МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Одеська політехніка»

ІНФОРМАТИКА ТА МАТЕМАТИЧНІ МЕТОДИ В МОДЕЛЮВАННІ

INFORMATICS AND MATHEMATICAL METHODS IN SIMULATION

Том 12, № 4

Volume 12, No. 4

Одеса – 2022
Odesa – 2022

Журнал внесений до переліку наукових фахових видань України (технічні науки) згідно наказу Міністерства освіти і науки України № 463 від 25.04.2013 р. Перереєстровано на категорію «Б» за фахами 121, 122, 125, 151 згідно наказу МОН України № 1473 від 26.11.2020 р.

Виходить 4 рази на рік

Заснований Одеським національним політехнічним університетом у 2011 році

Свідоцтво про державну реєстрацію КВ № 17610 - 6460Р від 04.04.2011р.

Published 4 times a year

Founded by Odesa National Polytechnic University in 2011

Certificate of State Registration КВ № 17610 - 6460Р of 04.04.2011

Друкується за рішенням редакційної колегії та Вченої ради Національного університету «Одеська політехніка»

Оригінал-макет виготовлено редакцією журналу

Адреса редакції: просп. Шевченка, 1, Одеса, 65044, Україна
Телефон: +38 048 705 8506
Web: www.immm.op.edu.ua (immm.opu.ua) E-mail: immm.ukraine@gmail.com

Editorial address: 1 Shevchenko Ave., Odesa, 65044, Ukraine
Tel.: +38 048 705 8506
Web: www.immm.op.edu.ua (immm.opu.ua) E-mail: immm.ukraine@gmail.com

ЗМІСТ/CONTENTS

DOI 10.15276/imms.v12.no4.269
УДК 681.518

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 269-279

ALGORITHM FOR AUTOMATIC DAMAGE IDENTIFICATION USING LOW-COST STRUCTURAL HEALTH MONITORING SYSTEMS

A.V. Basko, O.A. Ponomarova, Y.O. Prokopchuk

Prydniprovska State Academy of Civil Engineering and Architecture, Chernyshevskoho str., 24a, Dnipro, 49600; Ukraine; basko.artem@pgasa.dp.ua; pricmech@ukr.net; itk3@ukr.net

The analysis of existing structural health monitoring (SHM) systems and damage identification algorithms showed that monitoring systems with low-cost elements are currently used rarely, and damage identification algorithms for such systems are practically non-existent. Based on a previous analysis of existing SHM, as well as further research on feature extraction, which is mainly carried out using already recorded information for data processing. In addition, for the application of algorithms for identification, detection and assessment of possible damage to the object of monitoring. Therefore, it can be concluded that the use of existing SHM systems has a number of disadvantages related to the complexity of implementation and adjustment, as well as economic impracticality. Some studies are aimed at reducing the cost of the sensor node by using a cheaper elemental base. But data from the entire sensor network is still transmitted to the main data collection server. All this requires significant computing power to process such a large amount of data. It should be noted that with the increase of the sensor network, the speed of data transmission in it decreases. The proposed study presents the development of a modern method of automatic damage identification with the possibility of automatic assessment of the level of damage. In addition, the developed algorithm is used to obtain results in real time. Also, the algorithm is conceptually aimed at low-cost SHM systems, with the possibility of local computation directly on the sensor node. This will make it possible to compress a large amount of data as much as possible and transmit only useful information about the state of the object. The article describes in detail the algorithm for automatic data processing in real time, and provides an implementation in the C++ programming language. The results of the experiment using the proposed algorithm to check the efficiency and reliability of the work are also presented. The developed method can be used as an alternative to traditional methods, especially for low-cost SHM systems.

Keywords: damage identification algorithm, dynamic characterization, low-cost SHM system, structural response, wireless sensor network, structural health monitoring.

Introduction

Modern research in the field of structural health monitoring (SHM), as well as actually working systems installed on various architectural objects, such as bridges, buildings, wind turbines, nuclear power plants, showing the seriousness of their application [1-3]. Due to the complexity of the structures, and as a result of the use of various architectural forms and materials, influence of the environment, the life of the object may decrease and endanger nature and man. Therefore, such systems help to understand how structures react under operating [4-5]. This allows to prevent the negative consequences, caused by destructions and apply preventive procedures for periodic inspection and repair of possible damage.

A lot of data processing algorithms presented in a considerable amount research can be: the least mean squares (LMS) algorithm, the recursive least squares (RLS) algorithm and the Kalman filtering (KF) algorithm [6]. Also, one of the popular ones is the use of the generalized least square (GLS) method [7], and an algorithm based on the principle of fuzzy logic [8].

One of the popular solutions in signal processing is the use of machine learning technologies. Some methods have already shown their advantages in tasks related to text recognition, image identification, and text translation.

Bayesian analysis and decision tree are complex and rarely used due to lack of flexibility and high requirement of processing power [9]. The use of such machine learning methods as Artificial Neural Network (ANN) and Support Vector Machine (SVM) is described [10].

Different algorithms of damage recognition based on machine learning technology was described in detail. In addition, a number of characteristics were identified that a sensitive element must have for a successful process of damage identification [11].

The construction of the Low-Cost Wireless sensor has been described and the choice of components in the study carefully justified. In addition, signal processing for damage identification was presented [12].

An approach to interpreting both static and dynamic data was presented in the study. The main emphasis was placed on identifying daily and seasonal conditions, as well as possible trends associated with critical failures. The considered approaches were tested on the data set from the "Two Towers" of Bologna [13].

Among the wide variety of algorithms that researchers offer, we can distinguish the main requirements for their implementation:

- The need for precise synchronization of data from all sensors
- It is necessary to have a data set in an intact state of the monitored object.
- The impossibility of obtaining the result of calculations in real time
- The need to post-process and analyze data in MATLAB or other specialized software, which requires a qualified specialist.
- A large number of sensor nodes leads to an overload of the sensor network and it may not have time to transmit data.
- Large amount of memory for the accumulation of measurements.

For modern sensor networks, it is necessary to build systems based on dynamic calculation, local filtering, data compression, damage thresholds, and all this in automatic mode. All this will not only simplify the identification process, but also eliminate the human factor, increase the speed of the system's response to disturbing influences. Obviously, this will save money on the design of the sensor node and sensor network, as well as reduce the demands on the computational properties of the sensor node.

Data compression is one of the important requirements in modern web-based automatic damage identification. Many different data compression methods have been proposed in the scientific community [14].

Content statement of the problem. The purpose of the work is to develop an algorithm for automatic identification of damage to improve safety buildings and structures using cheap and low-power sensor nodes. The essence of the proposed automatic identification is the automatic adjustment of the algorithm using step-by-step vibration assessment for signal processing.

In addition, the work proposes an approach that allows you to automatically determine the levels of possible damage for their identification in real time.

To achieve the set goal, the following tasks are set before the work:

- analyze innovative approaches and algorithms for damage identification;
- to develop a simple and effective automatic algorithm for damage identification;
- using the MVS (Microsoft Visual Studio) environment and the C++ programming language, implement the proposed damage identification algorithm;
- determine the requirements that a sensor node must meet to implement an automatic algorithm;

▪ perform an experiment using the proposed algorithm and demonstrate the effectiveness of the algorithm;

In addition, this algorithm is very convenient to use when working with ADC values, since it is not tied to the data scale. Also, integer data is more preferred for microcontrollers because math operations on them are faster than on floating point numbers.

Algorithm of data processing. The proposed algorithm belongs to unsupervised algorithms, has advantages related to the speed of calculations and does not require initial data and long-term collection of initial data. Comparing all machine learning algorithms, it can be noted that unsupervised learning algorithms will have lower accuracy compared to supervised learning.

It is assumed that the proposed algorithm will work locally on the sensor node, and this is also the novelty of the study. And only data about the damaged state is transmitted to the central server, where the server can localize the problem.

Before considering the principle of operation of the proposed algorithm, it is necessary to determine a number of minimum technical characteristics that the sensor node must meet:

- microcontroller based on 32-bit architecture or higher;
- flash memory size not less than 512 Kbytes;
- support for one of the wireless protocols Zigbee, Zigbee Pro or LoRaWAN;
- acceleration measurement sensor sensitivity ± 4g or less
- built-in ADC with a data resolution of at least 16 bits per range and a sampling frequency of at least 200 Hz.

The developed approach can be divided into several main stages that has been described below. On the first stage to get the value of the current step S_i between two values, as the difference between the current and the previous value, as shown in equation (1).

$$S_i = |x_i - x_{i-1}|, \tag{1}$$

where x_i – is the current value and x_{i-1} – is the previous value of vibration.

Secondly, we should evaluate how our steps change. To estimate changes of steps we should follow next conditions, if $S_{i-1} > S_i$ we should follow equation (2), otherwise (3).

$$SE_i = \frac{S_i}{S_{i-1}} \cdot G, \tag{2}$$

$$SE_i = \frac{S_{i-1}}{S_i} \cdot G, \tag{3}$$

where SE_i – current step evaluation, S_i – current step, S_{i-1} – previous step, G – smoothing coefficient, the initial value was set to 0.5.

The third stage is estimating the smoothing coefficient GE, allows you to automatically evaluate the work of the smoothing coefficient. The estimation is performed by the one-dimensional measure G_koeff, which is set as a constant (it is recommended to set value from 0.1 to 0.3). For more noisy signal lower value, and higher value otherwise. So, we need to evaluate every N times, so if $S_{i-1} \leq$ G_koeff then use equation (4) otherwise use equation (5).

$$Temp\_GE = Temp\_GE + 1 \tag{4}$$

$$Temp\_GE = Temp\_GE \tag{5}$$

The GE smoothing coefficient is estimated every N values. Where N can be chosen to be F_{adc}, where F_{adc} is the sampling rate of the data. To calculate an estimated value of the smoothing coefficient, the equation (6) is used. When N values are reached, the temporary counter Temp_GE must be reset to zero.

$$GE = \frac{Temp\_GE}{N} \tag{6}$$

One of the important part of the algorithm is correction of the smoothing coefficient G. When the estimated value GE was got, we can adjust smoothing coefficient value. First, we need to define a constant GE_Limit value, it is value to which will tend to our estimated coefficient GE.

GE_Limit is chosen in the range from 0.7 to 0.9. Than the closer the value is to 1, the gain factor will tend to smooth out very much.

First, should to check if GE = GE_Limit, then the smoothing coefficient G will not change, otherwise we will check the following condition. If (|GE_Limit-GE|>0.1), the smoothing coefficient will be corrected the smoothing coefficient will be corrected in accordance with equation (7), else with equation (8).

$$G = G - \left(0.1 \cdot \frac{GE\_Limit - GE}{|GE\_Limit - GE|}\right) \quad (7)$$

$$G = G - \left(0.01 \cdot \frac{GE\_Limit - GE}{|GE\_Limit - GE|}\right) \quad (8)$$

At the final stage, the corrected value can be calculated using equation (9). It should be noted that the algorithm needs to make some certain number of periods to self-correct the smoothing coefficient. The number of periods will depend on the form of the data signal, it usually takes less than 10 periods.

$$XF_i = x_i \cdot SE_i + XF_{i-1} \cdot (1 - SE_i) \quad (9)$$

Automatic levels of damage identification. One of the advantages of the proposed method is the ability to automatically define several levels of damage identification. It should be noted that these levels will be located between the average and maximum values. In addition, damage levels are automatically retrained with changes in the input data.

For example, the first level will be more sensitive to small changes, and subsequent levels are more likely to detect critical damage.

In most of the presented systems, frequently some fixed threshold is chosen, which is greater than the arithmetic means by 5%. In the presented algorithm, the threshold metric is multidimensional and depends on the nature and rate of data change.

Further, the calculation of three levels of damage identification will be carried out. To start automatic learning damage levels, first of all need to wait a certain time. This must be done in order to train the coefficients of smoothing algorithm. Depending on the selected sample rate F_{adc}, this time might be chosen from 1 second to 1 minute. To initialize the first value of each of the levels, the algorithm chooses current value from the received data set, as shown in the equations (10-12), for each of the levels, respectively.

$$L1_i = x_i \quad (10)$$

$$L2_i = x_i \quad (11)$$

$$L3_i = x_i \quad (12)$$

Since the levels are located between the average and maximum values, so there is no any reasons to make many levels, as they will tend to the border of the maximum values.

The first level L1 is calculated relative to the average value XA of the received data. Thus, the new value updates the average value. To do this, we apply a filter that smoothes the data quite hard, as shown in the equation (13).

$$XA_i = x_i \cdot 0.01 + XA_{i-1} \cdot (1 - 0.01) \quad (13)$$

For the first level, the condition $x_{i-1} > XA_i$ must be met, if it is not met, then go to equation (15). Then we get the temporary value of the level, if $x_i < x_{i-1} < x_{i-2}$ then use equation (14), otherwise equation (15).

$$L1_{temp_i} = x_{i-1} \quad (14)$$

$$L1\_temp_i = L1\_temp_{i-1} \quad (15)$$

When the new temporary value of the level is received, then we can get the value of the current level using the following equation (16):

$$L1_i = L1\_temp_i \cdot k + L1_{i-1} \cdot (1 - k) \quad (16)$$

where k is the smoothing factor, it affects the level response speed, the choice of which depends on the sampling rate and data noise. To begin with, it is enough to choose $k=1/F_{adc}$.

The second level is calculated relative to the first level. For the second level, the condition $x_{i-1} > L1_i$ must be met, if it is not met, then go to equation (18). Then we get the temporary value of the level, if $x_i<x_{i-1}<x_{i-2}$ then use equation (17), otherwise equation (18).

$$L2_{temp_i} = x_{i-1} \quad (17)$$

$$L2\_temp_i = L2\_temp_{i-1} \quad (18)$$

When the new temporary value of the level is received, then we can get the value of the current level using the following equation (19):

$$L2_i = L2\_temp_i \cdot k + L2_{i-1} \cdot (1 - k) \quad (19)$$

Similarly, to the second level, the third level is calculated relative to the second level. For the second level, the condition $x_{i-1} > L2_i$ must be met, if it is not met, then we go to equation (21). Then we get the temporary value of the level, if $x_i<x_{i-1}<x_{i-2}$ then we use equation (20), otherwise equation (21).

$$L3_{temp_i} = x_{i-1} \quad (20)$$

$$L3\_temp_i = L3\_temp_{i-1} \quad (21)$$

When the new temporary value of the level is received, then we can get the value of the current level using the following equation (22):

$$L3_i = L3\_temp_i \cdot k + L3_{i-1} \cdot (1 - k) \quad (22)$$

Thus, for multilevel damage identification, it is sufficient to evaluate the excess of one of the levels by comparing the corrected value of XF_i and the levels $L1_i$, $L2_i$ and $L3_i$, which is displayed in equations (23-25).

$$XF_i > L1_i \quad (23)$$

$$XF_i > L2_i \quad (24)$$

$$XF_i > L3_i \quad (25)$$

Program code of proposed algorithm. For ease of understanding the operation of the calculation algorithm, the code was written in the C ++ programming language, which is shown in Fig. 1-2. The code below shows the implementation of calculations for data coming from one axis of the accelerometer. For complex structures where three-axis accelerometers are used, and maybe even more than one accelerometer, it will be rational to use an object-oriented programming approach.

```cpp
#include <cmath>
//sample rate
int Fadc = 200;
//x1-current value (xi)
//x2-previous value (xi-1)
//x3-before previous value (xi-2)
int x1 = 0, x2 = 0, x3 = 0;

//the functio "Get_x" gets current value x,
//implementation will depend on the specific sensor
int Get_x();

//S1-current value (Si)
//S2-previous value (Si-1)
int S1 = 0, S2 = 0;
//Step evaluation
float SE = 0.0f;
//smoothing coefficient G
float G = 0.5f;

//one-dimensional measure for the smoothing coefficient G
float G_koeff = 0.2f;
int Temp_GE = 0;
//estimated value of the work of the smoothing coefficient G
float GE = 0.0f;
//the value to which our estimated value tends GE
float GE_Limit = 0.8f;
//How many values need to get new G
int N = 0;
//Counter of current value N
int Ncur = 0;

//XF1-current adjusted value (XFi)
//XF2-previous adjusted value (XFi-1)
float XF1 = 0.0f, XF2 = 0.0f;

//XA1-current average (XAi)
//XA2-previous average (XAi-1)
float XA1 = 0.0f, XA2 = 0.0f;

//Current values for levels L1, L2, L3
float L1 = 0.0f, L2 = 0.0f, L3 = 0.0f;
//Previous values for levels L1, L2, L3
float L1_2 = 0.0f, L2_2 = 0.0f, L3_2 = 0.0f;
//Temporary values for calculating levels L1, L2, L3
float L1_temp = 0.0f, L1_temp2 = 0.0f;
float L2_temp = 0.0f, L2_temp2 = 0.0f;
float L3_temp = 0.0f, L3_temp2 = 0.0f;
```

a

```cpp
int main()
{
    L1 = L2 = L3 = XF1 = XF2 = x1 = x2 = x3 = Get_x();//get first value
    while(1)
    {
        x1 = Get_x();
        //Stage 1 "Getting a step"
        S1 = abs(x1 - x2);
        //Stage 2 "Step evaluation"
        if (S2 > S1)
        {
            SE = (S1 / S2)*G;
        }
        else
        {
            SE = (S2 / S1)*G;
        }
        //Stage 3 "Estimation of the smoothing coefficient"
        if (S2 <= G_koeff)
        {
            Temp_GE++;
        }
        Ncur++;
        if (Ncur >= N)
        {
            GE = Temp_GE / N;
            Temp_GE = 0;
            Ncur = 0;
        }
        //Stage 4 "Correction of the smoothing coefficient"
        if (GE != GE_Limit)
        {
            if (abs(GE_Limit - GE) > 0.1)
            {
                G -= 0.1*((GE_Limit - GE) / abs(GE_Limit - GE));
            }
            else
            {
                G -= 0.01*((GE_Limit - GE) / abs(GE_Limit - GE));
            }

        }
        //Step 5 "Get the adjusted value"
        XF1 = x1 * SE + XF2 * (1 - SE);
```

b

Fig. 1. Variable initialization (a) and main block of program (b)

```cpp
        //Stage 6 "Getting the values of the levels"

        //Level 1
        XA1 = x1 * 0.01 + XA2 * (1 - 0.01);
        if (x2 > XA1)
        {
            if (x1 < x2 && x2 < x3) {L1_temp = x2;}
            else {L1_temp = L1_temp2;}
        }
        else
        {
            L1_temp = L1_temp2;
        }
        L1_temp2 = L1_temp;
        L1 = L1_temp * 0.01 + L1_2 * (1 - 0.01);
        L1_2 = L1;

        //Level 2
        if (x2 > L1)
        {
            if (x1 < x2 && x2 < x3) {L2_temp = x2;}
            else {L2_temp = L2_temp2;}
        }
        else
        {
            L2_temp = L2_temp2;
        }
        L2_temp2 = L2_temp;
        L2 = L2_temp * 0.01 + L2_2 * (1 - 0.01);
        L2_2 = L2;

        //Level 3
        if (x2 > L2)
        {
            if (x1 < x2 && x2 < x3) {L3_temp = x2;}
            else {L3_temp = L3_temp2;}
        }
        else
        {
            L3_temp = L3_temp2;
        }
        L3_temp2 = L3_temp;
        L3 = L3_temp * 0.01 + L3_2 * (1 - 0.01);
        L3_2 = L3;
```

a

```cpp
        //Stage 7 "Identification of damage"
        if (XF1 > L1)
        {
            //Level 1 triggered
        }
        if (XF1 > L2)
        {
            //Level 3 triggered
        }
        if (XF1 > L3)
        {
            //Level 3 triggered
        }

        //Perform shift of values
        x3 = x2;
        x2 = x1;
        x1 = 0;
        S2 = S1;
        XF2 = XF1;
        XA2 = XA1;
    }
}
```

b

Fig. 2. Calculation of damage identification levels (a) and damage identification (b)

Experiments and results. In order to test and evaluate the quality of the proposed automatic damage identification algorithm, a set of acceleration data from the Hardanger Bridge [15] has been used. In this case, the data set was selected with a sampling rate of 200 Hz. It is important that there are no hardware filters for this sample. The data has the following time line from 0 to 50 seconds, the data is taken from the bridge in a calm state, and from 50 to 80 seconds, a transition is made to the storm wind data, which is shown in Fig. 3.

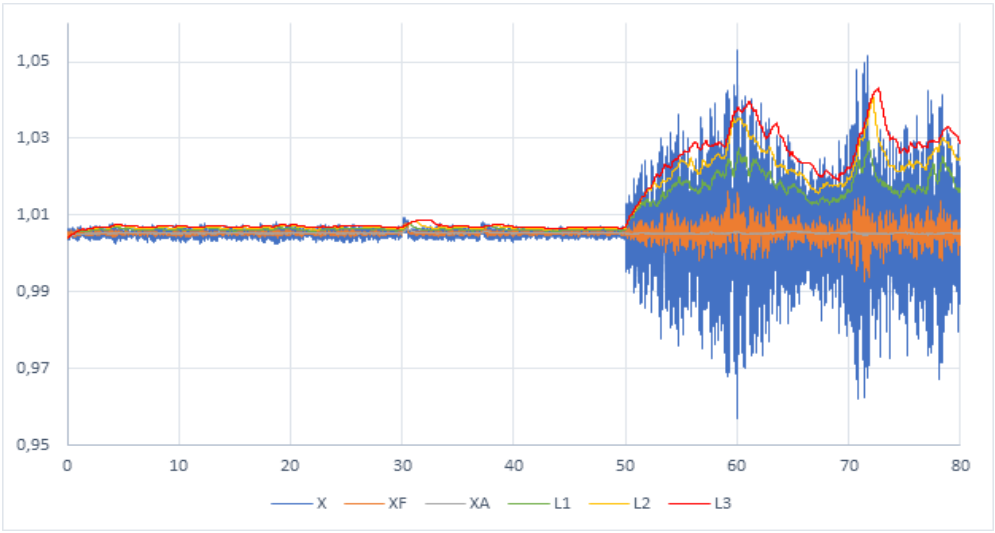


Fig. 3. Full set of training data

To understand the operation of the proposed algorithm, the test data will be displayed on a larger scale, while maintaining the time scale as shown in Fig. (4-7). In the first period of time, when the algorithm is not trained, we can create an increasing graph, which indicates the learning process, which can be seen in Fig. 4.

Fig. 4. Learning process of data

In Fig. 5, small disturbing signals are artificially created in order to show the sensitivity of the algorithm even to small signal changes with a pulse time of 500ms.

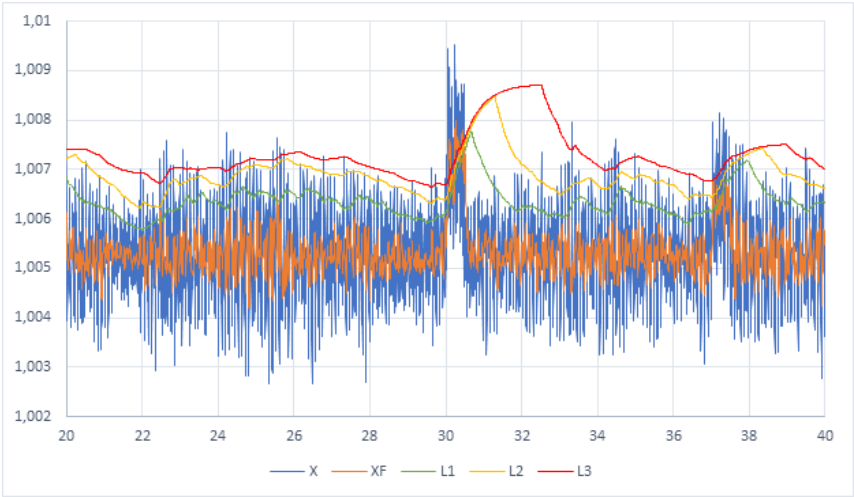


Fig. 5. Reaction on artificial disturbing signals

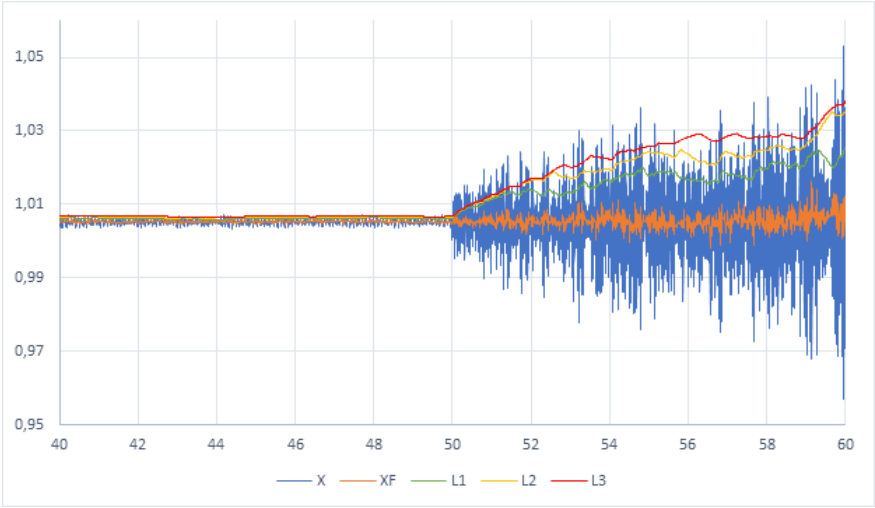


Fig. 6. Transition from quiet to stormy condition

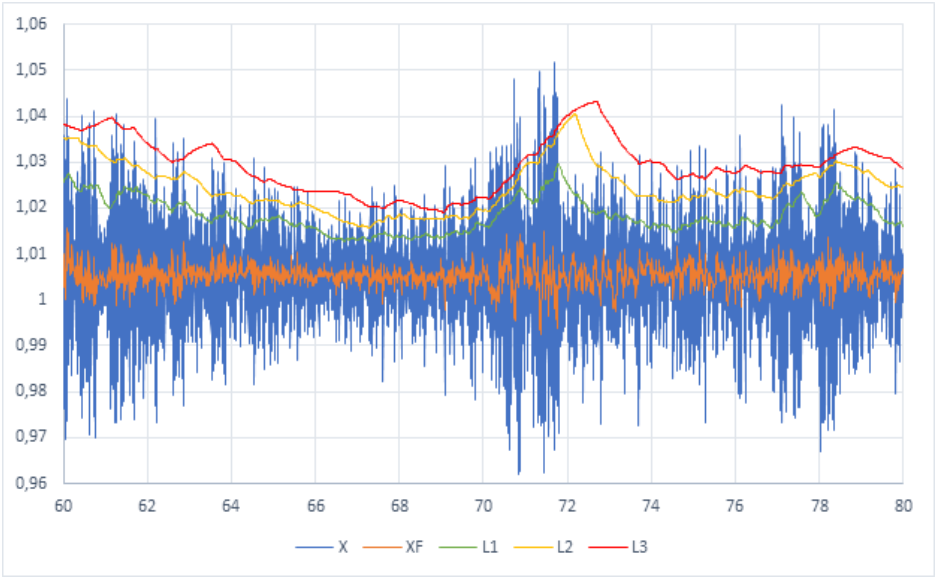


Fig. 7. Data from the bridge in a storm

The key feature of the proposed algorithm is the automatic calculation of damage identification levels. As can be seen from Fig. 8, the filtered data overlaps with all detected identification levels, which may indicate a high probability of critical damage.

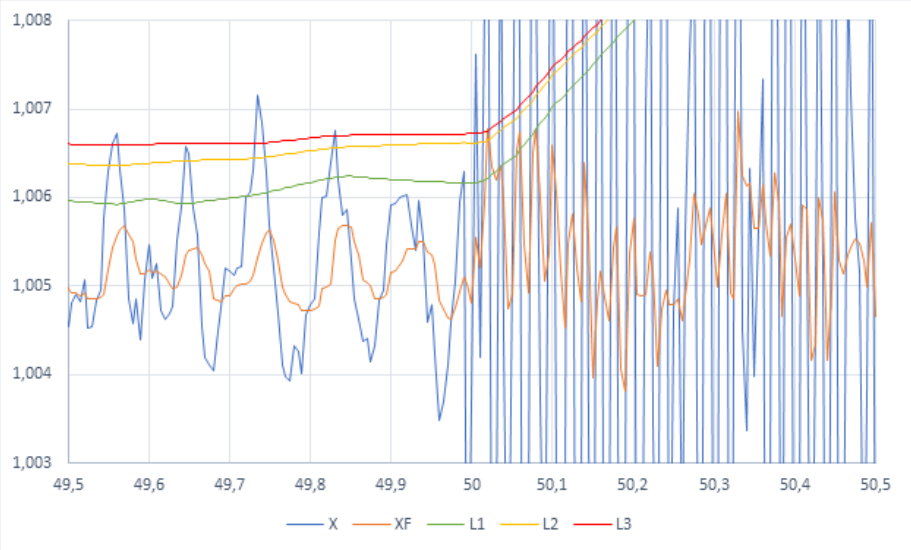


Fig. 8. Damage identification stage

Conclusion

An innovative solution in SHM, certainly was the use of wireless technologies. But existing systems are doing their best to improve by reducing their cost in order to popularize such method of monitoring. A decrease a cost of sensors by using a cheaper element base imposes restrictions on the use of data processing algorithms.

In the present study, a brief overview of the most commonly used data processing methods was made. The algorithm proposed in the study is a first step in the field of automatic identification of pressures. The described method assumes that the sensor will be automatically calibrated and adjusted without the participation of a trained specialist and a person.

Since this algorithm is the first attempt in the field of automatic identification, the next stages of development will be aimed at applying the algorithm to real systems and its subsequent improvement.

