RC4 Cryptographic Sequence on Variant Maps

Zhonghao Yang and Jeffrey Zheng

Abstract In modern cyberspace environment, big data streams are the most important issue in people's daily lives, each person produces a larger number of data streams every day from personal computer, cell phone, and kinds of wearable smart device. Security risks of storage and transmission of data streams may lead to personal privacy disclosure, it is important for network security to have useful tools facing challenges. Randomness testing provides useful tools to secure results of stream ciphers. Based on multiple statistical probability distributions, this chapter presents a visual scheme, variant maps, to measure a whole cryptographic sequence into multiple 1D and 2D maps. Mapping mechanism and sample cases are provided.

Keywords Random sequence · Big data · Variant map

1 Introduction

In modern cyberspace environments, more than 2.5 EB data streams per day are generated from global network environments [1]. Huge network companies managed massive data streams in PB every day [2]. The development of artificial intelligence fields makes it easier to extract valuable information from big data [3–5]. Big data

This work was supported by the Key Project on Electric Information and Next Generation IT Technology of Yunnan (2018ZI002), NSF of China (61362014) and Yunnan Advanced Overseas Scholar Project.

Z. Yang
Yunnan University, Kunming, China
e-mail: houseashley07@hotmail.com

J. Zheng (✉)
Key Laboratory of Quantum Information of Yunnan, Yunnan University, Kunming, China
e-mail: conjugatelogic@yahoo.com

J. Zheng
Key Laboratory of Software Engineering of Yunnan, Yunnan University, Kunming, China

J. Zheng (ed.), *Variant Construction from Theoretical Foundation to Applications*,
https://doi.org/10.1007/978-981-13-2282-2_19

and big data technology provide modern societies so much convenience to many places, and with several threats to network security [6, 7].

Stream ciphers are the most useful scheme to protect the security of data streams in both transmission and storage processes. Pseudorandom number sequences are generated by various algorithms based on recursive computational models, and true random number sequences are generated by different physical methods. The typical stream ciphers are RC4 and Salsa20. Stream ciphers can be built using block ciphers in OFB or CTR model. In this chapter, an RC4 stream cipher is selected to generate pseudorandom sequences for testing.

From a testing viewpoint, randomness tests focus on three aspects: probability, autocorrelation, and unpredictability. NIST 800-22 provides a list of randomness testing method based on p-value [8].

In this chapter, two types of 1D and 2D statistical probability maps are used to visualize a longer pseudorandom number sequence generated from an RC4 stream cipher.

2 Related Work

Variant map is an emerging technology proposed in 2010s to handle multiple 0–1 vectors in phase spaces on variant framework [9–11]. Different applications are explored for variant maps on ECG data sequences [12], bat echolocation call sequences [13], gene sequence [14], and cryptographic sequences [15–17].

3 Mapping Model

This chapter uses two mapping schemes on 1D and 2D statistical probability distributions as variant maps for an input N-length 0–1 sequence. The architectural diagram of the mapping model is shown in Fig. 1. It is composed of three components: segmentation, measurement, and visualization.

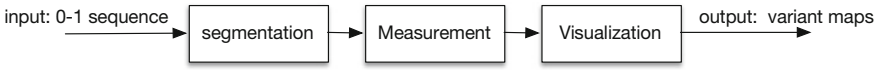

Fig. 1 Architecture of variant map for cryptographic sequence

Fig. 2 Measurement

3.1 Basic Symbol

(1) S: an input 0–1 sequence,
(2) s_i: the i-th segment of the input sequence,
(3) N: length of the input sequence,
(4) M: count of segments,
(5) m: length of a segment, and
(6) p: number of 1's elements in the segment.

3.2 Mapping Model

Three components can be described as follows.

- Segmentation

Input data is a 0–1 sequence S of length N. It can be divided into M segments and each segment has m elements.

$$\mathrm{M} = \left\lfloor \frac{N}{m} \right\rfloor$$

$$S = \{s_0, s_1, \ldots, s_i, \ldots, s_{M-1}\}, \quad 0 \leq i < M$$

- Measurement
  For each segment s_i of S, the following analysis is performed to obtain the one feature p_i of the segment, that is, the number of 1 of s_i, and $0 \leq p \leq m$. For example, for two segments $s_1 = 00011$ and $s_2 = 10110$, and two measurements are $p_1 = 2$ and $p_2 = 3$ (Fig. 2).

