Pseudorandom Sequences for Spread Spectrum Image Steganography

Alexandr Kuznetsov [1] [0000-0003-2331-6326], Oleksii Smirnov [2][0000-0001-9543-874X],
Anna Arischenko [1] [0000-0002-7498-3113], Iryna Chepurko [1] [0000-0003-2842-0181],
Alexander Onikiychuk [1] [0000-0003-3736-4660] and Tetiana Kuznetsova [1][0000-0001-6154-7139]

[1] V. N. Karazin Kharkiv National University, Kharkiv, Ukraine
kuznetsov@karazin.ua, annaarischenko@gmail.com,
i.chepurko@karazin.ua, onik4524a@gmail.com,
kuznetsova.tatiana17@gmail.com
[2] Central Ukrainian National Technical University, Kropivnitskiy, Ukraine,
dr.smirnovoa@gmail.com

Abstract. Digital steganography is a promising direction in the development of information protection methods. Information messages are hidden in redundant data (cover data), which are processed and transmitted in information and telecommunication systems. At the same time, the fact of existence of messages is being hidden, this allows providing their high safety of a steganosystem. We consider the pseudorandom sequences (signals), which are used for information-hiding in the cover images. The spread spectrum image steganography is used for the hiding, the essence of which is in modulation of information data by long pseudorandom (noise) sequences. Messages take the form of noise, and it is extremely difficult to detect such transmission. We investigate different ways of discrete signals generation and estimate the error rate in message restoration. It turns out that the way of the discrete signals generation influences on the error rate and we prove the choice of the most suitable signals. Moreover, we estimate distortions of the cover image as a result of data-hiding. The article mainly contains the results of experimental researches, which can be useful in justifying various ways of building direct spread spectrum steganographic systems.

Keywords: data-hiding, steganography, spread spectrum image steganography, pseudorandom sequences, spreading sequences.

1 Introduction

With the development of new information technologies, mobile gadgets, the Internet of Things, global systems and services, information security issues are becoming more relevant [1-3]. In addition to providing qualitatively new information services, the increase in processing speed and the volume of information, new threats and challenges are also emerging. In particular, according to leading information and analytical agencies, the number of cyber-attacks has increased sharply in recent years [4-6].

CybHyg-2019: International Workshop on Cyber Hygiene, Kyiv, Ukraine, November 30, 2019.

This forces to develop new ways of protecting information, including for the safe transmission of information messages through global information networks.

One of the current directions in the development of information protection technologies is steganography [7-9]. It combines methods and means of processing redundant data (also called cover data) to hide information messages in them. Cover data can be transmitted over open communication channels, for example, using e-mail. An outside observer, even analyzing cover dada, should not guess the presence of hidden information messages. This is the main difference between steganography and cryptography [10]. While cryptographic methods hide the semantic content of transmitted data, steganography hides the existence of messages. And this technique can be much more effective.

Various redundant files can be used as cover data, e.g. images, audio, texts, etc. [8, 9]. The most common case is the use of still images. This is due both to the high natural redundancy of realistic images and to their transmission frequency on the Internet. It is almost impossible to track, analyse, and even detect the fact of the hiding of information messages in still images stored and transmitted on the Internet. It is one of the important and interesting directions in the development of modern information security technologies.

Various steganographic techniques are used for data-hiding in cover images [7, 8]. In our view, the most interesting approach is the use of spread spectrum technique [11-23].

2 Related works

The spread spectrum technique traditionally is used in radio communication systems with a multiple access [25-28]. It is based on the modulation of information messages by so called spreading signals - long pseudorandom sequences which have random, noise-like form. In this case, the transmitted message becomes like noise and is very difficult to identify. Furthermore, applied correlation reception methods of complex noise-like signals allow correcting the arising errors, thereby increasing jamming resistance of the communication system.

In works [11-24] the spread spectrum technique is often used for information-hiding in digital cover images. For example, it was offered to use nonlinear modulation by pseudorandom sequences, elements of which are distributed according to the normal law with zero mean and one mean-squared error in [11-17]. Indeed, it is possible to hide the information messages at the acceptable level of distortions in the cover by interpreting images as noise in the communication channel [13].