References

1. Rice J. A., Mechitov K., Sim S. H., Nagayama T., Jang S., Kim R., Spencer B. F. J., Agha G., Fujino Y. Flexible smart sensor framework for autonomous structural health monitoring. *Smart Structures and Systems*. 2010. Vol. 6. P. 423–438, DOI: https://www.doi.org/10.12989/sss.2010.6.5_6.423
2. Araujo A., García-Palacios J., Blesa J., Tirado-Andrés F., Romero E., Samartín A., Nieto-Taladriz O. Wireless measurement system for structural health monitoring with high time-synchronization accuracy. *IEEE Transactions on Instrumentation and Measurement*. 2012. Vol. 61, Issue. 3. P. 801-810. DOI: https://www.doi.org/10.1109/TIM.2011.2170889
3. Fu Y., Hoang T., Mechitov K., Kim J., Zhang D., Spencer B. Sudden event monitoring of civil infrastructure using demand-based wireless smart sensors. *Sensors*. 2018. Vol. 18, Issue. 12. P. 1-17. DOI: https://www.doi.org/10.3390/s18124480
4. Hailing F., Khodaei Z.S., Ferri Aliabadi M.H. An event-triggered energy-efficient wireless structural health monitoring system for impact detection in composite airframes. *IEEE Internet of Things Journal*. 2019. Vol. 6, Issue. 1. P. 1183-1192. DOI: https://www.doi.org/10.1109/JIOT.2018.2867722

5. Wang P., Yan Y., Gui Y.T., Bouzid O., Ding Z. Investigation of wireless sensor networks for structural health monitoring. *Journal of Sensors.* 2012. Vol. 2012. P. 1-7. DOI: https://www.doi.org/10.1155/2012/156329
6. Kluga A., Kluga J. Dynamic Data Processing with Kaiman Filter. *Elektronika Ir Elektrotechnika*. 2011. Vol. 111, Issue. 5. P. 33–36. DOI: https://doi.org/10.5755/j01.eee.111.5.351
7. Yang Z., Gu Z., Zhang W. Dynamic data processing method based on generalized least square method. *Journal of Physics: Conference Series*. 2021. Vol. 2108, Issue. 1. P. 1–6. DOI: https://doi.org/10.1088/1742-6596/2108/1/012057
8. Gorski J., Dziendzikowski M., Dworakowski Z. Recommendation System for Signal Processing in SHM. *Artificial Intelligence and Soft Computing.* 2021. Vol. 12854. P. 328–337. DOI: https://doi.org/10.1007/978-3-030-87986-0_29
9. Gordan M., Razak H.A., Ismail Z., Ghaedi K. Recent Developments in Damage Identification of Structures Using Data Mining. *Latin American Journal of Solids and Structures.* 2017. Vol. 14, Issue. 13. P. 2373–2401. DOI: https://doi.org/10.1590/1679-78254378
10. Finotti R. P., Cury A. A., Barbosa F. D. S. An SHM approach using machine learning and statistical indicators extracted from raw dynamic measurements. *Latin American Journal of Solids and Structures*. 2019. Vol. 16, Issue. 2. P. 1–17 DOI: https://doi.org/10.1590/1679-78254942
11. Figueiredo E., Santos A. Machine Learning Algorithms for Damage Detection. *Vibration-Based Techniques for Damage Detection and Localization in Engineering Structures*. 2018. P. 1–39. DOI: https://doi.org/10.1142/9781786344977_0001
12. Caballero-Russi D., Ortiz A. R., Guzman A., Canchila C. Design and Validation of a Low-Cost Structural Health Monitoring System for Dynamic Characterization of Structures. *Applied Sciences.* 2022. Vol. 12, Issue. 6. P. 1–28. DOI: https://doi.org/10.3390/app12062807
13. Baraccani S., Palermo M., Azzara R. M., Gasparini G., Silvestri S., Trombetti T. Structural Interpretation of Data from Static and Dynamic Structural Health Monitoring of Monumental Buildings. *Key Engineering Materials*. 2017. Vol. 747. P. 431–439. DOI: https://doi.org/10.4028/www.scientific.net/kem.747.431
14. Aime J. O. O., Jean P. N., Guy-Germain A., Pierre E. Compression of Vibration Data by the Walsh-Hadamard Transform. *Journal of Engineering and Applied Sciences*. 2020. Vol. 15, Issue. 10. P. 2256–2260.
15. Fenerci A., Kvale K., Petersen Y., Ronnquist A., Oiseth O. Data Set from Long-Term Wind and Acceleration Monitoring of the Hardanger Bridge. *Journal of Structural Engineering*. 2021. Vol. 147, Issue. 5. P. 1–13. DOI: https://doi.org/10.1061/(asce)st.1943-541x.0002997

АЛГОРИТМ АВТОМАТИЧНОЇ ІДЕНТИФІКАЦІЇ ПОШКОДЖЕНЬ З ВИКОРИСТАННЯМ НЕДОРОГИХ СИСТЕМ МОНІТОРИНГУ СТАНУ КОНСТРУКЦІЙ

А.В. Басько, О.А. Пономарьова, Ю.О. Прокопчук

Придніпровська державна академія будівництва та архітектури, Chernyshevskoho str., 24a, Dnipro, 49600; Ukraine; basko.artem@pgasa.dp.ua; pricmech@ukr.net; itk3@ukr.net

Аналіз існуючих систем моніторингу стану конструкцій (SHM) та алгоритмів ідентифікації пошкоджень показав, що системи моніторингу з використанням маловартісних елементів на сьогодні використовується дуже мало, а алгоритми ідентифікації пошкоджень для таких систем практично відсутні. Ґрунтуючись на першочерговому аналізі існуючих SHM, а також на подальших дослідженнях щодо виділення ознак, які в основному здійснюється з використанням уже записаної інформації для обробки даних. Крім того для застосування алгоритмів ідентифікації, виявлення та оцінки можливих збитків об’єкту моніторинг. Отже можна зробити висновок, що використання існуючих систем SHM, має ряд недоліків пов'язаних зі складністю імплементації та налаштування, а також з економічною недоцільністю. Деякі дослідження спрямовані на зниження вартості сенсорного вузла за рахунок використання більш дешевої елементної бази. Але дані з усієї сенсорної мережі все одно передаються на головний сервер збору даних. Все це вимагає значної обчислювальної потужності для обробки такої великої кількості даних. Слід зауважити, що зі збільшенням сенсорної мережі швидкість передачі даних у ній знижується. Запропоноване дослідження представляє розробку сучасного методу автоматичної ідентифікації пошкоджень із можливістю автоматичної оцінки рівня пошкоджень. Крім того, розроблений алгоритм використовується для отримання результатів у реальному часі. Також алгоритм концептуально націлений на недорогі системи SHM, з можливістю локального обчислення безпосередньо на сенсорному вузлі. Це дозволить максимально стиснути великий об’єм даних і передавати лише корисну інформацію про стан об’єкта. У статті детально описано алгоритм автоматичної обробки даних в режимі реального часу, наведено реалізацію на мові програмування C++. Також представлені результати експерименту із застосуванням запропонованого алгоритму для перевірки ефективності та надійності роботи. Розроблений метод може бути використаний як альтернатива традиційним методам, особливо для недорогих систем SHM.

Ключові слова: алгоритм ідентифікації пошкоджень, динамічна характеристика, недорогі системи SHM, структурна реакція, бездротова сенсорна мережа, структурний моніторинг стану.

DOI 10.15276/imms.v12.no4.280
UDC 004.056.5

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 280-287

McELIECE CRYPTOSYSTEM BASED ON QUATERNARY HAMMING CODES

D.A. Isakov, A.V. Sokolov

National Odesa Polytechnic University
Ukraine, Odesa, 65044, Shevchenko Ave., 1, radiosquid@gmail.com

The operation of modern information protection systems is largely based on asymmetric cryptographic algorithms that allow encrypted information to be transmitted over an open communication channel without the need for prior key exchange. Modern asymmetric cryptographic algorithms are based on the use of such one-sided functions as factorization of large prime numbers and discrete logarithms, which require significant computational costs for their use, and are not resistant to promising quantum cryptanalysis attacks. Existing modifications of these cryptographic algorithms based on elliptic curves are also characterized by some significant drawbacks: many robust elliptic curves are currently patented, and algorithms on elliptic curves often require the use of powerful physical generators of truly random numbers. The solution to these problems is the use of the McEliece cryptographic system, which is based on the problem of decoding complete linear codes. Despite the prevailing performance of this system and its resistance to promising quantum cryptanalysis attacks, such a shortcoming as the large length of its public key has led to the fact that it is not often used in practice. In this paper, new families of Hamming (n,k)-codes in extensions of extended Galois fields are proposed and it is shown that on the basis of these codes a McEliece cryptosystem can be built, which is characterized by a much smaller size of the public key while providing a comparable number of generator matrices of the code so providing the comparable value of the protection levels number. The possibility of applying cascade Hamming codes on the extensions of extended Galois fields is shown, which makes it possible to obtain protection against the Sidelnikov attack. The proposed cryptosystem can be recommended for practical use in applications that require high speed (for example, mobile devices, IoT devices, and embedded systems), as well as significant cryptographic strength, including protection against promising quantum attacks.
Keywords: asymmetric cryptography, McEliece cryptographic system, Hamming codes, quaternary logic.

Introduction and statement of the problem

Asymmetric cryptographic algorithms are a crucial element of modern information protection systems, that largely determines their effectiveness and performance [1]. Today the use of asymmetric cryptography makes it possible to solve such basic tasks of information protection systems as key distribution, user authentication, confirmation of message authorship, protection of the software, etc.

Among the most widespread in practice [2] asymmetric cryptographic algorithms, it is necessary to note the Diffie-Hellman key distribution system, as well as RSA and El-Gamal full-fledged asymmetric cryptographic algorithms, which are based on the task of the complexity of factorization of large numbers and calculation of discrete logarithms. Despite the breakthrough nature of these cryptographic algorithms and their indisputably high importance in the tasks of ensuring the security of modern information technologies, they are characterized by some significant drawbacks, including high computational complexity of encryption/decryption algorithms, strict requirements for key information, vulnerability to promising quantum cryptanalysis attacks [3]. The computational complexity of these asymmetric cryptographic algorithms significantly limits their use on modern resource-constrained platforms: mobile devices, IoT (Internet of Things) devices, UAV (Unmanned Aerial Vehicles), and also leads to the fact that everywhere in practical

information transmission systems, asymmetric cryptographic algorithms are used once to exchange a key information, after which encryption of the main data arrays is performed using faster symmetric cryptographic algorithms.

The existence of effective quantum algorithms [4] for the factorization of big numbers and calculation of discrete logarithms, as well as the significant pace of development of quantum computers, is a quite real threat to the security of asymmetric cryptographic algorithms based on these computationally complex problems, which led to the appearance of modifications of the Diffie-Hellman, RSA, and El-Gamal cryptographic algorithms based on the elliptic curves [5]. Nevertheless, even these modifications are characterized by some significant drawbacks: the computational complexity remains high, not all elliptic curves provide a sufficient level of security, many robust elliptic curves are currently patented, algorithms on elliptic curves often require the use of powerful physical generators of truly random numbers, which leads to the complication and increase in the cost of the devices on which they are used.

As leading modern research shows, one of the powerful solutions of post-quantum cryptography, which is characterized by a significant reduction in computational complexity (and, therefore, is potentially suitable for use on resource-constrained devices), is crypto-code constructions [6], among which the McEliece cryptosystem is the most widely used solution. This cryptosystem is characterized by a lower level of computational complexity when compared to RSA and El-Gamal [7] cryptographic algorithms and a much higher level of resistance to quantum cryptanalysis attacks. Nevertheless, despite its advantages, the classical McEliece cryptographic system is characterized by a large amount of key information, which is necessary to achieve a sufficient level of cryptographic strength, which significantly limits its use in practice.

The *purpose* of this paper is to reduce the size of the public key in the McEliece cryptographic system while preserving its cryptographic strength using quaternary Hamming codes.

Quaternary Hamming codes based on the extensions of extended Galois fields

Hamming error-correcting codes are linear codes for the binary alphabet with code distance $d = 3$, which allow detecting double and correcting single errors [8]. Today, Hamming codes are quite well-researched and widely used in practice. However, despite the significant scientific results obtained in the theory of error-correcting coding for classes of binary linear codes, a large number of issues related to the application of non-binary codes remain unsolved. The conceptual idea of quaternary Hamming codes was introduced in [9].

The performed research shows that Hamming codes can be constructed not only for the extended Galois field $GF(2^2)$, but also for the extensions of the extended Galois fields $GF(q^r)$, $q = 2^u, u = 2,3,...$ which were introduced in [10, 11]. This fact allows us to talk about the existence of new families of Hamming codes in the extensions of extended Galois fields. For example, let the Galois field $GF(q^r)$, where r is the number of parity symbols in the Hamming code being constructed. Then the parameters of the code will be determined by the following ratios: the total number of codeword symbols $n = \frac{q^r - 1}{q - 1}$, the number of data symbols $k = \frac{q^r - 1}{q - 1} - r$, the number of errors detected by the code $f = 2$, the number of errors corrected by the code $t = 1$.

In this paper, we consider an extended Galois field $GF(2^2)$, the arithmetic of which can be constructed according to the only existing primitive irreducible polynomial $f(x) = x^2 + x + 1$ of degree $\deg(f(x)) = 2$. We provide the tables of addition and

multiplication in the Galois field $GF(2^2) = GF(4)$, the arithmetic of which is determined by the previously mentioned polynomial.

+	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

,

·	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	3	1	2

. (1)

Considering relation (1), we can talk about the existence of new families of Hamming codes based on the extension of extended Galois field $GF(4^r)$ the first of which has the following parameters: (5, 3), (21, 18), (85, 81) for $r = 2,3,4$. These codes can be specified using a parity-check matrix H, which is determined by the following relation

$$H = [H_{r,k} \mid I_{r,r}], \tag{2}$$

where $I_{r,r}$ is the matrix of order r whose columns have unit Hamming weight; $H_{r,k}$ is the matrix of size $r \times k$ that contains columns which are orthogonal in the Galois field $GF(4^r)$ to each other, as well as to the columns of the matrix $I_{r,r}$. Therefore, by construction, all columns of the matrix H are mutually orthogonal.

On the basis of the parity-check matrix, the generator matrix can be constructed by concatenating the matrix $I_{k,k}$ of order k whose columns have unit Hamming weight and transposed matrix $H_{r,k}$

$$G = [I_{k,k} \mid -H_{r,k}^T], \tag{3}$$

where T is the transposition symbol.

Consider the example of constructing a Hamming (5,3)-code based on the extension of the extended Galois field $GF(4^2)$. The Galois field $GF(4^2)$ contains 15 nonzero elements, each of which we can represent as a quaternary vector. At the same time, the specified elements form $(q^r - 1)/3$ classes, which contain linearly dependent vectors, in other words, those that can be obtained from each other by multiplying by a constant $a \in \{1,2,3\}$ in the Galois field $GF(4^2)$, i.e., according to the multiplication table (1). For our example, there are $(4^2 - 1)/3 = 5$ classes of linearly independent quaternary vector elements of the Galois field $GF(4^2)$ which are represented by different colors

$$V = \{01,02,03,10,11,12,13,20,21,22,23,30,31,32,33\}. \tag{4}$$

Choosing one vector from each class, we can construct the parity-check matrix H in accordance with (2)

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 0 & 1 \end{bmatrix}, \tag{5}$$

which completely defines the parity-check equations

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = s_1; \\ x_1 + 2x_2 + 3x_3 + x_5 = s_2. \end{cases} \tag{6}$$

where arithmetic operations are performed according to tables (1).

Let us also point out that the columns of the H matrix can be arranged in an arbitrary order, just as any vectors from the equivalent classes (4) can be chosen. In performing these operations, the correcting ability of the code does not change.

Note that decoding of information, as in the case of Hamming binary codes, can be performed on the basis of the syndrome method [8]. The specific value of the syndrome

$S = HC'^T = [s_1 \quad s_2]^T$ coincides with the column of the matrix H, which corresponds to the symbol where the error occurred, which is multiplied by the amplitude of the error. Thus, it becomes possible to correct the error that occurred after calculating the syndrome.

On the basis of the parity-check matrix, a generator matrix can be constructed in accordance with (3)

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 3 \end{bmatrix}, \tag{7}$$

which defines the encoding equations for the parity-check symbols

$$\begin{cases} x_4 = x_1 + x_2 + x_3; \\ x_5 = x_1 + 2x_2 + 3x_3. \end{cases} \tag{8}$$

Consider an example of the operation of the proposed Hamming correction code. Let an information message $A = [0 \quad 1 \quad 2]$ be given, which we will encode with the proposed Hamming code. Let's suppose that an error with a value of amplitude equal to 2 occurred in the second element of the codeword (which corresponds to the second information element), i.e., the error vector is equal to $e = [0 \quad 2 \quad 0 \quad 0 \quad 0]$, while the received codeword will have the form $C' = [0 \quad 3 \quad 2 \quad 3 \quad 3]$.

On the receiving side, we calculate the value of the syndrome, which will be equal to $S = [2 \quad 3]^T$. Since the obtained syndrome corresponds to the second column of the matrix multiplied by a constant $a = 2$, we conclude that the error occurred in the second element of the codeword, while the amplitude of the change was equal to 2. This allows us to reproduce the correct codeword C.

The algorithm for generation of parity-check matrices for quaternary Hamming codes

Note that in the case of using extensions of extended Galois fields, in particular, for the construction of Hamming quaternary codes, there is an urgent need to select linearly independent vectors in the complete quaternary vector code, i.e. to classify it into classes, each of which contains 3 linearly independent vectors. Solving this problem for large values of r can be quite computationally complex, as it involves sorting through a set of $q^r - 1$ elements.

To solve this problem, the proposed algorithm for generating the code of all possible vectors that does not contain their linear combinations can be applied:

Step 1. For the value $r = 1$, this code contains only one codeword $\{1\}$.

Step 2. For a given value of r, the code of linearly independent vectors is constructed as follows based on the code of linearly independent vectors of length $r - 1$.

Step 2.1. Concatenate the symbol "0" to each codeword of the code of linearly independent vectors of length $r - 1$ as the most significant element of the codeword.

Step 2.2. The rest of the codewords of the code of linearly independent vectors are constructed by concatenating the symbol "1" as the most significant element of the codeword to the complete code of quaternary vectors of length $r - 1$.

For example, with the help of the proposed algorithm, we will construct a code of linearly independent vectors for the value $r = 2$

$$\{0 \quad 1\}; \quad \{1 \quad 0\}; \quad \{1 \quad 1\}; \quad \{1 \quad 2\}; \quad \{1 \quad 3\}, \tag{9}$$

on the basis of which the code of linearly independent vectors for the value $r = 3$ can be built

$$\begin{array}{ccccccccccc} \{0 & 0 & 1\}; & \{0 & 1 & 0\}; & \{0 & 1 & 1\}; & \{0 & 1 & 2\}; & \{0 & 1 & 3\}; \\ \{1 & 0 & 0\}; & \{1 & 0 & 1\}; & \{1 & 0 & 2\}; & \{1 & 0 & 3\}; & & & \\ \{1 & 1 & 0\}; & \{1 & 1 & 1\}; & \{1 & 1 & 2\}; & \{1 & 1 & 3\}; & & & \\ \{1 & 2 & 0\}; & \{1 & 2 & 1\}; & \{1 & 2 & 2\}; & \{1 & 2 & 3\}; & & & \\ \{1 & 3 & 0\}; & \{1 & 3 & 1\}; & \{1 & 3 & 2\}; & \{1 & 3 & 3\}, & & & \end{array} \quad (10)$$

and so on.

We note that to build the parity-check matrix of the Hamming code, as its columns the linearly independent vectors of the code of linearly independent vectors can be taken in an unchanged form, or in the form of their linear combinations obtained by multiplying them by some constants $a_i, i = 1, 2, ..., {4^k}/{3} - 1$.

McEliece cryptographic system based on quaternary Hamming codes

The proposed quaternary Hamming codes could be the basis for the McEliece cryptographic system, while giving it specific advantages over existing variants of this cryptographic system based on other codes. Let's consider the algorithms of the McEliece cryptographic system based on quaternary Hamming codes in the mode of transmitting information from party B (Bob) to party A (Alice), accompanying it with specific examples.

The algorithm for generation of key information

Step 1. Alice chooses an (*n*,*k*)-linear code Υ that corrects one error. Then the generator matrix G of size $k \times n$ is calculated for the code Υ.

Step 2. In order to complicate the task of recovering the original code, Alice generates a random non-singular matrix S of size $k \times n$ over the alphabet $\{0,1\}$.

Step 3. Alice generates an arbitrary permutation matrix P of size $n \times n$.

Step 4. Alice computes the matrix $G' = SGP$ of size $k \times n$, which is considered as a public key. A private key is a set $\{S, G, P\}$.

As an example, let's choose the generator matrix (7) synthesized by us in Section 2, and also generate a random non-singular matrix S and a permutation matrix P

$$S = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}; \quad P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \quad (11)$$

We calculate the public key matrix

$$G' = SGP = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 3 & 0 & 1 \\ 1 & 0 & 2 & 1 & 0 \\ 0 & 1 & 2 & 0 & 1 \end{bmatrix}, \quad (12)$$

for the storage or transmitting of which in the binary form we will need $2kn$ bits, since the elements of this matrix are quaternary.

The algorithm for information encryption

We will describe the encryption algorithm in the form of specific steps.

Step 1. Bob represents his message *m* as sequences of quaternary characters x of length *k*.

Step 2. Bob generates a random vector *e* of length *n* with Hamming weight *t*.

Step 3. Bob calculates the ciphertext as $y = xG' + e$ and transmits it to Alice.

Suppose that after receiving the public key G' from Alice, Bob formed a message $x = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}$ and also chose an error vector $\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix}$. Bob can then encrypt his plaintext message

$$y=\begin{bmatrix}1 & 2 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 3 & 0 & 1\\ 1 & 0 & 2 & 1 & 0\\ 0 & 1 & 2 & 0 & 1\end{bmatrix}+\begin{bmatrix}0 & 1 & 0 & 0 & 0\end{bmatrix}=\begin{bmatrix}3 & 0 & 2 & 2 & 0\end{bmatrix}. \quad (13)$$

After encryption, the open message is transfered to Alice, who performs the following steps to decrypt the message using her private key.

The algorithm for information decryption

Step 1. Alice calculates the inverse matrix P^{-1}.

Step 2. Alice calculates $\hat{y}=yP^{-1}$.

Step 3. Alice uses the Υ code decoding algorithm to obtain from $\hat{y}$ the value of $\hat{x}$.

Step 4. Alice calculates $x=\hat{x}S^{-1}$.

For our example

$$P^{-1}=\begin{bmatrix}0 & 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\end{bmatrix}, \quad (14)$$

while $\hat{y}=\begin{bmatrix}0 & 2 & 0 & 3 & 2\end{bmatrix}$.

Applying the algorithm for quaternary decoding, we see that the syndrome is equal to $[1\ 1]$: due to permutations, the error has moved to the 1st symbol $eP^{-1}=\begin{bmatrix}1 & 0 & 0 & 0 & 0\end{bmatrix}$. Since the obtained syndrome corresponds to the 1-th column of the matrix H, and not to its linear combination, we conclude that the amplitude of the error was 1, i.e., the corrected codeword is $\hat{\hat{y}}=\begin{bmatrix}1 & 2 & 0 & 3 & 2\end{bmatrix}$, while $\hat{x}=\begin{bmatrix}1 & 2 & 0\end{bmatrix}$. We find that $x=[1\ \ 2\ \ 1]$, which corresponds to the original text.

Security of the system

In the case of using binary Hamming codes, as proposed in work [12] for the organization of the McEliece cryptographic system, the number of possible generator matrices G will be determined as $n!$ possible permutations of its rows due to the determined properties of the Hamming code. At the same time, the length of the open key will be kn. On the other hand, in the case of applying the quaternary Hamming code, we can choose each of the columns of the matrix in one of 3 ways from a set of its linear combinations, i.e., we get $(3n)!$ different matrices. At the same time, the length of the open key will be $2kn$.

For example, when using a binary Hamming (63,57)-code, the number of possible generator matrices will be $1.9826\cdot10^{87}$ while 3591 bits are required to store the public key, while a quaternary Hamming (21, 18)-code with the same number of generator matrices will require only 756 bits of the public key, i.e., in 4.75 times less. At the same time, due to the fact that the McEliece cryptographic system is based on Hamming codes in extensions of extended Galois fields it is able to simultaneously process a larger amount of input information, it potentially allows faster algorithmic implementations (compared to binary Hamming codes).

We also note that in the general case, the solution for systems of linear equations in extensions of the extended Galois fields is a more complicated task from a computational point of view, while for larger values of u, the specific Galois field isomorphism can be part of the secret key.

It is considered promising to use the McEliece cryptographic system based on the cascade quaternary Hamming codes to ensure security against the Sidelnikov attack. So, for example, 16 plaintext vectors encrypted by the McEliece cryptographic system based on the Hamming (21,18)-code can be re-encrypted by the McEliece cryptographic system

based on the Hamming (341,336)-code, which should ensure a high complexity of the relationship between the elements of the input and output data.

We note that more detailed security aspects of the proposed modification of the McEliece cryptographic system based on Hamming codes in the extensions of extended Galois fields is considered as subject of futher research.

Conclusions

Let's note the main results of the research performed:

1. New families of Hamming (n,k)-codes in the extensions of extended Galois fields $GF(q^r)$, where $q = 2^u, u = 2,3,\ldots$, are proposed. The properties of these codes as well as their ability to detect errors and correct errors have been established. The features of the application of the syndrome decoding algorithm for the developed families of Hamming codes are established.

2. An asymmetric McEliece cryptographic system based on Hamming codes in the extensions of extended Galois fields is proposed. It is shown that the use of extensions of extended Galois fields allows to provide a larger number of possible generator matrices of the Hamming code (i.e., a larger number of protection levels) with a shorter length of the public key. Thus, in the case of using a quaternary Hamming (21,18)-code instead of a binary Hamming (63,57)-code, with an equal number of possible generator matrices, the length of the public key will decrease by the factor of 4.75 times.

3. The proposed cryptographic system, based on the classical McEliece cryptographic system, is resistant to potential attacks of quantum cryptanalysis, has smaller values of the length of the public key, and therefore can be recommended for practical use in applications that require efficient and secure asymmetric cryptographic algorithms.

References

1. Al-Shabi M.A. A survey on symmetric and asymmetric cryptography algorithms in information security. *International Journal of Scientific and Research Publications (IJSRP)*. 2019. Vol. 9, No. 3. P. 576-589.
2. Schneier B. Applied Cryptography: Protocols, Algorithms, and Source Code in C. Wiley, 1996. 758 p.
3. Mavroeidis V. The impact of quantum computing on present cryptography. International *Journal of Advanced Computer Science and Applications*. 2018. Vol. 9, No. 3. P. 1-10.
4. Shor P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*. 1999. Vol. 41, No. 2. P. 303-332.
5. Washington L.C. Elliptic curves: number theory and cryptography. Chapman and Hall/CRC, 2008. 531 p.
6. Yevseiev S., Korol O., Kots H. Construction of hybrid security systems based on the crypto-code structures and flawed codes. *Eastern-European Journal of Enterprise Technologies*. 2017. Vol. 4, No. 9 (88). P. 4-21.
7. Tsyganenko A.S., Yevseiev S.P. Melnik K.V. McEliece and Niederreiter asymmetric crypto-code constructions in post-quantum cryptography. ITSEC: the 8th Intern. Sci. Conf. Kyiv: NAU, 2018. P. 26-27.
8. Mazurkov M.I. Fundamentals of the theory of information transmission. Odesa: Science and Technology, 2005. 168 p.
9. Yiyi H., Junbiao D. Quaternary checksum, redundancy and Hamming code. Preprint at Researchgate, 2022. P. 1-8.
10. Mazurkov M. I., Konopaka E. A. Families of linear recurrent sequences based on complete sets of isomorphic Galois fields. *Proceedings of universities. Radioelectronics.* 2005. No. 11. P. 58-65.

11. Mazurkov M.I., Sokolov A.V. Nonlinear transformations based on complete classes of isomorphic and automorphic representations of field GF(256). *Radioelectronics and Communications Systems.* 2013. Vol. 56, No. 11. P. 513-521.
12. Kabeya T.C. McEliece's Crypto System based on the Hamming Cyclic Codes. International *Journal of Innovative Science and Research Technology*. 2019. Vol. 4, No 7. P. 293-296.

КРИПТОСИСТЕМА McELIECE НА ОСНОВІ ЧЕТВІРКОВИХ КОДІВ ГЕМІНҐА

Д.А. Ісаков, А.В. Соколов

Національний університет «Одеська політехніка»
Україна, Одеса, 65044, пр-т Шевченка, 1, radiosquid@gmail.com

Застосування сучасних систем захисту інформації у великій мірі базується на асиметричних криптографічних алгоритмах, що дозволяють передавати зашифровану інформацію відкритим каналом зв'язку без необхідності попереднього обміну ключами по додатковому закритому каналу. Сучасні асиметричні криптографічні алгоритми засновані на використанні таких односторонніх функцій як факторизація великих простих чисел та дискретне логарифмування, які вимагають значні обчислювальні затрати для свого застосування, є нестійкими до атак квантового криптоаналізу. Існуючі модифікації цих криптоалгоритмів на основі еліптичних кривих також не позбавлені суттєвих недоліків: багато «вдалих» еліптичних кривих на сьогодні запатентовані, алгоритми на еліптичних кривих часто потребують застосування потужних фізичних генераторів істинно випадкових чисел. Вирішенням зазначених проблем є застосування криптосистеми МакЕліса, яка базується на проблемі декодування повних лінійних кодів. Незважаючи на переважаючу швидкодію даної системи та її стійкість до атак квантового криптоаналізу, такий її недолік, як великі довжини її відкритого ключа, призвів до того, що вона нечасто застосовується на практиці. У даній роботі запропоновано нові сімейства (n,k)-кодів Гемінґа над розширеннями розширених полів та показано, що на основі даних кодів може бути побудована криптосистема МакЕліса, що характеризується значно меншим розміром відкритого ключа при сумірному рівні кількості генераторних матриць коду. Показана можливість застосування каскадних кодів Гемінґа над розширеннями розширених полів, що дозволяє отримати захист від атаки Сідельникова. Запропонована криптосистема може бути рекомендована до практичного використання у застосунках, що потребують великої швидкодії (наприклад, мобільних пристроях, пристроях IoT, вбудовуваних системах), а також значної криптостійкості, у тому числі, захищеності від перспективних квантових атак.

Ключові слова: асиметрична криптографія, криптографічна система МакЕліса, коди Гемінґа, четвіркова логіка.

DOI 10.15276/imms.v12.no4.288
УДК 519.85.004.42

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 288-295

A PROPOSAL OF STORAGE FOR CHROMOSOMAL DATA

O. Pysarchuk[1], Yu. Mironov[2]

[1]Igor Sikorsky Kyiv Polytechnic Institute. Peremogy Ave., 37, Kyiv, Ukraine, 3056, PlatinumPA2212@gmail.com
[2]National Aviation University, Guzar Lubomir Ave., Kyiv, Ukraine, 3056, 03058, yuriymironov96@gmail.com[2]

This article considers a problem of chromosomal pathologies detection. Chromosomal pathologies are dangerous and pose a great threat for family planning. To address them, the karyotyping process is conducted. Currently this process is manual or semi-manual, despite high effort and error cost. So, there is a need for automation of the process. This process (and automation algorithm) can be separated into different stages with various objectives. However, the shared element among these stages is format that allows to store and manage data efficiently. The goal of this paper is to propose such format. The paper revises peculiarities of karyotyping process and briefly describes steps of the pathology detection algorithm. It also considers common formats for bioinformatics data. However, their efficiency is debatable, since data stored in these formats is redundant for the task at hand. After that, a new custom data format is proposed. This format represents main entities involved in the process of anomaly recognition. Several fragments of algorithm are considered, and their complexity is estimated combined with proposed data format. As a result, this paper proposes a new data storage format used in a chromosome abnormalities recognition algorithm, and metrics that can be used to make measurable improvement over the proposed format.