Calculating all segments of S, a set of p measurements are determined.

$$\{p_0, \ldots, p_i, \ldots, p_{M-1}\} = \{p_i\}_{i=0}^{M-1}, \quad 0 \leq i < M$$

- Visualization

From the generated sequence of measurements, two types of diagrams can be created: The first one is a 1D map, 1DP sorted from $\{p_i\}_{i=0}^{M-1}$ directly shown in Fig. 3a. The second one is a 2D map, 2DP sorted from a pair of measurements $\{p_i, p_{i+1}\}_{i=0}^{M-1}$

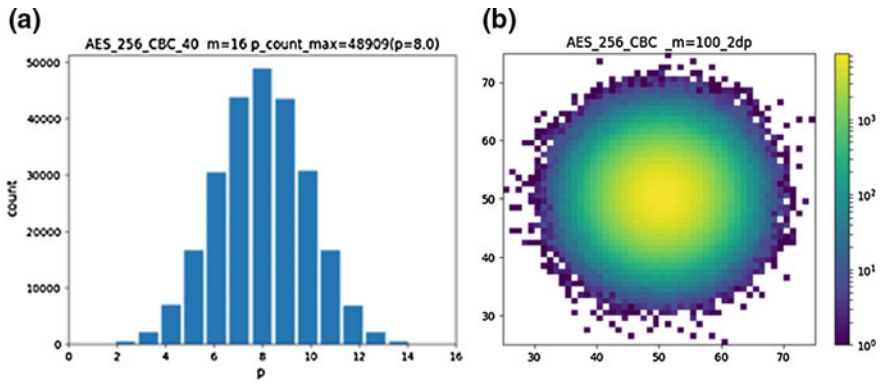

Fig. 3 Two maps; **a** 1DP; **b** 2DP

created from $\{p_i\}_{i=0}^{M-1}$ shown in Fig. 3b. This mapping scheme is one of Markov chain models.

4 Random Sequence Data Sources

In this chapter, a pseudorandom generator is based on an AES block cipher on the OFB mode. A total amount of 120 MB cryptographic sequences has been generated.

5 Mapping Results

The input sequence is mapped with a list of various lengths on different segmentations. Three sets of various m lengths are selected and two types of relevant 1DP and 2DP maps are shown in Fig. 4a–c, for (a) $m = \{8, 16, 32, 64, 128, 256\}$, (b) $m = \{80, 100, 120, 140, 160\}$, and (c) $m = \{126, 127, 128, 129, 130\}$. Four enlarged 2DP maps are shown in Fig. 5 for $m = \{126, 127, 128, 129\}$ and two enlarger 2DP maps are shown in Fig. 6 for $m = \{128, 130\}$, respectively.

6 Result Analysis

In Fig. 4, both 1DP and 2DP maps are illustrated. When the input sequence is larger enough to $m \times 2^m$, the results of 1DP maps are corresponding to binomial distributions. It is interesting to see significant changes when various lengths of segments are applied.

