Application of Arithmetic Coding Methods in Cryptographic Information Protection Systems

Dmytro Havrylov[a], Olexandr Shaigas[b], Oksana Stetsenko[b], Yurii Babenko[b], and Valerii Yroshenko[b]

[a] *Kharkiv National University of Radio Electronics, 14 Nauky ave., Kharkiv, 61103, Ukraine*
[b] *Taras Shevchenko National University of Kyiv, 60 Volodymyrska str., Kyiv, 01033, Ukraine*

Abstract

The article analyzes recent publications, which indicated the rapid creation of data by different types of media. It is noted that one-fifth of the data created is critical and needs protection. To reduce the amount of data stored, it is recommended to use encoding methods without constructing an encoding table. That is why arithmetic and adaptive arithmetic coding from the point of view of the possibility of application in systems of cryptographic protection of information for systems of critical infrastructure are considered in detail. As a result of research for the guaranteed protection of the information, it is offered to use the technology of consecutive cryptographic protection of the information (after coding) with an application of adaptive arithmetic coding. For systems with streaming data processing (technology of selective protection of information), it is proposed to use arithmetic coding. Under the key information in the arithmetic coding algorithm, we understand the weight of each element, the rest of the information does not require additional protection.

Keywords

Information security, arithmetic coding, adaptive arithmetic coding, big data.

1. Introduction

Despite the active development of technologies and principles of media construction aimed at increasing the amount of data that can be stored (while reducing the physical volume and weight), the need for algorithms to reduce the volume is still an urgent problem that needs to be addressed. This effect persists due to the simultaneous development of data entry technologies. These technologies include a photo and video cameras from different companies, which allow you to obtain multidimensional data of high quality (resolution) and depth of image construction, which in turn significantly increases the amount of information needed for storage and subsequent reproduction on display media. At the same time, the use of high-tech devices is observed not only at the level of states or enterprises, but also at the level of personal users, who in turn can provide access and/or cover photos, videos, etc. on social networks, blogs, cloud storage or own sites. As a result of this feature, the user initiates requirements for administrators, providers, owners of the above information resources from the standpoint of resource requirements (storage) and performance (computing power) by closing the circle "Information resource (storage)—Quality (volume)—Information resource (refuge)." After all, the market for information services has grown so much that a dissatisfied user (long download time, insufficient functionality, or even unsatisfactory site design) has the opportunity to easily and quickly (within a few minutes) change the service provider, which will harm the latter. The multitude of users united in the organization, the enterprise, the state has led to the emergence of the concept of Big Data.

Under Big Data we understand work with the information of considerable volume and various on structural, the semantic structure of the data which is updated every second and is in various sources

Cybersecurity Providing in Information and Telecommunication Systems, January 28, 2021, Kyiv, Ukraine
EMAIL: havrylov_d@ukr.net (A.1); skons@ukr.net (B.2); S1981@ukr.net (B.3); babenkomahalych@gmail.com (B.4); Eroshenko59@gmail.com (B.5)
ORCID: 0000-0002-3344-7808 (A.1); 0000-0001-8172-4948 (B.2); 0000-0001-8420-7989 (B.3); 0000-0002-8115-3329 (B.4); 0000-0003-3175-6444 (B.5)

© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-2923/paper14.pdf

(storages) for the purpose of maintenance, an increase of efficiency of work, creation of new products and increase of competitiveness. The consulting company Forrester gives a brief statement: “Big data combines techniques and technologies that make sense of the data with maximum practicality.”

2. Analysis of Existing Research

According to analysts at IDC [10] and other researchers, in the coming years, the bulk of the data will be provided not by users but by companies. Industry and other sectors of the economy will account for up to 60% of all data in the world. For example, in 2015, companies generated a third of all global data.

The authors of these studies agree that in the future the quality of data will be much more important than their quantity. “Not all data is equally important, and without context, they are not useful at all. In this period of change, leadership will belong to the organizations that will be able to identify the most critical subgroups of information with the maximum impact on the desired area of activity and focus on them,” said in a report by IDC analyst.

The publications also noted that by 2025 the amount of data created in the world will exceed 150 Zettabytes (10^{21} bytes) per year (Fig. 1). At the same time, one-fifth of all data by 2025 will be considered critical. That is, this is the information on which will depend the lives and safety of people, the preservation of capital by companies, the reputation of countries, the international situation, world peace, and the existence of the planet Earth as a whole.

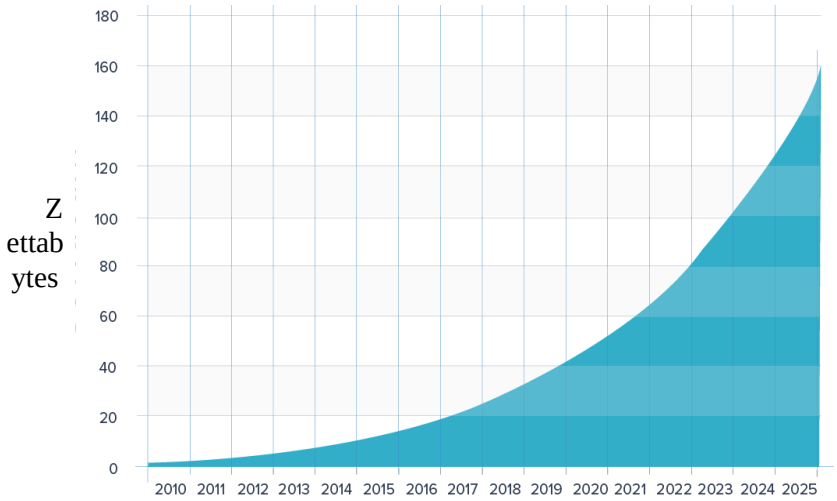


Figure 1: The amount of data created in the world with a forecast until 2025

However, in the coming years, the gap between the amount of data that needs to be protected and the information actually protected will only grow, increasing the amount of data even more. By 2025, up to 90% of all information must be protected in one way or another. However, in fact, less than half of all information will be protected.