In this article we investigate different options for spreading signals generation, as well as their influence on the quality characteristics of the steganosystem. In particular, we evaluate the validity of transmitted data, by estimating the bit error rate (BER) in restored messages. What is more, we estimate the amount of distortions in the cover image. For this purpose, we calculate the mean squared error (MSE) between original image and the one received after the information message was hidden in it.

Characteristics that are considered (BER and MSE) allow comparing different options for generating spectrum spreading signals. We show, that the changing of the signal forming rule can significantly affect BER. Indeed, in spread spectrum radio communication systems, natural noise is not correlated with spreading sequences in the communication channel. However, if information is hidden in digital images, this may not be the case. Neighboring pixels in natural scenes are highly correlated and such communication can infringe basic estimates, which justify the correct data restoration. We investigate some ways of spreading signals generation and prove the choice of the best alternative.

3 Spread Spectrum Image Steganography

Data transmission in spread spectrum radio communication systems can be simplified as a relation [24-28]:

$$N = I + P\sum_{i=1}^{k} b_i \varphi_i \ , \tag{1}$$

where every information bit $b_i \in \{-1,1\}$ is multiplied by a spreading pseudorandom sequence φ_i from the set (ensemble) of weakly correlated discrete signals:

$$\forall \varphi_i \in \varphi = \{\varphi_0, \varphi_1, ..., \varphi_{M-1}\},$$

$$\forall i \neq j : \rho(\varphi_i, \varphi_j) \approx 0,$$

- P is a power gain of discrete signals;
- k is a number of bits of the information message, simultaneously transmitted in the communication channel (in code division systems this value can describe subscriber capacity with multiple access);
- I is natural noise in the communication system;
- $\rho(\varphi_i, \varphi_j)$ is cross-correlation coefficient of the sequences φ_i and φ_j;
- N is received signal at the receiving side (additive mix of the useful signal and noise)

Information is restored by means of correlation reception. For this purpose correlation coefficient is calculated (scalar product of vectors) [24-28]:

$$\rho\left(N, \varphi_j\right) = I\varphi_j + \varphi_j P\sum_{i=1}^{k} b_i \varphi_i \ .$$

The natural noise I and the noise signal φ_i are statistically independent (uncorrelated) in communication systems, i.e.

$$\rho\left(I,\varphi_j\right)=I\varphi_j\approx 0\,.$$

Different noise signals are also uncorrelated with each other, i.e.

$$\forall j\neq i:\varphi_j\varphi_i\approx 0\,.$$

Then

$$\rho\left(N,\varphi_j\right)\approx Pb_j\varphi_j\varphi_j\,,$$

and the value b_j can be defined by the sign $\rho\left(N,\varphi_j\right)$:

$$b_j=sign\left[\rho\left(N,\varphi_j\right)\right].\tag{2}$$

The following assumptions are used to hide the information message in the cover image [11-17]. The digital image I is interpreted as noise in the communication channel, and we assume, that

$$\rho\left(I,\varphi_j\right)=I\varphi_j\approx 0\,.$$

Information bits are modulated by spreading sequences:

$$\sum_{i=1}^{k}b_i\varphi_i\,,$$

after that, as in (1), the enhanced result is added to the cover image.

The rule (2) is also used here to restore information bits. As before, we suppose $\forall j\neq i:\varphi_j\varphi_i\approx 0$. But the assumption

$$\rho\left(I,\varphi_j\right)=I\varphi_j\approx 0$$

may not be executed. Indeed, separate pixels of realistic images are highly correlated. In this case, the result

$$\rho\left(I,\varphi_j\right)=I\varphi_j$$

depends on statistical properties of spreading sequences, i.e. the way of the set generation $\varphi=\{\varphi_0,\varphi_1,...,\varphi_{M-1}\}$

In this article we consider different ways of discrete signals generation and investigate the efficiency of their use for data hiding in cover images. We estimate the bit error rate when restoring data by the rule (2). BER is the number of bit errors N_{error} divided by the total number of transferred bits N_{total} [29]:

$$BER = \frac{N_{error}}{N_{total}} . \tag{3}$$

BER is a unitless performance measure, often expressed as a percentage [29]. We estimate BER in the absolutes, i.e. directly by (3).