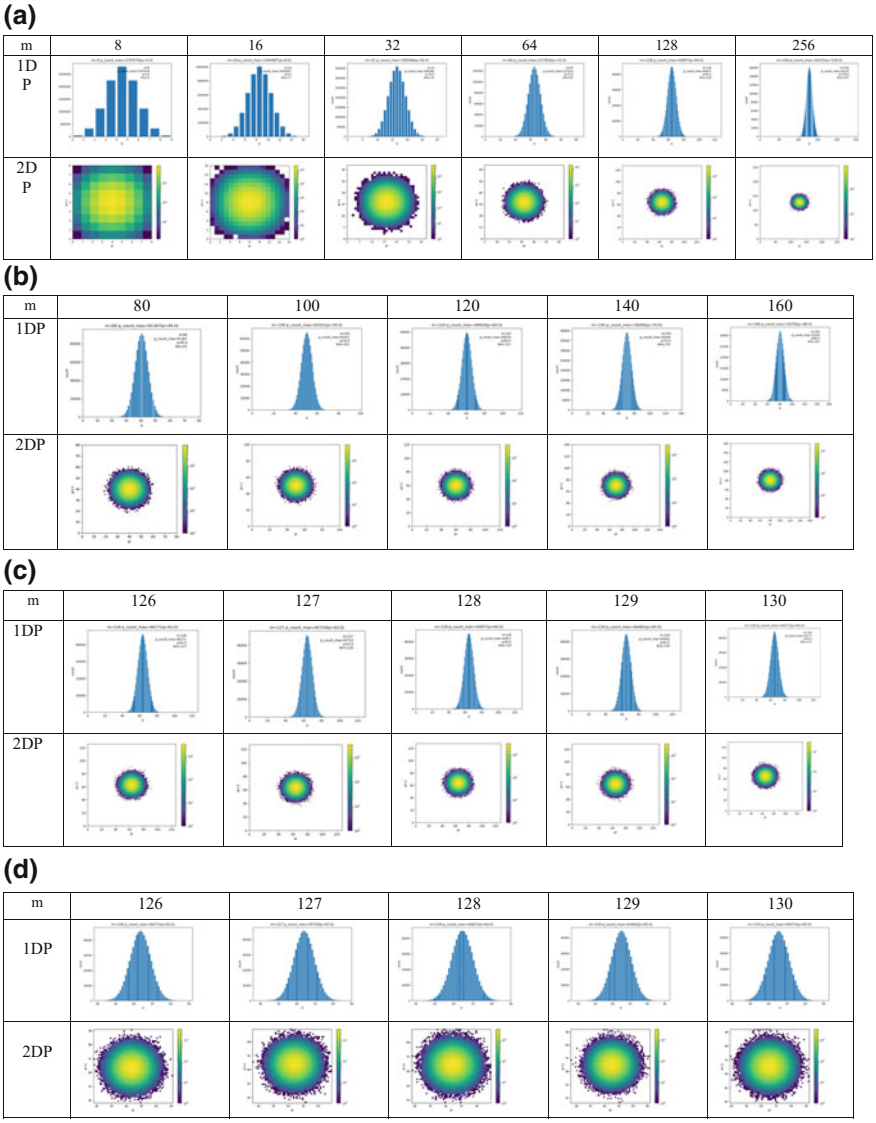

Fig. 4 1DP and 2DP maps. **a** $m = \{8, 16, 32, 64, 128, 256\}$; **b** $m = \{80, 100, 120, 140, 160\}$; **c** $m = \{126, 127, 128, 129, 130\}$; **d** enlarged 1dp and 2dp, $m = \{126, 127, 128, 129, 130\}$

For various 2DP maps in Figs. 4, 5, and 6, 2D distributions are represented as pseudocolor to illustrate relevant 3D structures. From smaller maps to enlarged maps,

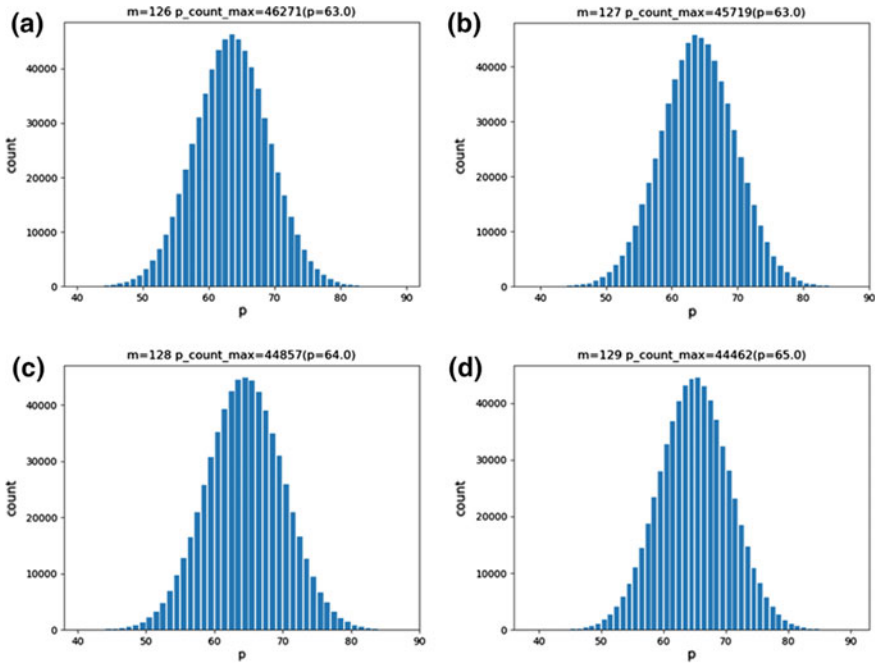

Fig. 5 Enlarger 1DP maps. **a** $m = 126$; **b** $m = 127$; **c** $m = 128$; **d** $m = 129$

many interesting features can be identified and significant symmetric or nonsymmetric properties could be identified. Enlarger maps can see further refined patterns in detail.

7 Conclusion

Mapping model in this chapter is a focus on a single sequence for two types of 1DP and 2DP maps. 1DP maps are corresponding to classical statistical maps and 2DP maps are represented as various Markov chains. Further researches and experiments are required to explore useful tools on cryptographic sequences in detail (Figs. 7 and 8).