The researchers also note that a significant amount of data will come from the devices that surround us every day in accordance with the concept of the Internet of Things, so we can conclude that:

First, by 2025, 75% of the world's population will have permanent access to the Internet. The amount of data sent to the world by type of media is shown in Fig. 2.

Secondly, the number of smart gadgets and home robots that will produce so-called metadata—service information that machines will exchange with each other for coordinated work.

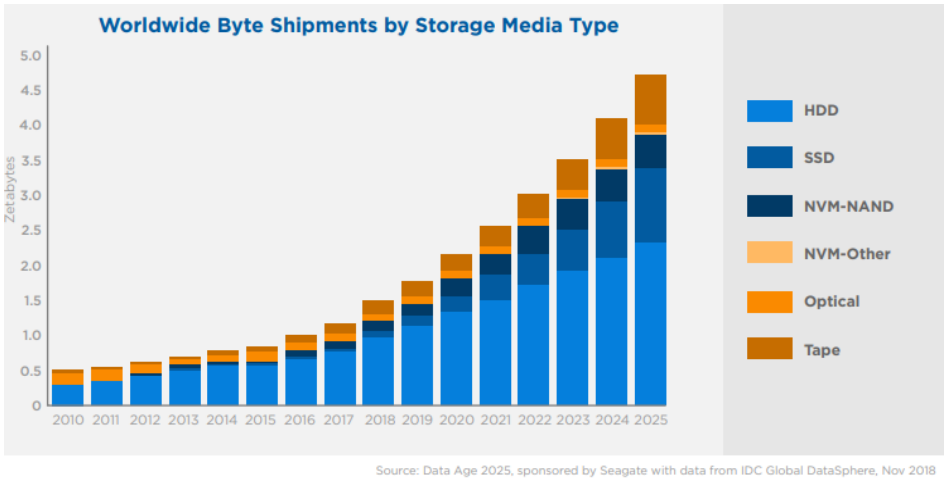


Figure 2: The amount of data sent to the world by type of media

Compared to today, everyone will interact 20 times more often with the Internet or with devices with Internet access. If now the average number of interactions is a little more than 1400, then by 2025 we will face the network more than 4900 times a day (Fig. 3).

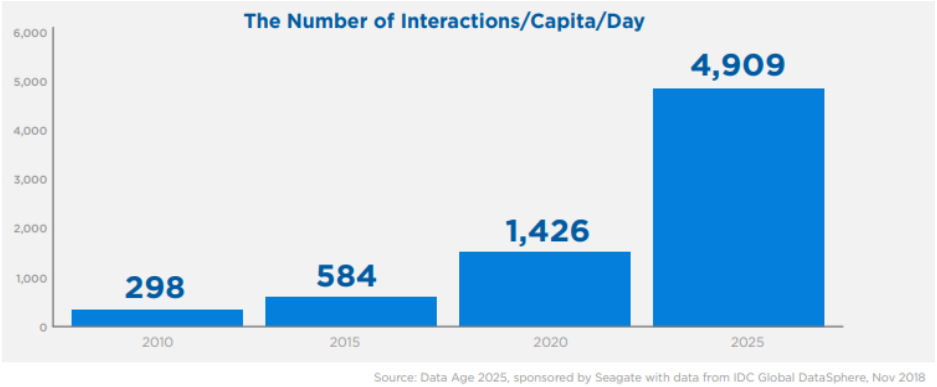


Figure 3: Number of user interactions with the Internet (once/day)

Therefore, looking for ways to solve the problem of storing large amounts of information, it was found that there are different approaches to solving the problem of reducing the code sequence (data volume) for processing, storage, and/or transmission of information. Analysis of coding methods based on the construction of Huffman tables, RLE, Lempel-Ziv, arithmetic coding, and adaptive arithmetic coding indicated both positive and negative properties of each method. Given that the main criteria for selecting the algorithm are the ability to reduce the amount of data (without increasing in exceptional cases) when entering the data processing with the probability of occurrence of elements close to equal probability and the time of the algorithm. Empirically, using an object-oriented Java programming language, statistics were collected from about 10,000 objects, which included numbers in binary and decimal representations, data arrays (matrices, blocks), words and sentences, and selected for more detailed study of arithmetic coding and adaptive arithmetic coding.

The purpose of this work is to determine the suitability of arithmetic coding and adaptive arithmetic coding for use in technologies for cryptographic protection of information in critical infrastructure.

3. The Main Part of the Research

The choice of methods based on the principles of arithmetic coding as research is due to the fact that these methods are exceptional because they do not build code tables, like most coding methods, and in the case of adaptive arithmetic coding have the ability to current coding. As a result of coding, we get

a real number within the interval (0,1) (the beginning of the working interval; the end of the working interval), which allows us to unambiguously reproduce the coded sequence.