It should be noted, that in our investigations we estimated BER without the use of error-correcting coding. This case has also been considered in other works, for example the table 2 of [14] shows similar results.

MSE is used to estimate the distortions of the cover image [29-31]. The MSE value for a monochrome $m \times n$ image I is determined by:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{i,j} - N_{i,j}]^2 , \tag{4}$$

where N is noisy approximation of a cover image I , as in (1).

We used various 256×256 images as given data (similar to following works [13, 14, 16, 17]). The results represent averages, taken from several various images.

4 Results

Consider several options for spreading sequences generation φ_i in (1). For each case, we will rate BER and MSE. These values characterize errors in the restored message and distortions in the cover image.

4.1 Nonlinear sequences with Gaussian distribution

The first case we are considering has been described in works [13, 14, 16, 17]. Each spreading sequence was proposed to be formed using the ratios:

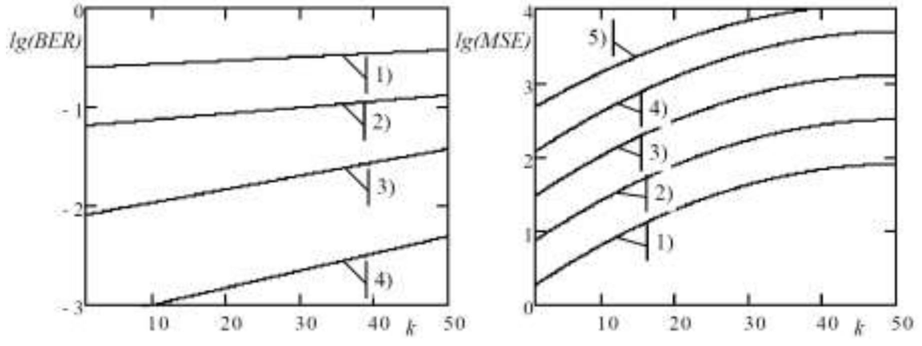
$$(\varphi_i)_j = \begin{cases} \Phi^{-1}((u_i)_j), b_i = -1; \\ \Phi^{-1}((u'_i)_j), b_i = 1, \end{cases} \tag{5}$$

where

$$(u'_i)_j = \begin{cases} (u_i)_j + 0.5, u_i < 0.5; \\ (u_i)_j - 0.5, u_i \geq 0.5, \end{cases} \tag{6}$$

- $(u_i)_j$ - a random value uniformly distributed on the interval (0,1);
- Φ^{-1} represents the inverse cumulative distribution function for a standard Gaussian random variable.

The obtained results for different P are shown in Fig. 1.



a) BER(k) dependencies for cases:
1) P = 8; 2) P = 16; 3) P = 32; 4) P = 64

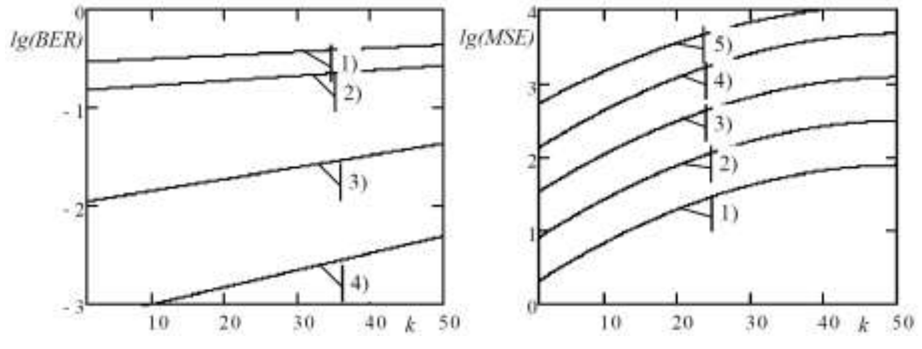
b) MSE(k) dependencies for cases:
1) P = 1; 2) P = 2; 3) P = 4; 4) P = 8; 5) P = 16

Fig. 1. Empirical dependences BER(k) and MSE(k) for discrete sequences from [13, 14, 16, 17]