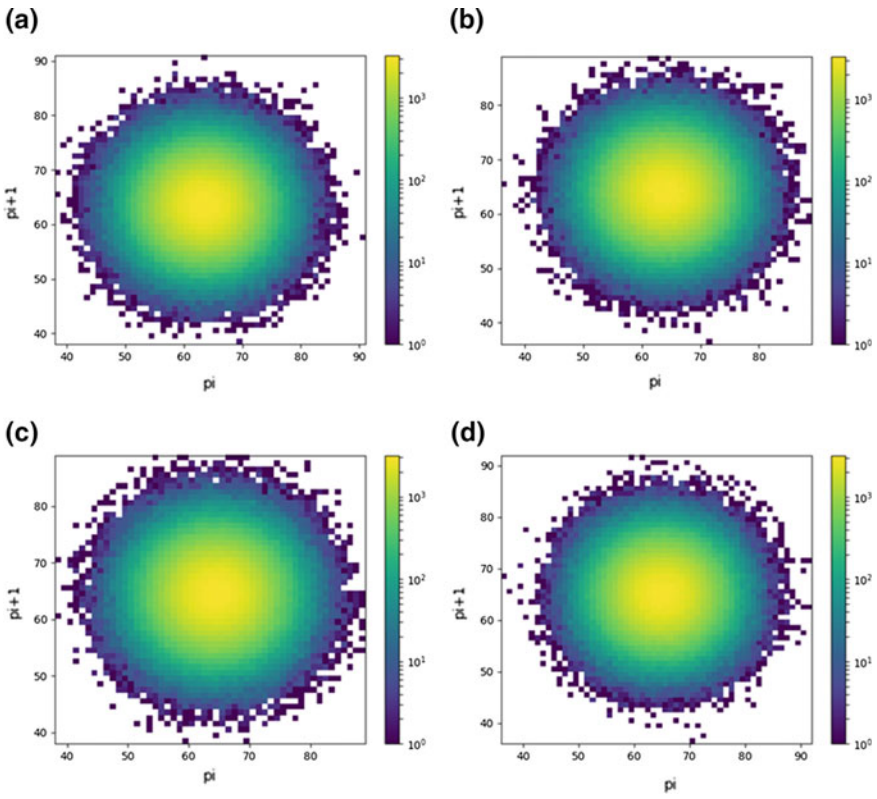

Fig. 6 Enlarged 2DP maps. **a** $m = 126$; **b** $m = 127$; **c** $m = 128$; **d** $m = 129$