The algorithm of arithmetic coding (Fig. 5) and decoding (Fig. 6) allows finding a discrepancy with the algorithm of adaptive coding (Fig. 7) and decoding (Fig. 8), which is that the weights of the symbols coming to the input of the encoder, for arithmetic coding are formed before the coding process and must be transmitted to the decoder. At the same time, with adaptive arithmetic encoding, the value of the weights of each symbol is formed in the encoding process, without the need to transmit these service data to the decoder.

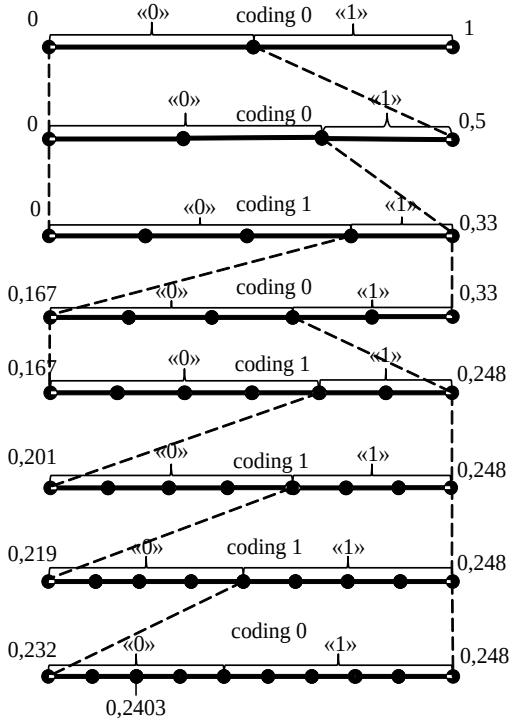


Figure 4: Construction of adaptive arithmetic code

The weight of the element is based on the formula:

$$\eta_i = \eta_i^{(0)} + \eta_i^{(1)} + ... + \eta_i^{(w)}, \quad (1)$$

where η_i is the sum of the weights on the i^{th} step, $\eta_i^{(0)}$ is the weight "0" on the i^{th} step, $\eta_i^{(1)}$ is the weight "1" on the i^{th} step, $\eta_i^{(w)}$ is the weight w on the i^{th} step;

The value of the segment in the i^{th} step is by the formula:

$$\rho_i = \frac{h_i - l_i}{\eta_i}, \quad (2)$$

The final step is to find the code number by determining the arithmetic mean between the beginning and end of the working interval of the last encoded character in the message:

$$Z = \frac{l_i + h_i}{2}, \quad (3)$$

where Z is the code number of the message.

Graphically, the process of adaptive arithmetic coding of the bit sequence "0 0 1 0 1 1 1 0" will take the form of Fig. 4. As a result, we obtain the coded number 24_2 = 1 1 0 0 0 with a compression ratio of 1.6.