We also investigated the efficiency of data hiding during discrete signals generation φ_i according to a simplified scheme:

$$(\varphi_i)_j = \Phi^{-1}((u'_i)_j),\ (u'_i)_j = \begin{cases}(u_i)_j + 0.5, u_i < 0.5;\\ (u_i)_j - 0.5, u_i \geq 0.5.\end{cases} \quad (7)$$

The results for BER and MSE for this variant are shown in Fig. 2:



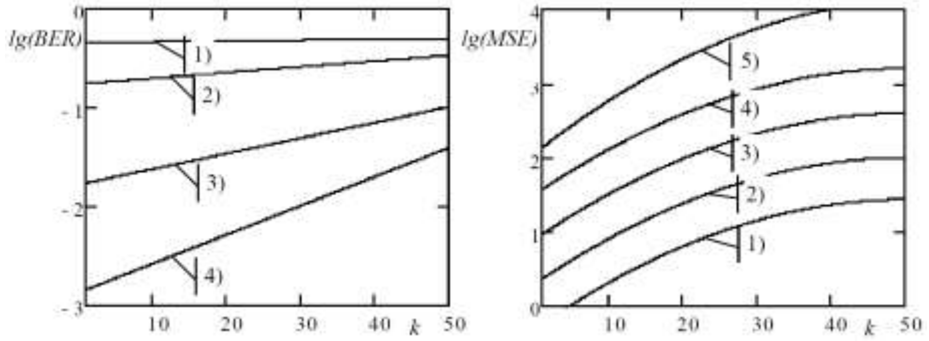
a) BER(k) dependencies for cases:
1) P = 8; 2) P = 16; 3) P = 32; 4) P = 64

b) MSE(k) dependencies for cases:
1) P = 1; 2) P = 2; 3) P = 4; 4) P = 8; 5) P = 16

Fig. 2. Empirical dependences BER(k) and MSE(k) for discrete sequences formed by (7)

4.2 Discrete sequences with the uniform distribution on the interval (-1,1)

As an alternative, we realized another way of discrete sequences generation, when their elements are distributed according to uniform law on the interval (-1,1). The results of the experimental exploring are shown in Fig. 3.



a) BER(k) dependencies for cases:
1) P = 8; 2) P = 16; 3) P = 32; 4) P = 64

b) MSE(k) dependencies for cases:
1) P = 1; 2) P = 2; 3) P = 4; 4) P = 8; 5) P = 16

Fig. 3. Empirical dependences BER(k) and MSE(k) for discrete sequences with the uniform distribution on the interval (-1,1)

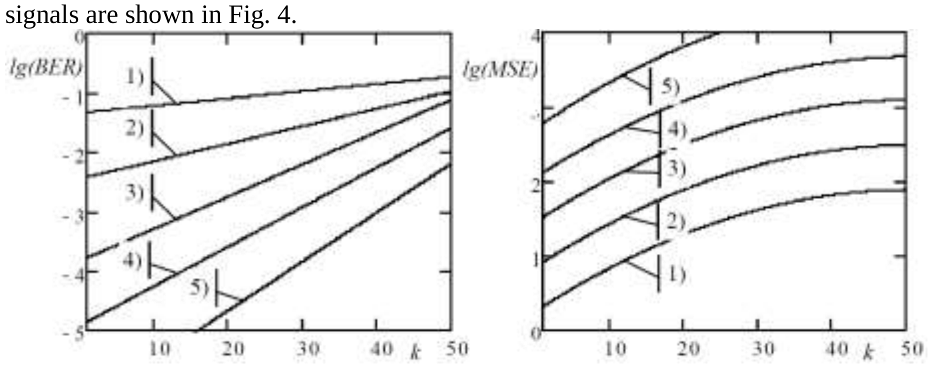
4.3 Walsh-Hadamard Signals

Another way of the formation of the set $\varphi = \{\varphi_0, \varphi_1, ..., \varphi_{M-1}\}$ that we investigated was to use Hadamard matrices. These matrices are formed by a recurrent rule:

$$H_{2^i} = \begin{bmatrix} H_{2^{i-1}} & H_{2^{i-1}} \\ H_{2^{i-1}} & -H_{2^{i-1}} \end{bmatrix}, \ H_1 = [1].$$

Rows (or columns) of matrices H_{2^i} are mutually orthogonal, i.e. their scalar product is zero. The set of discrete signals $\varphi = \{\varphi_0, \varphi_1, ..., \varphi_{M-1}\}$, composed of such rows (or columns), is called Walsh-Hadamard sequences [24].