Keywords: Algorithm complexity, data analysis, domain-driven design

Introduction

Congenital diseases have a significant impact on survivability rate of newborn children, including early deaths, miscarriages and stillbirth [1, 2]. Apart from this, further quality of life of surviving children may also be affected. A major subcategory of congenital diseases are chromosomal diseases [1].

These risks and issues are addressed by reproductive medicine. It introduces a wide variety of methods that handle various problems related to family planning and reproductive health. To solve the tasks of reproductive medicine, techniques of other branches of biology is used – for instance, cytogenetics. One of the main techniques for diagnosing chromosomal diseases is karyotyping [3].

Karyotyping is a process that implies an analysis of biological materials of parents/fetus, getting a visual depiction of their chromosomes and comparing chromosomes to ideograms – schematic depiction of "ideal" chromosomes [4]. Ideal chromosomes, called ideograms (fig. 1.A), are actually just a reference schematic representation used "by eye" to identify real chromosomes (fig. 1.B).

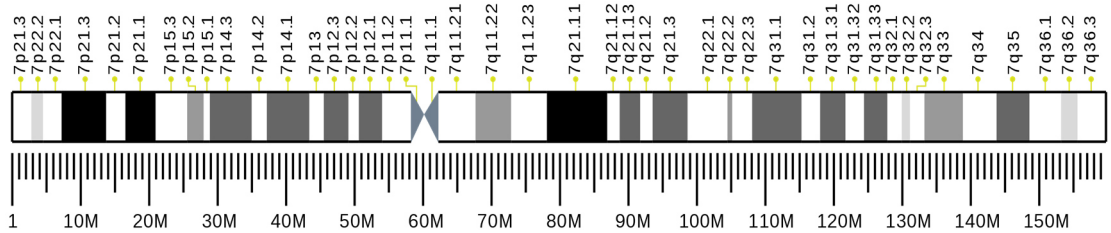


Fig. 1.A. Ideogram of human chromosome 7

Fig. 1.B. Pair of human chromosomes 7

While being time consuming and prone to human error, karyotyping is conducted manually, with limited means of automation. There are some solutions that present partial automations (like Lucia Karyo [5]), but they mostly provide useful utilities than true means of automation.

Proposing a system that offers automation of chromosomal recognition is a large undertaking that can be divided into multiple separate units of research in various fields: Computer Vision, algorithms, decision making systems [6]. Full automation of chromosomal recognition includes:

- Extracting data from visual depiction of live chromosomes;
- Extracting data from ideograms;
- Compare "parsed" chromosomes and ideograms with respect to certain rules;
- Make a decision – a basis for medical diagnosis;

In order to perform all these actions efficiently, it is crucial to store and manipulate data in an efficient manner. Therefore, it is necessary to design a data format that will complement the algorithm.

Related Papers

The general problem of programmatic chromosome pathologies recognition is not entirely new – there are even ready commercial solutions for reproductive laboratories to use. However, the common issue is that they do not create a comprehensive pipeline that would be able to get input image and produce a diagnosis. The aforementioned Lucia Karyo [5] can serve as an example – while it is helpful and effective for improving efficiency of karyotyping, it mostly consists of a database and specific image editing features. Moreover, no open data formats designed specifically for chromosomes have been found.

The global task under consideration is related to the field of bioinformatics, and there are several well-known formats for storing biological data. The prominent examples are Variant Call Format [7] and Stockholm Format [8], but there is also a considerable variety of accepted ways to store bioinformatic data [9].

Variant Call Format is a format that stores information about positions in the genome. One of its main features is brevity – it only stores differences from the norm. VCF focuses on the unit of measurement that is not applicable to the problem considered by this article, but the idea of storing only differences might potentially grant a boost to storage format and algorithm efficiency. It should also be noted that VCF operates a flat array of data, while chromosome pathology might need some data nesting to represent itself in full.

Stockholm format, like VCF, operates genome sequence alignment. However, it has vastly different syntax and shares richer metadata features. Nonetheless, using Stockholm format for the purpose of this article seems as redundant as using VCF.

However, these formats are mostly centered around genetic information, which is redundantly precise for the task at hand. Therefore, there might be a need to develop custom format centered around chromosomes instead of genes.

Goal

The goal of the paper is to propose a data storage format designed for efficient storage of chromosomal data, that would be fit to handle both recognized “live” chromosomal data and ideogram data.

Such a format would be useful for algorithms focused on chromosome management.

Main Body

In order to formulate a proposal of a data format, it is necessary to revise the process of karyotyping domain and the general algorithm of its automation.

The goal of karyotyping process is to gather information about chromosomes of a person and compare them to the “reference” versions, known to be normal and healthy. By comparing actual and reference chromosome, mismatches can be found and diagnosis can be made. The process can be roughly separated into three stages.

The first stage of karyotyping process includes gathering of the material and preprocessing. This process is out of scope of the research since it it heavily relies on specific medical hardwar. This stage results in a metaphase plate – a random arrangement of chromosomes depicted on fig. 2.

Metaphase plate includes all the chromosomes of a person, plus possible noise and obstacles.

Fig. 2. Metaphase plate

The second stage implies recognition of metaphase plate contents. Metaphase plate contains all the chromosomal data of a person, but it is not arranged and insufficient to state a diagnosis. In this stage, a medical specialist has to manually inspect a metaphase plate image, remove noises, obstacles and categorize each chromosome.

A healthy karyotype consists of 22 “numbered” pairs and XX/XY pair for female/male patients respectively. Therefore, there are 24 types of chromosomes – 22

"numbered" ones, X and Y. Each of these 24 "types" has a distinctive visual pattern, consisting of specific combination of black and white stripes (bands) of specific lengths. So, in order to categorize a chromosome, specialist has to compare it to a schematic representation of an ideal chromosome called "ideogram" (fig. 1.A)

Categorized chromosomes are arranged in pairs, forming a karyogram – an intermediate result of a karyotype (fig. 3).

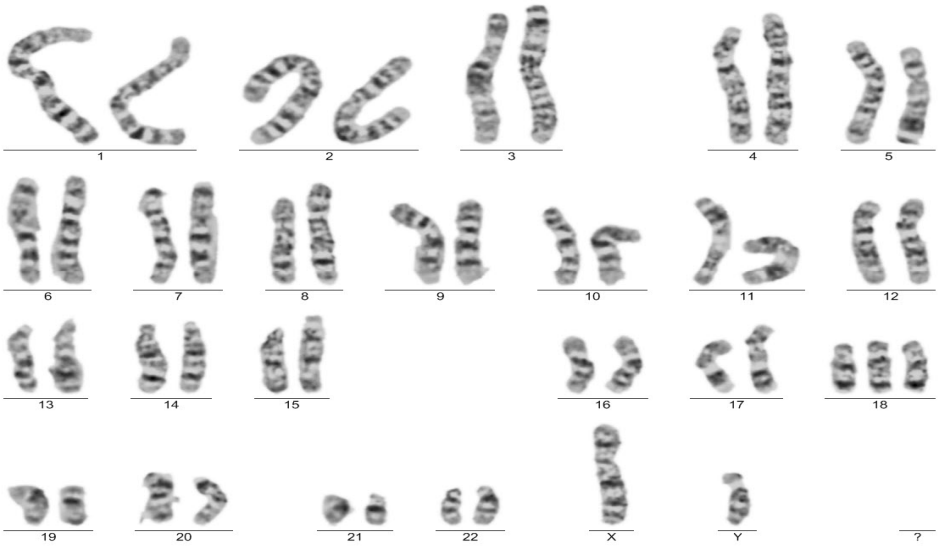


Fig.3. Karyogram with trisomy 18

Karyogram is a reference image for more convenient visual analysis, allowing to detect anomalies. There are several major groups of anomalies to detect:

- Quantitative anomalies. Chromosomes commonly go in groups of two, but sometimes there are three (trisomy) or more chromosomes of the same type. Notorious examples are Down syndrome (trisomy 21) [10], Edwards syndrome (trisomy 18) [11], but there are more syndromes caused by redundant number of chromosomes. Detecting a bigger-than-expected number of healthy chromosomes is a sign of quantitative anomaly;
- Structural anomalies. A malformed chromosome is something that is called a structural anomaly. Common examples of structural anomalies are duplication (same part of a chromosome repeating twice), deletion (part of a chromosome is missing) and translocation (part of a chromosome is moved) [12]. Detecting a partial match of a chromosome may help detecting a structural anomaly, but many checks are required to confirm the exact nature of such anomaly;

The third stare results in the actual diagnosis. Medical specialist analyses a combination of detected anomalies and states the final result.

Having overviewed the process of karyotyping, it is possible to revise a general algorithm for its partial automation. The algorithm can be broken into separate atomic parts which could be addressed separately. Each part addresses specific problem, can be improved independently or mocked. When combined, these parts form a pipeline that takes chromosome image as input and a suggested diagnosis as an output.

This paper is focused on a data format used in such an algorithm, but brief algorithm overview is necessary for the context.

The first stage remains unautomated, since it depends on advanced medical hardware. The metaphase plate (fig. 2) is considered to be the input data for the algorithm.

The second stage starts from an image and its goal is to recognize its content. As an input image, anything depicting a set of chromosomes can be used, such as metaphase plates and karyograms. In realistic scenarios, metaphase plates would be used, but user could also upload a karyogram to validate manually conducted karyotyping.

In any case, this stage starts from noise removal from the image (which is basically anything that does not have features of a chromosome). After that, a contour detection is conducted, identifying tangible objects inside an image. The detected contours are filtered again, removing objects that do not resemble chromosomes by a set of criteria. This way, algorithm would be able to continue working with a definitive set of visual objects.

The next step of this stage is extracting features from each specific visual object. This step is iterative and considers one object at a time, attempting to detect information that could later be used to determine

During this step, the algorithm loops through the set of objects and extracts features from each of them, moving data into a custom format considered by this article. Having extracted image data into this format, it is possible to compare parsed chromosomes to ideograms (also described into same custom format).

This comparison, as well as identifying chromosomal diseases, constitute the third step. This would require a separate decision-making mechanism [13], that would also account for variable chromosome sizes and would detect chromosome being a mutated version of a certain ideogram. And, having a set of detected chromosomes and their abnormalities, it is possible to determine a suggested diagnosis. In order to map abnormalities to a distinct diagnosis, a knowledge base is required. So, it would be logical to keep a "dictionary" with possible chromosome states and corresponsive diagnoses. It should be noted, that not each chromosome abnormality has a distinctive name assigned to it – but it can be surely stated that if there is any mismatch with healthy karyotype, an anomaly takes place.

Since algorithm is focused on chromosome storage and comparison, designing efficient data format is essential for its effectiveness. This format should be able to represent a set of chromosomes, where each chromosome consists of bands with specific color and size. Fig. 4 shows a class diagram with suggested data format.

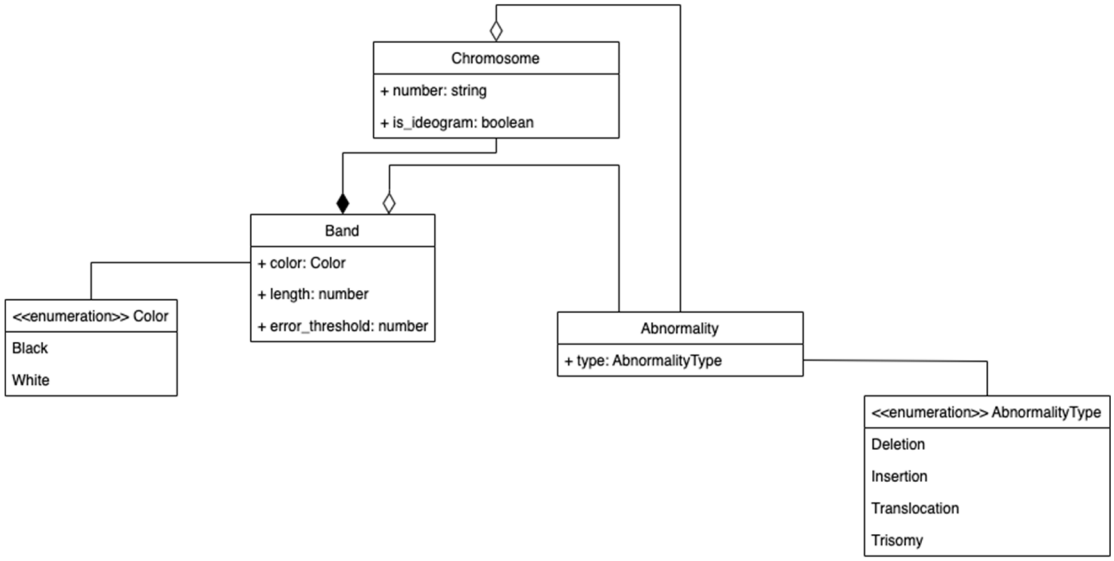


Fig. 4. Class diagram depicting proposed data storage

The *Chromosome* is the main object of the storage. It stores information of a single chromosome. The *number* property stores a suggested number of chromosome – a value from range 1…24, X or Y. *is_ideogram* flag shows what data is actually stored in this instance – a real chromosome or ideogram. This way all the data used in algorithm is kept in a uniform format.

Each chromosome has nested list of *Bands* – a colored segments of chromosome. Each segment has *color* and *length.* Bands could be a result of feature detection from image, or just a serialization of ideogram for more straightforward comparison. Since chromosomes could be of a variable size, it is proposed to store the relative length. This

way it would be possible to compare chromosomes by their internal features despite size differences.

There is a hypothetical issue comes from a fuzzy nature of data – it is improbable that there would be exact match between ideal band and a patient band. To address this issue, *error_threshold* field is added to a band. It stores maximum allowed deviation between ideal bend and chromosome bend.

The aforementioned objects are created as a part of chromosome feature extraction. After that, parsed chromosomes are compared against ideograms, bend by bend, to find a perfect match. On this stage, quantitative anomalies could be detected – this happens when there is an improper number of valid chromosomes. All the cromosomes that do not have a perfect match along with ideograms, are compared to ideograms bend by bend to find the closest match. When found, a structural abnormality is registered – this means that a part of chromosome is missing or duplicated.

Anomalies are stored into the *Abnormality* entity. This entity has a *type* and references to chromosomes and bends. In case of quantitative anomaly, reference to a chromosome is sufficient. In case of structural anomaly, affected bands are referenced.

In a stage of diagnosis making, a list of abnormalities is traversed and compared to a list of known medical conditions. If there is a match – a diagnosis is proposed. If there are no matches, but there are still abnormalities – it means the patient has an unidentified pathology, which is not uncommon. If there is no supposed diagnosis and there are no abnormalities – the patient's chromosomes are considered healthy.

Having described the nature of data format, it is necessary to evaluate its efficiency of executing core algorithm operations on proposed data format. Suppose there is a set of N chromosomes and a set of M ideograms. Each chromosome n consists of B_n bands, and ideogram m consists of B_m bands. In this case:

Table 1

Operations complexity

	Operation	Complexity
1.	Identification of a single chromosome. A chromosome is iteratively compared to each ideogram *m.*	*O(M)*
2.	Identification of quantitative abnormality. Each chromosome *n* is compared to each ideogram *m.* After that, each detected chromosome is looped to find unwanted duplicates.	*O(N*M+N)*
3.	Identification of a deletion. Each chromosome *n* is compared to each ideogram *m*. If a match is detected and it is partial, it is supposedly a deletion.	*O(N*M)*

The provided tests are fairly simple, but they should be sufficient to demonstrate the nature of data, serve as a reference for further research and efficiency measurement. Moreover, lack of open formats designed to store chromosomes, it is important to establish a baseline for improvements and comparison.

Results and Discussion

Unfortunately, not much could be reused from the research of existing storage formats of biological data. Formats like VCF or Stockholm formats are too detailed and store too much information about chromosomes to be efficient for the task at hand.

Due to the lack of research dedicated to designing data storage of chromosomes, it is hard to state that the proposed storage format offers outstanding efficiency. However, it represents application data domain in a straightforward and non-redundant way, introducing code entities matching domain objects. This would allow to reliably represent complexity of operations that are performed during cytogenetic analysis, and would allow to introduce measurable algorithms for its automation.

Conclusions

Automation of chromosomal diseases recognition is a relevant problem. Currently it is done in a manual or semi-manual way – and due to its high effort consumption and high cost of error it requires automation. This paper proposes a data format used for storing data for automation algorithm.

The essence of the problem has been revised, describing the nature of karyotyping process as well as general automation algorithm. Existing data formats used in domain of biocybernetics are revised, and none of them proved to be efficient for the task at hand.

After considering the requirements that should be addressed, a new custom data format has been suggested. This format stores both chromosomes and ideograms as a collection of bands, allowing easier comparison because of similar nature of data. It also stores detected abnormalities in separate objects, mostly for caching reasons.

A simple set of operations has been described and its potential complexity has been evaluated. This would allow to make measurable improvements in further research, establishing a metrics baseline.

References

1. Wapner RJ. Genetics of stillbirth. *Clinical Obstetrics and Gynecology*. 2010. Vol. 53(3), P.628-34. URL: https://doi.org/10.1097/GRF.0b013e3181ee2793
2. Hay S.B. ACOG and SMFM guidelines for prenatal diagnosis: Is karyotyping really sufficient? *Prenatal Diagnostics*. 2018. Vol. 38(3). P.184-189. URL: https://doi.org/10.1002/pd.5212
3. O'Connor. Karyotyping for chromosomal abnormalities. *Nature Education.* 2008. URL: https://www.nature.com/scitable/topicpage/karyotyping-for-chromosomal-abnormalities-298/
4. Cytogenetics L., Karyo L. URL: https://www.lucia.cz/en/products/lucia_karyo
5. Pysarchuk O., Mironov Y. A Proposal of Algorithm for Automated Chromosomal Abnormality Detection. *Modeling, Control and Information Technologies: Proceedings of International Scientific and Practical Conference*. 2021. Vol. 5, P.83–86. URL: https://doi.org/10.31713/MCIT.2021.26
6. Samtools team. The Variant Call Format Specification. 2022. URL: https://samtools.github.io/hts-specs/VCFv4.3.pdf;
7. Sonnhammers E. Stockholm format. URL: https://sonnhammer.sbc.su.se/Stockholm.html
8. Wikipedia, Wikimedia Foundation. Biological sequence formats. URL: https://en.wikipedia.org/wiki/Category:Biological_sequence_format
9. Antonarakis S.E. Down syndrome. *Nature Reviews Disease Primers*. 2020. Vol.6(1). P:9. URL: https://doi.org/10.1038/s41572-019-0143-7
10. Outtaleb F.Z. Trisomy 18 or postnatal Edward´s syndrome: descriptive study conducted at the University Hospital Center of Casablanca and literature review. *Panafrican Medical Journal*. 2020. Vol.37, P.309. URL: https://doi.org/10.11604/pamj.2020.37.309.26205;
11. Clancy S., Shaw K. DNA deletion and duplication and the associated genetic disorders. *Nature Education*. 2008. Vol. 1(1). P.23. URL: https://www.nature.com/scitable/topicpage/dna-deletion-and-duplication-and-the-associated-331/
12. Pysarchuk O., Mironov Y. Chromosome Feature Extraction and Ideogram-Powered Chromosome Categorization. *Advances in Computer Science for Engineering and Education. ICCSEEA 2022. Lecture Notes on Data Engineering and Communications Technologies*. Springer, Cham. 2022. URL: https://doi.org/10.1007/978-3-031-04812-8_36;

O. Pysarchuk, Yu. Mironov

ФОРМАТ ЗБЕРІГАННЯ ХРОМОСОМНИХ ДАНИХ

О.О. Писарчук[1], Ю.Г. Міронов[2]

[1] Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». 03056, м. Київ, Солом'янський район, пр-т Перемоги, 37. , PlatinumPA2212@gmail.com
[2]Національний авіаційний університет, 03058, Київ, пр. Гузара Любомира, yuriymironov96@gmail.com

Публікація розглядає задачу автоматизованого розпізнання хромосомних патологій. Хромосомні патології представляють значну небезпеку при плануванні вагітності. Для протидії цій проблемі задіюється процес каріотипування. На даний момент такий процес проводиться вручну чи в напівавтоматичному режимі, не дивлячись на великі часові затрати та високу ціну помилки. Отже, існує потреба в автоматизації даного процесу. Процес каріотипування (як і алгоритм його автоматизації) можна поділити на кілька етапів з різними цілями. Але у цих етапів є спільний елемент – формат, в якому зберігаються та обробляються дані. Метою даної статті є пошук такого формату. В статі оглянуто особливості процесу каріотипування та коротко описані кроки алгоритму з розпізнавання хромосомних патологій. Також розглянуто поширені в сфері біоінформатики формати даних. Нажаль, їхня ефективність у застосуванні до представленої задачі є сумнівною. Публікація пропонує новий формат даних для вирішення проблеми. В форматі представлені основні сутності задіяні в процесі розпізнавання аномалій. Розглянуто кілька фрагментів алгоритму, і розглянута їхня ефективність в комбінації з форматом даних. Результатом даної статі є запропонований формат для зберігання та обробки даних, що може бути використаний в процесі автоматичного розпізнавання хромосомних аномалій. Отримані заміри можна використовувати для подальшого покращення результативності алгоритму та ефективності формату зберігання даних.

Ключові слова: складність алгоритму, аналіз даних, domain-driven design

DOI 10.15276/imms.v12.no4.296
УДК 004.822

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 296-305

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ТА ОБРОБКА ТЕКСТОВИХ ДОКУМЕНТІВ НА ОСНОВІ МУЛЬТИАГЕНТНОГО ПІДХОДУ

О.В. Барабаш., В.П. Колумбет

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», проспект Перемоги, 37, Київ, 03056, Україна
E-mail: bar64@ukr.net , kvplinux@gmail.com

Ефективність керування підприємством, установою певною мірою залежить від того, наскільки розумно в ньому організований документообіг. Адже, документообіг та управлінська діяльність тісно пов’язані одне з одним. Від того, наскільки оперативно здійснюється рух, опрацювання документів та їх передавання на виконання, залежить швидкість отримання інформації, необхідної для прийняття управлінського рішення. Для підвищення ефективності роботи з електронними документами за рахунок їх автоматизованого інтелектуального аналізу запропоновано комплексний підхід до розробки підсистеми керування електронними документами в CASE-системах, що допускають динамічне налаштування за мінливих умов експлуатації та потреб користувачів. Для аналізу документів використовуються агентний та онтологічний підходи. Онтології дозволяють у явному вигляді представити семантику та структуру документа. Використання агентів дозволяє спростити процес аналізу, зробити його розширюваним і масштабованим. Результати інтелектуального пошуку та обробки документів, одержуваних з гетерогенних джерел, можуть бути використані не тільки для автоматичної класифікації та каталогізації документів в інформаційній системі у зручній для користувача формі, а й для зниження трудомісткості виконання етапу аналізу предметної області інформаційної системи, її проектування, а також для інтелектуалізації процесів створення звітних документів на основі інформації, розміщеної у базі даних системи. Дослідження надає можливість створення CASE-технології, призначеної для створення інформаційної системи, що динамічно налаштовуються та мають унікальні можливості адаптації до мінливих умов експлуатації на основі «зворотного зв'язку» і інтелектуального аналізу документів.

Ключові слова: мультиагентні системи, онтологія, інтелектуальний пошук, агент, аналіз документів, інформаційні системи, CASE-технології, адаптивні системи.

Вступ

Наразі розроблено велику кількість CASE-систем, що автоматизують найбільш трудомісткі етапи розробки інформаційних систем пов'язаних з програмуванням бізнес-операцій та створенням інтерфейсу. Значно тривалим та трудомістким стає етап аналізу предметної області, який зазвичай автоматизується CASE-системами. Таким чином, одним із перспективних напрямків розвитку CASE-систем є автоматизація цього процесу. У CASE-системах, орієнтованих на створення інформаційної системи з динамічною адаптацією під час їх використання, де стадія аналізу предметної області «розтягується» на весь час функціонування системи, це завдання стає особливо актуальним. Якщо врахувати, що на стадії експлуатації таких систем завдання реінжинірингу покладено на користувачів-фахівців у предметних областях, але не в області інформаційних технологій, то засоби автоматизації аналізу стають найважливішими компонентами. Іншими словами, якщо ставити завдання динамічного налаштування інформаційної системи до мінливих умов, то основою реалізації засобів її динамічної адаптації будуть засоби реструктуризації даних у базі даних

(БД) інформаційної системи. А ці засоби дозволяють вносити зміни до моделі даних на основі результатів аналізу предметної області, нормативно-довідкових та розпорядчих документів, що регламентують діяльність у зазначеній області. Звідси випливає необхідність підтримки в динамічно адаптованих системах одного з найскладніших і трудомістких етапів розробки інформаційної системи – етапу аналізу. Джерелом інформації для аналізу можуть бути документи різного виду, так як діяльність будь-якої бізнес-системи будується саме з урахуванням нормативних документів. Підтримка бізнес-операцій засобами інформаційної системи вимагає відображення в моделі даних системи вимог, закріплених у нормативно-довідкових даних, розпорядчих документах, у вигляді обмежень, що накладаються на дані (атрибути, властивості об'єктів предметної області, інформація про які зберігається в БД, а також зв'язки між ними) та операції, що виконуються над ними [1].

В результаті аналізу має бути побудована система взаємопов'язаних документів:

належать до визначених напрямів діяльності бізнес-системи (до визначених понять, об'єктів предметної області);

відображають зв'язки між цими поняттями (з кожним поняттям може бути пов'язаний документ або сукупність документів, зв'язки між документами відображають зв'язки між поняттями);

містять нормативну інформацію, яка також може бути виділена на основі аналізу змісту документів.

На основі побудованої системи взаємопов'язаних документів можна частково автоматизувати процес аналізу змін предметної області та внесення змін до моделі предметної області інформаційної системи [2]. Таким чином, система управління документами стає не лише «надбудовою» над інформаційною системою та її БД, що дозволяє отримувати результати обробки даних та зберігаються в БД інформаційної системи, у зручній для користувачів формі, а й стає основою засобів розробки інформаційної системи – засобів реструктуризації даних.

Огляд останніх досліджень та публікацій.

Питання керування електронними документами та використання мультиагентного та онтологічному підходу є досить актуальним. Цій тематиці було присвячено ряд сучасних робіт зарубіжних та вітчизняних вчених. Серед них зокрема варто виділити: Буров Є.В., Микіч Х.І., Верес О.М., Литвин В.В., Сорока М.Ю., Касілов О.В., Крамська К.І., Омельяненко В.А., Рогушина Ю.В. Їх роботи присвячені дослідженню та розвитку мультиагентних систем, та реалізації онтологічного підходу.

В роботі [3] представлено мультиагентний підхід до розробки інтелектуальних систем управління проектами. Для багатокритеріальної оцінки варіантів з використанням апарату нечітких множин обрано вид функцій загальної корисності та корисності локальних критеріїв. Даний метод не дозволяє описати статичні і динамічних ситуації, а, отже, не займається питаннями їх аналізу.

В статті [4] досліджено процес обробки знань у когнітивній інформаційній системі, керованій моделями. Це дозволяє засобами системного підходу суттєво підвищити рівень інформаційної формалізації знань, здійснюючи індивідуальний підбір методів та засобів політики інформаційної безпеки на підприємстві на основі побажань підприємця та експертних оцінок. Даний метод реалізується у складі набору компонентів для розробки мультиагентних систем. Проте він не призначений для аналізу формалізованих знань.

В роботі [5] досліджено та розроблено методи й засоби ідентифікації проблемних ситуацій на базі онтологій із використанням механізмів логічного виведення, які застосовано в інтелектуальних системах підтримки прийняття рішень для завдань тестування програмного забезпечення. Розглянуто актуальну

проблему тестування програмного забезпечення із використанням онтологічного моделювання для своєчасного виявлення помилок та поліпшення якості розроблюваного програмного продукту. Але, даний метод не дозволяє описати статичних і динамічних ситуацій, а, отже, не займається питаннями їх аналізу.

В статті [6] автори досліджували лінгвістичні антології, їх проектування та використання в навчальних інтелектуальних системах. Представлено модель лінгвістичної онтології для предметних сфер. У запропонованій моделі описано набір відносин лінгвістичної онтології, спеціально підібраний для опису аналізованої предметної сфери. Разом із тим, даний метод не дозволяє описати статичний і динамічний процеси, а, отже, не займається питаннями їх аналізу.

Отже, на підставі проведеного аналізу виявлено, що існуючі методи не повною мірою вирішують завдання аналізу й обробки даних з одночасним використанням агентного та онтологічного підходу. Вони не враховують динаміку процесів, не приділяють достатньої уваги аналізу «вузьких місць», не використовують повноцінно інформацію з онтологічної моделі в умовах динамічних процесів з використанням мультиагентного підходу.

Мета і завдання дослідження

Запропоновано та досліджено автоматизований підхід для інтелектуального аналізу з метою підвищення ефективності роботи з електронними документами. Реалізація інтелектуального пошуку та обробки документів, одержуваних з гетерогенних джерел, які можуть бути використані не тільки для автоматичної класифікації та каталогізації документів в інформаційній системі у зручній для користувача формі. Зниження трудомісткості виконання етапу аналізу предметної області інформаційної системи, її проектування. Інтелектуалізація процесів створення звітних документів на основі інформації, розміщеної у базі даних системи.

Виклад основного матеріалу

Для підвищення ефективності обробки, електронний документ вимагає наявності метаданих, що описують структуру та семантику даних. Одним із можливих підходів до опису інформації, закладеної в документі, є підхід на основі онтологій. Під онтологією розуміється база знань спеціального типу, яка може читатися та розумітися не тільки розробниками а й користувачами, та дозволяє поліпшити взаємодію розробників та програмних агентів. Онтологічний підхід має такі переваги:

- зручність сприйняття людиною;
- відсутність вимог до кваліфікації користувача при розробці онтології;
- можливість опису одного документа різними онтологіями.

Для вирішення описаного вище завдання було обрано онтологічний підхід, згідно якого необхідно описати структуру і зміст документа. Відповідно до запропонованого підходу, онтологія використовується для опису семантики даних документа та його структури. Враховуючи специфіку розв'язуваних у даній роботі завдань, конкретизуємо поняття онтології: вважатимемо, що онтологія – це специфікація деякої предметної області, яка містить в собі словник термінів (понять) предметної області і множину зв'язків між ними. Ці зв’язки описують, як ці терміни співвідносяться між собою у конкретній предметній області.