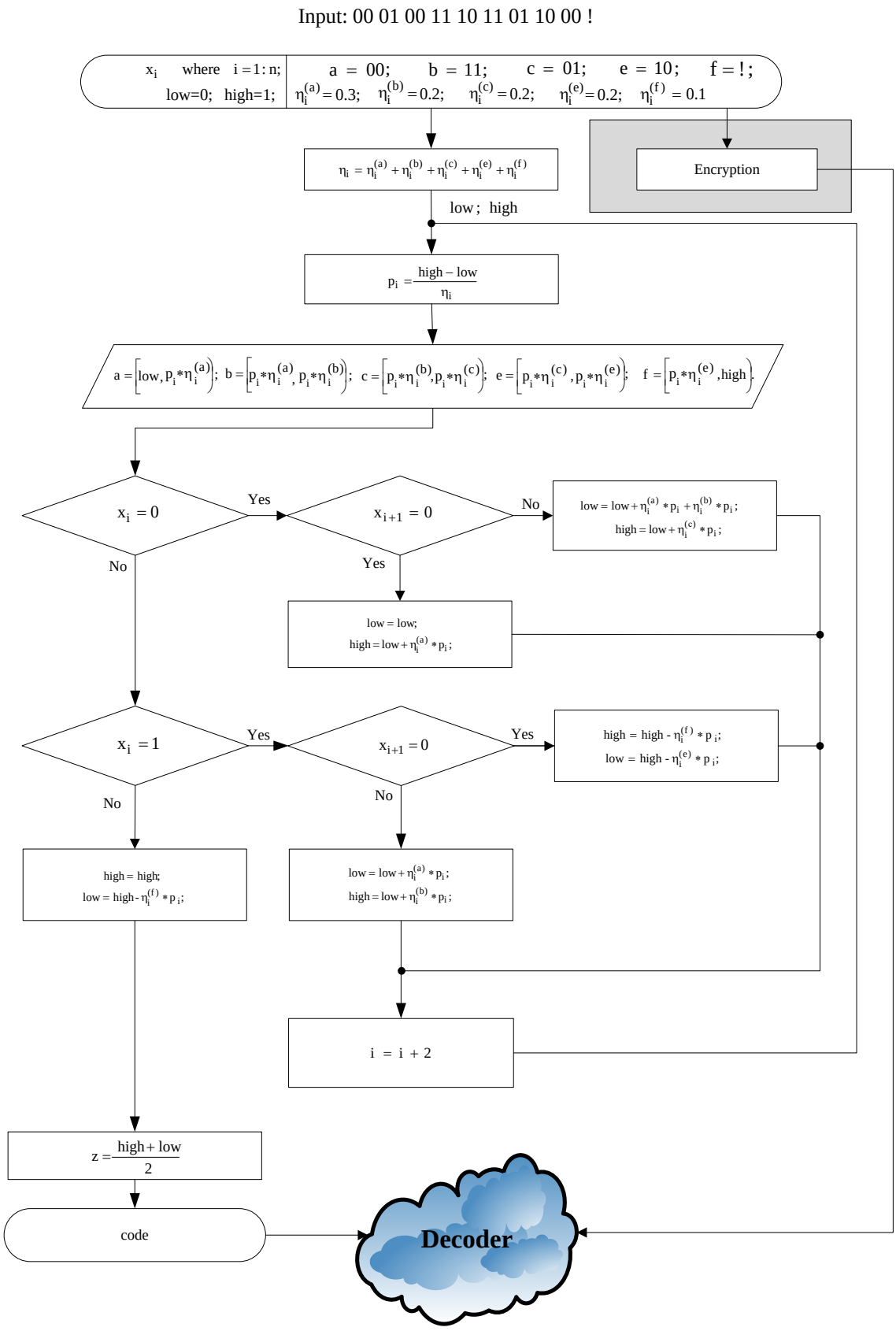
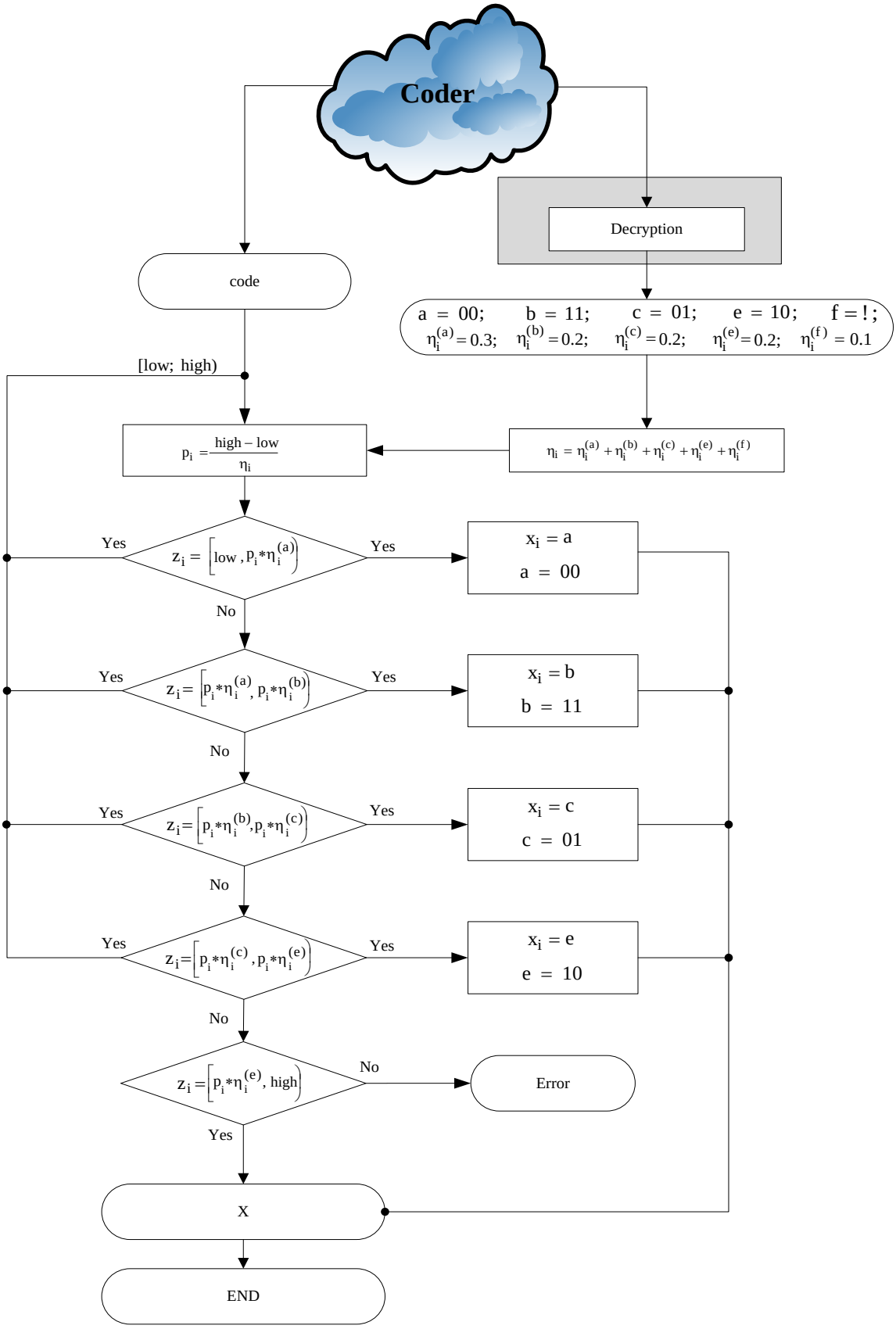


Figure 5: Example of arithmetic coding for a binary sequence with a stop marker "!"



Output: 00 01 00 11 10 11 01 10 00

Figure 6: Example of arithmetic coding decoder operation for binary sequence

Input: 00 01 00 11 10 11 01 10 00 !

