The Technique for Testing Short Sequences as a Component of Cryptography on the Internet of Things

Svitlana Popereshnyak [0000-0002-0531-9809]

Taras Shevchenko National University of Kyiv, 24, Bohdana Havrylyshyna str., Kyiv, 04116, Ukraine
spopereshnyak@gmail.com

Abstract. An article dedicated to on topical issues related to Internet of Things security. IoT data protection solutions must span edge to cloud, provide scalable encryption and key management, and not impede data analysis. The available approaches to testing random or pseudorandom sequences show low flexibility and versatility in the means of finding hidden patterns in the data. It is revealed that for sequences of length up to 100 bits there are not enough existing statistical packets. The classification of the main problems of information security in Internet of Things is given. The complexity of using classical cryptographic algorithms for information security in Internet of Things is considered. The paper proposed a methodology for testing pseudorandom sequences, obtained an explicit form of the joint distribution of numbers of 2-chains and numbers of 3-chains of various options random bit sequence of a given small length. Examples, tables, diagrams that can be used to test for randomness of the location of zeros and ones in the bit section are presented. As a result of the implementation of this technique, an information system will be created that will allow analyzing the pseudorandom sequence of a small length and choosing a quality pseudorandom sequence for use in a particular subject area.

Keywords: Internet of Things, Algorithms, multidimensional Statistics, Random Sequence, s-chains, Cryptography, Pseudorandom Sequence, Statistical Testing.

1 Introduction

Organizations have only just begun discovering and benefiting from the opportunities provided by the Internet of Things (IoT). The ability to capture and analyze data from distributed connected devices offers the potential to optimize processes, create new revenue streams, and improve customer service. However, the IoT also exposes organizations to new security vulnerabilities introduced by increased network connectivity and devices that are not secured by design. And advanced attackers have demonstrated the ability to pivot to other systems by leveraging vulnerabilities in IoT devices.

IoT devices collect some volumes of data, some of which will require protection based on sensitivity or compliance requirements. IoT data protection solutions must

span edge to cloud, provide scalable encryption and key management, and not impede data analysis. The available approaches to testing random or pseudorandom sequences show low flexibility and versatility in the means of finding hidden patterns in the data.

Random sequences have found the widest application from the gaming computer industry to mathematical modeling and cryptology.

We list some areas of their usage:

1. Modeling. In computer simulation of physical phenomena. In addition, mathematical modeling uses random numbers as one of the tools of numerical analysis.
2. Cryptography and information security. Random numbers can be used to test the correctness or effectiveness of algorithms and programs. Many algorithms use the generation of pseudo-random numbers to solve applied problems (for example, cryptographic encryption algorithms, the generation of unique identifiers, etc.).
3. Decision making in automated expert systems. The use of random numbers is part of decision-making strategies. For example, for the impartiality of the choice of examination paper by a student in an exam. Randomness is also used in the theory of matrix games.
4. Optimization of functional dependencies. Some mathematical optimization methods use stochastic methods to search for extremums of functions.
5. Fun and games. Accident in games has a significant role. In computer or board games, chance helps to diversify the gameplay.

There are various approaches to the formal definition of the term “randomness” based on the concepts of computability and algorithmic complexity [1].

By implementing some algorithm, software generators produce numbers (although not obvious) depending on the set of previous values, so the received numerical sequences are not truly random and are called pseudo-random sequences (PRS). At the moment, more than a thousand software PRS generators are known, which differ in algorithms and values of parameters. Statistical properties are significantly different from the number sequences that are generated by them.

The presented and not presented results allow us to characterize the state of modern technologies of designing the PRS (focusing on the most progressive of them by the following basic provisions [1-13].

2 Features of Information Protection in the Internet of Things

“Things” today are not only personal items of ordinary consumers, but also various equipment that is actively used in many fields of activity – trade, transport, medicine, construction, banking, sports, etc. It follows that the Internet of things is most often heterogeneous network, i.e. devices of various classes and types are combined and interact with each other.

