个商用分组密码标准。分组长度为 128 bit，密钥长度也为 128 bit。

SOSEMANUK 算法是一种序列密码算法，该算法是欧洲流密码工程胜选算法之一，其密钥长度为256 bit，初始向量长度为 128 bit。

Triviurm 算法是一种基于硬件的序列密码算法，是欧洲流密码工程 eSTREAM 胜选算法之一，支持多种长度的密钥和初始向量。本文使用经典的长度，密钥长度为 80 bit，初始向量长度为 80 bit。

AES 算法是一种分组密码算法，作为美国新的分组密码标准，相对于其他传统加密算法，AES 在扩散性、混淆性和数据加解密效率等方面具有较为明显的优势。AES 算法的分组长度为 128 bit，支持 128 bit、192 bit 和 256 bit 的密钥长度，分别记为 AES-128、AES-192、AES-256，相应的迭代轮数分别为 10 轮、12 轮和 14 轮。本文使用的是经典的 AES-256。

1.2 随机检测方法

NIST 随机性检测标准是美国国家标准与技术研究院(National Institute of Standards and Technology)用于检测比特序列与真随机序列之间偏差的方法集。主要包括：单比特频数检验(Frequency)、块内频数检验(Block Frequency)、游程检验(Runs)、最大游程检验(Longest Run)、二元矩阵秩检验(Rank)、离散傅里叶变换检验(FFT)、非重叠匹配检验(NonOverlapping Template)、重叠匹配检验(Overlapping Template)、全局通用统计检验(Universal)、线性复杂度检验(Linear Complexity)、序列串行检验(Serial)、近似熵检验(Approximate Entropy)、累加和检验(Universal)、随机偏移检验(Random Excursions)以及随机偏移变量检验(Random Excursions Variant)。其具体应用过程可参考文献[12]中的研究内容。

1.3 密钥和初始向量的构造方式

为了降低密钥和初始向量的选取对密文序列的检测判断产生影响，本文采用 5 种随机方式复合 2 种影射方式，共 10 种方式构造[5,12]。

(1)随机方式

①反馈移位寄存器

n 位线性反馈移位寄存器的逻辑功能如图 1 所示，0 和 1 是特征为 2 的素域 GF(2)的两个元素。其中 $F(x_0, x_1, \cdots, x_{n-1})=\sum_{i=0}^{n-1} a_i x_i$，$a_i \in \mathrm{GF}(2)$。

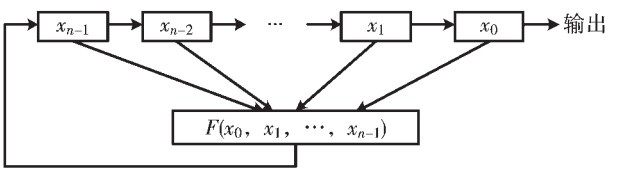

图 1 n 位反馈移位寄存器

②random 随机方式

通过 rand( )函数设置参数作为种子，调用 rand( )函数，它会依据内部状态返回一个随机数，同时更新内部状态，得到密文序列。

③AES 分组加密算法方式

使用固定的密钥初始化 AES 分组加密算法，初始明文设为全 0，加密后得到的密文作为伪随机数序列输出；将上一次的密文输出作为新一次的明文输入，如此迭代加密产生密文序列。

④线性反馈移位寄存器联合 AES 方式

线性反馈移位寄存器方式产生伪随机数序列，将其作为 AES 分组加密算法的明文输入，加密后产生密文序列。

⑤随机方式 rand 联合 AES 方式

使用 rand 方式产生伪随机数序列，将其作为 AES 分组加密算法的明文输入，加密后产生密文序列。

(2)影射方式

本文约定 32 B、16 B 或 10 B 长密钥的二进制表示中，若 1 仅出现为 0、1、2 或 3 次，则称此密钥为弱密钥。32 B 或 10 B 长初始向量的二进制表示中，若 1 仅出现为 0、1、2 或 3 次，则称此初始向量为弱初始向量。

①弱方式

5 种随机方式中的任一种方式产生 64 bit 伪随机数，将其影射至所有弱密钥和弱初始向量之中，生成伪随机的弱密钥或弱初始向量。

②全域投影方法

5 种随机方式中任一种方式产生64 bit 伪随机数[12]。

2 SPSS 数据分析

SPSS(Statistical Product and Service Solutions)统计产品与服务解决方案是常用统计软件[13]。

2.1 样本数据

NIST 随机性检测标准中 15 种检验方法的失败个数作为因变量共 15 列，增加密码算法种类标识列作为因子。7 种算法、5 种随机方式和 2 种影射方式总计得到 70 行数据。