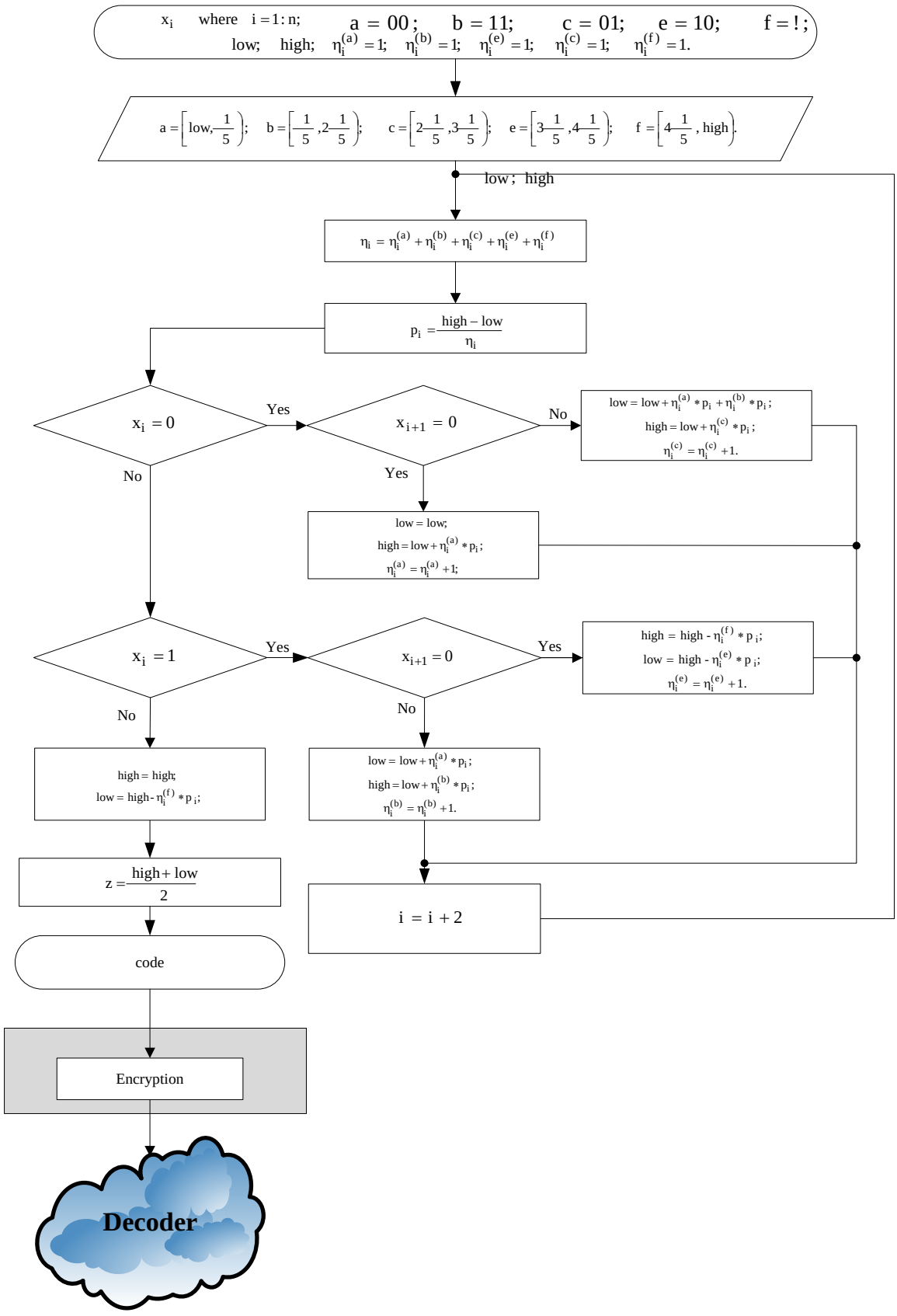
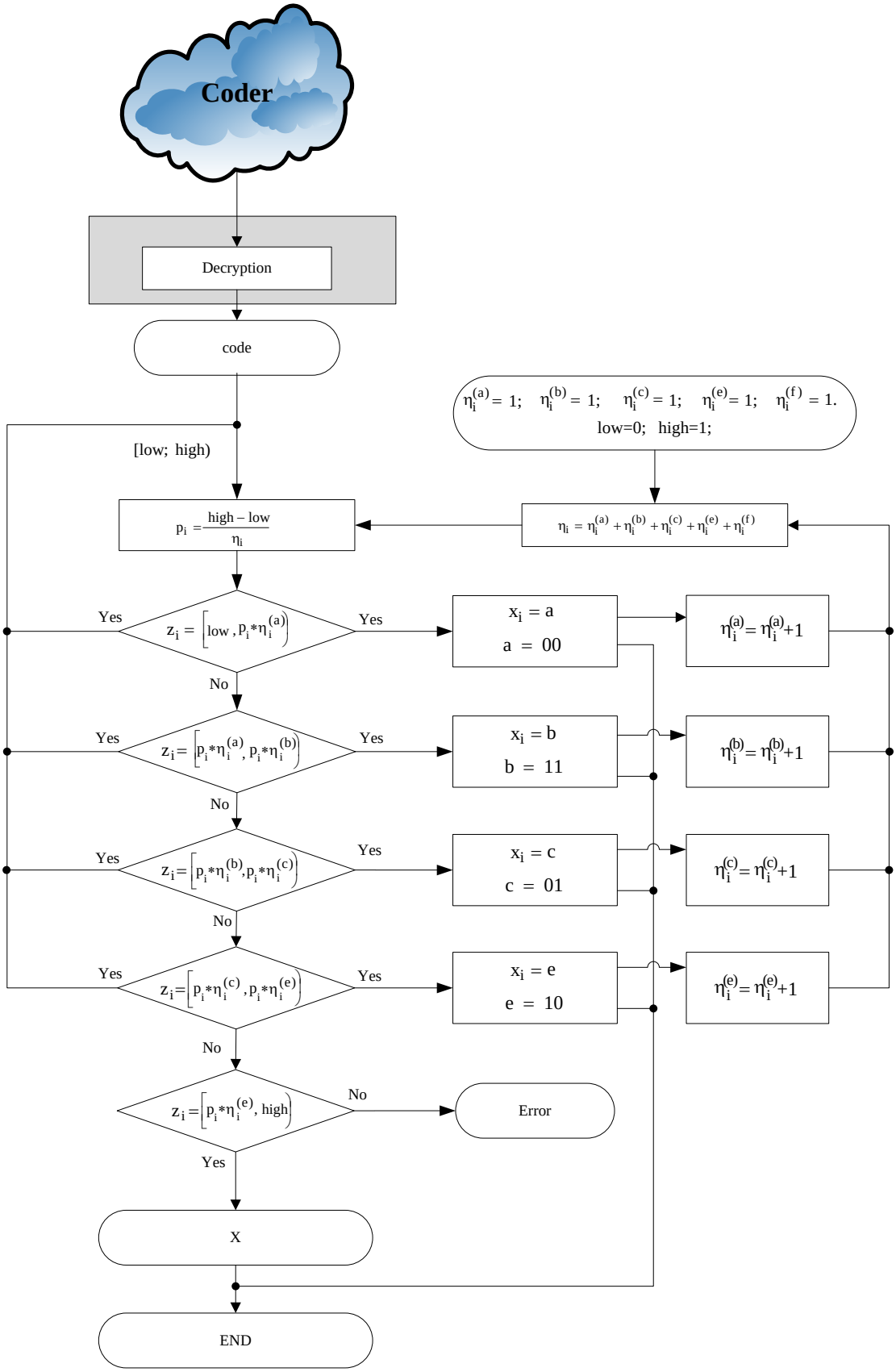