Recommendations for protecting information in the Internet of things are aimed at improving the security of devices, networks and data.

First of all, IoT devices, as a rule, due to their portability and mobility, are physically accessible to cybercriminals, and can be stolen to gain access to confidential data and establish communication with other network devices. To prevent this threat,

it is necessary to provide physical protection, for example, by using protective covers on devices or cases that provide for restrictions on direct access to devices. In addition to direct access, devices can provide remote access to update configuration data or software. To protect against this, it is necessary to provide that the software ports to be closed and apply strong passwords at the level of downloading and updating firmware, which will prevent access to the device if it is compromised.

At the same time, on the other hand, many IoT devices are becoming vulnerable to cyberattacks because their software is not updated in a timely manner. To minimize such risks, it is recommended to implement an automatic update by default, because, even if software updates are released on time, consumers do not always install them manually immediately after the release.

Attention should also be paid to the organization of data storage on the devices themselves, because often this information is related to the user's personal data, financial transaction data and data on critical objects of various fields of activity.

Safety must be ensured both throughout the entire period of the product's functioning and after its decommissioning. Cryptographic keys must be stored in non-volatile memory of the device in not open form. In addition, disposal of decommissioned devices may be envisaged.

To protect networks, first of all, methods of "strong authentication" should be provided, including, for example, two-factor authentication, assignment of "hard" specified unique identification and authentication data, as well as the use of modern secure protocols [14]. Cryptographic algorithms must be adapted to the Internet of things.

In order to minimize the risks of denial of service attacks against devices, it is recommended to provide bandwidth limits for the network of Internet of things devices, both at the software and hardware levels. In case of detection of suspicious traffic, devices should provide the ability to signal about that with the subsequent analysis of the identified threat.

Data protection is primarily ensured through the use of cryptographic methods adapted to the features of devices with limited opportunities. If the device is compromised, it should be possible to urgently erase key information used in cryptographic operations.

Devices of the Internet of things should transmit and process only the information that is necessary for the implementation of their main functions - as a rule, this is the collection of information about the state of their environment or about the user. It follows that it is necessary to pay attention to the information circulating in the network, minimizing the risk of leakage of confidential information.

In addition to the heterogeneity of networks, a feature of the Internet of things is also that the devices have different computing resources, bandwidth and support different technologies and protocols. The lack of common standards and protocols remains a serious problem in building a network of "things". Also, many "things" have limited power capabilities and must support energy-saving modes.

The listed features of the Internet of things impose restrictions when building a security system in such a network. The usual methods of protecting information in wireless networks may not be enough, or they may not be applicable due to the restrictions imposed by the Internet of things.

The main methods for ensuring security, as in traditional networks, remain encryption, identification / authentication, and the implantation of physical security measures.

The security system should be designed to provide protection for devices and gateways, the transmission network, as well as applications that are deployed to ensure the functioning of the devices.

Encryption is a widely used, effective and quite flexible solution for ensuring in the confidentiality of information and for creating a security system. However, any encryption, and especially strong one, requires an increase in productivity and additional computing resources, which is not always possible in the conditions of the Internet of things.

As for authentication, the researchers proposed a fairly large number of approaches that could be implemented to solve security problems [14, 15]. One common method is two-factor authentication. For example, one-time password authentication (OTP). With this approach, after providing the credentials, the user or device must also present a one-time password generated by the key distribution center, thereby confirming its authenticity. This method does not require additional computing resources or storage from the devices, but it is not applicable for devices that, for example, simply cannot support the ability to enter the received one-time password. The same problem is relevant for the authentication method, the second factor of which is the hardware identifier.

Other studies suggest using the concept of "digital memories" for authentication, which would solve the problem of users remembering complex passwords. However, this method imposes resource limits on devices.

The proposed methods also include authentication using cryptography based on elliptic curves. Despite the fact that in this case the necessary basic parameters of elliptic curves are not calculated by the devices themselves, after the calculation, a sufficiently large amount of data must be transferred, which may be limited by the network bandwidth [14].