2.2 单因素方差分析准备工作

单因素方差分析的前提条件主要包括正态性和方差齐性。

理论上多元方差分析要求各因变量服从多元正态分布，但目前常见的统计软件还无法实现多元正态性检验，所以在实际应用中，弱化为考察每一个变量是否服从正态分布[13-15]。

由于样本量超过 50，选用柯尔莫戈洛夫-斯米诺夫检验，每一个变量的正态性检验结果如表 1 所示。在 0.05 的显著性水平下，只有累加和检验不服从正态分布。进一步，通过非参数检验得到累加和检验为单峰对称分布，因此可以认为近似正态分布。

表 1 柯尔莫戈洛夫-斯米诺夫检验

观测变量	统计	自由度	显著性
单比特频数检验	0.076	70	0.200*
块内频数检验	0.058	70	0.200*
累加和检验	0.138	70	0.002
游程检验	0.098	70	0.090
最大游程检验	0.070	70	0.200*
二元矩阵秩检验	0.047	70	0.200*
离散傅里叶变换检验	0.055	70	0.200*
非重叠匹配检验	0.083	70	0.200*
重叠匹配检验	0.047	70	0.200*
全局通用统计检验	0.068	70	0.200*
近似熵检验	0.053	70	0.200*
随机偏移检验	0.075	70	0.200*
随机偏移变量检验	0.061	70	0.200*
序列串行检验	0.106	70	0.051
线性复杂度检验	0.064	70	0.200*

注：* 表示真显著性的下限。

方差齐性检验中，仅单比特频数检验基于平均值的显著性等于 0.046，其余均大于 0.05，如表 2 所示。因此可以认为方差相等，满足单因素方差分析使用的前提要求。

2.3 单因素方差分析

选择密码算法各类标识作为控制变量，15 种检验方法的失败个数作为观测变量。

(1)零假设：控制变量不同水平下观测变量各总体的均值无显著性差异，即 7 种密码算法关于 15 种检验方法得到的失败数构成的 15 维向量的均值相等。

(2)单因素方差分析部分结果如表 3 所示，仅二元矩阵秩检验的显著性等于 0.48，其余均大于 0.05。

表 3 ANOVA 部分结果

		平方和	自由度	均方	F	显著性
单比特频数检验	组间	1 821.486	6	303.581	0.499	0.807
	组内	38 353.500	63	608.786		
	总计	40 174.986	69			
二元矩阵秩检验	组间	11 256.200	6	1 876.033	2.266	0.048
	组内	52 162.600	63	827.978		
	总计	63 418.800	69			

(3)在显著性水平 0.05 下，概率值 P 比 0.05 大，则零假设成立，认为控制变量不同水平下观测变量的总体均值基本相同，控制变量各水平的效应同时为 0，控制变量的不同水平对观测变量的影响并不显著。

3 结论

本文通过 5 种随机方式和 2 种影射方式构造密钥和初始向量，生成对称密码算法 10 类密文，使用 15 种 NIST 随机性检验方法统计失败次数，再利用 SPSS 软件进行单因素方差分析。实验结果表明，7 种对称密码算法在 0.05 显著性水平下的检验结果在统计学上无显著差异。因此，此统计检验可作为评价密码算法好坏的指标之一。

参考文献

[1] 李洪超.基于密文特征的密码算法识别研究[D].西安：西安电子科技大学，2018.
[2] 王瑛，张文科，罗影，等.加密流量检测与态势预警平台研究[J].信息安全与通信保密，2020(2)：98-105.
[3] 吴杨，王韬，邢萌，等.基于密文随机性度量值分布特征的分组密码算法识别方案[J].通信学报，2015，36(4)：146-155.
[4] 吴杨，王韬，李进东.分组密码算法密文的统计检测新方法研究[J].军械工程学院学报，2015(3)：58-64.
[5] 冯秀涛.祖冲之序列密码算法[J].信息安全研究，2016，2(11)：1028-1041.
[6] ZUC 算法研制组. ZUC-256 流密码算法[J].密码学报，

表 2 方差齐性检验部分结果

		莱文统计	自由度 1	自由度 2	显著性
单比特频数检验	基于平均值	2.292	6	63	0.046
	基于中位数	1.919	6	63	0.091
	基于中位数并具有调整后自由度	1.919	6	52.858	0.095
	基于剪除后平均值	2.231	6	63	0.051

(下转第 50 页)