Figure 7: Example of adaptive arithmetic coding for a binary sequence with a stop marker "!"



Output: 00 01 00 11 10 11 01 10 00

Figure 8: Example of adaptive arithmetic coding decoder operation for binary sequence

The analysis of Fig. 5 and 6 allows us to draw the following conclusions on the possibility of applying these methods in cryptographic protection systems: due to the importance of the weight values of the element *w* in the i^{th} step, we can conclude that this information is key. Removing or replacing the weight values of the element *w* in the i^{th} step will make it impossible to correctly reproduce the encoded sequence. It should be noted that the value of the weight of the element *w* at each step is equal to:

$$\eta_1^{(w)} = \eta_2^{(w)} = \eta_3^{(w)} = \cdots = \eta_i^{(w)}, \quad (4)$$

This feature makes it possible to assume that the use of cryptographic data for the values of the weight of the element *w* can increase the degree of data security; due to the need to pass to determine the weight of the element *w*, the application of the method of arithmetic coding increases the time for data processing.

The analysis of Fig. 7 and 8 allows us to draw the following conclusions on the possibility of applying these methods in cryptographic protection systems:

- In comparison with the method of arithmetic coding to reduce the time for data processing during coding due to the lack of a passage to determine the weight of the element *w*.
- In comparison with the method of arithmetic coding, this method usually has a lower compression ratio.
- Due to the fact that by the end of the processing of coded elements it is impossible to determine the value of the weight of the element *w* application of the method of adaptive arithmetic coding in cryptographic information protection technologies is reduced to sequential data processing.

For the general analysis, the estimation of efficiency of to-do methods of arithmetic coding and adaptive arithmetic coding in the system of cryptographic protection of the information is carried out (Fig. 9).

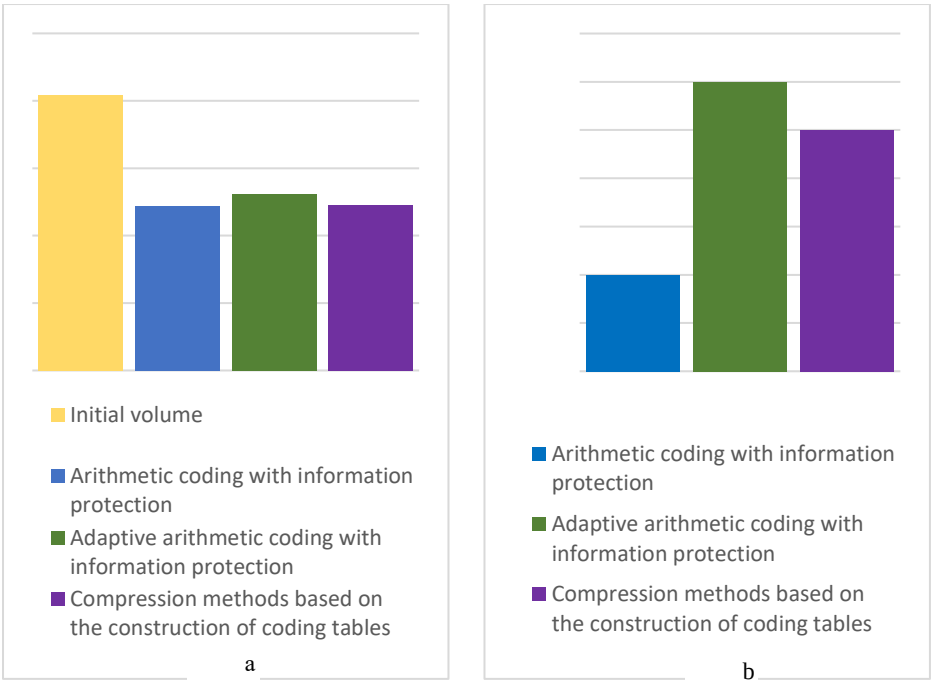


Figure 9: Evaluation of the effectiveness of the methods under study by the amount of data(a) and by the processing time (b)

4. Conclusions

The paper considers in detail the arithmetic and adaptive arithmetic coding from the standpoint of the possibility of application in systems of cryptographic protection of information for critical infrastructure systems.