Thus, the various existing authentication methods are applicable to a single network and to a separate class of devices. The application of uniform methods and means is complicated by the lack of standardization and heterogeneity of such networks.

3 Problem Statement

Before responsible using in mathematical modeling and cryptology, PRS should be tested. Unfortunately, for many PRS tests, there are some limitations:

- checked out only one of the probable ones properties that are characterize PRS;
- not fix family alternatives;
- do not have theoretical ones ratings power.
- do not give a correct an estimate of chance sequences provided a little sample.

Problems small and large samples refer to the main problems that arise in practical application methods analysis data. Let's be use the next classification samples by number [16], based on requirements presented in the program criteria:

- very small sampling - from 5 to 12,
- small sampling - from 13 to 40,
- sampling average the number - from 41 to 100,
- big ones sampling - from 101 and more.

The minimum size of the sample limits not so much the algorithm of calculating the criterion, but the distribution of its statistics. For a row algorithms with too much small ones numbers sample normal approximation distribution of statistics criterion will be under question.

During the research, the localization of the local sections of the bit sequence was conducted to detect the dependencies in the location of its elements by using the exact distributions of the corresponding statistics. In the work an explicit form of the joint distribution of the numbers of 2-chains and numbers of 3-chains of various variants in a random sequence was obtained. This joint distribution allows more accurate comparison of the use of one-dimensional statistics, to analyze the bit sequence small length by chance.

4 Joint Distribution of number of 2-chains and number of 3-chains of a provided type in binary sequence

Consider a sequence of random variables

$$\gamma_1, \gamma_2, \dots, \gamma_n, \tag{1}$$

where $\gamma_i = \{0, 1\}$, $i = 1, 2, \dots, n$, $n > 0$.

Subsequences $\gamma_j, \gamma_{j+1}, \dots, \gamma_{j+s-1}$, sequences (1) are called s-chains, $j = 1, 2, \dots, n - s + 1, s = 1, 2, \dots, n$.

Denote $\eta(t_1, t_2, \dots, t_s)$ the number of s-chains in the sequence (1) that coincide with $t_1, t_2, \dots, t_s$, where $t_i = \{0, 1\}$, $i = 1, 2, \dots, s$.

Theorem. Let sequence (1) consist of n, $n > 0$ independent identically distributed random variables; $P\{\gamma_i = 1\} = p$, $P\{\gamma_i = 0\} = q$, $p + q = 1$, $i = 1, 2, \dots, n$ and $k_1, k_2, k_3, t,$ – integer numbers such that $k_1 \geq 0, k_2 \geq 0, n \geq k_1, k_3 \geq 0, t, t_1 \epsilon \{0, 1\}$. Then

$$P\{\eta(t_1^* t_1) = k_1, \eta(t^* 0 t^*) + \eta(t^* 1 t^*) = k_2\} = \sum_{m_1=k_1}^{m-k_1} p^{m_1} q^{m_0} \chi(m_t, k_1, k_2) \tag{2}$$

$$\chi(m_t, k_1, k_2) = \begin{cases} 1, if\ m_t = k_1 = k_2 = 0, \\ \psi(m_t, k_1, k_2), \text{elsewhere} \end{cases},$$

$$\psi(m_t, k_1, k_2) = \sum \sum_{i \in \{k_1, k_1+1\}} C_{i-1}^{\delta_{t^*}} C_i^{\delta_t - m_t + 2i} Z(m_t - i;\ m_t - i - \delta_t) C_{m_{t^*} - i + 1}^{k_1 - \delta_{t^*}},$$

where is the symbol $\sum$ denotes addition over all non-negative integers δ_t and δ_{t^*} such that $\delta_t + \delta_{t^*} = k_2$,

$$\mathrm{Z}(a,b) \stackrel{\text{def}}{=} \begin{cases} C_{a-1}^{b-1}, & \text{if } a \geq b \geq 1; \\ 1, & \text{if } a = b = 0; \\ 0, & \text{elsewhere.} \end{cases}$$