The results of the experimental exploring of BER and MER for Walsh-Hadamard signals are shown in Fig. 4.



a) BER(k) dependencies for cases:
1) P = 1; 2) P = 2; 3) P = 4; 4) P = 8; 5) P = 16

b) MSE(k) dependencies for cases:
1) P = 1; 2) P = 2; 3) P = 4; 4) P = 8; 5) P = 16

Fig. 4. Empirical dependences BER(k) and MSE(k) for Walsh-Hadamard sequences

5 The Discussion of Results and Conclusions

According to the results of experimental studies, all considered methods of discrete signals generation are almost equivalent by the distortion of the cover image. It can be explained by the close range of possible values of sequences and a similar way of data hiding. The signals with the nonlinear modulation rule, which are proposed in [13, 14, 16, 17], have a little benefit. Walsh-Hadamard signals look worst of all by MSE (but this loss is small and almost invisible on the logarithmic scale).

The first three ways of discrete signals generation are almost identical by the BER minimization criterion. Even with a high gain P, these spreading sequences generation techniques do not allow getting small BER values. For example, when P = 64 the error rate is approximately 10^{-3} and higher, it implies the mandatory use of error-correcting code. Nonlinear sequences from [13, 14, 16, 17] have a little benefit among the first three ways. However, the greatest gain in BER reduction is from the use of Walsh-Hadamard signals. It can be seen from the Fig. 4, that even at P = 16 low BER values of about 10^{-5} and below are already achieved. It gives huge opportunities on the practical construction of steganographic systems for information hiding in cover images.

A promising area for our future researches is the development of an adaptive rule for the formation of pseudorandom spreading sequences. For example, if a rule for discrete signals generation from the set $\varphi = \{\varphi_0, \varphi_1, ..., \varphi_{M-1}\}$ takes into account the statistical properties of the cover image, then the error rate (BER) can be significantly reduced, and it is possible to achieve error-free information recovery. In our view, the use of new classes of pseudorandom sequences, for example from our previous works [32-36], is also promising.

The results can be used in various computer science applications. In particular, for modernization of various cryptographic algorithms, optimization of calculations, modeling and telecommunications [37-43].

References

1. M. Blowers, Ed., "Evolution of Cyber Technologies and Operations to 2035", Advances in Information Security, 2015. DOI:10.1007/978-3-319-23585-1.
2. A. I. Awad and M. Fairhurst, Eds., "Information Security: Foundations, Technologies and Applications," Apr. 2018. DOI:10.1049/pbse001e.
3. T. R. Peltier, "Information Security Policies, Procedures, and Standards," Apr. 2016. DOI:10.1201/9780849390326.
4. F. Z. Leccisotti, R. Chiesa, and D. De Nicolo, "Analysis of Possible Future Global Scenarios in the Field of Cyber Warfare", Cyber Security and Threats, pp. 1584-1608. DOI:10.4018/978-1-5225-5634-3.ch077.
5. T. A. Johnson, Ed., "Cybersecurity", Apr. 2015. DOI:10.1201/b18335.
6. S. Sen and C. Jayawardena, "Analysis of Cyber-Attack in Big Data IoT and Cyber-Physical Systems - A Technical Approach to Cybersecurity Modeling", 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-7.
7. "Digital Watermarking and Steganography", 2008. DOI:10.1016/b978-0-12-372585-1.x5001-3.