Для побудови ієрархії понять онтології застосовуються такі базові типи відносин:

- “is_a” (“примірник – клас”);
- "part_of" ("частина – ціле");
- "synonym_of" ("синонім»);.

Слід врахувати, що ці типи відносин є базовими і залежать від онтології, але необхідно надати користувачеві можливість додавання нових відносин, які враховували специфіку описуваної предметної області.

Крім відносин онтологія включає два типи вершин. До першого типу віднесемо вершини, що описують структуру документа. Наприклад: таблиця, дата, посада тощо. (це загальні поняття, які не залежать від конкретної предметної області). Іншим типом будуть вершини, що містять поняття документа. Перший тип вершин будемо називати структурними вершинами, другий тип – семантичними вершинами. На рис. 1 структурні вершини мають темний відтінок, а семантичні вершини зображені світлішим відтінком.

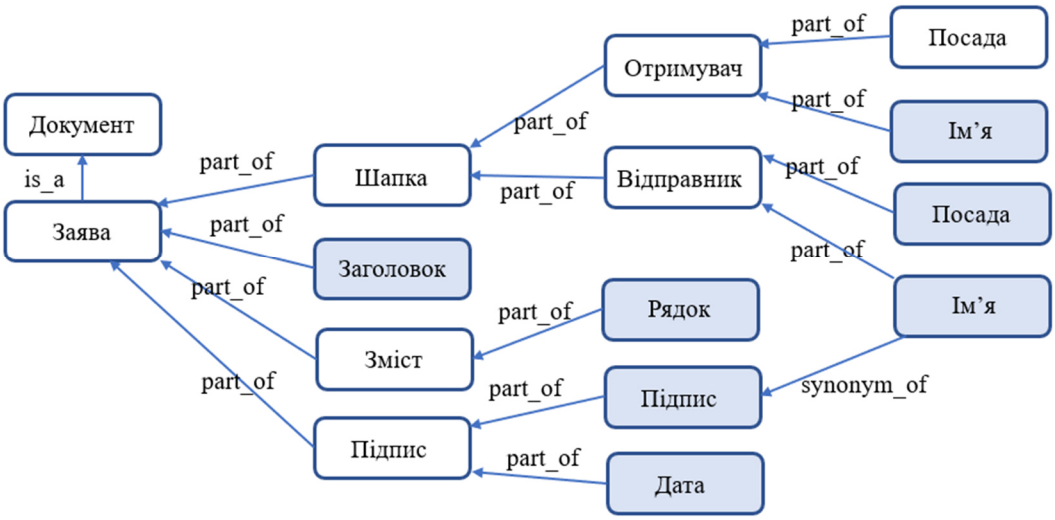


Рис. 1. Приклад онтології, що описує клас документів «Заява»

Фактично у цьому контексті онтологія – це ієрархічна понятійна основа аналізованої предметної області. Онтологія документа використовується для аналізу документа, завдяки їй з документа можна отримати необхідні дані: відомо, де шукати дані і як вони можуть бути інтерпретовані.

Якщо представляти документ з використанням онтологій, то завдання зіставлення онтології та наявного документа зводиться до пошуку понять онтології в документі. Як наслідок, системі необхідно відповісти на запитання: чи описує дана онтологія документ чи ні. На останнє запитання можна відповісти ствердно, якщо в процесі зіставлення в документі були знайдені всі поняття, включені в онтологію. Перш ніж здійснювати пошук вершин, що містять поняття документа, необхідно провести пошук вершин, що описують структуру документа. Таким чином, вихідне завдання зводиться до пошуку в тексті документа загальних понять на основі формальних описів.

У наведеному прикладі на рисунку 2 вершини онтології розбиті на дві площини, що враховується при зіставленні документу та його онтології.

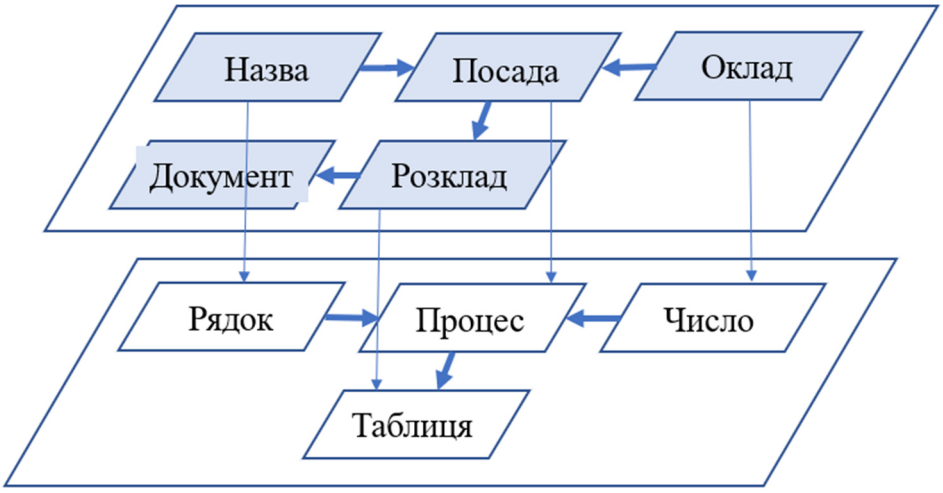


Рис.2. Приклад розбиття вершин онтології для документа на дві площини

Агентний підхід до аналізу документів

До процесу пошуку документів висувається низка вимог:

- висока швидкість обробки великих обсягів даних;
- відмовостійкість;
- масштабованість;
- налаштованість на потреби користувачів та мінливі умови.

Для вирішення проблеми виділення загальних понять з урахуванням формальних описів пропонується агентний підхід [7]. Тут під агентом розуміється система, спрямована на досягнення певної мети, здатна взаємодії з середовищем та іншими агентами [8]. Такий підхід задовольнятиме вимогам до процесу пошуку, якщо при побудові системи будуть реалізовані всі переваги мультиагентних систем.

При використанні даного підходу для кожної вершини онтології, що містять загальні поняття, створюється агент, який проводить пошук даного конкретного поняття. Для визнання агента інтелектуальним необхідною умовою є наявність у нього бази знань. Таким чином, щоб визначити агентів діючих у системі, необхідно обрати спосіб для опису бази знань (БЗ), характер взаємодії із середовищем та взаємодії.

Однією з найважливіших властивостей агентів є соціальність чи здатність до взаємодії [9]. Як було зазначено раніше, для кожної вершини онтології, що містить загальне поняття (семантична вершина), створюється агент.

Даний агент націлений на вирішення двох завдань:

1) весь наявний список шаблонів поняття він розбиває на окремі компоненти і запускає спрощених агентів для пошуку структурних вершин;

2) збирає результати з усіх списків, отриманих агентами нижчого рівня.

Згадані вище агенти нижчого рівня є рефлекторними. Вони отримують шаблон. Метою цих агентів є знаходження у тексті фрагментів, що підпадають під цей шаблон.

Важливим питанням є комунікація агентів. Механізми комунікації агентів поділяються на безпосередні та опосередковані. Прикладом реалізації безпосередньої комунікації може бути модель взаємодії «замовник – підрядник» [10]. Механізм опосередкованої комунікації реалізується за допомогою архітектури «дошки оголошень»:

- модель «замовник – підрядник». Дана модель передбачає розподіл усієї множини агентів системи на два класи – клас замовників і клас підрядників. Суть даної моделі взаємодії полягає у вирішенні різних завдань шляхом спрямування їх

на виконання найбільш підходящим для цього агентам. За розподіл завдань відповідальні агенти-замовники. Потенційні підрядники аналізують виставлені замовниками заявки, аналізують їх щодо можливості реалізації і, у випадку позитивного результату аналізу, подають заявку замовнику;

- модель «дошка оголошень». Архітектура її заснована на моделі класної дошки, де представлено поточний стан системи, у межах якої оперують агенти. Агенти постійно аналізують інформацію на дошці, намагаючись знайти застосування своїм можливостям. Якщо в певний момент часу агент виявляє можливість внесення свого вкладу в процес вирішення поточних завдань, він залишає на дошці інформацію про початок роботи в даному напрямку, а після закінчення роботи поміщає результат на дошку.

Враховуючи особливості розв'язуваного завдання, реалізовано комбінацію двох моделей комунікації «замовник – підрядник» та «дошки оголошень».

Архітектуру мультиагентної системи та процес аналізу документа представлено на рисунку 3.

Подання бази знань агентів

Одним із найважливіших питань у системі є питання подання БЗ агента. На даний момент представлення БЗ агента можливе трьома різними способами: з використанням онтологій, за допомогою регулярних виразів та на базі продукції.

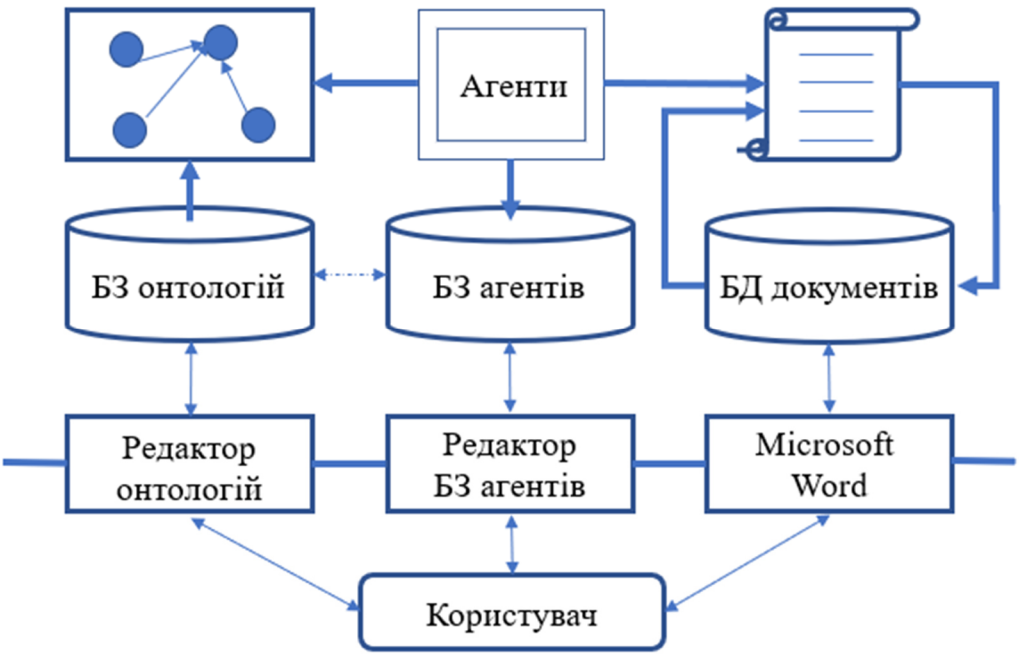


Рис. 3. Архітектура мультиагентної системи

Подання знань агента за допомогою онтології – найбільш виразний спосіб, який використовує всі переваги явного уявлення знаннями (рис. 4). Перевагою цього способу є те, що для «доказу» вершини онтології ми можемо застосувати різні засоби. Наприклад, це може бути простий збіг ключової фрази або звернення до БД інформаційної системи [11]. Онтології дозволяють описати різні ситуації у разі, якщо не вдається знайти точну відповідність. В якості прикладу можна знайти узагальнююче або конкретизуюче поняття, тощо.

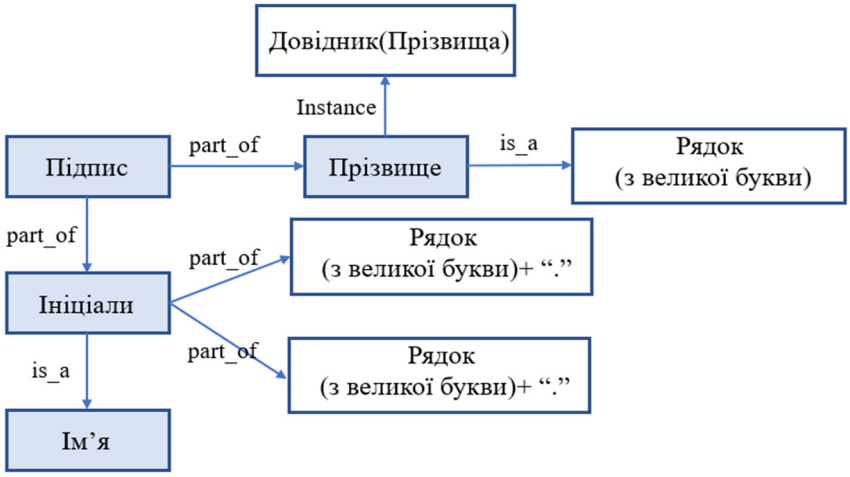


Рис. 4. Подання знань агента за допомогою онтології

Вміст аналізуємого документа представлено у вигляді спеціальної об'єктної моделі, за основу якої була взята об'єктна модель документа Microsoft Word. Для доступу до цієї об'єктної моделі розроблені API-функції, що дозволяють оперувати однаковими поняттями під час роботи з документами у різних форматах. До складу API-функцій включені функції синтаксичного розбору додатків, функції для обчислення різних метрик між поняттями, функції для вилучення інформації про структуру документа. Якщо для пошуку поняття вершини потрібні додаткові дії, вони можуть бути описані за допомогою скрипта з використанням вищезазначених API-функцій [12]. У скрипті також можна використовувати звернення до об'єктної моделі самої інформаційної системи.

Другим підходом є підхід з використанням регулярних виразів. Останні дозволяють легко враховувати різні форми слова та працювати з великими обсягами інформації [13]. Однак необхідно враховувати, що іноді, особливо для некваліфікованих користувачів, завдання правильної побудови регулярного виразу стає досить складним. З метою її спрощення в системі передбачається наявність спеціального редактора, що дозволяє працювати з регулярними виразами природною мовою. Наприклад, еквівалентом до "\d{5}" є "п'ятизначне число" і т.д. Крім того, бажано реалізувати функції побудови регулярного виразу "за зразком". Це означає, що за прикладами, наведеними користувачем, можлива автоматична побудова регулярного виразу. Наприклад, користувач запропонував дві дати: «1.10.22» та «15.07.2018». Система має побудувати регулярний вираз, який би відповідав обом форматам подання дат:

«(\d{1,2}).(\d{1,2}).((\d{4})|(\d{2}))».

Недоліком регулярних виразів є те, що при пошуку вони не дозволяють враховувати місцезнаходження шуканого слова/фрази. Для усунення цього недоліку можливе спільне використання регулярних виразів та правил продукційного типу, які є третім способом подання БЗ агента.

Продукції переважно використовуються для аналізу структури документа. Запроваджено спеціальні поняття, які можуть бути використані при заданні умов [14]. Наприклад, правило, що містить заголовки в тексті, може бути сформульовано наступним чином:

Якщо *(шрифт абзацу відрізняється від абзацу до та після абзацу) та (абзац вирівняний по центру),* ***то*** *цей абзац є заголовком*.

Таким чином, для некваліфікованих користувачів, наявність спеціального редактора та функції побудови регулярного виразу "за зразком" дозволяє працювати з регулярними виразами природною мовою та звичним інструментарієм, що спрощує правильну побудову регулярного виразу.

Результат та обговорення

В інформаційному пошуку для порівняння якості результатів було введено дві характеристики: точність та повнота. Подібні характеристики можна запровадити і для системи зіставлення документа та онтології. Під точністю (*T*) розумітимемо частку правильно проведених відповідностей документа та онтології по відношенню до всіх зроблених системою відповідностей [15]. Під повнотою (*S*) – частку правильно проведених відповідностей по відношенню до всіх відповідностей документа та онтології.

Нехай *N* – число існуючих відповідностей між документом та онтологією, *K* – число проведених системою зіставлень, *H* – число правильно проведених системою зіставлень. Тоді:

$$T = \frac{H}{K} \quad \text{та} \quad S = \frac{H}{N}.$$

Зазвичай, ці два критерії «конфліктують». Тому практично стовідсоткова точність і повнота недосяжні. Роботи з оцінки поки що не проводилися, наступним етапом дослідження стане оцінка величин *T* та *S* під час проведення експериментів на реальних документах.

Засоби аналізу документів можуть бути використані як для зниження трудомісткості роботи користувачів з документами, так і для підтримки вирішення завдання аналізу предметної області розробниками [16]. В даному випадку пропонується глибока інтеграція функціональних підсистем, що включають засоби розробки та засоби, з якими працюють «кінцеві користувачі». Це дає можливість створення CASE-технології, призначеної для створення інформаційної системи, що динамічно налаштовується має унікальні можливості адаптації до мінливих умов експлуатації на основі «зворотного зв'язку» та інтелектуального аналізу документів.

Висновки

В роботі запропоновано та досліджено автоматизований підхід для інтелектуального аналізу з метою підвищення ефективності роботи з електронними документами. Досліджено та реалізовано інтелектуальний пошук та обробку документів, одержуваних з гетерогенних джерел, які можуть бути використані для автоматичної класифікації та каталогізації документів в інформаційній системі, а також для подання їх користувачеві у зручній формі. Доведено зниження трудомісткості виконання етапу аналізу предметної області інформаційної системи, її проектування, а також підвищення рівня інтелектуалізації процесів створення звітних документів на основі інформації, розміщеної у базі даних системи. Використання агентів дозволяє спростити процес аналізу, зробити його таким, що розширюється і масштабується.

Список літератури

1. Плескач В.Л., Рогушина Ю.В. Агентні технології. Монографія. Київ: Нац. торг.-екон. ун-т, 2005. 344 с.
2. Гладун А.Я., Рогушина Ю.В. Онтологічній підхід до проблем підвищення якості розроблення національних стандартів України. *Стандартизація, сертифікація, якість*. 2016. № 2. С. 19 – 28.
3. Омельяненко В.А. Мультиагентний підхід до розробки інтелектуальних систем управління проектами. *Управління проектами та розвиток виробництва*: *зб. наук. пр. СНУ ім. Володимира Даля*, 2016. № 3 (59). С. 5–13.
4. Буров Є.В. Опрацювання знань у когнітивній інформаційній системі, керованій моделями. *Східно-Європейський журнал передових технологій*. 2009. № 6 (7 (42)). С. 40 – 49.

5. Буров Є.В., Микіч Х.І., Верес О.М., Литвин В.В. Система ідентифікації проблемних ситуацій тестування програмного забезпечення. *Вісник Національного університету «Львівська політехніка» «Інформаційні системи та мережі».* Львів, Видавництво Львівської політехніки, 2019. № 6. С. 30 – 40.
6. Ткаченко О., Ткаченко К., Ткаченко О. Лінгвістичні онтології: проектування та використання в навчальних інтелектуальних системах. *Цифрова платформа: інформаційні технології в соціокультурній сфері.* 2021. Т. 4, № 1. С. 97 – 111.
7. https://doi.org/10.31866/2617-796X.4.1.2021.236950
8. Сорока М.Ю. Метод адаптації поведінки агентів в інтелектуальній навчальній системі підготовки диспетчерів управління повітряним рухом. *Системи управління, навігації та зв'язку.* 2020. Т. 2, № 60. С. 17 – 20.
9. https;//doi.org/10.26906/SUNZ.2020.2.017
10. Budaev D., Amelin K., Voschuk G. Realtime task scheduling for multi-agent control system of UAV's group based on network-centric technology. *Int. Conference on Control, Decision and Information Technologies* (CoDIT'2016). 2016. P. 378 – 381.
11. Barabash O., Kolumbet V. Multiagent approach to computer management in a heterogeneous distributed computer environment. *Системи управління, навігації та зв'язку*. 2022. Т. 1, № 67. С. 38 – 42.
12. https;//doi.org/10.26906/SUNZ.2022.1.038
13. Kolumbet, V., Svynchuk, O. Multiagent methods of management of distributed computing in hybrid clusters. *Сучасні інформаційні системи.* 2022. Т. 6, № 1. С. 32 – 36.
14. https://doi.org/10.20998/2522-9052.2022.1.05
15. Філатова Т.В., Журан О.А., Івченко І.Ю. Проєктування інформаційних систем та рішення інформатизації переходу компаній в онлайн сферу. *Інформатика та математичні методи в моделюванні*. 2021 Том 11, № 4. С. 365 – 372.
16. https://doi.org/10.15276/imms.v11.no4.365
17. Радуш В.В., Лебедєва О.Ю., Кушніренко Н.І., Зоріло В.В. Моделювання організаційних заходів для створення політики безпеки організації з використанням бізнес-процесів. *Інформатика та математичні методи в моделюванні*. 2021. Т. 11, № 3. С. 239 – 246. https://doi.org/10.15276/imms.v11.no3.239
18. Кобозєва А.А. Розробка інструментальних засобів автоматизованого проектування комплексу технічних засобів інформаційної системи. *Кібербезпека в Україні: правові та організаційні питання.* Матеріали міжнародної науково-практичної конференції, 26 листопада 2020 р. Одеса. 2021. С. 86 – 89.
19. Кобозева А.А. Основы общего подхода к решению проблемы обнаружения фальсификации цифрового сигнала. *Электромашиностроение и электрооборудование*. 2009. № 72. С. 35 – 41.
20. Касілов О.В., Крамська К.І. Моделі і метод синтезу агентної інформаційно-пошукової системи. *Сучасні інформаційні системи.* 2020. Т. 4, № 2. С. 94 – 99.
21. https://doi.org/10.20998/2522-9052.2020.2.14
22. Буров Є.В. Ефективність застосування онтологічних моделей для побудови програмних систем. *Математичні машини і системи*. 2013. № 1. С. 44 – 55.

INTELLECTUAL ANALYSIS AND PROCESSING OF TEXT DOCUMENTS BASED ON A MULTI-AGENT APPROACH

Oleg Barabash, Vadym Kolumbet

National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute",
37 Peremogy Avenue, Kyiv, 03056, Ukraine, E-mail: bar64@ukr.net , kvplinux@gmail.com

The effectiveness of managing an enterprise or institution to some extent depends on how intelligently the document flow is organized in it. After all, document flow and management activities are closely related to each other. The speed of obtaining the information necessary for making a management decision depends on how quickly the movement, processing of documents and their transfer to execution is carried out. In order to improve the efficiency of working with electronic documents due to their automated intellectual analysis, a comprehensive approach to the development of a subsystem for managing electronic documents in CASE-systems, which allow dynamic adjustment under changing operating conditions and user needs, is proposed. Agent and ontological approaches are used to analyze documents. Ontologies allow you to explicitly represent the semantics and structure of a document. The use of agents allows you to simplify the analysis process, make it extensible and scalable. The results of the intelligent search and processing of documents obtained from heterogeneous sources can be used not only for automatic classification and cataloging of documents in the information system in a user-friendly form, but also for reducing the complexity of the stage of analysis of the subject area of the information system, its design, as well as to intellectualize the processes of creating reporting documents based on information placed in the system database. The research provides an opportunity to create CASE technology, designed to create an information system that is dynamically adjusted and has unique adaptation capabilities to changing operating conditions based on "feedback" and intelligent document analysis.

Keywords: multi-agent systems, ontology, intelligent search, agent, document analysis, information systems, CASE technologies, adaptive systems.

DOI 10.15276/imms.v12.no4.306
УДК 681.514:621.029

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 306-314

РОЗРОБКА ДОДАТКУ ДЛЯ ШИФРУВАННЯ ДАНИХ ЗІ ЗБЕРЕЖЕННЯМ ФОРМАТУ

Т.І. Горбатюк, О.Ю. Лебедєва

Національний університет «Одеська Політехніка», просп. Шевченка, 1, Одеса, 65044, Україна; e-mail: infsec2011@gmail.com

Розглядається розробка додатку для шифрування даних зі збереженням формату. Наводяться основні визначення, такі як особиста інформація, конфіденційна ідентифікаційна інформація. За останні кілька років зростання кількості витоків персональних даних призвело до втрати цілісності даних для сотень мільйонів записів користувачів. Такі атаки спрямовані як проти великих компаній, так і проти малих підприємств. Частка кібератак з метою крадіжки персональних даних у підприємств забезпечена через халатність компаній щодо коректного шифрування та маскування персональних даних працівників та клієнтів. У роботі розкривається поняття шифрування даних із збереженням формату, як процес шифрування даних таким чином, щоб вихідні дані залишалися в тому самому форматі, що й вхідні дані. У шифруванні із збереженням формату розглядається два режиму роботи – FF1 та FF3. При аналізі існуючих режимів шифрування зі збереженням формату, був виділений режим FF3-1, який був спеціально створений для вирішення проблеми шифрування на невеликих доменах тобто областях даних. Режим FF3-1 був реалізований у програмному додатку. У роботі також розкривається поняття псевдонімізації та методи досягнення її. Псевдонімізація може бути досягнута за допомогою різних методів, таких як маскування даних, шифрування або токенізація. Процеси маскування даних змінюють значення даних залишаючи їх вхідний формат. Мета полягає в тому, щоб створити версію, яка не піддається розшифровці або зворотній інженерії. В роботі було обрано метод маскування даних. У зв'язку з цим основним завданням є втілити методи які б допомагали компаніям шифрувати та маскувати дані з мінімальними вкладами в реструктуризацію та збереження функціональності на льоту. Для того, щоб досягти цілі чіткої псевдонімізації, щоб факт підміни реальних даних був непомітним, вирішено модернізувати підхід до маскування та шифрування для полів типу ім’я, прізвище та email адреса. В роботі удосконалено методи псевдонімізації та маскування полів даних: типу ім’я, прізвище та email адреса та застосування методу FPE в режимі FF3-1 для полів типу номер телефону та кредитної карти. Удосконалений метод псевдонімізації та маскування полів даних реалізовано у програмному додатку.

Ключові слова: персональна інформація, шифрування зі збереженням формату, псевдонімізація, маскування даних.

Вступ

Особиста інформація (Personal Identifiable Information PII) – це будь-які дані, які можуть бути використані для ідентифікації конкретної особи. Приклади включають повне ім’я особи, телефонні та номери соціального страхування, фізичні або імейл адреси, тощо.

Особиста інформація містить не лише очевидні посилання на особистість людини, а і інформаційні фрагменти, які в поєднанні з іншими наборами даних розкривають особу, тобто класифікується як ідентифікаційна інформація.

Конфіденційна ідентифікаційна інформація включає будь-який набір даних, який містить повне ім’я, адресу або фінансову інформацію. Неконфіденційна ідентифікаційна інформація – це будь-які загальні дані, доступні з загальнодоступних ресурсів (таких як профілі в соціальних мережах), наприклад поштовий індекс або дата народження.

Доступ до ідентифікаційної інформації без авторизації становить значний ризик як для окремих осіб, так і для компанії. Індивідуальна шкода може включати крадіжку особистих даних, шахрайство або шантаж. Водночас компанія може зазнати шкоди через порушення цілісності даних, наприклад втрати довіри громадськості, юридичної відповідальності або великих штрафів.

Ідентифікаційну інформацію можна оцінити шляхом визначення рівня її впливу на конфіденційність ідентифікаційної інформації. Рівні впливу на конфіденційність ідентифікаційної інформації варіюються від низького, помірного або високого, щоб вказати на потенційну шкоду, яка може бути завдана особі чи організації, якщо дані будуть скомпрометовані.

Під час внутрішніх тестів на проникнення в мережу консультанти з безпеки часто знаходять конфіденційну інформацію, яка незахищено зберігається на файлових серверах. Це включає чутливі дані про особу (повне ім’я, телефони, адреси, як фізичні так і електронні), номери кредитних карток, паспортів, сторонніх додатків для входу в платіжні системи, аналітику та інші веб-маркетингові та бізнес-портали. Це один із головних ризиків, виявлених під час незахищених практик зберігання інформації.

Оскільки кожне підприємство, незалежно від його розміру, підлягає під закони про ідентифікаційну інформацію, нормативні акти та інші повноваження, пов’язаних із захистом ідентифікаційної інформації, воно повинне дотримуватися чіткого підходу щодо безпеки та конфіденційності даних, щоб захистити ідентифікаційну інформацію, яку зберігає і використовує у своєму середовищі [1]. Часто перепоною для компаній, що оперують персональними даними людей є комплексність в реалізації шифрування даних, які вони зберігають. Це пов’язане з тим, що часто вони використовують застарілі системи, чиї кодові бази та моделі даних не можна змінити або надто обтяжливі та ризиковані для оновлення. Наприклад, база даних, яка зберігає конфіденційні дані про охорону здоров’я, яка була введена в дію майже 20 років тому і продовжує працювати. Більшість організацій схильні до ризиків, пов’язаних з реструктуризацією такої ключової виробничої системи [2].

Постійно зростаюча кількість випадків витоку даних, пов’язаних із особистою (ідентифікаційною) інформацією, призвела до втрат акціонерів на мільярди доларів, штрафів на мільйони доларів і збільшення ризику крадіжки особистих даних осіб, чиї конфіденційні дані були розкриті. Порушення безпеки даних є небезпечним як для окремих осіб так і організацій. У зв'язку з цим актуальним завданням є втілити методи які б допомагали таким компаніям шифрувати та маскувати дані з мінімальними вкладами в реструктуризацію та збереження функціональності на льоту.

Мета статті та постановка задач

Метою є розробка додатку для шифрування персональних даних зі збереженням формату шляхом удосконалення методу псевдонімізації та маскування даних.

Для досягнення поставленої мети необхідно розв’язати наступні задачі:

1. Проаналізувати принципи роботи та визначити переваги та недоліки сучасних методів шифрування зі збереженням формату та методи псевдонімізації даних.

2. Удосконалити метод псевдонімізації та маскування даних.

3. Реалізувати додаток для шифрування даних зі збереженням формату на основі розробленого покращеного методу псевдонімізації та маскування даних.

Об’єктом дослідження є удосконалення методу псевдонімізації та маскування даних.

Предметом дослідження є методи шифрування зі збереженням формату та

псевдонімізації даних.

Отримані результати можуть бути використані для підвищення стійкості до кібератак підприємств та покращення захисту персональних даних, якими вони користуються і зберігають.

Основна частина

Шифрування із збереженням формату (FPE) – це процес шифрування даних таким чином, щоб вихідні дані (зашифрований текст) залишалися в тому самому форматі, що й вхідні дані (відкритий текст). Значення «формату» різне. Зазвичай використовуються лише кінцеві набори символів; цифровий, буквений або буквено-цифровий» [3].

FPE дає змогу розгортати широко поширене шифрування без критичних наслідків для застарілих систем. Додаткам, яким не потрібен доступ до конфіденційних даних, надаються шифртексти FPE. Якщо програмі потрібен доступ до даних, у робочий процес можна вставити функцію дешифрування, щоб надати доступ до відкритого тексту.