$$\mathrm{P}\{\eta(t_1^* t_1) = k_1, \eta(t^* t^* t^*) = k_2\} = \sum_{m_1=k_1}^{m-k_1} p^{m_1} q^{m_0} C_{m_{t^*}}^{k_1} \varphi(m_t, k_1, k_2), \qquad (3)$$

$$\varphi(m_t, k_1, k_2) = \sum_{i \in \{k_1,\, k_1+1\}} C_i^{m_t - k_2 - i}\, \mathrm{Z}(m_t - i, m_t - i - k_2),$$

$$\mathrm{P}\{\eta(t_1^*\, t_1) = k_1, \eta(t^* t t^*) = k_2\} = \sum_{m_1=k_1}^{m-k_1} p^{m_1} q^{m_0}\, \phi(m_t, k_1, k_2) \qquad (4)$$

$$\phi(m_t, k_1, k_2) = \begin{cases} 1, & if\ \ m_t = k_1 = k_2 = 0, \\ \sum_{i \in \{k_1,\, k_1-1\}} C_i^{k_2}\, \mathrm{Z}(m_t;\ i+1) C_{m_{t^*}-i}^{k_1-k_2}, & \text{elsewhere} \end{cases}.$$

5 Results and Discussion

As a result of applying this technique for testing pseudo-random sequences for two-dimensional statistics (relations (2) - (4)), you can build a bubble diagram with which you can get the probability of the distribution of zeros and ones in a given sequence.

Consider examples of bubble diagrams for a bit sequence of small length *n*, *n* = 16.

5.1 Graphic Illustration of the Use of Equality (2)

Fig. 1 gives a bubble chart in which the first parameter (horizontal axis) is the value k_1, the second parameter (vertical axis) is the value k_2 , and the third parameter (the bubble size) is the probability of the event occurring $\{\eta(t_1^* t_1) = k_1, \eta(t^* 0 t^*) + \eta(t^* 1 t^*) = k_2\}$, presented in percent.

After analyzing Fig. **1** it can be concluded that for the analysis of the sequence of chains of small and medium length (from 13 to 100 elements), one-dimensional statistics do not always give the correct result. For example, if we consider the sequence where the parameter k_1 = 4, then we can draw a conclusion with a high degree of probability of randomness of the sequence with these characteristics, however, if we pay attention when k_1 = 4 and k_2 = 0 it can be argued that this sequence is non-random, therefore as shown in Fig. 1 we have $\mathrm{P}\{\eta(t_1^* t_1) = k_1, \eta(t^* 0 t^*) + \eta(t^* 1 t^*) = k_2\} = 1{,}30\%$. What also shows the lack of use of one-dimensional statistics for the analysis of short and medium bit sequences.

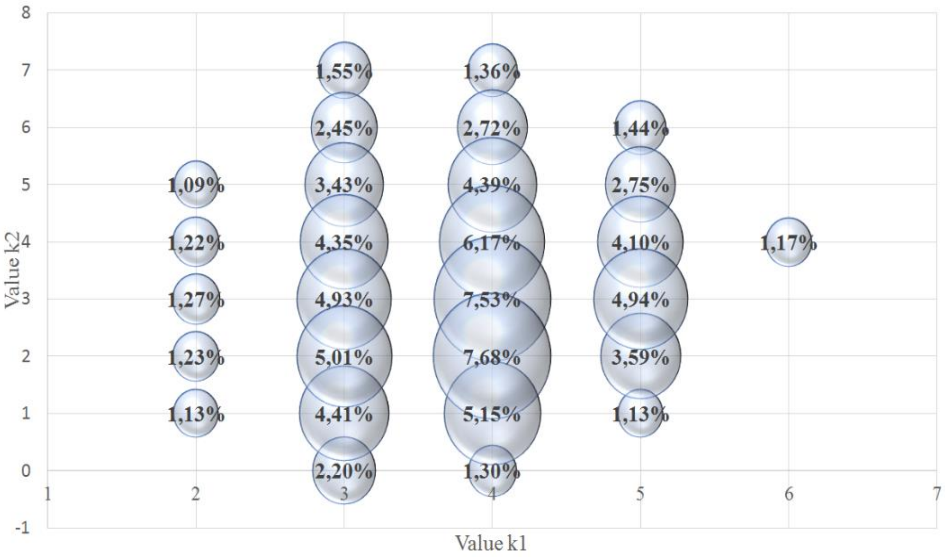

Fig. 1. Bubble chart of sequence with the length 13 for (2).