Based on this, arithmetic coding methods can be used in cryptographic protection systems to reduce the original amount of data. Moreover, the analysis indicated that adaptive coding allows obtaining a higher compression ratio and fewer data processing times than adaptive arithmetic coding. Thus, we believe that arithmetic coding is a more acceptable solution for use in cryptographic information security systems compared to adaptive arithmetic coding and compression methods based on the construction of coding tables.

5. References

[1] S. Ramakrishnan, et al., Cryptographic and Information Security Approaches for Images and Videos. CRC Press, 2018, 962 p.. doi: 10.1201/9780429435461.

[2] Announcing the ADVANCED ENCRYPTION STANDARD (AES). Federal Information Processing Standards Publication, 197 (2001).

[3] DSTU 7624:2014: Information Technology. Cryptographic protection of information. Symmetric block transformation algorithm. Order of the Ministry of Economic Development of Ukraine № 1484 (29.12.2014).

[4] DSTU GOST 28147:2009: Information processing system. Cryptographic protection. Cryptographic transformation algorithm GOST 28147-89 (22.12.2008).

[5] F. Dufaux, T. Ebrahimi, Toward a Secure JPEG. Applications of Digital Image Processing XXIX, Vol. 6312, 2006. doi: 10.1117/12.686963.

[6] M. Farajallah, Chaos-based crypto and joint crypto-compression systems for images and videos, 2015. URL: https://hal.archives-ouvertes.fr/tel-01179610.

[7] T. Honda, Y. Murakami, Y. Yanagihara, T. Kumaki, T. Fujino, Hierarchical image-scrambling method with scramble-level controllability for privacy protection, in: IEEE 56th International Midwest Symposium on Circuits and Systems (MWSCAS), 2013, pp. 1371-1374. doi: 10.1109/MWSCAS.2013.6674911.

[8] Information technology – JPEG 2000 image coding system: Secure JPEG 2000. International Standard ISO/IEC 15444-8; ITU-T Recommendation T.807, 2007, 108 p.

[9] Sh. Ji, X. Tong, M. Zhang, Image encryption schemes for JPEG and GIF formats based on 3D baker with compound chaotic sequence generator, 2012. URL: https://arxiv.org/abs/1208.0999.

[10] JPEG Privacy & Security Abstract and Executive Summary, 2015. URL: https://jpeg.org/items/ 20150910_privacy_security_summary.html.

[11] R. L. Rivest, A. Shamir, L. M. Adleman, A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM, (2) 21, 1978, pp. 120–126. doi: 10.1145/359340.359342

[12] R. Sharma, S. Bollavarapu, Data Security using Compression and Cryptography Techniques. International Journal of Computer Applications, Vol. 117, No. 14, 2015, pp. 15-18. doi: 10.5120/20621-3342.

[13] V. B. Vasiliev, I. N. Okov, Yu. N. Strezhik, A. A. Ustinov, N. V. Shvetsov, Video data compression and protection in UAV information exchange radio channels, in: Scientific and practical conference on Prospects for the development and use of complexes with unmanned aerial vehicles, 924 State Center for Unmanned Aviation of the Ministry of Defense of the Russian Federation, 2016, pp. 202–204.

[14] K. Wong, K. Tanaka, DCT based scalable scrambling method with reversible data hiding functionality, in: 4th International Symposium on Communications, Control and Signal Processing (ISCCSP), 2010, pp. 1-4. doi: 10.1109/ISCCSP.2010.5463307.

[15] L. Yuan, P. Korshunov, T. Ebrahimi, Secure JPEG Scrambling enabling Privacy in Photo Sharing, in: 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2015, pp. 1-6. doi: 10.1109/FG.2015.7285022.

[16] K. M. Faraoun, A parallel block-based encryption schema for digital images using reversible cellular automata. Engineering Science and Technology, Vol. 17, 2014, pp. 85–94. doi: 10.1016/j.jestch.2014.04.001.