У шифруванні із збереженням формату розглядається три режими роботи – FF1, FF2 та FF3, які в цілому називаються FFX. (NIST «Спеціальна публікація (SP) 800-38G, Рекомендації щодо режимів роботи блокового шифру: методи шифрування зі збереженням формату»).

FFX використовує кілька раундів функції Feistel над відкритим текстом разом із ключем для створення зашифрованого тексту. Функція Feistel розділяє відкритий текст на дві частини, переставляє текст, щоб змінити його вигляд, а потім змінює ліву половину тексту на праву і навпаки. Метод FF1 використовує 10 раундів функції Фейстеля, а FF3 використовує 8 раундів, FF2 так і не отримав схвалення NIST.

FF3-1 – це схема шифрування, що настроюється, створена для вирішення проблеми шифрування на невеликих доменах. Було введено налаштування (tweaks) для синтетичного збільшення доменного простору. З введенням налаштувань зловмисник тепер повинен мати правильне шифрування відкритого тексту та правильне налаштування, щоб через кодову книгу повернутися від шифрування до відкритого тексту.

При аналізі існуючих режимів шифрування зі збереженням формату, був виділений режим FF3-1 [4], який був реалізований у програмному додатку.

У європейскому загальному регламенту захисту даних (GDPR) псевдонімізація визначається як «обробка персональних даних таким чином, що дані більше не можуть бути пов’язані з певним суб’єктом даних без використання додаткової інформації». Таким чином відбувається обмін особистими даними з неідентифікуючими даними, і для відтворення вихідних даних потрібна додаткова інформація.

Метод псевдонімізації, як визначено в GDPR, – це будь-який метод, який гарантує, що дані не можна використовувати для ідентифікації особистості. Це вимагає видалення прямих ідентифікаторів і, бажано, уникнення кількох ідентифікаторів, які в поєднанні можуть ідентифікувати особу. Крім того, ключі шифрування або інші дані, які можна використовувати для повернення до початкових значень даних, слід зберігати окремо та надійно.

Псевдонімізація може бути досягнута за допомогою різних методів, таких як маскування даних, шифрування або токенізація. Вона зазвичай використовується як техніка для захисту особистих даних у застарілих виробничих системах від несанкціонованого доступу, де інші методи безпеки непридатні.

Маскування даних – це спосіб створити фальшиву, але реалістичну версію організаційних даних. Мета полягає в тому, щоб захистити персональні та конфіденційні дані, одночасно забезпечуючи функціональну альтернативу, коли

реальні дані не потрібні, наприклад, під час навчання користувачів, демонстрацій продажів або тестування програмного забезпечення.

Процеси маскування даних змінюють значення даних залишаючи їх вхідний самий формат. Мета полягає в тому, щоб створити версію, яка не піддається розшифровці або зворотній інженерії. Є кілька способів змінити дані, включаючи перетасування символів, заміну слів або символів і шифрування (рисунок 1).

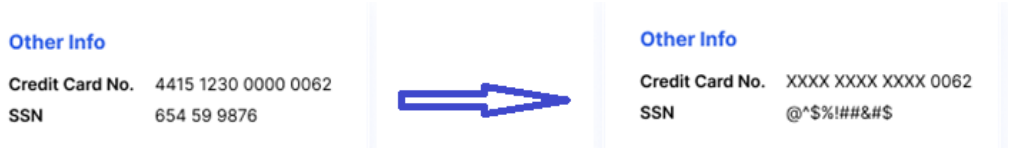


Рис.1. Приклад маскування даних

Існує чимало методів маскування даних. Розглянемо такі як перетасування, обнуління, варіація значень та перестановка.

Якщо існує, ідентифікаційний номер, такий як 76498 у робочій базі даних, можна замінити на 84967 у тестовій базі даних. Це метод перетасовування окремих знаків у значенні даних. Він дуже простий у реалізації, але його можна застосувати лише до деяких типів даних і він менш безпечний.

Під час перегляду неавторизованим користувачем дані виглядають відсутніми або «нульовими», це метод обнуління. Але це робить дані менш корисними для розробки та тестування.

Вихідні значення даних замінюються функцією, наприклад різницею між найменшим і найвищим значенням у ряді. Наприклад, якщо клієнт придбав кілька продуктів, ціну покупки можна замінити діапазоном між найвищою та найнижчою сплаченою ціною. Це метод варіації значень і він може надати корисні дані для багатьох цілей, не розкриваючи вихідний набір даних.

В методі перестановки, за винятком того, що значення даних змінюються в межах одного набору даних. Дані переставляються в кожному стовпці за допомогою випадкової послідовності; наприклад, перемикання між реальними іменами клієнтів у кількох записах клієнтів. Вихідний набір виглядає як реальні дані, але він не відображає справжню інформацію для кожної особи чи запису даних.

Був проведений аналіз продуктів з імплементованим методами псевдонімізації [5] та шифруванням зі збереженням формату та виділені їх сильні та слабкі сторони. Після аналізу сильних та слабких сторін продуктів на ринку, було виявлено, що реалізовані продукти пропонують лише типи маскування за допомогою спеціальних знаків (“*”, “#”, або “0”) та шифрування полів типу ім’я, прізвище та email адреса в рамках змішування значень в межах документа, або перемішування букв в межах самих значень, що слабо підходить під загальну концепцію псевдонімізації даних. Тому було вирішено покращити методи псевдонімізації та маскування полів даних: типу ім’я, прізвище та email адреса та розумне застосування методу FPE в режимі FF3-1 для полів типу номер телефону та кредитної карти.

Для того, щоб досягти цілі чіткої псевдонімізації, щоб факт підміни реальних даних був непомітним, вирішено модернізувати підхід до маскування та шифрування для полів типу ім’я, прізвище та email адреса.

Для псевдонімізації поля “ім’я” та “прізвище” потрібно замінювати реальні значення на інші, яке будуть іншими реальними іменами і прізвищами. Щоб виконати поставлену задачу, необхідно створити словники з унікальними людськими (в межах словника) іменами та прізвищами, достатньо великі, щоб алгоритм випадково вибраного числа міг обрати значення за порядковим номером

для подальшої заміни справжнього імені і прізвища користувача. Потім в окремий словник з ключами для імен та прізвищ буде записано реальне ім'я/прізвище користувача (значення) і порядковий номер фейкового імені/прізвища (ключ) для подальшого процесу де-псевдонімізації. Наприклад: якщо за алгоритмом було згенероване число 23, то у словнику з фейковими іменами шукаємо ім'я за порядковим номером 23. Заміняємо реальне ім'я користувача на псевдонім "Jonathan" та зберігаємо реальне ім'я окремо в словник ключів поряд з згенерованим числом 23. Такий же процес для заміни значення прізвища.

Процес де-псевдонімізації буде відбуватись у зворотньому напрямку. Підгружаючи файл із зашифрованими та замаскованими даними, програма розділить файл на колонки і зчитає значення в колонці "ім'я", і знайде значення "Jonathan" у словнику фейкових імен. Відповідно до знайденого імені, за алгоритмом буде взято порядковий номер цього імені у словнику фейкових імен та потім у словнику ключів для імен за ключем порядкового номеру, числа 23, знайдемо реальне ім'я користувача, яке буде значенням у цьому словнику для ключа 23. Такий же процес для де-псевдонімізації поля прізвища. Таким чином, частина де-псевдонімізації полів "ім'я" та "прізвище" буде завершена.

Було також розроблене та імплементоване рішення для захисту словників даних за допомогою AES (режим CBC) з ключем шифрування в 128-біт [6].

Щодо удосконаленого методу псевдонімізації для поля "email" адреси, то процес передбачає відокремлення домену (domain) адреси від частини юзернейму (local-part) користувача, тому що маскувати чи шифрувати домен email адреси частіше за все не є виправданим. Процес псевдонімізації email адреси буде спиратися на замасковані ім'я та прізвище, щоб у вихідному (замаскованому і зашифрованому) датасеті дані виглядали як справжні. Для цього візьмемо першу букву замаскованого імені і ціле замасковане прізвище та поєднаємо їх через крапку. Наприклад, для замаскованого юзера з ім'ям Jonathan та прізвищем Richardson частина юзернейма після алгоритму об'єднання буде виглядати наступним чином: J.Richardson. Наступним кроком буде частина генерації 8 значного випадкового числа (наприклад, 74925207), і приєднання цього номеру до згенерованої частини юзернейму email адреси. Також, це 8 значне згенероване число (74925207) буде служити ідентифікатором, тобто ключем, для зберігання реального юзернейму email адреси як значення у словнику ключів для email адрес для їх подальшої де-псевдонімізації. Тобто, замаскований юзернейм email адреси виглядатиме наступним чином: J.Richardson74925207. Потім до згенерованого юзернейму додається символ @ та домен реальної email адреси і це буде замасковане значення email адреси.

Процес де-псевдонімізації для поля email адреси виглядатиме наступним чином: підгружаючи файл із зашифрованими та замаскованими даними, програма розділить файл на колонки і зчитає значення в колонці "email" адрес, і знайде значення "J.Richardson74925207@gmail.com". Згідно подальших кроків, алгоритм зіскануює значення email адреси і виявить у ній 8-значне число (74925207), потім у словнику ключів для email адрес буде виконано пошук по ключам і для знайденого ключа 74925207 поряд буде значення реального юзернейму email адреси. Наступним кроком стане приєднання (через символ @) незамаскованого домену email адреси до знайденого значення (реального юзернейму) у словнику ключів для email адрес. Таким чином, частина де-псевдонімізації поля "email" адреса буде завершена.

Для повноцінного маскування та шифрування зі збереженням формату для поля "номер телефону" було обрано наступний алгоритм дій. За основу було взято номер телефону українського формату, де можливі два типи збереження з кодом країни (+380) та без нього. Оскільки в табличних даних, де зберігаються номери

телефонів, (в csv файлах) частіше за все, не підтримується форматування для цифрових значень, що починається з 0 або з +, було вирішено не враховувати шифрування таких випадків. А отже, маємо значення довжиною в 12 і 9 символів відповідно.

Для того, щоб забезпечити гарантовано надійне маскування поля номера телефону і 12 символів, було вирішено відділити частину номера телефону, що є кодом країни (2 символа) і також код мобільного оператора (3 символа), тобто, перші 5 символів і не піддавати їх шифруванню, а шифрувати методом зі збереженням формату FPE в режимі FF3-1. Тобто, наприклад, коли програма зчитує мобільний номер з поля у вигляді “380671234567”, то значення “38067” записується як окреме значення і зберігається окремо, а значення “1234567” буде шифруватись і перетворюватись у інше цифрове значення такої ж довжини (7 цифр). Після того до зашифрованого значення приєднуємо значення з кодом країни і кодом оператора і зберігаємо як повноцінне зашифроване і замасковане значення. Під час шифрування методом зі збереженням формату FPE в режимі FF3-1 нам потрібні два значення key та tweak, які будуть збережені у словнику для ключів мобільних телефонів. При розшифруванні береться номер по порядку і проходиться по словнику ключів розшифровуючи значення за значенням, таким чином відбувається процес розшифрування.

Зазвичай, люди не вникають у глибокий аналіз номерів карток. Для них достатньо всього побачити 16 цифр щоб ідентифікувати, що це саме номер картки. Тому було вирішено шифрувати усі 16 цифр картки, щоб уникнути можливих ризиків методом зі збереженням формату FPE в режимі FF3-1. Тобто, наприклад, коли програма зчитує поле “credit card” і дані карток з нього, то шифрування зі збереженням формату буде виконуватись над 16 цифрами, де в результаті шифрування на виході буде інше зашифроване 16-значне значення.

У розробленому програмному додатку пропонується два режиму на вибір для корисувача: зашифрувати, та розшифрувати файли. Вікно при виборі режиму шифрування зображено на рисунку 2, де зліва показано опції розділювача для підвантаженого .csv файлу з реальними даними користувачів, та кнопка підвантаження файлу. Праворуч зображений treeview з підвантаженими колонками з файлу користувача та даними в ньому, а також динамічно згенеровані labels, що відповідають назвам колонок, та опції dropdown lists (рис.3), де наявні назви методів якими користувач може зашифрувати або замаскувати відповідну до label-а колонку.

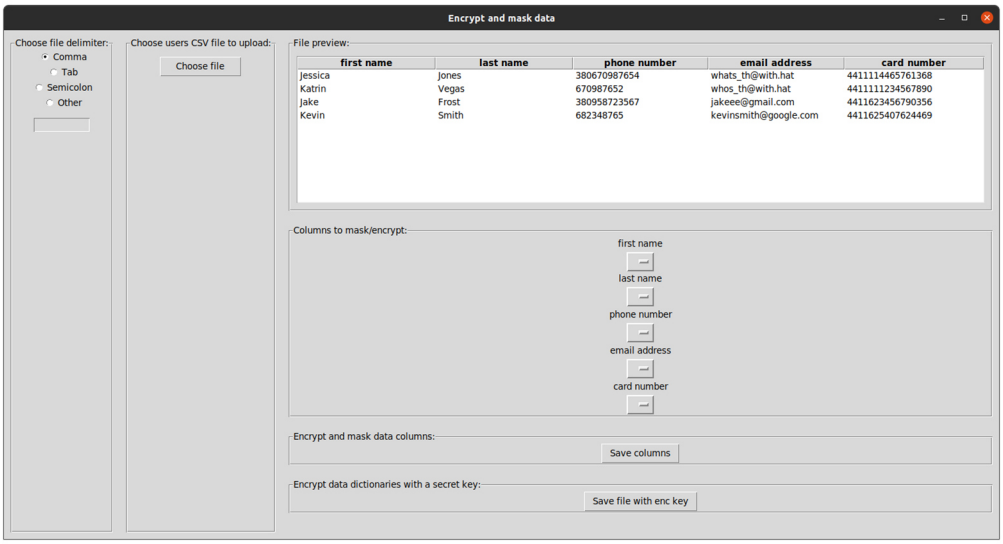


Рис. 2. Вікно програми в режимі шифрування

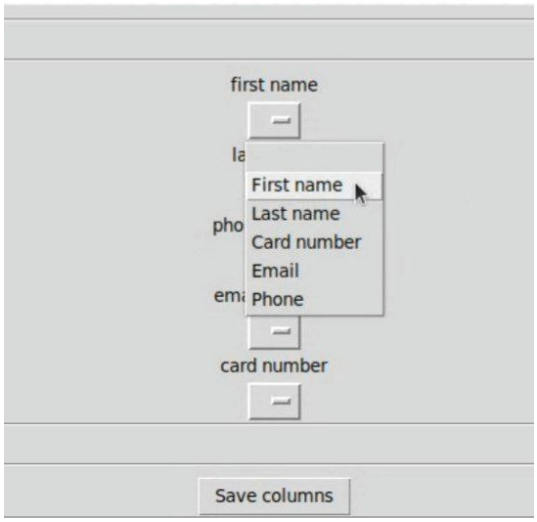


Рис. 3. Опції вибору у dropdown list-ах

Потім користувач натискає на кнопку “Save columns” та зберігає encrypted.csv файл разом з файлами словників з ключами відповідно до вибраних опцій в dropdown lists. Наступним кроком користувачу потрібно натиснути кнопку “Save file with enc key”, що зашифрує словники (за допомогою AES (режим CBC) з ключем шифрування в 128-біт) з ключами та збереже один файл ключа під назвою mykey.key для майбутнього розшифрування цих словників.

Для порівняння вхідних і вихідних (замаскованих та зашифрованих даних наведені наступні скріншоти датасетів (рис. 4) та (рис. 5) відповідно.

	A	B	C	D	E
1	first name	last name	phone number	email address	card number
2	Jessica	Jones	380670987654	whats_th@with.hat	4411114465761368
3	Katrin	Vegas	670987652	whos_th@with.hat	4411111234567890
4	Jake	Frost	380958723567	jakeee@gmail.com	4411623456790356
5	Kevin	Smith	682348765	kevinsmith@google.com	4411625407624469
6					

Рис. 4. Вхідний файл з реальними даними користувачів users_w_comma.csv

	A	B	C	D	E
1	first name	last name	phone number	email address	card number
2	Paul	Sanders	380677766448	P.Sanders86077122@with.hat	7698650865550020
3	Jeffrey	Conley	679419326	J.Conley29851818@with.hat	6764499424904407
4	Sue	Acevedo	380952469074	S.Acevedo47794405@gmail.com	5127538074400576
5	Padraig	Kline	683828986	P.Kline74270084@google.com	4794620070515788
6					

Рис.5. Вихідний файл з замаскованими та зашифрованими даними користувачів encrypted.csv

Наступним пунктом є дешифрування та де-псевдонімізація даних. Для цього користувач при запуску програми вибирає опцію розшифрування. У оновленому вікні (рис. 6) він підвантажує попередьо зашифрований файл encrypted.csv та ключ для розшифрування словників, потім натискає на кнопку “Decrypt users data” та після оповіщення про успішне розшифрування натискає на кнопку “Save decrypted file”.

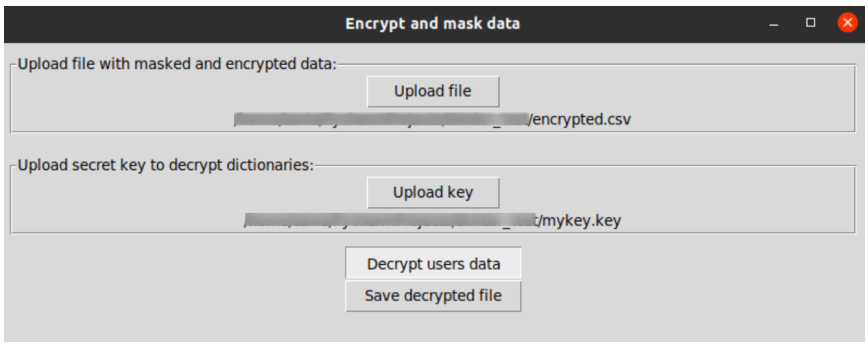


Рис. 6. Вікно програми в режимі дешифрування

Після цього файл буде збережено у вибраній директорії і таким чином процес дешифрування та де-псевдонімізації даних завершено.

Метод шифрування зі збереженням формату та псевдонімізації даних у розробленому додатку було реалізовано за допомогою використання наступних бібліотек Python: codecs, string, cryptography.fernet >> Fernet, pandas, csv, random, os, binascii, fff3 >> FF3Cipher, Crypto.Cipher >> AES, numpy, pandas >> read_csv.

Висновки

Під час розробки додатку для шифрування персональних даних зі збереженням формату були імплементовані удосконалені методи псевдонімізації та маскування даних. Були розв'язані поставлені на початку задачі та був розроблений додаток для шифрування даних зі збереженням формату та удосконаленими методами псевдонімізації на високорівневій мові програмування Python з використанням допоміжних бібліотек. Розроблене програмне забезпечення задовольняє шифрування персональних даних зі збереженням формату та удосконаленими методами псевдонімізації полів даних: типу ім'я, прізвище та email адреса та імплементоване розумне застосування методу FPE в режимі FF3-1 для полів типу номер телефону та кредитної карти.

Результати даної роботи можуть бути можуть бути використані для підвищення стійкості до кібератак підприємств та покращення захисту персональних даних, якими вони користуються і зберігають.

Список літератури

1. PII Protect Cybersecurity. How to Secure your Data URL: https://thecyphere.com/blog/pii-protect/
2. What is format-preserving encryption (FPE) and its benefits? URL: https://www.ubiqsecurity.com/what-is-format-preserving-encryption-fpe-and-its-benefits%EF%BF%BC/
3. Recommendation for Block Cipher Modes of Operation: Methods for Format-Preserving Encryption URL: https://csrc.nist.gov/publications/detail/sp/800-38g/rev-1/draft
4. Recent Cryptanalysis of FF3 URL: https://csrc.nist.gov/news/2017/recent-cryptanalysis-of-ff3
5. Pseudonymization according to the GDPR [definitions and examples] URL: https://dataprivacymanager.net/pseudonymization-according-to-the-gdpr/
6. Advanced Encryption Standard (AES) URL: https://www.techtarget.com/searchsecurity/definition/Advanced-Encryption-Standard

DEVELOPMENT OF AN APPLICATION FOR DATA ENCRYPTION WITH FORMAT PRESERVATION

T. Horbatiuk, O. Lebedieva

National Odesa Polytechnic University,
1, Shevchenko Ave., Odesa, 65044, Ukraine; e-mail: infsec2011@gmail.com

The paper describes the development of an application for data encryption with format preservation. Basic definitions such as personal information, identifying information are provided. Over the past few years, a growing number of personal data breaches have resulted in the loss of data integrity for hundreds of millions of user records. Such attacks are aimed at both large companies and small businesses. The cyber-attacks aimed at stealing personal data from enterprises and it is ensured due to the negligence of companies regarding the conscious encryption and masking of personal data of employees and customers. The paper reveals the concept of format-preserving data encryption, as the process of encrypting data in such a way that the output data remains in the same format as the input data. Two modes of operation are considered in format-preserving encryption - FF1 and FF3. When analyzing the existing FPE modes with format preservation, the FF3-1 mode was selected due to the ability to solve the problem of encryption on small data domains. FF3-1 mode was implemented in a software application. The work also reveals the concept of pseudonymization and methods of achieving it. Pseudonymization can be achieved using various methods such as data masking, encryption or tokenization. Data masking process change the value of data while leaving its input format. The goal is to create an output that cannot be deciphered or reverse engineered. The method of data masking was implemented. Thus, the main task was to implement methods that would help companies encrypt and mask data with minimal contributions to restructuring and preserving functionality on the fly. In order to achieve the goal of clear pseudonymization, so that the fact of replacing real data is imperceptible, it was decided to modernize the approach to masking and encryption for fields such as first name, last name, and email address. The methods of pseudonymization and data fields masking: first name, last name, and email address, and the use of the FPE method in FF3-1 mode for fields such as phone number and credit card were implemented. An improved method of pseudonymization and masking of data fields is implemented in the software application.

Keywords: personal information, format-preserving encryption, pseudonymization, data masking.

DOI 10.15276/imms.v12.no4.315
УДК 519.68.177:629.022

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 315-326

РОЗРОБКА КОМП'ЮТЕРНИХ ПРОГРАМ ОПТИМАЛЬНОГО ПРОЄКТУВАННЯ ПАСИВНОЇ ПІДВІСКИ КОЛІСНОГО РОБОТА

Н.М. Єршова, Н.В. Ткачук, С.А. Ренгач

Придніпровська державна академія будівництва та архітектури
Дніпро, 49005, вул.Архітектора Олега Петрова, 24а, E-mail:
nersova107@gmail.com, tkachuk.nadya7@gmail.com, docker1280@gmail.com

Зараз широко використовуються мобільні роботи. На основі аналізу великої кількості можливих варіантів підвіски дійшли висновку про доцільність застосування колісних роботів з динамічною підвіскою. Для проектування систем, що перебувають під впливом випадкових обурень, розроблено метод стохастичного динамічного програмування. Метод дуже ефективний, оскільки дозволяє визначити закон оптимального управління параметрами підвіски робота. Відповідно до цього методу проектування підвіски робота має виконуватися поетапно: проектується пасивна підвіска на основі методу динамічного програмування для безперервних детермінованих систем та аналізу динамічних показників робота; перевіряється доцільність введення в пасивну підвіску пристроїв активного управління параметрами за алгоритмом Калмана-Бьюсі і приймається остаточне рішення про принцип дії підвіски; визначається закон оптимального управління параметрами підвіски робота, якщо прийнято рішення про проектування активної підвіски. В цьому випадку потрібно менше витрат зовнішньої енергії для зменшення шкідливих коливань. У існуючій науково-технічній літературі розглядаються математичні моделі роботів з активною підвіскою і неясно, на якій основі прийнято параметри підвіски вихідної системи. В даній роботі розглядається перший етап проектування підвіски колісного робота. Комп'ютерних програм для оптимального проектування пасивної підвіски колісного робота не існує. Отже, метою даної роботи є розробка комп'ютерних програм оптимального проектування пасивної підвіски колісного робота. В основу алгоритму пошуку проектних рішень закладено матричний метод динамічного програмування Р. Беллмана для неперервних динамічних систем, що є науковою новизною роботи. Створена комп'ютерна програма «ROBOT» призначена для вибору оптимальних параметрів підвіски за умови, що максимальні значення прискорення та коефіцієнта вертикальної динаміки центру мас робота не перевищують допустимих значень. Оцінка динамічних властивостей робота виконується за комп'ютерною програмою «DINAM», результати якої видаються у вигляді графіків та таблиці розрахунків. В якості моделі-аналога прийнято параметри колісного робота TIGER. На основі комп'ютерного моделювання, варіантного моделювення і оптимізації отримані параметри пасивної підвіски робота, що забезпечують йому найкращі динамічні властивості порівняно з моделлю-аналогом. Дана робота є важливим внеском в галузі комп'ютерних наук і програмної інженерії.

Ключові слова: оптимальне проектування підвіски колісного робота, комп'ютерна програма, матричний метод динамічного програмування

Вступ

В даний час широко використовуються мобільні роботи. Залежно від робочого середовища розрізняють п'ять типів мобільних роботів: наземні, підземні, літаючі, плаваючі та космічні. Наземні мобільні роботи, у свою чергу, залежно від способу пересування поділяються на такі класи: колісні, крокуючі, гібридні та спеціалізовані. Найбільшого поширення набули колісні мобільні наземні роботи. За способом управління роботою коліс розрізняють такі групи колісних роботів: з жорстко закріпленими колесами; автомобільна група (поворот

здійснюється лише за рахунок задніх коліс); з довільним незалежним керуванням поворотом кожного колеса вліво чи вправо; група роботів, здатних переміщатися у будь-яких напрямках. Друга та третя групи забезпечують найкращу динаміку машини. На основі аналізу великої кількості можливих варіантів підвіски зроблений висновок про доцільність застосування колісних роботів з динамічною підвіскою автомобільної групи управління.

Для проектування систем, що перебувають під впливом випадкових обурень, розроблено метод стохастичного динамічного програмування. Метод дуже ефективний, оскільки дозволяє визначити закон оптимального управління параметрами підвіски робота.

Відповідно до методу стохастичного динамічного програмування проектування підвіски робота виконується поетапно [1]:

– проектується пасивна підвіска на основі методу динамічного програмування для безперервних детермінованих систем та аналізу динамічних показників робота;

– перевіряється доцільність введення в пасивну підвіску пристроїв активного керування параметрами за алгоритмом Калмана-Бьюсі та приймається остаточне рішення про принцип дії підвіски;

– визначається закон оптимального управління параметрами підвіски робота, якщо прийнято рішення про проектування активної підвіски.

Задачі оптимального проектування зручно вирішувати за допомогою матричного методу динамічного програмування Р. Беллмана. Для реалізації алгоритмів пошуку проектних рішень необхідно створювати відповідні комп'ютерні програми.

Аналіз останніх досліджень і публікацій

В останні роки роботи прискорили темпи заміни людей у виконанні завдань, орієнтованих на людину. Оскільки робот стає розумнішим, основна його місія не обмежується транспортуванням, а поширюється на маніпуляції. Одним із прикладів є гуманоїдний робот на колесах, який має гуманоїдний корпус на верхній частині робота, при цьому робот може як і пересуватися, так і виконувати маніпуляції [2, 3].

Проведені роботи над колісною мобільною платформою для стійкості можна розділити на два типи. У першому випадку стабільність забезпечується пасивним способом. До них відносяться збільшення ваги колісної мобільної платформи [4], а також збільшення площі опорного полігону [5]. У будь-якому випадку центр мас робота розташований більш стабільно в межах опорного полігону. Пасивне управління легко здійснити, якщо діапазон його застосування обмежений механічною конструкцією. В цих дослідженнях не використується пасивне управління пружно-дисипативними параметрами підвіски робота.

Другий підхід, навпаки, – активне управління. Цей підхід гарантує стабільність шляхом активного управління розташуванням центра мас робота. До цього підходу відносяться численні публікації про стійкість гуманоїдних роботів [6, 7, 8]. Активне управління колісною мобільною платформою досліджувалось скоріше в контексті активної підвіски легкових автомобілів [9, 10]. Хоча його ефективність очевидна, активна підвіска не була впроваджена в колісних роботах. В роботах [11, 12] викладено принципи активного управління параметрами пасивної підвіски на основі методу стохастичного динамічного програмування.

У роботі [13] представлено гнучке активне управління колісним роботом TIGER за допомогою пружного приводного механізму. Робот (рис. 1) управляє обладнаною підвіскою та мінімізує кути крену та нахилу робота, коли він знаходиться на нерівній місцевості. Для управління розроблена повна модель руху робота та сенсорна модель інерційного вимірювального блока. Стан робота

оцінюється за допомогою спостерігача, де фільтр Калмана виконує роль компенсатора. Потім розробляється і використовується лінійний квадратичний регулятор для здійснення оптимального управління за допомогою зворотного зв'язку.

Рис. 1. Колісний робот TIGER

Робот складається з корпусу, з'єднаного з колесами через еластичний приводний механізм. Колесо підтримується на кузові за допомогою пружинно-амортизаційної системи послідовно з лінійним редукторним двигуном постійного струму.

Коли лінійний двигун постійного струму рухається, відносна відстань між тілом робота та колесом змінюється.

Основною функцією тіла робота є перенесення всіх компонентів робота, включаючи батареї, електричні та електронні компоненти, різне корисне навантаження та опорна система підвіски.

Корпус робота повинен бути міцним і легким. Корпус кузова розроблений за принципом ферми для забезпечення жорсткості, невелика вага, проста збірка та виготовлення. Найбільш новими частинами конструкції робота є підвіска з пружним приводним механізмом.

Лінійний двигун постійного струму підключений до демпфера через коромисло, щоб активно управляти висотою підвіски та ізолювати тіло робота від зовнішніх перешкод, які можуть бути спричинені рельєфом та іншими нерівностями.

Нижній важіль підвіски розроблено з різними місцями кріплення амортизатора, для зміні початкової конфігурації підвіски відносно тіла робота.

Динамічна модель пружного приводного механізму представлена системою з двома ступенями свободи, як показано на рисунку 2.

Шина представлена у вигляді маси m_u, що спирається на землю через циліндрову пружину з жорсткістю K_t.

Амортизатор зображено як пружинно-демпферний блок, K_s, C_s, що підтримує тіло робота масої m_s з активною силою F_a, де K_s - жорсткість

пружини, C_s - коефіцієнт опору гідравлічного гасителя коливань. z - лінійне переміщення центрів мас уздовж вертикальної осі.