An approach to testing using n-dimensional statistics allows us to rely on a deeper justification of the randomness of generated sequences.

5.2 Graphic Illustration of the Use of Equality (3)

In Fig. **2** shows the use of the relation (3) for a small sample. n, $n = 16$, and some values k_1 and k_2.

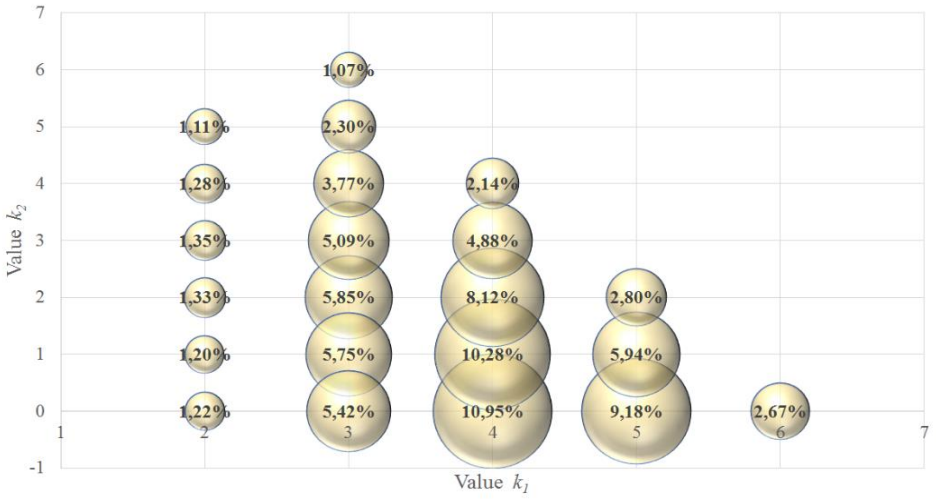

Fig. 2. Bubble chart of sequence with the length 16 for formula (3).

Fig. **2** gives a bubble chart in which the first parameter (horizontal axis) is the value k_1, the second parameter (vertical axis) is the value k_2, and the third parameter (bubble size) is the probability of the event occurring $\{\eta(\, t_1^* t_1) = k_1, \eta(\, t^* t^* t^*) = k_2\,\}$, which is represented as a percentage.

5.3 Graphic Illustration of the Use of Equality (4)

In Fig. 3 shows the use of relation (4) for a small sample n, $n = 16$, and some values k_1 and k_2.

Fig. 3 gives a bubble chart in which the first parameter (horizontal axis) is the value k_1, the second parameter (vertical axis) is the value k_2, and the third parameter (bubble size) is the probability of the event occurring $\{\eta(t_1^*\, t_1) = k_1, \eta(\, t^* t t^*) = k_2\,\}$, which is represented as a percentage.