8. F. Shin, “Digital Watermarking and Steganography,” December 2017. DOI:10.1201/9781315219783.
9. N. F. Johnson and S. Jajodia, "Exploring steganography: Seeing the unseen," in Computer, vol. 31, no. 2, pp. 26-34, Feb. 1998. DOI: 10.1109/MC.1998.4655281 .
10. I. V. S. Manoj, “Cryptography and Steganography,” International Journal of Computer Applications, vol. 1, no. 12, pp. 63–68, Feb. 2010. DOI:10.5120/257-414.
11. A. Z. Tirkel, C. F. Osborne and R. G. Van Schyndel, "Image watermarking-a spread spectrum application," Proceedings of ISSSTA'95 International Symposium on Spread Spectrum Techniques and Applications, Mainz, Germany, 1996, pp. 785-789 vol.2. DOI:10.1109/ISSSTA.1996.563231.
12. J. R. Smith and B. O. Comiskey, “Modulation and information hiding in images,” Lecture Notes in Computer Science, pp. 207-226, 1996. DOI:10.1007/3-540-61996-8_42.
13. L. M. Marvel, C. G. Boncelet, R. Jr., and Charles T., “Methodology of Spread-Spectrum Image Steganography,” Jun. 1998. DOI:10.21236/ada349102.
14. L. M. Marvel, C. G. Boncelet and C. T. Retter, "Spread spectrum image steganography," in IEEE Transactions on Image Processing, vol. 8, no. 8, pp. 1075-1083, Aug. 1999. DOI:10.1109/83.777088.
15. M. Kutter, “Performance Improvement of Spread Spectrum Based Image Watermarking Schemes through M-ary Modulation,” Lecture Notes in Computer Science, pp. 237–252, 2000. DOI:10.1007/10719724_17.
16. F. S. Brundick and L. M. Marvel, “Implementation of Spread Spectrum Image Steganography,” Mar. 2001. DOI:10.21236/ada392155.
17. Patent No.: US 6,557,103 B1, Int.Cl. G06F 11/30. Charles G. Boncelet, Jr., Lisa M. Marvel, Charles T. Retter. Spread Spectrum Image Steganography. Patent No.: US 6,557,103 B1, Int.Cl. G06F 11/30. – № 09/257,136; Filed Feb. 11, 1999; Date of Patent Apr. 29, 2003.
18. Fan Zhang, Bin Xu and Xinhong Zhang, “Digital image watermarking algorithm based on CDMA spread spectrum”, 2006 12th International Multi-Media Modelling Conference, Beijing, 2006, pp. 4 pp.-. DOI: 10.1109/MMMC.2006.1651359.
19. T. T. Nguyen and D. Taubman, “Optimal linear detector for spread spectrum based multidimensional signal watermarking”, 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, 2009, pp. 113-116. DOI: 10.1109/ICIP.2009.5414121.
20. E. Nezhadarya, Z. J. Wang and R. K. Ward, “Image quality monitoring using spread spectrum watermarking”, 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, 2009, pp. 2233-2236. DOI: 10.1109/ICIP.2009.5413955.
21. S. Ghosh, P. Ray, S. P. Maity, H. Rahaman, Spread Spectrum Image Watermarking with Digital Design, 2009 International Advance Computing Conference, 2009, pp. 868-873.
22. H. O. Altun, A. Orsdemir, G. Sharma and M. F. Bocko, “Optimal Spread Spectrum Watermark Embedding via a Multistep Feasibility Formulation”, in IEEE Transactions on Image Processing, vol. 18, no. 2, pp. 371-387, Feb. 2009. DOI: 10.1109/TIP.2008.2008222.
23. A. Samčović and M. Milovanović, “Robust digital image watermarking based on wavelet transform and spread spectrum techniques”, 2015 23rd Telecommunications Forum Telfor (TELFOR), Belgrade, 2015, pp. 811-814. DOI: 10.1109/TELFOR.2015.7377589.
24. U. Madhow, Fundamentals of Digital Communication. Cambridge: Cambridge University Press, 2008. DOI:10.1017/cbo9780511807046.
25. Shakhovska N., Holoshchuk R., Fedushko S., Kosar O., Danel R., Repka M. The sequential associative rules analysis of patient's physical characteristics. CEUR Workshop Proceedings. Vol. 2255: IDDM 2018, 2018. P. 82–92.
26. V. P. Ipatov, “Spread Spectrum and CDMA”, Mar. 2005. DOI:10.1002/0470091800.