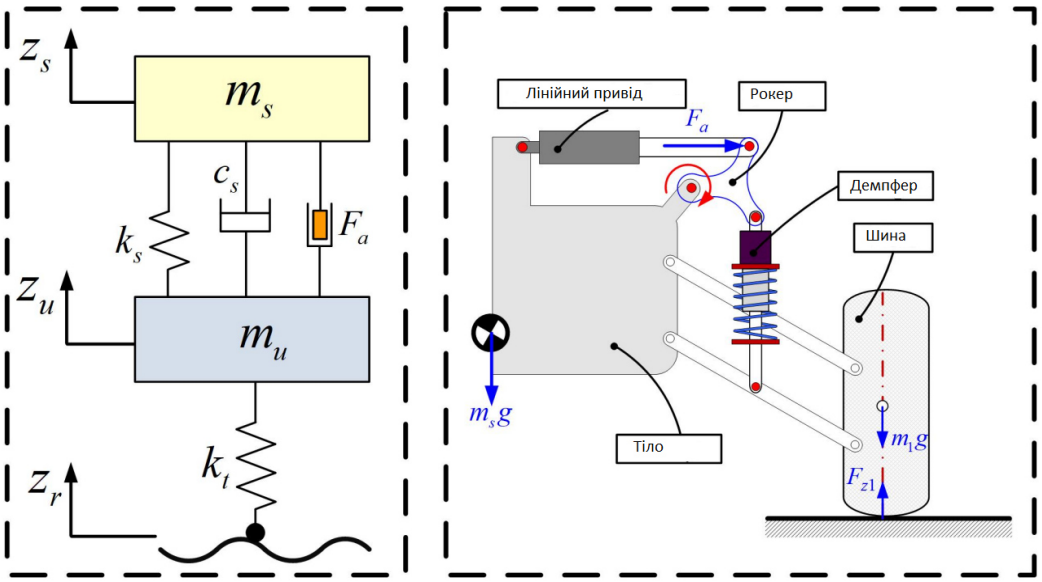


Рис. 2. Динамічна модель підвіски робота TIGER

Рівняння динаміки управління системою мають наступний вигляд:

$$m_s \ddot{z}_s = K_s(z_u - z_s) + C_s(\dot{z}_u - \dot{z}_s) + F_a;$$

$$m_u \ddot{z}_u = -K_s(z_u - z_s) - C_s(\dot{z}_u - \dot{z}_s) + K_t(z_r - z_u) - F_a.$$

Для оптимального управління розроблено лінійну динамічну модель робота з сімома ступенями свободи, як показано на рис. 3. Параметри робота наведені в табл. 1.

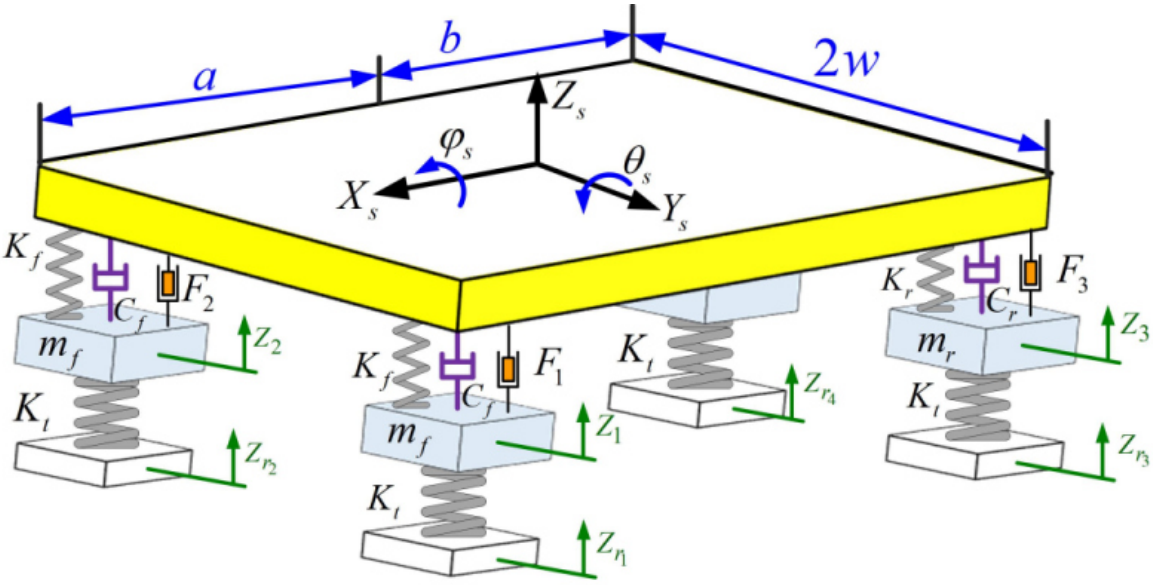


Рис. 3. Повна модель робота TIGER

На основі аналізу літературних джерел можна визначити задачі, які потрібно вирішити для покращення динамічних властивостей колісних роботів:

– створити методику оптимального проектування пасивної підвіски робота;

– виконати розробку комп'ютерних програм оптимального проектування пасивної підвіски робота;
– шляхом варіантного моделювання визначити раціональні значення коефіцієнта опору гасителя коливань шини.

Таблиця 1

Параметри підвіски робота TIGER

Назва параметра	Позначення	Значення
Підресорена маса	m_s	1,21 т
Передня непідресорена маса	m1,2	0,05 т
Задня непідресорена маса	m3,4	0,05 т
Жорсткість передньої пружини	K_f	20 кН/м
Жорсткість задньої пружини	K_r	20 кН/м
Жорсткість пружин шини	K_t	220 кН/м
Коефіцієнт опору переднього гасителя коливань	C_f	3 кНс/м
Коефіцієнт опору заднього гасителя коливань	C_r	3 кНс/м

Мета роботи

Отже метою даної роботи є розробка комп'ютерних програм оптимального проектування пасивної підвіски колісного робота. В основу алгоритма закла матричний метод динамічного програмування Р. Беллмана для безперервних динамічних систем.

Оптимальне проектування пасивної підвіски колісного робота

Диференціальне рівняння вертикальних коливань механічної системи «робот-дорога», що представлена найпростішою моделлю (рис. 4), записується у вигляді:

$$m\ddot{z} + b\dot{z} + cz = c\eta + b\dot{\eta}, \qquad (1)$$

де m - підресорена маса робота; c, b - відповідно жорсткість підвіски і коефіцієнт опору гасителів коливань, встановлених в підвісці робота; z - лінійне переміщення центру мас робота уздовж вертикальної осі; η - амплітуда нерівності дороги. Дорога і шини вважаються абсолютно жорсткими.

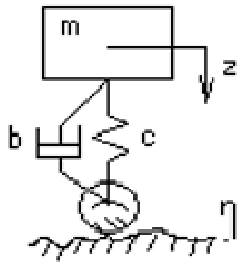


Рис. 4. Модель механічної системи «робот-дорога»

В методі динамічного програмування Р. Беллмана підвіска розглядається як управляюча функція, що регулює переміщення підресореної маси робота запропонованим образом. Вводиться сила, діюча з боку підвіски на масу (рис. 5).

Ставиться задача – визначити розрахункові формули параметрів проектування: жорсткості підвіски і коефіцієнта опору гасителів коливань [14].

Диференціальне рівняння коливального процесу має вигляд

$$m\ddot{z} = -F \qquad (2)$$

або

$$\ddot{z} = -u, \qquad (3)$$

де $u = F / m$ - невідома синтезуюча функція, тобто це прискорення.

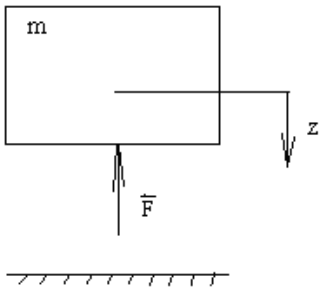


Рис. 5. Розрахункова схема найпростішого колісного робота

Рівняння (3) представимо в нормальній формі Коші, приймаючи $z = x_1, \quad \dot{z} = x_2$.

$$\begin{aligned} \dot{x}_1 &= x_2; \\ \dot{x}_2 &= -u. \end{aligned} \tag{4}$$

Напишемо систему (4) в матричній формі

$$\dot{X} = AX + BU, \tag{5}$$

де $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} z \\ \dot{z} \end{bmatrix}; \quad U = [u]; \quad A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$

Квадратичний функціонал якості, що характеризує витрату енергії на зменшення шкідливих коливань має вигляд

$$J = \int_0^\infty (X' PX + U' GU) dt \tag{6}$$

або

$$J = \int_0^\infty (\alpha x_1^2 + \gamma x_2^2 + \mu u^2) dt \,. \tag{7}$$

Отже, матриці P і G функціоналу мають вигляд: $P = \begin{bmatrix} \alpha & 0 \\ 0 & \gamma \end{bmatrix}; \quad G = [\mu].$

Формулювання задачі синтезу: знайти фізично здійсненну синтезуючу функцію рівняння (5), яка задовольняє обмеженням і доставляє мінімум функціоналу (7). Крім того, параметри проектування повинні забезпечити коливальний процес в системі з малими амплітудами переміщення і прискорення центру мас робота.

Необхідною умовою оптимальності є вирішення нелінійного алгебраїчного рівняння Ріккаті

$$P + A'S + SA - SBG^{-1}B'S = 0 \tag{8}$$

або в розгорненому вигляді

$$\begin{bmatrix} \alpha & 0 \\ 0 & \gamma \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} + \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} \begin{bmatrix} 1/\mu \end{bmatrix} \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} = 0.$$

Звідки отримаємо систему нелінійних алгебраїчних рівнянь для визначення елементів симетричної матриці помилки оцінки S.

$$\begin{cases} \alpha - S_{12}^2 / \mu = 0; \\ \gamma + 2S_{12} - S_{22}^2 / \mu = 0; \\ S_{11} - S_{12} S_{22} / \mu = 0. \end{cases} \tag{9}$$

З системи (9) визначаємо

$$S_{12} = \pm\sqrt{\alpha\mu}; \quad S_{22} = \pm\sqrt{\mu(\gamma + 2S_{12})}; \quad S_{11} = S_{12} S_{22} / \mu.$$

Оскільки матриця S повинна бути позитивно визначена, то з можливих значень S_{11}, S_{12}, S_{22} обираємо для подальших розрахунків тільки позитивні значення.

Синтезуюча функція визначається матричним виразом:

$$U = -G^{-1}B'SX = -\left[\frac{1}{\mu}\right]\begin{bmatrix}0 & -1\end{bmatrix}\begin{bmatrix}S_{11} & S_{12}\\ S_{12} & S_{22}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \frac{1}{\mu}(S_{12}x_1 + S_{22}x_2) = \frac{1}{\mu}(S_{12}z + S_{22}\dot{z}).$$

Після підстановки елементів S_{12}, S_{22} в синтезуючу функцію маємо

$$u = ((\gamma + 2\sqrt{\alpha\mu})/\mu)^{1/2}\dot{z} + (\alpha/\mu)^{1/2}z. \quad (10)$$

В рівняння (2) підставимо $F = mu$ і перенесемо складові рівняння в ліву частину

$$m\ddot{z} + m((\gamma + 2\sqrt{\alpha\mu})/\mu)^{1/2}\dot{z} + m(\alpha/\mu)^{1/2}z = 0. \quad (11)$$

Задаємося структурою проектованого об'єкту з класу лінійних систем. Припустимо, що в підвіску робота будуть встановлені циліндрові пружини і гідравлічні гасителі коливань, сила опору яких пропорційна першому ступеню швидкості переміщень. Вільні коливання моделі-аналога описуються диференційним рівнянням

$$m\ddot{z} + b\dot{z} + cz = 0, \quad (12)$$

де c, b – відповідно жорсткість підвіски і коефіцієнт опору гасителів коливань.

Рівняння (11) і (12) описують одну і ту ж систему, тобто параметри пружно-дисипативних зв'язків можна визначити шляхом порівняння коефіцієнтів при z і $\dot{z}$

$$c = m(\alpha/\mu)^{1/2};\ b = m((\gamma + 2\sqrt{\alpha\mu})/\mu)^{1/2}. \quad (13)$$

Отримали аналітичну залежність для параметрів проектування. Розв'язок отримується швидко і без особливих труднощів, але немає упевненості в тому, що при підстановці довільних значень α, γ, μ отримаємо фізично виконану підвіску, яка забезпечує роботу потрібні динамічні властивості в заданому діапазоні швидкостей руху. Без розв'язку цих проблем задача оптимізації позбавлена практичного сенсу. Отже, кожна отримана сукупність параметрів проектування повинна бути фізично виконана. Це досягається обгрунтованим вибором вагових коефіцієнтів квадратичного функціоналу якості.

Вибір вагових коефіцієнтів квадратичного функціоналу якості

Елементи матриць P і G зазвичай вибирають методом проб і помилок, що суттєво ускладнює синтез систем за даними критеріями. При вирішенні цієї проблеми необхідно виходити з основного призначення проектованої системи. Створюється підвіска робота, одне з призначень якої забезпечити коливальний процес в системі з малими амплітудами переміщення і прискорення центру мас робота, але наявність аперіодичного закону руху небажана. Щоб виключити резонансні явища в системі в період руху, частота коливань маси робота не повинна перевищувати 2 Гц.

Для отримання фізично реалізуємої підвіски досліджуємо залежність основних динамічних показників робота від вагових коефіцієнтів, представлену формулами:

$$\nu^2 = \sqrt{\alpha/\mu};\quad \xi^2 = (b/b_{kp})^2 = 0{,}5 + \frac{\gamma}{4\sqrt{\alpha\mu}}, \quad (14)$$

де $\nu^2 = c/m$ – квадрат частоти власних коливань маси робота; b/b_{kp} – ступень демпфування коливального процесу; m – маса робота; c – жорсткість підвіски; $b_{kp} = 2\sqrt{mc}$ – критична величина коефіцієнта опору гасителів коливань.

Аналіз формул свідчить про те, що ступінь демпфування знаходиться в прямій залежності від вагового коефіцієнта γ, причому при γ=0 ступінь демпфування $\xi \approx 0{,}7$; при γ>0 – ξ>0,7, що відповідає аперіодичному закону

руху маси. При проектуванні підвіски коефіцієнт γ може приймати негативні значення. Частота власних коливань залежить від коефіцієнтів α і μ . Оскільки частота коливань величина позитивна, то і вагові коефіцієнти α і μ повинні бути позитивними. Зона їх допустимих значень визначається реальним діапазоном частот власних коливань системи. Частота коливань маси робота повинна бути в межах 1…2 Гц, отже, значення α і μ повинні відповідати цьому діапазону. Звідси можна вивести залежністі між ваговими коефіцієнтами.

$$\alpha \le \mu v^4 = 12{,}56^4 \mu . \tag{15}$$

Вимога позитивної визначеності матриці S встановлює додаткову залежність між ваговими коефіцієнтами

$$\gamma > -\sqrt{\alpha\mu} . \tag{16}$$

Алгоритм пошуку проектних рішень

При проектувані підвіски робота потрібно вибрати таку сукупність параметрів проектування, яка забезпечить коливальний процес в системі з малими амплітудами переміщення і прискорення центру мас робота. Отже, алгоритм пошуку проектних рішень повинен містити блок розрахунку максимальних значень прискорення і коефіцієнта вертикальної динаміки центру мас робота.

Алгоритм пошуку проектних рішень приведено на рис. 6.

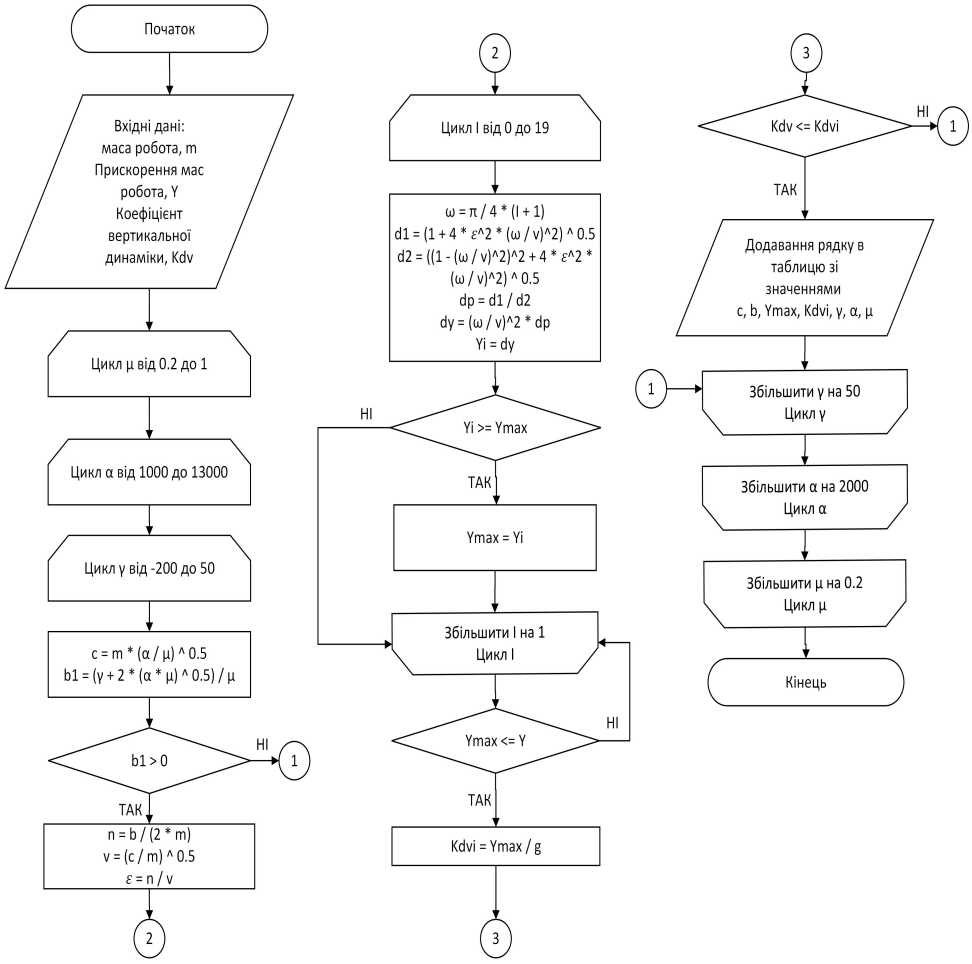
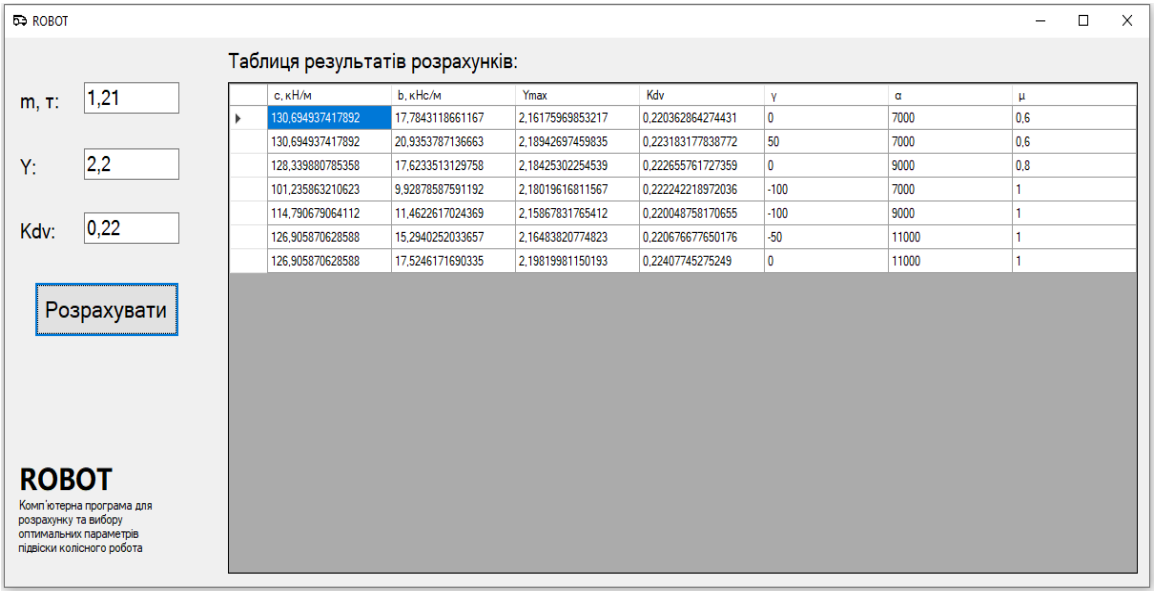


Рис. 6. Алгоритм пошуку проектних рішень

Створена комп'ютерна програма «ROBOT» оптимального проектування пасивної підвіски робота. Результати розрахунку для робота TIGER приведені в таблиці 2.

Таблиця 2

Оптимальни варіанти параметрів підвіски робота TIGER



Тестування комп'ютерної програми виконано шляхом комп'ютерного моделювання в середовищі SimlnTech [15], розрахунком за формулами в комп'ютерній програмі DINAM і середовищі Excel.

Для оцінки якості динамічних властивостей робота створена комп'ютерна програма DINAM, результати якої видаються у вигляді таблиці і графіків (рис. 7).

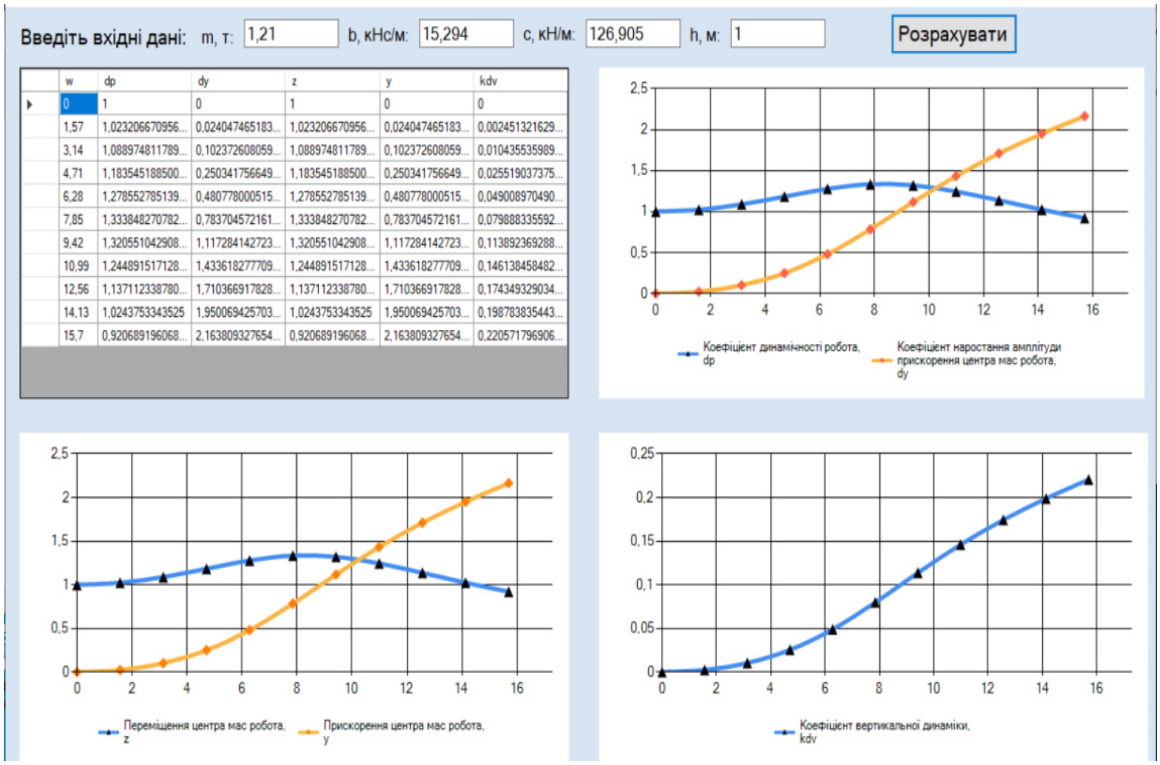


Рис. 7. Результати роботи програми DINAM

Для прийняття рішення була виконана за допомогою програми DINAM оцінка якості динамічних властивостей робота при різних варіантах параметрів підвіски, що наведені в таблиці 3.

Таблиця 3

Оцінка якості динамічних властивостей робота TIGER

Варіанти	b	c	d_p	$Y_{\max}$	K_{dv}
Модель-аналог	12	80	1,34	2,646	0,269
Варіантне моделювання	6,475	80	1,85	2,003	0,204
Оптимальний варіант 1	9,928	101,235	1,55	2,179	0,222
Оптимальний варіант 2	11,462	114,790	1,48	2,158	0,219
Оптимальний варіант 3	15,294	126,905	1,33	2,163	0,220
Оптимальний варіант 4	17,524	126,905	1,27	2,196	0,223

Існує поняття показник коливальності M - відношення максимального значення коефіцієнта динамічності d_p до його значення при $\omega = 0$ $M = d_{p\max} / d_{p0}$. Показник коливальності характеризує схильність системи до коливань. Чим вище M, тим менш якісна система за інших рівних умов. Вважається допустимим, якщо $1{,}1 \le M \le 1{,}5$. Робот TIGER має b=12 кНс/м і M=1,34, тобто відповідає допустимій нормі. Але максимальне значення амплітуди прискорення при ω=15,7 с$^{-1}$ дорівнює 2,646, що не допустимо. При виконані оптимізації допустимі значення $Y_{\max}$=2,2 м/с^2 і K_{dv}=0,22. Тоді для реалізації можна прийняти параметри оптимальних варіантів 2 і 3.

Висновки

1. У роботі представлена розробка комп'ютерної програми «ROBOT» для оптимального проектування пасивної підвіски колісного робота та комп'ютерна програма «DINAM» для оцінки його динамічних властивостей. Тестування комп'ютерних програм виконано шляхом комп'ютерного моделювання в середовищі SimlnTech і розрахунком за формулами в середовищі Excel.

2. На основі комп'ютерного моделювання, варіантного моделювання та оптимізації отримані параметри пасивної підвіски робота, що забезпечують йому найкращі динамічні властивості порівняно з моделлю-аналогом.

3. Комп'ютерне моделювання та оптимізацію потрібно виконати на першому етапі проектування, що дозволить створювати сучасні колісні роботи і скоротити їх термін проектування.

Список літератури

1. Єршова Н. М. Свідоцтво про реєстрацію авторського права на твір наукового характеру «Методология поэтапного проектирования системы подвешивания транспортных экипажей» № 69687 від 13.01.2017.
2. Petersen R., Philippsen L., Philippsen R. Implementation and stability analysis of prioritized whole-body compliant controllers on a wheeled humanoid robot in uneven terrains. *Autonomous Robots*, 2010. Vol. 35(4),P.301–319.
3. Rojas S., Shen H., Griffiths H., Li N., Zhang L. Motion and gesture compliance control for high performance of a wheeled humanoid robot. *ASME 2017 International Mechanical Engineering Congress and Exposition*. American Society of Mechanical Engineers, 2017. P. V04AT05A058 – V04AT05A058.

4. Lauwers T.B, Kantor G.A, Hollis R.L. A dynamically stable singlewheeled mobile robot with inverse mouse-ball drive. *Robotics and Automation. Proceedings 2006 IEEE International Conference*. IEEE, 2006. P 2884–2889.
5. Xin S., You Y., Zhou C., Fang C., Tsagarakis N. A torque-controlled humanoid robot riding on a two-wheeled mobile platform. *Intelligent Robots and Systems. IEEE/RSJ International Conference on*. IEEE, 2017. P.1435–1442.
6. Henze B., Roa M.A., Ott C. Passivity-based whole-body balancing for torque-controlled humanoid robots in multi-contact scenarios. *The International Journal of Robotics Research*, Vol. 35(12). P.1522–1543, 2016.
7. Kuindersma S., Deits R., Fallon M., Valenzuela A., Dai H., Permenter F., Koolen T., Marion P., Tedrake R. Optimization-based locomotion planning, estimation, and control design for the atlas humanoid robot. *Autonomous Robots*. 2016. Vol. 40(3). P:429–455.
8. Kudruss M., Naveau M., Stasse O., Mansard N., Kirches C., Soueres P., Mombaur K. Optimal control for whole-body motion generation using center-of-mass dynamics for predefined multi-contact configurations. *Humanoid Robots (Humanoids), IEEE-RAS 15th International Conference*. IEEE, 2015. P. 684–689.
9. Cori´c M., Deur J., Xu L., Tseng H.E., Hrovat D. Optimisation of active suspension control inputs for improved vehicle ride performance. *Vehicle system dynamics*, 2016. Vol. 54(7). P.1004–1030.
10. Sun W., Gao H., Yao B.. Adaptive robust vibration control of fullcar active suspensions with electrohydraulic actuators. *IEEE Transactions on Control Systems Technology*. Vol. 21(6). P.2417–2422, 2013.
11. Ershova N., Bondarenko I., Shibko O., Velmagina N.. Development of the procedure for verifying the feasibility of designing an active suspension system for transport carriages. *Eastern-European Journal of Enterprise Technologies.* 2018. № 3/7 (93). P. 53-63.
12. Єршова Н. М., Калашников К. О. Свідоцтво про реєстрацію авторського права на твір наукового характеру «Принцип адаптивного управления движением робота с динамической подвеской», № 97129 від 08.04.2020 р.
13. Attia T. Design and Development of a Novel Reconfigurable Wheled Robot for Off-RoadApplications. Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degrel of Doctor of Philosophy in Mechanical Engineering. Blacksburg, Virginia, 2018. 142 p.
14. Ершова Н.М. Современные методы теории проектирования и управления сложными динамическими системами: монография. Днепропетровск: ПГАСА, 2016. 272 с.
15. Карташов Б. А., Шабаев Е. А., Козлов О. С., Щекотуров А. М. Среда динамического моделирования SimlnTech: практикум по моделированию технических систем автоматического регулирования. ДМК Пресс, 2017. 424 с.