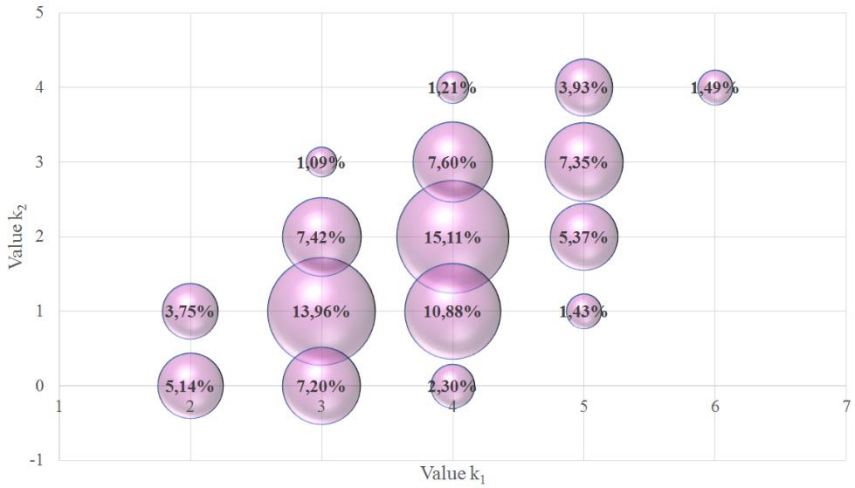

Fig. 3. Bubble chart of sequence with the length 16 for formula (4).

In this paper, the exact compatible distributions of some statistics (0, 1) -sequences of length $1 < n < \infty$ are given. For a bit sequence of small length *n*, *n* = 16, the tables containing the numerical values of the corresponding distribution are given. These tables, as well as the proposed graphic representations, can be used to test the hypothesis of the randomness of the arrangement of zeros and units.

6 The Results of the Comparison the NIST Statistical Test Suite and Test of PRS of Small Length using Multidimensional Statistics

Consider the well-known examples that are given in [17, 18]. Let us analyze the submitted sequences for the corresponding tests, where:

- P is the probability of sequence randomness according to the selected criterion from the first column,
- P_1 is the probability obtained using relation (2),
- P_2 is the probability obtained using relation (3),
- P_3 is this is the probability obtained using relation (4).

Table 1. The results of the comparison.

Test	**Input Size Recommendation**	**length**	**Sequences**	**P**	**P_1**	**P_2**	**P_3**
Frequency (Monobit) Test	n>=100	10	1011010101	0,527	0,021	0,049	0,021
Frequency Test within a Block	n>=100	10	0110011010	0,801	0,097	0,212	0,129
Runs test	n>=100	10	1001101011	0,147	0,097	0,212	0,129
Binary Matrix Rank Test	n>= 38000	N=20 M = Q = 3	01011001001 010101101	0,741	0,112	0,289	0,245
Discrete Fourier Transform (Spectral) Test	n>=1000	N=10	0001010011	0,109	0,106	0,212	0,129
Non-overlapping Template Matching Test	N >= 200	N=20, 2 blocks of length 10	10100100101 110010110	0,344	0,098	0,176	0,105
Maurer's "Universal Statistical" Test	n>= 380000	N=20	01011010011 101010111	0.767	0,112	0,289	0,245
Serial test	n>=100	N=10	0011011101	0,907	0,025	0,212	0,028
Approximate Entropy test	n>=100	N=10	0100110101	0,261	0,021	0,049	0,021
Cumulative Sums (Cusum) Test	n>=100	N=10	1011010111	0,411	0,097	0,212	0,129

Test	Input Size Recommendation	length	Sequences	P	P_1	P_2	P_3
Random Excursions Test	$n>10^6$	N=10	0110110101	0,502	0,003	0,049	0,003
Random Excursions Variant Test	$n>10^6$	N=10	0110110101	0,683	0,003	0,049	0,003

As can be seen from the table, the use of two-dimensional statics gives a more accurate result for short sequences. And also, according to [16], the recommended minimum sequence length n is greater than 100 bits.

7 Conclusions

An analysis of the effectiveness of pseudorandom sequence generators is an urgent issue of cybersecurity in the use of more advanced methods of encryption and information security. The available techniques show low flexibility and versatility in the means of finding hidden patterns in the data. To solve this problem, it is suggested to use algorithms based on multidimensional statistics.

The approach to testing using multidimensional statistics allows you to rely on a deeper justification of the randomness of the generated sequences. This area is promising for scientific research.

The paper proposed a methodology for testing a sequence and obtained a correct view of the joint distribution of the numbers of 2-chains and the numbers of 3-chains of various variants in a random bit sequence of a given small length.

These algorithms and scheme of work for verification statistical tests of randomness sequences (proposed in chapter II) combine all the advantages of statistical methods and are the only alternative for the analysis of sequences of small and medium length.