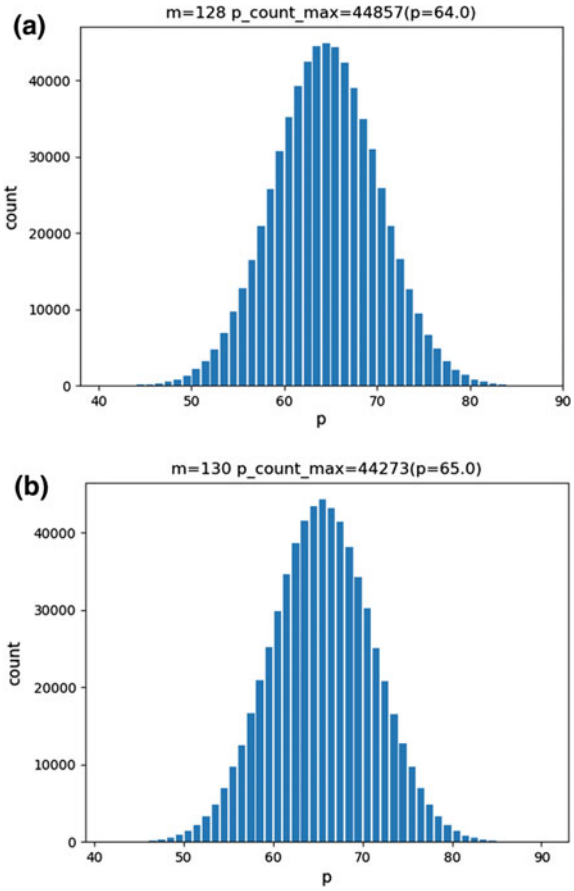

Fig. 7 Enlarger 1DP maps. **a** $m = 128$; **b** $m = 130$

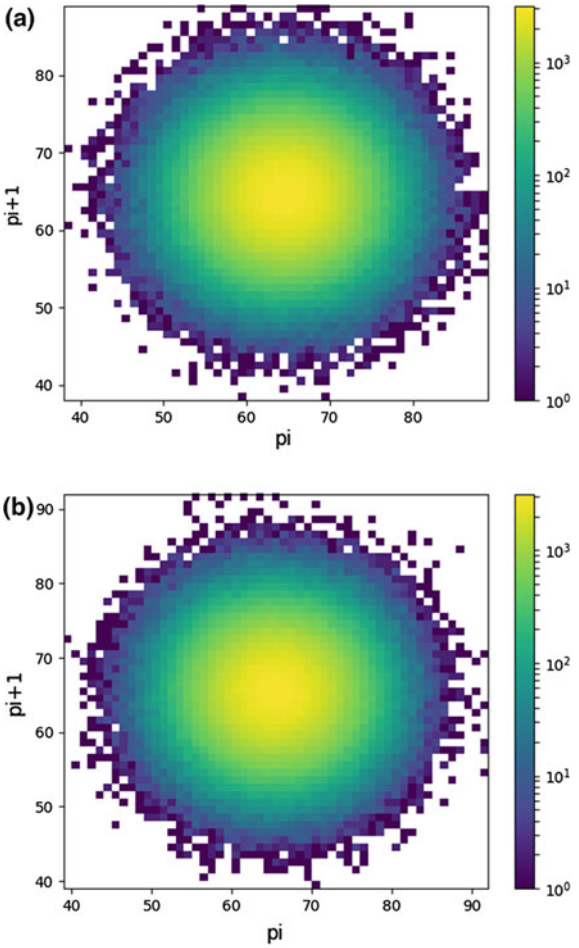

Fig. 8 Enlarger 2DP maps. **a** $m = 128$; **b** $m = 130$

References

1. A. Anwaar et al., Big data for development: applications and techniques. Big Data Analytics **1**(1), 2 (2016)
2. V. Mayer-Schönberger, K. Cukier, *Big Data: a revolution that will transform how we live, work, and think* (Eamon Dolan/Houghton Mifflin Harcourt, 2013)
3. M.M. Najafabadi et al., Deep learning applications and challenges in big data analytics. Journal of Big Data **2**(1), 1 (2015)
4. R. Fang, S. Pouyanfar, Y. Yang et al., Computational health informatics in the big data age: a survey. ACM Comput. Surv. **49**(1), 12 (2016)
5. M.D. Assunção et al., Big data computing and clouds: trends and future directions. J. Parallel Distrib. Comput. **79–80**, 3–15 (2015)
6. L. Xu, C. Jiang, J. Wang et al., Information security in big data: privacy and data mining. IEEE Access **2**, 1149–1176 (2014)
7. L. Lerman, G. Bontempi, O. Markowitch, A machine learning approach against a masked AES. J Cryptographic Eng **5**(2), 123–139 (2015)
8. L.E. Bassham III et al., *SP 800-22 Rev. 1a. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications* (Nist Special Publication 2010)
9. J.Z.J. Zheng, C.H. Zheng, A framework to express variant and invariant functional spaces for binary logic. Front. Electr. Electron. Eng. China **5**, 163 (2010)
10. J.Z.J. Zheng, C.H.H. Zheng, T.L. Kunii, *A Framework of Variant Logic Construction for Cellular Automata* (InTech, Shanghai, 2011)
11. J. Zheng, C. Zheng, Variant measures and visualized statistical distributions. Acta Photonica Sinica **40**, 1397 (2011)
12. Y. Ji et al., Variant maps on normal and abnormal ECG data sequences. Biol. Med. **8**(6), 1 (2016)
13. D.M. Heim, O. Heim, P.A. Zeng, J. Zheng, Successful creation of regular patterns in variant maps from bat echolocation calls. Biol. Syst. Open Access **5**, 166 (2016)
14. J. Zheng, W. Zhang, J. Luo, W. Zhou, V. Liesaputra, Variant map construction to detect symmetric properties of genomes on 2D distributions. J. Data Min. Genomics Proteomics **5**, 1 (2014)
15. J. Zheng, J. Luo, J. Zhou, Pseudo DNA sequence generation of non-coding distributions using variant maps on cellular automata. Appl. Math. **5**, 153 (2014)
16. W.Z. Yang, J. Zheng, Variant Pseudo-Random Number Generator. Hakin9 Extra **6**, 28–31 (2012). http://hakin9.org/hakin9-extra-62012/.
17. J. Zheng et al., Variant map system to simulate complex properties of DNA interactions using binary sequences. Adv. Pure Math. **5**(7A), 5–24 (2013)