Н.М. Єршова, Н.В. Ткачук, С.А. Ренгач

DEVELOPMENT OF COMPUTER PROGRAMS FOR THE OPTIMAL DESIGN OF THE PASSIVE SUSPENSION OF A WHEELED ROBOT

N.M. Yershova, N.V. Tkachuk, S.A. Renhach

Prydniprovska State Academy of Civil Engineering and Architecture
24 a, St. Architect Oleg Petrov, Dnipro, 49005, Ukraine. E-mail:
nersova107@gmail.com, tkachuk.nadya7@gmail.com, docker1280@gmail.com

Currently, mobile robots are widely used. Based on the analysis of a large number of possible suspension options, they came to the conclusion about the feasibility of using wheeled robots with dynamic suspension. For the design of systems under the influence of random disturbances, a method of stochastic dynamic programming has been developed. The method is very effective, as it allows determining the law of optimal control of robot suspension parameters. According to this suspension design method, the work should be performed in stages: a passive suspension is designed based on the method of dynamic programming for continuous deterministic systems and the analysis of the robot's dynamic indicators; the expediency of introducing active control parameters of the Kalman-Busy algorithm into the passive suspension is checked and a final decision is made about the suspension's operating principle; the law of optimal management of robot suspension parameters is determined, if a decision is made to design an active suspension. In this case, less external energy is needed to suppress harmful vibrations. In the existing scientific and technical literature, mathematical models of robots with active suspension are considered, and it is not clear on what basis the suspension parameters of the original system were adopted. This work considers the first stage of designing the suspension of a wheeled robot. There are no computer programs for optimal design of the passive suspension of a wheeled robot. Therefore, the purpose of this work is to develop computer programs for the optimal design of the passive suspension of a wheeled robot. R. Bellman's matrix method of dynamic programming for continuous dynamic systems, which is a scientific novelty of the work, is the basis of the design solution search algorithm. The created computer program "ROBOT" is designed to select optimal suspension parameters, provided that the maximum values of acceleration and the coefficient of vertical dynamics of the center of mass of the robot do not exceed the permissible values. The evaluation of the dynamic properties of the robot is performed using the computer program "DINAM", the results of which are issued in the form of graphs and calculation tables. The parameters of the TIGER wheeled robot are taken as an analog model. On the basis of computer modeling, variant modeling and optimization, the passive suspension parameters of the robot were obtained, which provide it with the best dynamic properties compared to the analog model. This work is an important contribution to the field of computer science and software engineering.

Keywords: optimal design of wheeled robot suspension, computer program, matrix method of dynamic programming

DOI 10.15276/imms.v12.no4.327
УДК 681.514:621.029

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 327-336

МЕТОД ВИЯВЛЕННЯ РОЗМИТТЯ ТА МУЛЬТИПЛІКАТИВНОГО ШУМУ ЯК ОБРОБКИ ЦИФРОВОГО ЗОБРАЖЕННЯ НА ОСНОВІ АНАЛІЗУ КОЕФІЦІЄНТІВ ДИСКРЕТНОГО КОСИНУСНОГО ПЕРЕТВОРЕННЯ

Є.О. Корольова, Б.О. Розумяк, В.В. Зоріло, О.Ю. Лебедєва, Д.А. Маєвський

Національний університет «Одеська політехніка», пр.Шевченка, 1, Одеса, 65044
vikazorilo@gmail.com

Очевидним є той факт, що без інформаційної безпеки не може бути ніякої безпеки. Перебування в тому чи іншому інформаційному просторі змінює свідомість, змушує людей чинити злочини різного ступеня тяжкості, або виправдовувати ті підтримувати злочинців. Важливим інструментом в порушенні інформаційної безпеки є пропаганда, яка не гребує ніякими методами для досягнення мети, в тому числі використовує підробці фото-, відео- та аудіо-файли для здійснення інформаційно-психологічного впливу на суспільство. Методи боротьби з подібними випадками носять всебічний, комплексний характер. Поміж них важливу роль посідають методи встановлення істинності цифрового контенту або доведення того, що та чи інша інформація є фейком. При підробці цифрових файлів, зокрема, цифрових зображень, часто використовують методи пост обробки, такі як розмиття, мультиплікативний шум тощо. Виявлення даних видів пост обробки наразі не є вирішеною задачею, хоч і має певні успіхи. Метою цієї роботи є підвищення ефективності виявлення обробки цифрового зображення шляхом розробки методу виявлення розмиття та методу виявлення мультиплікативного шуму. Проведено обчислювальний експеримент, в результаті якого встановлено, що зручним та ефективним інструментом для виявлення розмиття за Гаусом та мультиплікативного шуму є дискретне косинусне перетворення. Розмиття зменшує значення високочастотних коефіцієнтів дискретного косинусного перетворення, в той час коли мультиплікативний шум їх збільшує. При цьому необроблене зображення має високочастотні коефіцієнти в межах певної норми. В роботі вдалося виділити порогові значення, що дозволяють зафіксувати відхилення від норми в той чи інший бік, на основі чого розроблено метод виявлення розмиття та метод виявлення мультиплікативного шуму, засновані на аналізі високочастотних коефіцієнтів дискретного косинусного перетворення

Ключові слова: інформаційна безпека, цифрова криміналістика, цифрові зображення, розмиття, мультиплікативний шум.

Сучасний світ – це світ технологій, в тому числі інформаційних і цифрових. Захищеність інформаційного простору суспільства є показником безпеки не лише психологічної, соціальної, а й фізичної. Захист інформації від порушень цілісності посідає важливе місце в комплексних системах захисту інформації та в цифровій криміналістиці. Важливо розуміти, що розвиток технологій, засобами яких можна здійснити порушення цілісності інформації, сягнув такого рівня, коли й непрофесійні користувачі здатні виконати те чи інше спотворення інформаційного файлу.

Важливе місце для сучасної людини виконує цифрова фотографія. Ми передаємо через цифрові фото безліч важливої інформації: тексти, зображення сцен, предметів, людей, подій. Часто зображення використовують для документування чогось важливого: речових доказів, порушень правил дорожнього руху, важливих історичних подій та осіб тощо. В таких випадках фальсифікація цифрових зображень може мати небажані наслідки для системи інформаційної безпеки. Найпоширеніші види підробки та обробки цифрових

зображень: накладання шуму (Гаусів, мультиплікативний, сіль-перець); розмиття; підвищення різкості; клонування; фотомонтаж.

Методи виявлення підробок цифрових зображень можна розділити на дві групи: активні і пасивні методи виявлення порушень цілісності цифрових зображень.

Для активної ідентифікації підробок потрібні попередньо витягнуті або вбудовані дані. Умовно методи, які використовуються для активного захисту медіа, можна поділити на два види: методи на основі цифрового підпису і методи на основі цифрових водяних знаків [1-2].Активні методи мають спільний основний недолік – вони мають бути застосовані до того, як зображення буде піддано порушенням цілісності різного роду. Щоб подолати цей недолік, існує багато методів пасивного захисту.

Пасивні методи використовують безпосередньо зображення для перевірки його цілісності [3]. Фальсифікація може порушити базову статистичну властивість через неузгодженість шуму, розмиття зображення, підвищення різкості зображення [4], копіювання-переміщення (клонування) або фотомонтаж [5] та домальовування об’єктів на зображенні засобами графічних редакторів [6].

Методи, направлені на виявлення цих підробок, часто використовують особливості якоїсь конкретної обробки і не є універсальними.

Метою цієї роботи є підвищення ефективності виявлення обробки цифрового зображення шляхом розробки методу виявлення розмиття та методу виявлення мультиплікативного шуму.

На сьогодні один із найефективніших методів виявлення розмиття цифрового зображення заснований на аналізі швидкості росту сингулярних чисел блоків матриці цифрового зображення[7]. Один із кроків метода – додаткова перевірка експертним розмиттям, що потребує виконання експертом попередніх маніпуляцій над зображенням. Це вносить певні незручності в процес експертизи, тому актуальним лишається пошук нових інструментів для вирішення даної задачі.

В даній статті розглянуто вплив розмиття на власні значення і на коефіцієнти дискретного косинусного перетворення блоків матриці цифрового зображення. Цифрові зображення, як і будь-який сигнал, можна розкласти на високі, середні та низькі частоти. Низьким частотам головним чином відповідають фонові частини зображення, а високим частотам – контури. Оскільки розмиття впливає на високі частоти, а саме зменшує їх, було прийняте рішення аналізувати тільки ті власні значення, що відповідають високим частотам, оскільки в ранніх публікаціях було встановлено відповідність найменших власних значень саме високим частотам.

Для експерименту сформуємо базу з 100 зображень формату без втрат (tif) та 100 зображень у форматі з втратами (jpg). Розмиття будемо виконувати у графічному редакторі AdobePhotoshop засобами фільтру «Розмиття за Гаусом» з радіусом 1 піксель. Після обробки всі зображення збережено у форматі без втрат. При такому значені радіуса розмиття візуально не помітно, що можна побачити на рисунку 1. Побудувавши графік середніх значень перших 5 власних значень по всім блокам ЦЗ для 100 оригінальних і 100 розмитих зображень у форматі JPG, було виявлено, що саме перші 4 власні значення зазнали значного зменшення при розмитті, що підтверджує цю відповідність (рисунок 2).

Рис. 1. Приклад розмитого зображення з радіусом розмиття 1 піксель

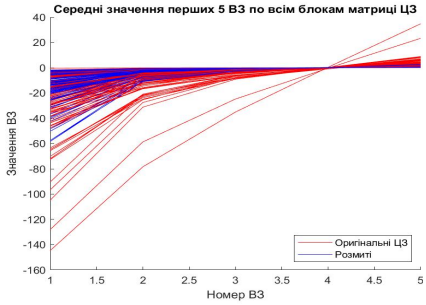


Рис. 2. Графік середніх значень п’яти найменших власних значень по всім блокам ЦЗ

Експеримент показав, що, використовуючи власні значення, знайти поріг для відокремлення оброблених зображень від необроблених вкрай важко або неможливо так, щоб цей метод був ефективніший за його аналог на основі використання сингулярного розкладання. Використання власних значень дало велику кількість помилок другого роду.

Проаналізуємо вплив розмиття на високочастотні коефіцієнти ДКП, які знаходяться в правому нижньому куті квадратної матриці. Матрицю зображення розіб’ємо стандартним чином на блоки 8×8. Для кожного блоку застосуємо ДКП та виділимо коефіцієнти, що відповідають високим частотам (9 коефіцієнтів в нижньому правому куті). Знайдемо середнє арифметичне значення для відповідних коефіцієнтів по всім блокам цифрового зображення, в результаті чого отримаємо 9 значень, які поставимо у відповідність цифровому зображенню. На рисунках 2 та 4 можемо бачити отримані результати для зображень без втрат і з втратами відповідно. Кожний графік відповідає одному зображенню та є інтерполяційним сплайном для 9 усереднених високочастотних коефіцієнтів ДКП.

Згідно результатів, отриманих при аналізі коефіцієнтів ДКП до та після розмиття (червоний та синій колір відповідно), можна зробити висновок, що зменшення високочастотних коефіцієнтів ДКП при розмитті ЦЗ у форматі з втратами, буде менше, ніж у випадку аналізу ЦЗ, збереженого без втрат. Це відбувається за рахунок того, що, при збережені ЦЗ у форматі із втратами абсолютні значення високочастотних коефіцієнтів ДКП зменшуються як результат стиснення. Тому пошук порогового значення для зображень з втратами може бути складнішим, ніж для зображень без втрат, але можливим.

Отримавши графіки середніх значень високочастотних коефіцієнтів ДКП, було вирішено обрати амплітуду графіка, як основну характеристику ЦЗ при розробці методу виявлення розмиття.

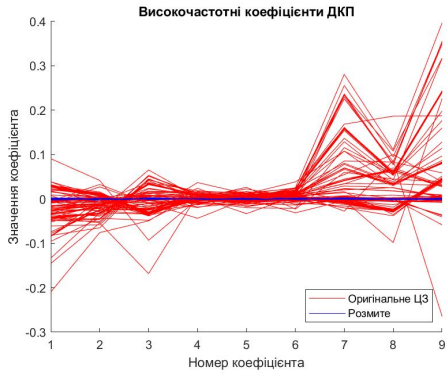


Рис. 3. Графіки середніх значень ДКП по всім блокам матриці ЦЗ для зображень у форматі TIFF

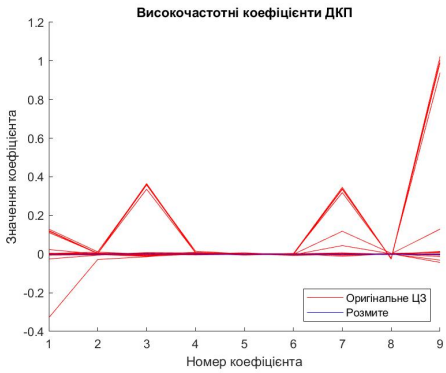


Рис.4. Графік середніх значень ДКП по всім блокам матриці ЦЗ для зображень у форматі JPG

Було проведено аналіз амплітуд графіків середніх значень високочастотних ДКП по всім блокам матриці ЦЗ. За допомогою цього аналізу було встановлено, що, використовуючи високочастотні коефіцієнти ДКП можна виявити розмиття ЦЗ. Але, якщо в зображенні була значна доля фонової частини, ефективність такого метода зменшується за рахунок помилок другого роду.

Для покращення результатів було проаналізовано середню амплітуду графіків високочастотних коефіцієнтів ДКП по всім блокам, а також блок з максимальною амплітудою графіку високочастотних коефіцієнтів ДКП. Останнє дало кращий результат та дозволило ефективно виявляти розмиття незалежно від долі фонової частини зображення.

Всі експерименти проведено над зображеннями у колірній моделі RGB, в контексті вирішуваної задачі немає різниці, яку колірну компоненту аналізувати, адже для матриць R, G, B результати обробки будуть якісно однакові. Також незначно вплине на результат переведення зображення в полу тонове, що і було виконано.

Знайдемо порогове значення для виявлення розмитих зображень. При використанні амплітуди середніх значень високочастотних коефіцієнтів ДКП по всім блокам було побудовано гістограми для 100 зображень у форматі TIFF до та після розмиття (рисунок 5 і 6 відповідно).

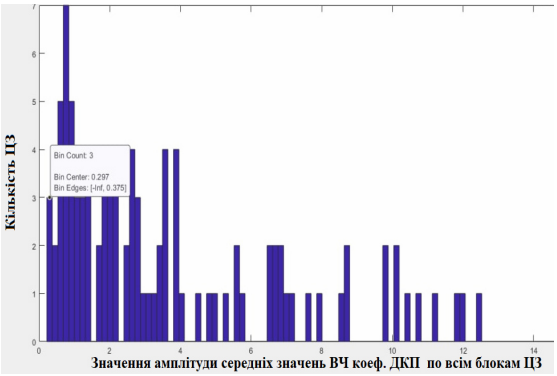


Рис. 5. Гістограма амплітуд середніх значень високочастотних коефіцієнтів ДКП по всім блокам оригінальних ЦЗ у форматі TIFF

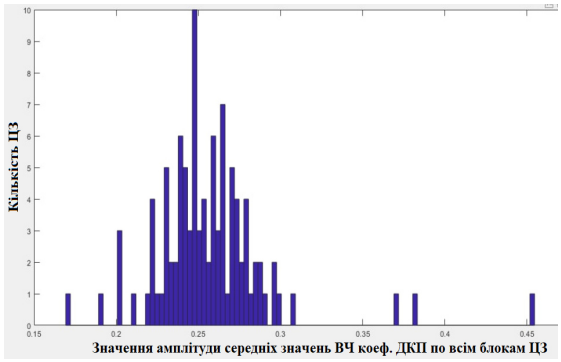


Рис. 6. Гістограма амплітуд середніх значень високочастотних коефіцієнтів ДКП по всім блокам розмитих ЦЗ у форматі TIFF

Виходячи з результатів, отриманих при побудові гістограм, можна бачити, що є деяке розмежування значень амплітуд. Експериментально встановлено, що використання в якості порогового значення «0.3» дає змогу у більшості випадків встановити, чи є зображення обробленим. Помилки першого роду склали 4%.

Але, як вже було зазначено, коли в зображені значна доля фонової частини або якщо воно містить багато неосвітлених чи, навпаки, засвічених зон, висока ймовірність виникнення хибної тривоги (рисунок 7, 8). Враховуючи значення порогу, метод визначив дане зображення як розмите через його недоліки.

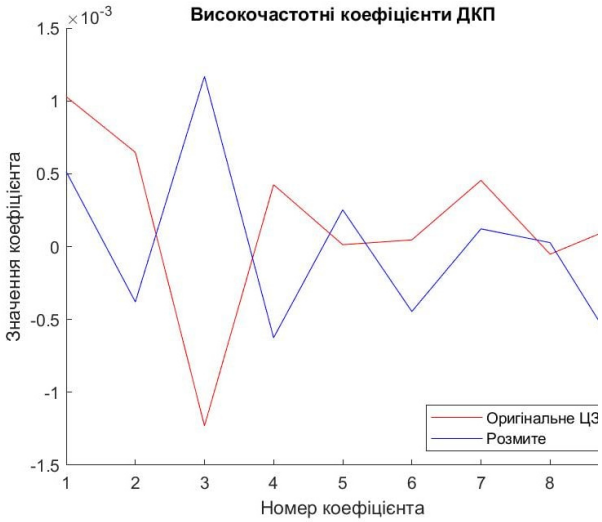


Рис. 7. Графік середніх значень високочастотних коефіцієнтів ДКП

Рис. 8. Приклад ЦЗ зі значною долею фонової частини

Знайдемо максимальну амплітуду для кожного оригінального та розмитого зображення та побудуємо гістограми (рисунок 9-12).

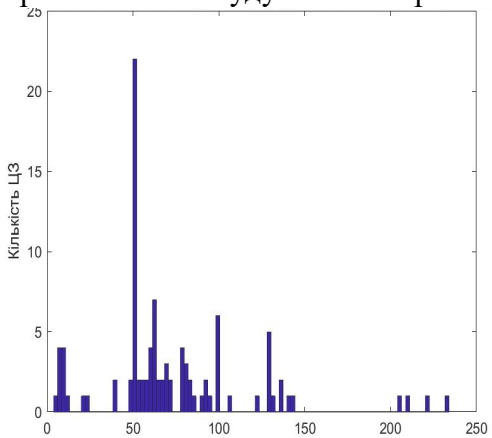


Рис. 9. Гістограма значень блоку з максимальною амплітудою високо частотних коефіцієнтів ДКП для оригінального ЦЗ в форматі JPG

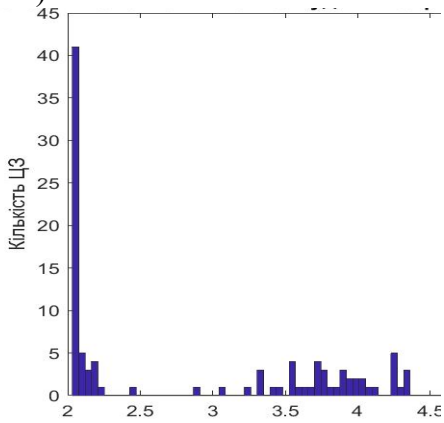


Рис.10. Гістограма значень блоку з максимальною амплітудою високо частотних коефіцієнтів ДКП для розмитого ЦЗ в форматі JPG

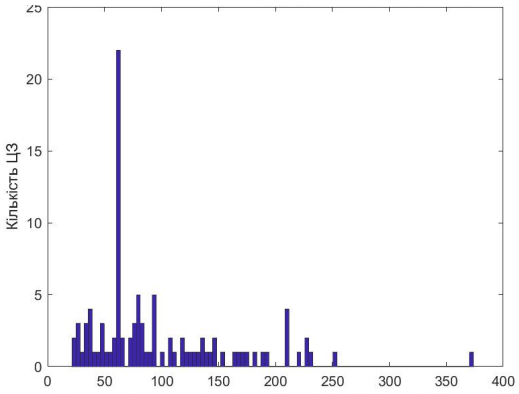


Рис. 11. Гістограма значень максимальних амплітуд високочастотних коефіцієнтів ДКП по всім блокам оригінальних ЦЗ в форматі TIF

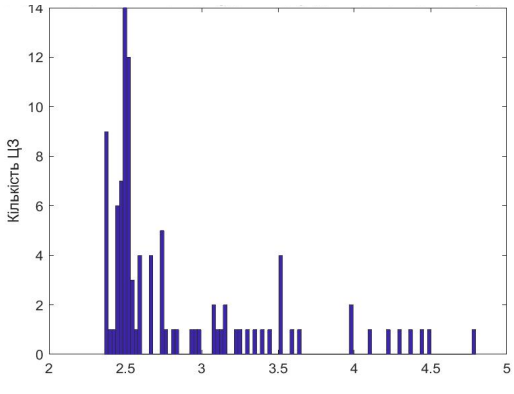


Рис.12. Гістограма значень блоку з максимальною амплітудою високо частотних коефіцієнтів ДКП для розмитого ЦЗ в форматі TIF

Експериментально встановлено, що при пороговому значенні «5» вдається досягти високої ефективності метода. Як і очікувалось, зображення форматів TIFF та JPG дали різни результати, але це незначно вплинуло на кінцевий результат. Для обох форматів запропонований поріг дав найкращий результат.

Нехай A – m×n-матриця цифрового зображення. Алгоритм розробленого метода виявлення розмиття ЦЗ складається з наступних кроків.

Крок 1. Розбити матрицю A стандартним чином на блоки 8×8.

Крок 2. Знайти для кожного блоку коефіцієнти ДКП.

Крок 3. Виділити високочастотні коефіцієнти, які знаходяться під побічною діагоналлю.

Крок 4. Знайти амплітуду графіка високочастотних коефіцієнтів ДКП кожного блоку.

Крок 5. Знайти блок з максимальним значенням амплітуди серед всіх блоків ЦЗ.

Крок 8. Порівняти значення максимальної амплітуди із пороговим значенням «5»:

Якщо амплітуда нижче порога – ЦЗ розмите, *інакше* – не розмите.

Згідно отриманих даних було виявлено, що аналізуючи власні значення матриці ЦЗ важко виявити розмиття зображення, не маючи оригінальне ЦЗ, або без проведення експертного розмиття. Також, на основі результатів, був розроблений метод виявлення розмиття цифрових зображень на основі аналізу високочастотних коефіцієнтів ДКП.

Розроблений метод дозволяє з високою ймовірністю вірно виявити розмиття ЦЗ. Для порогового значення «5» кількість помилок першого та другого роду для ЦЗ формату JPG – 1%. Для зображень формату TIFF помилок першого та другого роду не було виявлено.

Протилежний ефект на високочастотні коефіцієнти викликає накладання на зображення мультиплікативного шуму. Даний вид обробки також часто застосовується до зображень під час приховання слідів їх підробки. Високочастотні коефіцієнти реагують на мультиплікативний шум збільшенням. Даний ефект також можна спостерігати і при аналізі власних і сингулярних чисел блоків матриці цифрового зображення, однак високі значення найменших сингулярних або власних значень може свідчити як про наявність мультиплікативного шуму, так і про високу якість цифрового зображення та велику кількість чітких контурів, тому кількість помилок 2 роду або хибних тривог може складати більше 30%. Розглянемо, як будуть змінюватись усереднені значення високочастотних коефіцієнтів під впливом мультиплікативного шуму (рис. 13)

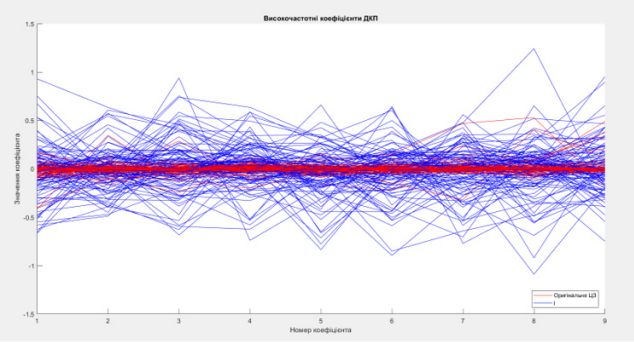


Рис.13. Середні значення високочастотних коефіцієнтів ДКП оригінальних (червоні лінії) та збурених зображень (сині лінії)

Експеримент показав, що при застосуванні мультиплікативного шуму аналізовані коефіцієнти ДКП змінюються настільки, що стає можливим відділити оброблені зображення від необроблених. Знайдемо порогове значення для відокремлення оброблених зображень від необроблених на основі отриманих під час експерименту результатів.

На рисунку 14 зображено гістограму амплітуд середніх значень високих частот блоку 8×8 для оригінальних зображень, на рисунку 15 – зображено гістограму амплітуд середніх значень високих частот для блоку 8×8 зашумлених зображень з параметром дисперсії 0,01.

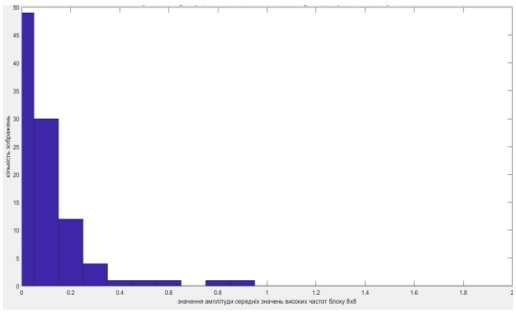


Рис.14. Гістограма амплітуд середніх значень високих частот для блоку 8×8 оригінальних зображень

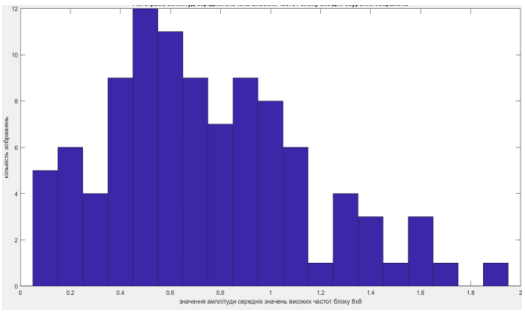


Рис. 15. Гістограма амплітуд середніх значень високих частот для блоку 8×8 зашумлених зображень

Далі необхідно знайти середні значення амплітуд високих частот в кожному блоці 8×8 для оригінальних та збурених зображень, які показані на рисунку 16 та рисунку 17.

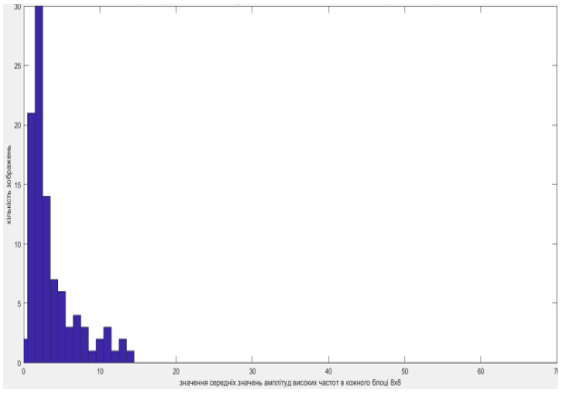


Рис. 16. Гістограма середніх значень амплітуд високих частот в кожному блоці 8×8 для оригінальних зображень

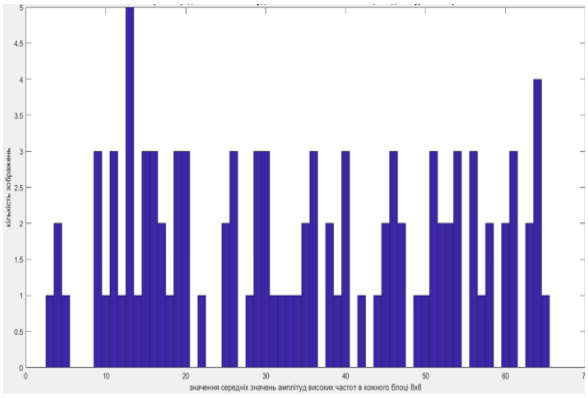


Рис. 17. Гістограма середніх значень амплітуд високих частот в кожному блоці 8×8 для зашумлених зображень

Такі ж характеристики було отримано для зображень, оброблених мультиплікативним шумом з дисперсією 0.001 та 0.005.

Візуально ми бачимо, що для оброблених зображень коефіцієнти ДКП значно більші, ніж для необроблених, тому задача полягає в тому, щоб визначити порогове значення для виявлення мультиплікативного шуму. Для цього потрібно отримати усереднені показники, які можна поставити у відповідність зображенню і використати в якості порогового значення.

Розглянемо коефіцієнти ДКП, вони можуть бути додатні і від'ємні. Знайдемо середні арифметичні значення для високочастотних коефіцієнтів ДКП і знайдемо модуль різниці між максимальним та мінімальним значенням – показник 1. Розглянемо також як альтернативу показник 2: спочатку знайдемо

модуль різниці максимального та мінімального значення високочастотних коефіцієнтів в блоках матриці цифрового зображення, після чого знайдемо середнє арифметичне отриманих різниць. Будемо використовувати дані показники в якості порогового значення. Експериментально встановлено, що певні значення даних показників дійсно дозволяють розрізняти оброблені зображення від необроблених. В таблиці 1 показано результати їх використання.

Таблиця 1

Результати розробленого методу

Збурення (величина дисперсії)	ACC		Помилки I роду		Помилки II роду		Порогове значення	
	Показник №1	Показник №2	Показник №1	Показник №2	Показник №1	Показник №2	Показник №1	Показник №2
0.001	0.89	0.855	5%	14%	17%	15%	0.1457	6.3424
0.005	0.88	0.935	7%	4%	17%	9%	0.1457	9.1088
0.01	0.955	0.965	4%	4%	5%	3%	0.2378	11.8751

Як можемо бачити, дані показники дають різні результати за різних параметрах мультиплікативного шуму. Однак за характеристикою ACC Показник 1 кращий.

Характеристика ACC (Accuracyof CC) – кількісний показник ефективності. Він визначає точність виявлення порушення цілісності. Формула розрахунку ACC представлена нижче:

$$ACC = (TP + TN)/(TP + FN + TN + FP),$$

де TP (TruePositive) – число правильно виявлених ЦЗ, цілісність яких була порушена (істино позитивний результат);

TN (TrueNegative) – число правильно виявлених оригінальних ЦЗ (істинонегативний результат);

FP (FalsePositive) – число оригінальних ЦЗ, помилково прийнятих за такі, цілісність яких була порушена (хибно позитивний результат (хибна тривога) або помилка II роду);

FN (FalseNegative) – число ЦЗ, цілісність яких була порушена, помилково визнаних оригінальними (хибно негативний результат або помилка I роду).

Ефективність розробленого методу по виявленню мультиплікативного шуму була перевірена по даному показнику та показана у таблиці 1.

Крім того на даному етапі досліджень, коли експеримент було проведено для мультиплікативного шуму х дисперсією 0,001, 0,005 та 0,01 і для кожної дисперсії отримано різні значення Показника 1 (далі П1), рекомендується включити в алгоритм всі три випадки. Розробка універсального порогового значення є предметом подальших досліджень.

На основі отриманих результатів розроблено метод виявлення мультиплікативного шуму, основні кроки алгоритму якого наступні.

Нехай F – m×n матриця яскравості пікселів цифрового зображення.

Крок 1. Розбити матрицю F на непересічні блоки 8×8.

Крок 2. Знайти для кожного блоку коефіцієнти ДКП.

Крок 3. Виділити з кожного блоку по 9 коефіцієнтів ДКП, що відповідають високим частотам.

Крок 4. Знайти середні арифметичні значення відповідних коефіцієнтів ДКП по всім блокам цифрового зображення.