To implement the proposed approach, a PRS software test package is being developed, which will include tests using multidimensional statistics, which are well recommended for testing a small length PRS. The complex is based on software products developed in C++, Python, for analyzing PRS. Currently, more than 20 PRS tests have been implemented, and the test database is being updated.

As a result of the implementation of this technique, an information system will be created that will allow analyzing the PRS of a small length and choosing a quality PRS for use in a particular subject area.

References

1. Popereshnyak, S.: Analysis of pseudorandom small sequences using multidimensional statistics. In: The 3rd IEEE International Conference on Advanced Information and Communication Technologies (AICT), pp. 5.4.1-5.4.4. IEEE Press, Ukraine (2019).

2. Sulaiman, S., Muda, Z., Juremi, J.: The new approach of Rijndael key schedule. In: The 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), pp. 23-27. IEEE Press, Kuala Lumpur (2012).
3. McLoone, M., McCanny, J. V.: High-performance FPGA implementation of DES using a novel method for implementing the key schedule. In: IEE Proceedings - Circuits, Devices and Systems, **150(5)**. pp. 373-378. IEEE Press (2003).
4. Nejad, F. H., Sabah, S., Jam, A. J.: Analysis of avalanche effect on advance encryption standard by using dynamic S-Box depends on rounds keys. In: The 2014 International Conference on Computational Science and Technology (ICCST), pp. 1-5. IEEE Press, Kota Kinabalu (2014).
5. Liu, H., Jin, C. Lower bounds of differential and linear active S-boxes for 3D-like structure. The Computer Journal, **58(4)**, 904-921 (2015).
6. Bhaskar, C. U., Rupa, C.: An advanced symmetric block cipher based on chaotic systems. In: The 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1-4. IEEE Press, Vellore (2017).
7. Ferguson, N., Schneier, B.: Practical Cryptography, John Wiley & Sons (2003).
8. Menezes, A.J., Oorschot, P.C., Vanstone, S.A.: Handbook of Applied Cryptography, CRC Press, (1997).
9. Tran, B. N., Nguyen, T. D., Tran, T. D.: A new S-box structure to increase complexity of algebraic expression for block cipher cryptosystems. In: International Conference on Computer Technology and Development, pp. 212-216. IEEE Press, Kota Kinabalu (2009).
10. Busireddygari, P.; Kak, S.: Pseudorandom tableau sequences, In: 51st Asilomar Conference on Signals, Systems, and Computers, pp. 1733 – 1736. IEEE Press (2017).
11. Gurugopinath, S., Samudhyatha, B.: Multi-dimensional Anderson-Darling statistic based goodness-of-fit test for spectrum sensing. In: Seventh International Workshop on Signal Design and its Applications in Communications (IWSDA). pp. 165-169. Bengaluru, India. (2015).
12. Wang, H., Yang, E.-H., Zhao, Z., Zhang, W.: Spectrum sensing in cognitive radio using goodness of fit testing. In: IEEE Transactions on Wireless Communications, **8(11)**, pp. 5427-5430, IEEE Press (2009).
13. Teguig, D., Nir, V. Le, Scheers, B.: Spectrum sensing method based on goodness of fit test using chi-square distribution" Electronics Letters, **50(9),** pp. 713-715 (2014).
14. Wilson, P.: Inter-Device Authentication Protocol for the Internet of Things. University of Victoria, pp. 4-10 (2017).
15. Crossman M. A., Liu H. Study of authentication with IoT. In: 2015 IEEE International Symposium, pp. 1–7. IEEE Press (2015).
16. Gaydyshev I.P. Programmnoye obespecheniye analiza dannykh, AtteStat. Rukovodstvo pol'zovatelya. **13**. (2012).
17. Moody, D.: Post-quantum cryptography: NIST's plan for the future. In: Proceedings of the Seventh International Conference on Post Quantum Cryptography. IEEE Press, Japan, (2016). https://pqcrypto2016.jp
18. Special Publication 800-22. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. http://csrc. nist.gov