27. “Introduction to CDMA Wireless Communications”, 2007. DOI:10.1016/b978-0-7506-5252-0.x5001-7.
28. “The Generalized CDMA”, CDMA: Access and Switching, pp. 1–28. DOI:10.1002/0470841699.
29. S. Hara and R. Prasad, “DS-CDMA, MC-CDMA and MT-CDMA for mobile multi-media communications”, Proceedings of Vehicular Technology Conference - VTC, Atlanta, GA, USA, 1996, pp. 1106-1110 vol.2. DOI: 10.1109/VETEC.1996.501483.
30. “Probability Theory of Bit Error Rate”, Optical Bit Error Rate, 2009. DOI:10.1109/9780470545430.ch7.
31. J. Korhonen and J. You, “Peak signal-to-noise ratio revisited: Is simple beautiful?” 2012 Fourth International Workshop on Quality of Multimedia Experience, Yarra Valley, VIC, 2012, pp. 37-38. DOI: 10.1109/QoMEX.2012.6263880.
32. “Data Compression,” 2007. DOI:10.1007/978-1-84628-603-2.
33. Yu.V. Stasev, A.A. Kuznetsov, A.M. Nosik, “Formation of pseudorandom sequences with improved autocorrelation properties”, Cybernetics and Systems Analysis, vol. 43, Issue 1, pp. 1-11, January 2007. DOI: 10.1007/s10559-007-0021-2.
34. N. Naumenko, Yu. Stasev, A. Kuznetsov, “Methods of synthesis of signals with prescribed properties.” Cybernetics and Systems Analysis, vol. 43, Issue 3, pp. 321-326, May 2007. DOI: 10.1007/s10559-007-0052-8.
35. O.Karpenko, A.Kuznetsov, V.Sai, Yu.Stasev. “Discrete Signals with Multi-Level Correlation Function”, Telecommunications and Radio Engineering, vol. 71, 2012 Issue 1. pp 91-98. DOI: 10.1615/TelecomRadEng.v71.i1.100.
36. A. Kuznetsov, S. Kavun, V. Panchenko, D. Prokopovych-Tkachenko, F. Kurinniy and V. Shoiko, “Periodic Properties of Cryptographically Strong Pseudorandom Sequences”, 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 2018, pp. 129-134. DOI:10.1109/INFOCOMMST.2018.8632021.
37. A. Kuznetsov, O. Smirnov, D. Kovalchuk, A. Averchev, M. Pastukhov and K. Kuznetsova, “Formation of Pseudorandom Sequences with Special Correlation Properties”, Advanced Information and Communications Technologies (AICT), 2019, pp. 395-399.
38. Chornei R., Hans D. V., and Knopov P. (2005). “Controlled Markov fields with finite state space on graphs”. Stochastic Models, 21(4), 847-874. DOI:10.1080/15326340500294520.
39. Gnatyuk S., Okhrimenko T., Azarenko O., Fesenko A., Berdibayev R. “Experimental Study of Secure PRNG for Q-trits Quantum Cryptography Protocols”, Proceedings of the 2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT 2020), Kyiv, Ukraine, May 14, 2020, pp. 183-188.
40. Runovski K., & Schmeisser H. (2004). On the convergence of fourier means and interpolation means. Journal of Computational Analysis and Applications, 6(3), 211-227.
41. Hu Z., Gnatyuk S., Okhrimenko T., Tynymbayev S., Iavich M. High-speed and secure PRNG for cryptographic applications, International Journal of Computer Network and Information Security, Issue 12 (3), pp. 1-10, 2020.
42. Bondarenko S., Liliya B., Oksana K., & Inna G. (2019). Modelling instruments in risk management. International Journal of Civil Engineering and Technology, 10(1), 1561-1568.
43. Hu Z., Gnatyuk S., Kovtun M., Seilova N. “Method of searching birationally equivalent Edwards curves over binary fields”, Advances in Intelligent Systems and Computing, Vol. 754, pp. 309-319, 2019.
44. Tkach B. P., Urmancheva L. B. (2009). Numerical-analytic method for finding solutions of systems with distributed parameters and integral condition. Nonlinear Oscillations, 12(1), 113-122. DOI:10.1007/s11072-009-0064-6.