《电子技术应用》http://www.chinaaet.com

[2] SMRITI, CHHAGAN C. Double threshold-based energy detection spectrum sensing scheme by considering the sensing history in confusion region[C]//2018 5th International Conference on SPIN, 2018: 518-521.
[3] DHANANJAYA S, YUVARAJU B N. A novel method in matched filter spectrum sensing to minimize interference from compromised secondary users of cognitive radio networks[C]//2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques(ICEECCOT). IEEE, 2020.
[4] GHOSH D, BAGCHI S. Cyclostationary feature detection based spectrum sensing technique of cognitive radio in nakagami-m fading environment[C]//Computational Intelligence in Data Mining-Volume 2, 2015.
[5] Peng Qihang, Zeng Kun, Wang Jun, et al. A distributed spectrum sensing scheme based on credibility and evidence theory in cognitive radio context[C]//IEEE International Symposium on Personal. IEEE, 2006.
[6] GOHAIN P B, CHAUDHARI S, KOIVUNEN V. Evidence theory based cooperative energy detection under noise uncertainty[C]//GLOBECOM 2017-2017 IEEE Global Communications Conference. IEEE, 2017.
[7] ZENG Y, LIANG Y C. Maximum-minimum eigenvalue detection for cognitive radio[C]//IEEE International Symposium on Personal. IEEE, 2007.
[8] 王颖喜, 卢光跃. 基于最大最小特征值之差的频谱感知技术研究[J]. 电子与信息学报, 2010(11): 2571-2575.
[9] 徐家品, 杨智. 基于随机矩阵特征值比的频谱感知改进算法[J]. 电波科学学报, 2015, 30(2): 282-288.
[10] TULINO A M, VERDÚ S. Random matrix theory and wireless communications[J]. Communications and Information Theory, 2004, 1(1): 1-182.
[11] DEMPSTER A P. Upper and lower probabilities induced by a multivalued mapping[J]. Annals of Mathematical Stats, 1967, 38(2): 325-339.
[12] SHAFER G. A mathematical theory of evidence[J]. Technometrics, 1976, 20(1): 106.
[13] ERYIGIT S, GUR G, BAYHAN S, et al. Energy efficiency is a subtle concept: fundamental trade-offs for cognitive radio networks[J]. IEEE Communications Magazine, 2014, 52(7): 30-36.
[14] 王云川, 许晓荣, 姚英彪, 等. 一种能效优先的认知无线电模仿主用户攻击防御策略设计与性能分析[J]. 电信科学, 2017, 33(8): 100-106.
[15] 张平, 李建武, 冯志勇, 等. 认知无线网络基础理论与关键技术研究[J]. 电信科学, 2014(2): 1-13.

(收稿日期: 2020-11-19)

作者简介:
石新(1996-), 男, 硕士研究生, 主要研究方向: 信息与信号处理、无线通信等。
刘顺兰(1965-), 女, 教授, 主要研究方向: 信息与信号处理、无线通信等。
张无际(1995-), 男, 硕士研究生, 主要研究方向: 信息与信号处理、无线通信等。

扫码下载电子文档

(上接第 45 页)

2018, 5(2): 167-169.
[7] 张静. LTE 核心加密算法 SNOW 3G 的安全分析[J]. 信息通信, 2016(1): 208-209.
[8] 高家奇, 李斌勇, 廖怀凯, 等. 高级加密 AES 算法研究及性能分析[J]. 网络安全技术与应用, 2019, 226(10): 31-33.
[9] 姚思, 陈杰. SM4 算法的一种新型白盒实现[J]. 密码学报, 2020, 7(3): 358-374.
[10] 谢端强, 李恒, 李瑞林, 等. 对 Sosemanuk 算法改进的猜测决定攻击[J]. 国防科技大学学报, 2012, 34(6): 79-83.
[11] 刘鹏焜, 陈恭亮, 李建华. Trivium 算法在随机访问条件下的应用研究[J]. 通信技术, 2017, 50(1): 133-139.
[12] 王超, 温涛, 段冉阳. NIST 随机性检测方法研究[J]. 信息技术与网络安全, 2018, 37(11): 5-8, 15.
[13] 邹祎. SPSS 软件单因素方差分析的应用[J]. 价值工程, 2016, 35(34): 219-222.
[14] 王超, 范国浩, 付宝仁. ZUC 算法随机性检测研究[J]. 信息技术与网络安全, 2018, 37(11): 9-11, 15.
[15] Fan Yutao, Su Guiping. A new testing method of randomness for true random sequences[C]//2014 IEEE 5th International Conference on Software Engineering and Service Science, Beijing, 2014: 537-540.

(收稿日期: 2021-01-25)

作者简介:
朱玉倩(1994-), 通信作者, 女, 硕士研究生, 主要研究方向: 信息安全, E-mail: zhuyuqian18@mails.ucas.ac.cn。
王超(1982-), 男, 博士, 高级工程师, 主要研究方向: 密码学、统计学。
张艳(1990-), 女, 本科, 助理工程师, 主要研究方向: 信息安全。

扫码下载电子文档

版权声明