[17] S. Auer, A. Bliem, D. Engel, A. Uhl, A. Unterweger, Bitstream-based JPEG Encryption in Real-time, in: International Journal of Digital Crime and Forensics (2013). doi: 10.4018/jdcf.2013070101.
[18] H. Kobayashi, H. Kiya, Bitstream-Based JPEG Image Encryption with File-Size Preserving, in: IEEE 7th Global Conference on Consumer Electronics (GCCE), 2018, pp. 1-4. doi: 10.1109/gcce.2018.8574605.
[19] K. Minemura, Z. Moayed, K. Wong, X. Qi, K. Tanaka, JPEG image scrambling without expansion in bitstream size, in: 19th IEEE International Conference on Image Processing, 2012, pp. 261-264. doi: 10.1109/ICIP.2012.6466845.
[20] A. Phatak, A Non-format Compliant Scalable RSA-based JPEG Encryption Algorithm. International Journal of Image, Graphics and Signal Processing, Vol. 8, No. 6, 2016, pp 64-71. doi: 10.5815/ijigsp.2016.06.08.
[21] Ch.-L. Tsai, Ch.-J. Chen, W.-L. Hsu, Multi-morphological image data hiding based on the application of Rubik's cubic algorithm, in: IEEE International Carnahan Conference on Security Technology (ICCST), 2012, pp. 135-139. doi: 10.1109/CCST.2012.6393548.
[22] K.-W. Wong, Image encryption using chaotic maps. Intelligent Computing Based on Chaos, Vol. 184, 2009, pp. 333–354. doi: 10.1007/978-3-540-95972-4_16.
[23] Yu. Wu, S. Agaian, J. Noonan, Sudoku Associated Two Dimensional Bijections for Image Scrambling, in: IEEE Transactions on multimedia, 2012, 30 p. URL: https://arxiv.org/abs/1207.5856v1.
[24] Y. Yang, B. B. Zhu, S. Li, N. Yu1, Efficient and Syntax-Compliant JPEG 2000 Encryption Preserving Original Fine Granularity of Scalability. EURASIP Journal on Information Security, Vol. 2007, Article ID 56365, 2008, 13 p. doi: 10.1155/2007/56365.
[25] V. Barannik, N. Barannik, Yu. Ryabukha, D. Barannik, Indirect Steganographic Embedding Method Based On Modifications of The Basis of the Polyadic System, in: 15th IEEE International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET'2020), 2020, pp. 699-702. doi: 10.1109/TCSET49122.2020.235522.
[26] V. Barannik, V. Barannik, Binomial-Polyadic Binary Data Encoding by Quantity of Series of Ones, in: 15th IEEE International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET'2020), 2020, pp. 775-780. doi: 10.1109/TCSET49122.2020.235540.
[27] V. Barannik, T. Belikova, P. Gurzhii, The model of threats to information and psychological security, taking into account the hidden information destructive impact on the subconscious of adolescents, in: 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT), 2019, pp. 656-661. doi: 10.1109/ATIT49449.2019.9030432.
[28] V. V. Barannik, M. P. Karpinski, V. V. Tverdokhleb, D. V. Barannik, V. V. Himenko, M. Aleksander, The technology of the video stream intensity controlling based on the bit-planes recombination, in: 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), 2018, pp. 25-28. doi: 10.1109/IDAACS-SWS.2018.8525560.
[29] V. V. Barannik, Yu. N. Ryabukha, O. S. Kulitsa, The method for improving security of the remote video information resource on the basis of intellectual processing of video frames in the telecommunication systems, Telecommunications and Radio Engineering, Vol. 76, No 9, 2017, pp. 785-797. doi: 10.1615/TelecomRadEng.v76.i9.40.
[30] V. Barannik, S. Shulgin, The method of increasing accessibility of the dynamic video information resource, in: 2016 13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), 2016, pp. 621-623. doi: 10.1109/TCSET.2016.7452133.
[31] V. Barannik, D. Tarasenko, Method coding efficiency segments for information technology processing video, in: 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), 2017, pp. 551-555. doi: 10.1109/INFOCOMMST.2017.8246460.
[32] Ch.-Ch. Chen, W.-J. Wu, A secure Boolean-based multi-secret image sharing scheme. Journal of Systems and Software, Vol. 92, 2014, pp. 107-114. doi: 10.1016/j.jss.2014.01.001.

[33] T.-H. Chen, Ch.-S. Wu, Efficient multi-secret image sharing based on Boolean operation. Signal Processing, Vol. 91, Iss. 1, 2011, pp. 90-97. doi: 10.1016/j.sigpro.2010.06.012.
[34] M. Deshmukh, N. Nain, M. Ahmed, An (n, n)-Multi Secret Image Sharing Scheme Using Boolean XOR and Modular Arithmetic, in: IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), 2016, pp. 690-697. doi: 10.1109/aina.2016.56.
[35] M. Naor, A. Shamir, Visual Cryptography, in: Proceedings of the Advances in Cryptology – EUROCRYPT'94. Lecture Notes in Computer Science, Vol. 950, 1995, pp. 1-12. doi: 10.1007/bfb0053419.
[36] Ch.-N. Yang, Ch.-H. Chen, S.-R. Cai, Enhanced Boolean-based multi secret image sharing scheme. Journal of Systems and Software, Vol. 116, 2016, pp. 22-34. doi: 10.1016/j.jss.2015.01.031.
[37] P. Korshunov, T. Ebrahimi, Using warping for privacy protection in video surveillance, in: 18th International Conference on Digital Signal Processing (DSP), 2015, pp. 1-6. doi: 10.1109/ICDSP.2013.6622791.
[38] A. N. Alimpiev, V. V. Barannik, S. A. Sidchenko, The method of cryptocompression presentation of videoinformation resources in a generalized structurally positioned space, Telecommunications and Radio Engineering, Vol. 76, No 6, 2017, pp. 521-534. doi: 10.1615/TelecomRadEng.v76.i6.60.(2017).
[39] G. Nattress, Chroma Sampling: An Investigation, 2008. URL: http://www.nattress.com/Chroma_Investigation/chromasampling.htm.
[40] D. A. Kerr, Chrominance Subsampling in Digital Images, 2012. URL: http://dougkerr.net/Pumpkin/articles/Subsampling.pdf.