Крок 5. Знаходимо П1 як різницю максимального та мінімального середніх значень високочастотних коефіцієнтів:

якщо П1>0,1457*, то* вважати цифрове зображення обробленим;

інакше вважати цифрове зображення необробленим.

На основі отриманих результатів розроблено метод виявлення мультиплікативного шуму як порушення цілісності цифрового зображення. Ефективність даного методу в термінах помилок 1 і 2 роду складає 7% і 17% відповідно. Метод має потенціал для подальшого вдосконалення.

Список літератури

1. Tymoshenko D. How the photo fakes work at war. Radio Svoboda. July 05, 2018. URL: https://www.radiosvoboda.org/a/donbass-realii/29341812.html
2. Rodyhin K., Iermakova I. The kinds of media photo content manipulations in the context of information and semantic warfare. *Visnyk Lvivskoho universytetu. Seriia Zhurnalistyka.* 2020. No. 47. P. 200–214.
3. Guglielmi G. Peer-reviewed homeopathy study sparks uproar in Italy. *Nature*. 2018. No. 562. P.173-174.
4. Brainard J, You J. What a massive database of retracted papers reveals about science publishing's 'death penalty'. *Science*. 2018. No.10. P.189-194.
5. Kumar B.S., Karthi S., Karthika K., Cristin R. A Systematic Study of Image Forgery Detection. *Journal of computational and theoretical Nanoscience*. 2018. Vol.15(8). P.2560–2564.
6. Кобозева А.А. Основы общего подхода к решению проблемы обнаружения фальсификации цифрового сигнала. *Електромашинобудування та електрообладнания*. 2009. Вип. 72. С. 35-41.
7. Зоріло В.В., Головко Ю.О., Якименко І.З., Гураль І.В. Модифікований метод виявлення розмиття цифрового зображення. *Сучасні комп'ютерні інформаційні технології: матеріали Всеукраїнської конференції з міжнародною участю АСІТ '2017*. Тернопіль: ТНЕУ, 2017. С. 203-204.

Є.О. Корольова, Б.О. Розумяк, В.В. Зоріло, О.Ю. Лебедєва, Д.А. Маєвський

METHOD OF DETECTION OF BLUR AND MULTIPLICATIVE NOISE AS DIGITAL IMAGE PROCESSING BASED ON ANALYSIS OF DISCRETE COSINE TRANSFORMATION COEFFICIENTS

V.V. Zorilo, Ye.O. Koroliova, B.O. Rozumiak, D.A. Mayevsky

National Odesa Polytechnic University, Shevchenko str, 1, Odesa, 65044
vikazorilo@gmail.com

It is obvious that there can be no security without information security. Staying in one or another in formation space changes consciousness, forces people to commit crimes of varying degrees of severity, or to justify supporting criminals. An important tool in the violation of information security is propaganda, which does not disdain any methods to achieve the goal, including using fake photo, video, and audio files to realize informational and psychological influence on society. The methods of dealing with such cases are comprehensive and complex in nature. Among them, the methods of establishing the truth of digital content or proving that this or that in formation is fake play an important role. When forging digital files, in particular, digital images, post-processing methods such as blurring, multiplicative noise, etc. are often used. Identifying these types of post-processing is currently not a solved task, although it has certain successes. The purpose of this work is to improve the detection efficiency of digital image processing by developing a blur detection method and a multiplicative noise detection method. A computational experiment was conducted, as a result of which it was established that the discrete cosine transformation is a convenient and effective tool for detecting Gaussian blur and multiplicative noise. Blur ring reduces the value of high-frequency DCP coefficients, while multiplicative noise increases them. At the same time, the raw image has high-frequency coefficients within a certain norm. In the work, it was possible to single out threshold values that allow recording deviations from the norm in one direction or another, on the basis of which a method for detecting blur ring and a method for detecting multiplicative noise, based on the analysis of high-frequency coefficients of DCP, was developed.
Keywords: information security, digital image forensics, blurring, multiplicative noise.

DOI 10.15276/imms.v12.no4.337
УДК 004.056.53

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 337-345

ВИЯВЛЕННЯ ТА ВИПРАВЛЕННЯ ПОМИЛОК У ЗАХИЩЕНИХ СИСТЕМАХ ЗБЕРІГАННЯ ДАНИХ НА ОСНОВІ ОБЧИСЛЕННЯ ПРОЕКЦІЙ ЧИСЛА

С.В. Кулина

Західноукраїнський національний університет, вул. Львівська, 11,
м. Тернопіль, 46020, Україна; e-mail: sersks@wunu.edu.ua

Система розподіленого зберігання даних включає в себе окрім самих пристроїв на яких буде зберігатися інформація, також і канали зв'язку по яких її пересилають. При обміні інформацією на значні відстані виникає ряд нових проблем, які потребують вирішення. Аналіз сучасного стану та розвитку систем передачі та зберігання даних свідчить про те, що є необхідність застосування різноманітних технічних та програмних методів для захисту від спотворення, втрати чи крадіжки інформації. Використання лише одного з них не є ефективним, тому зазвичай використовують їх поєднання. Одним із поширених методів захисту даних від спотворень є використання коригуючих кодів, серед яких важливе місце займають коди на основі системи залишкових класів (СЗК). Завдяки представленню інформації у вигляді залишків від ділення даних на кожен елемент системи взаємно простих модулів досягається зменшення розрядності чисел з якими проводяться операції. У роботі проаналізовано подання інформації у вигляді залишків для реалізації методу розподіленого зберігання даних, завдяки чому можна спростити реалізацію систем збору та передачі інформації. Запропоновано використати для захисту даних розширення початкового діапазону обчислення за рахунок введення надлишкових модулів, завдяки чому ймовірність виявити помилку становить 100%. Проаналізовано переваги та недоліки виявлення та виправлення помилок з використанням коригуючих кодів у СЗК за допомогою методів обчислення синдрому та за допомогою обчислення проекції числа. Кількість інформаційних модулів у проектованих системах може змінюватися у залежності від поставленої задачі, тому у проведених дослідження було визначено, що при збільшенні числа інформаційних модулів та комбінацій спотворених символів кількість невиявлених помилок у відсотковому поданні практично залишається без змін. Запропоновано при проектуванні систем збору інформації використовувати СЗК без додаткового виправлення помилок, оскільки є можливість підібрати набір модулів таким чином, що максимальна кількість помилок що виникатиме не впливатиме на роботу системи.

Ключові слова: Коригуючі коди, система залишкових класів, китайська теорема про залишки, розподілене зберігання даних, методи зворотного перетворення, виправлення помилок.

Вступ

Основою будь-якої системи розподіленого зберігання даних окрім пристроїв на яких буде зберігатися інформація є також канали зв'язку по яких вона пересилається. При цьому варто враховувати, що передача інформаційних пакетів між пристроями однієї мережі є відносно захищеною та менше піддається впливу завад. Проте при обміну інформацією за межами цієї мережі значно зростає ймовірність спотворення, втрати чи крадіжки інформації. В залежності від каналу обміну інформацією використовуються різноманітні технічні рішення, завдяки яким зменшується ймовірність втрати чи спотворення повідомлення.

Проте використання лише технічних рішень не є ефективним, тому поряд з цими методами захисту інформації також застосовують різноманітні алгоритмічні рішення – коригуючі коди, методи криптографічного захисту та багато іншого [1-5]. На даний час дослідження захисту від спотворень, втрати чи несанкціонованого

доступу до інформації при передачі між користувачами та в межах однієї системи є важливою галуззю наукових досліджень [6-7].

Одним із таких методів захисту даних є використання коригуючих кодів на основі системи залишкових класів (СЗК). СЗК дозволяє окрім паралельної обробки даних також і захистити інформацію від спотворення, втрати чи крадіжки шляхом введення додаткових перевірочних модулів, що є значно ефективнішим ніж використання резервного копіювання чи дзеркального збереження масивів даних [8].

Подання чисел у вигляді залишків також дає змогу спростити реалізацію систем збору та передачі інформації, а також дозволяє вирішувати клас задач, що складніше реалізуються у позиційних системах числення [9-10].

Значний теоретичний внесок у розвиток системи залишкових класів та її застосування в системах збору та обробки інформації зробили такі вчені: І. Я. Акушський, Д. І. Юдицький, В. М. Амербаєв, В. А. Торгашев, А. О. Коляда, В. А. Краснобаєв, Я. М. Николайчук, М. І. Червяков, О. А. Фінько, А. Омонді (A. Omondi), Б. Премкумар (B. Premkumar), Дж. Кардаріллі (G. Cardarilli).

Дослідження, що сприяли розвитку мережного кодування даних належать іншим відомим науковцям, серед яких: С. Г.Бунін, В. О. Романов, В. А. Романюк, I. Акиілдіз (I. Akyildiz), Р. Ахлсведе (R. Ahlswede), К. Фрагоулі (C. Fragouli) та ін.

Мета і задачі дослідження

Метою роботи є підвищення надійності систем зберігання даних шляхом використання коригуючих кодів системи залишкових класів.

Для досягнення поставленої мети необхідно вирішити наступні задачі: - розглянути спосіб прямого перетворення СЗК на основі китайської теореми про залишки; - проаналізувати існуючі шляхи виправлення помилок в СЗК за допомогою методу обчислення синдрому та методу проекцій; - провести експериментальні дослідження виявлення помилок у інформаційному повідомленні при одному перевірочному модулі різної розрядності; - провести експериментальні дослідження виявлення помилок у інформаційних повідомленнях при зміні кількості інформаційних модулів.

Надлишкова система залишкових класів

В СЗК інформацію представляють у вигляді залишків $(x_1, x_2, ..., x_i, ...x_k)$ від ділення даних на кожен елемент системи взаємно простих модулів $(p_1, p_2, ..., p_i, ...p_k)$, де k кількість модулів системи [11].

Числове значення даних X, подане в позиційній системі числення, повинно лежати в діапазоні $0 < X < P$, де

$$P = \prod_{i=1}^{k} p_i \,. \tag{1}$$

Пошук залишків відбувається за формулою:

$$x_i = X(\mathrm{mod}\, p_i). \tag{2}$$

Згідно виразу (2) всі числа з діапазону значень P можна представити у вигляді матриці:

$$B=\begin{vmatrix} p_1 & p_2 & ... & p_k \\ 0 & 0 & ... & 0 \\ 1 & 1 & ... & 1 \\ ... & ... & ... & ... \\ p_1-1 & p_1-1 & ... & p_1-1 \\ 0 & p_1 & ... & p_1 \\ ... & ... & ... & ... \\ x_{i,1} & x_{i,2} & ... & x_{i,k} \\ ... & ... & ... & ... \\ p_1-1 & p_2-1 & ... & p_k-1 \end{vmatrix}$$

Розглянемо систему з 3 модулів: $p_i=[3,5,7]$. Загальний діапазон даної системи модулів становитиме: $P=3\cdot5\cdot7=105$, а будь-яке значення X можна знайти згідно формули обчислення залишків (2). Для прикладу, нехай X=10, тоді представлення цього числа у СЗК матиме вигляд (1,0,3).

СЗК дозволяє захистити інформацію при розширенні початкового діапазону обчислення за рахунок введення надлишкових модулів. Основною перевагою цього є можливість виявлення помилки в одному символі з ймовірністю 100% при використанні одного додаткового модуля [12]. При роботі з надлишковими модулями їх називають перевірочними та вводять поняття робочого та загального діапазонів.

Перевірочні символи в коригуючих кодах СЗК обчислюються за формулою [11]:

$$x_{k+a}=X\left(\mathrm{mod}\,p_{k+a}\right),$$

де k – кількість інформаційних модулів, а a кількість перевірочних модулів.

Для виявлення помилок використовують поняття загального $H=\prod_{i=1}^{k+a}p_i$ та робочого (формула 1) діапазонів.

Для відновлення позиційного представлення числа X використовується формула зворотного перетворення на основі китайської теореми про залишки [12]:

$$X=\left(\sum_{i=1}^{k+a}x_i\cdot B_i\right)\mathrm{mod}M, \qquad (3)$$

де B_i – базисні числа СЗК, які обчислюються згідно рівняння:

$$B_i=\frac{M}{p_i}\cdot t_i\equiv1(\mathrm{mod}\,p_i),$$

де t_i – набір коефіцієнтів, які забезпечують ортогональність перетворень та задовольняють умову: $0<t_i<p_i$.

Виправлення помилок методом обчислення синдрому

На даний час основними методами виявлення та виправлення помилок з використанням коригуючих кодів на основі СЗК є виправлення помилок за допомогою обчислення синдрому та за допомогою обчислення проекції числа [13].

Згідно із алгоритмом виявлення помилок за допомогою обчислення синдрому відновлення позиційного числа відбувається на основі формули 3 окремо за кожним модулем робочого та перевірочного діапазонів [14]:

$$Y=\left(\sum_{i=1}^{k}x_i\cdot B_i\right)\mathrm{mod}\,P,$$

$$E = \left(\sum_{i=k+1}^{k+a} x_i \cdot B_i \right) \bmod P_c ,$$

де $P_c = \prod_{i=k+1}^{k+a} p_i$ і задовольняється умова $P_c > P$,

Саме обчислення синдрому відбувається за формулою:

$$s = (Y - E) \bmod P_c .$$

Приклад. Розглянемо використання системи з трьома інформаційних та двома перевірочними модулями:

– система модулів: $p_i = [3,5,7,11,13]$;

– робочий діапазон даної системи модулів буде становити: P=105;

– перевірочний діапазон: $P_c = 143$.

Іншим способом реалізації системи з таким самим робочим діапазоном може бути використання набору модулів з трьома інформаційними і одним перевірочним $p_i = [3,5,7,107]$, що дозволить задовільнити умову $P_c > P$. Проте, як ми можемо бачити розрядність перевірочного модуля значно зростає, що ускладнює реалізацію відновлення.

Проте основою виявлення помилок за допомогою обчислення синдрому є створення таблиці синдрому, у якій кожному значенні помилки відповідає лише одне значення синдрому. Кількість елементів у такій таблиці буде рівне P_c та дозволяє зменшити кількість необхідних обчислень, проте залишається проблема відновлення позиційного числа при зростанні кількості символів у повідомленні [14].

Також варто враховувати те, що при роботі з багаторозрядними модулями таблиця буде збільшуватися у геометричній прогресії, що значно збільшить час та ускладнить вибірку необхідного значення синдрому.

Виправлення помилок методом обчислення проекції числа

Проаналізувавши існуючий метод обчислення проекцій числа, який базується на визначенні, якщо з числа $X = (x_1, x_2, \ldots, x_k)$, вилучити модуль p_i, то саме значення X не зміниться а також проекції цього числа за всіма модулями будуть рівні.

Для виявлення помилки необхідно обчислити позиційне представлення числа X, при цьому, якщо отримане число у діапазоні $[0, \mathrm{P_k})$, то помилки немає, або пошкоджень зазнали залишки за двома і більше модулями.

У випадку якщо значення $X > P$, то сталася помилка у одному з залишків. Для її виправлення необхідно обчислити проекцій числа X за всіма модулями p_i . Якщо значення проекції $X_i < P$, то помилка відбулася за модулем p_i .

У загальному випадку при обчисленні проекцій X_i необхідно розраховувати базисні числа B_i для кожної з проекцій, тобто при використанні системи з шести модулів необхідно розрахувати та зберегти крім базового набору для всіх шести модулів ще і додатково шість наборів по шість модулів для систем проекцій.

У запропонованому спрощеному методі для знаходження проекцій числа X_i використовуються базисні числа B_i, які були розраховані для числа X, а формула зворотного перетворення матиме вигляд:

$$X_j = \left(\sum_{i=1}^{n} x_i \cdot B_i \right) \bmod \frac{P_n}{p_j} \qquad (4)$$

де *j* – номер проекції.

Як видно з формули (4) для обчислення проекцій використовуються ті самі базисні числа, що і для відновлення початкового значення числа X. Це забезпечує спрощення алгоритму виправлення помилок на основі коригуючих кодів СЗК за рахунок зменшення в n раз кількості операцій множення при зворотному перетворенні.

Приклад. Розглянемо приклад використання системи з чотирма інформаційних та двома перевірочними модулями.

Система модулів: $p_i = [5,7,11,13,17]$.

Загальний діапазон для даної системи модулів буде становити: $P_n = 765765$.

Робочий діапазон: $P_k = 3465$.

Базисні числа $M_i = [306306, 656370, 680680, 556920, 412335, 450450]$.

Нехай число X=73, яке у СЗК з обраними модулями матиме вигляд (3, 3, 1, 7, 8, 5) отримало спотворення в третьому розряді, тоді спотворене повідомлення буде мати вигляд (3, 3, 0, 7, 8, 5). Обчислені згідно формули 4 значення Хпоч наведені в таблиці 1.

Таблиця 1

Обчислення значень Хпоч згідно номера проекції

Хпоч	p_1	p_2	p_3	p_4	p_5	p_6	j
85158		3	0	7	8	5	1
85158	3		0	7	8	5	2
73	3	3		7	8	5	3
15543	3	3	0		8	5	4
26253	3	3	0	7		5	5
40113	3	3	0	7	8		6

Як видно з таблиці 1 при обчисленні значення Хпоч, у всіх випадках окрім вірного значення Хпоч $> P_k$. Це наочно показує, що при втраті або пошкодженні значення за одним із модулів, ми можемо відновити втрату маючи два перевірочних символи. На основі проведених обрахунків запропонований метод обчислення проекцій забезпечує спрощення обчислень за рахунок зменшення в шість разів кількості операцій множення при зворотному перетворенні. Це досягається завдяки використанню початкових базисних чисел.

Експериментальні дослідження виявлення помилок у послідовностях інформаційних символів

Умову перевірки на спотворення залишку можна використати для виявлення помилок у двох і більше символах. Для цього були проведені експериментальні дослідження, результатом яких було визначення відсотки виявлення помилок в послідовностях інформаційних символів. При проведених дослідженнях був використаний набір модулів необхідний для передачі 24 бітного інформаційного повідомлення та перевірочний символ різної розрядності, а саме: набір інформаційних модулів p_1=257, p_2=263, p_3=269 та перевірочний модуль починаючи з p_4=271. Оскільки незначне збільшення перевірочного модуля суттєво не впливало на обчислення, то кожне наступне значення перевірочного модуля було більше на один розряд.

У таблиці 2 розрахований загальний діапазон, робочий діапазон та кількість комбінацій з помилками в трьох символах для даних наборів модулів. Не залежно від значення перевірочного модуля робочий діапазон не змінювався і становив 18181979, а кількість комбінацій залишалася рівною 17975296.

Таблиця 2

Обчислення кількісного значення невиявлених помилок

Значення p_4	Загальний діапазон	Робочий діапазон	Кількість комбінацій	Кількість невиявлених помилок	Кількість невиявлених помилок, %
271	4927316309	18181979	17975296	66327	0, 36899
509	9254627311	18181979	17975296	35315	0, 19646
1021	18563800559	18181979	17975296	17606	0, 09795
2039	37073055181	18181979	17975296	8817	0, 04905
4093	74418840047	18181979	17975296	4393	0, 02444
8191	148928589989	18181979	17975296	2195	0, 01221
16273	295875344267	18181979	17975296	1105	0, 00615
32742	595441630271	18181979	17975296	549	0, 00305
65521	1191301446059	18181979	17975296	274	0, 00152

Графічне відображення залежності відсотка невиявлених помилок від перевірочного модуля наведено на рисунку 1.

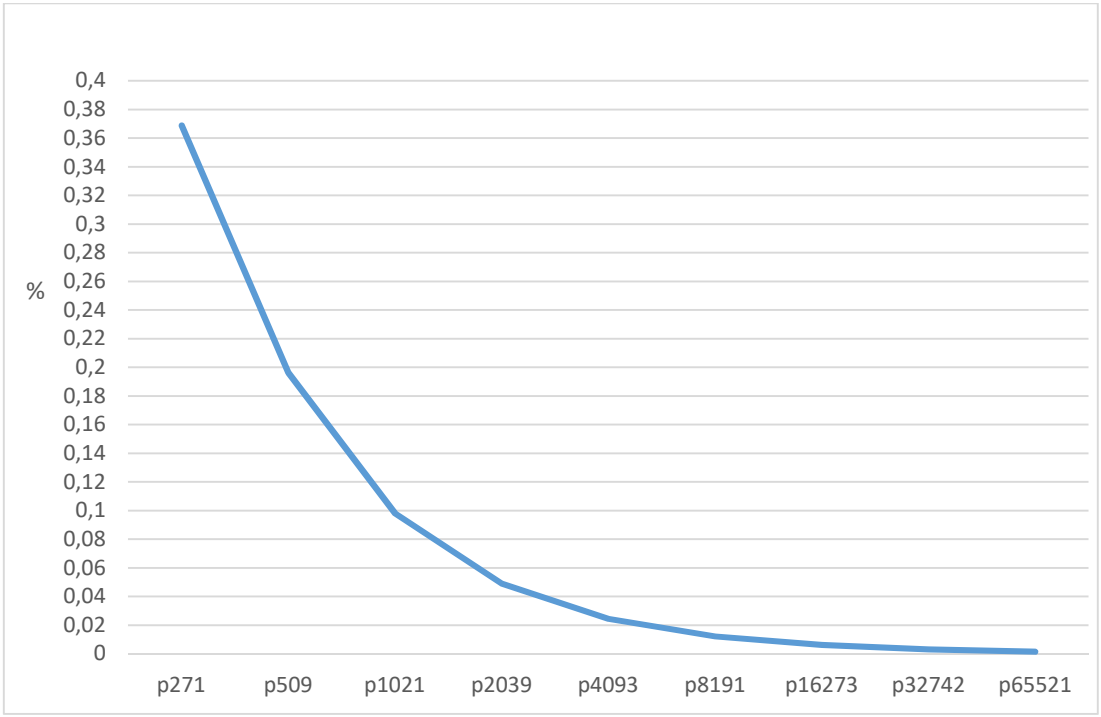


Рис. 1. Залежність кількості (%) невиявлених помилок від перевірочного модуля

За результатами проведеного експериментального дослідження можна зробити висновок, що при збільшенні значення перевірочного модуля на один розряд, кількість невиявлених помилок в трьох символах у заданому діапазоні зменшується в 1,88-2,02 рази. Також з таблиці 2 видно, що збільшення перевірочного модуля з 271 до 65521 дозволяє зменшити кількість невиявлених помилок приблизно у 240 разів.

Зазвичай при зростанні розрядності повідомлення зростає ймовірність спотворення більшої кількості символів. Для обчислення цього значення було проведено експериментальне дослідження залежності кількості невиявлених помилок. Результати розрахунків при спотворенні усіх інформаційних символів та при збільшенні їх кількості наведені в таблиці 3.

Таблиця 3

Обчислення кількісного значення виявлених помилок у всіх модулях окрім перевірочного

Кількість модулів, (n, k)*	Загальна кількість комбінацій	Кількість невиявлених помилок	Кількість невиявлених помилок, %
(3, 2)	67072	251	0,374
(4, 3)	17975296	66327	0,369
(5, 4)	4853329920	17521051	0,361
(6, 5)	1339519057920	4755290656	0,355

* - де n – загальна кількість модулів; k – кількість інформаційних модулів.

Графічне відображення залежності % невиявлених помилок від кількості модулів наведено на рисунку 2.

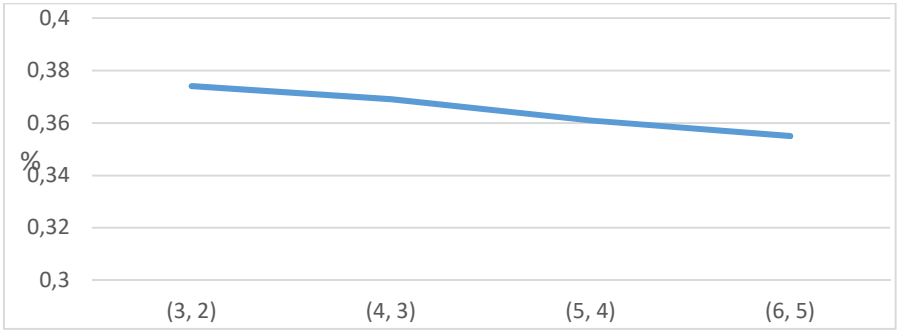


Рис.2. Залежність % невиявлених помилок від кількості модулів

Як видно з рисунку 2 зі збільшенням числа інформаційних модулів та комбінацій спотворених символів кількість невиявлених помилок у відсотковому поданні практично залишається без змін.

На основі результатів дослідження можна зробити висновок, що збільшення кількості помилок при зміні кількості інформаційних модулів прямо пропорційне збільшенню робочого діапазону.

Висновки

Поєднання технічних та програмних методів дозволяє більш ефективно захистити інформацію при передачі по каналам зв'язку. Використання коригуючих кодів на основі СЗК, як одного із методів програмного захисту, дозволяє нам завдяки рівноцінності інформаційних та перевірочних символів відновити втрачену інформацію. Використання методу обчислення синдрому вимагає створення та зберігання таблиці синдромів і тоді виправлення однієї помилки здійснюється шляхом зчитування відповідного значення з таблиці. Зберігання великого масиву даних може негативно позначитись на системах обробки інформації. Іншим недоліком є те, що необхідно задовільнити умову $P_c > P$, проте перевагою даного методу є швидкість виправлення помилок. На відміну від методу пошуку синдрому метод проекцій не потребує збереження масивів даних, проте після виявлення спотворення необхідно провести додаткові обчислення, кількість яких залежать від кількості модулів.

Для обчислення ймовірності виникнення помилок при передачі інформації були проведені експериментальні дослідження, які показали, що при збільшенні значення перевірочного модуля на один розряд, кількість невиявлених помилок в інформаційних символах зменшується в 1,88-2,02 рази.

Кількість інформаційних модулів у залежності від поставленої задачі може змінюватись. У проведених дослідження було визначено, що при збільшенні числа інформаційних модулів та комбінацій спотворених символів кількість невиявлених

помилок у відсотковому поданні практично залишається без змін. При ймовірності помилки менше 0,4 % можливе використання приведених комбінацій модулів для проектування систем захищеного зберігання інформації.

Список літератури

1. Ляшук О. М. Mhed – високоефективний метод захисту даних на основі багатошарового гібридного шифрування. *Вісник НТУУ "КПІ"*. Серія Радіотехніка, Радіоапаратобудування. 2014. №56. С. 144–151.
2. Polotai O., Kukharska N., Lagun A. The steganographic approach to data protection using arnold algorithm and the pixel-value differencing method. *7 Proceedings of the 2020 IEEE 3rd International Conference on Data Stream Mining and Processing, DSMP 2020*. 2020. P. 174–177.
3. Клименко С.В., Малайчук В.П., Селіванов Ю.М., Петренко О.М., Астахов Д.С. Система передачі інформації із застосуванням інтерактивного блокового криптографічного алгоритму TWOFISH. *Актуальні проблеми автоматизації та інформаційних технологій*. 2021. С. 62–71.
4. Barannik V., Havrilov D., Shulgin S., Parkhomenko M., Yroshenko V. The possibility of using the arithmetic coding method in the systems of cryptographic information protection. *Ukrainian Scientific Journal of Information Security*. 2020. Vol. 26. issue 3. P. 156-167.
5. Цаволик Т.Г. Коригуючі коди в системі залишкових класів зі спеціальними модулями. *Вимірювальна та обчислювальна техніка в технологічних процесах*. Технічні наук и. 2016. № 3. С. 100–104.
6. Kaganyuk, O., Chernyashchuk N., Burchak I. Аналіз апаратних та програмних методів захисту інформації від несанкційного доступу до інформаційного каналу передачі даних. *COMPUTER-INTEGRATED TECHNOLOGIES: EDUCATION, SCIENCE, PRODUCTION 47*. 2022. С. 76–82.
7. Горбенко І.Д., Гріненко Т.О. Захист інформації в інформаційно-телекомунікаційних системах. *Харківський національний ун-т радіоелектроніки*. - Х.: ХНУРЕ, 2004. С. 364-368.
8. Яцків, В.В., Кулина С.В. Метод надійного зберігання даних на основі надлишкової системи залишкових класів. *Вісник Хмельницького національного університету*. 2019. С. 98-104.
9. Волинський О. Систематизація характеристик теоретико-числових базисів та їх застосування для побудови високопродуктивних спецпроцесорів. *Вісник Тернопільського національного технічного університету*. 2011. Том 16. №3. С. 183–189.
10. Alrajeh, N.A., Marwat, U., Shams, B., Shah, S.S.H. Error correcting codes in wireless sensor networks: an energy perspective. *Applied Mathematics & Information Scienc-es.* 2015. P. 809-818.
11. Omondi A., Premkumar B. Residue Number System: Theory and Implementation. Imperial College Press. 2007. Vol. 2. 296 P.
12. Yuke Wang. Residue-to-Binary Converters Based On New Chinese Remainder Theorems. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*. Vol. 47. No. 3. March 2000. P. 197–205.
13. Яцків В.В. Виявлення та виправлення багатократних помилок на основі модулярних коректуючих кодів. *Інформаційні технології та комп'ютерна інженерія*. 2015. Том 33. № 2. С. 77–82.
14. Яцків В.В. Теоретичні основи створення і структурна організація компонентів безпровідних сенсорних мереж підвищеної ефективності : дис.... д-ра техн. наук : 05.13.05 / Тернопільський нац.екон. ун-т. Тернопіль, 2016. 280 с.

С.В. Кулина

ERROR DETECTION AND CORRECTION IN PROTECTED DATA STORAGE SYSTEMS BASED ON THE CALCULATION OF NUMBER PROJECTIONS

S.V. Kulyna

West Ukrainian National University, 11 11 Lvivska st.,
Ternopil, 46020, Ukraine; e-mail: sersks@wunu.edu.ua

The system of distributed data storage includes, in addition to the devices themselves, on which information will be stored, as well as communication channels through which it is forwarded. When transmitting information to considerable distances, there are a number of new problems that need to be solved. An analysis of the current state and development of data transmission and storage systems shows that it is necessary to use various technical and software methods of information protection from distortion, loss or theft. The use of only one of them is not effective, so they usually use their combination. The most widely used data protection methods are the methods based on the use of correction codes, particularly the residue number system (RNS) codes. Due to the presentation of information in the form of remainders from the division of data into each element of the system of mutually simple modules, a reduction in the number of digits with which operations are carried out is achieved. The simplification of information collection and transmission system due to the presentation of information in the form of residuals is investigated in this paper. The initial range of calculations expansion implemented by the introduction of additional backup modules is suggested for data protection, that allows to increase the probability of error detection to 100%. The advantages and disadvantages of detecting and correcting errors with the use of corrective codes in RNS using the methods of calculating the syndrome and calculating the number projection is analyzed. The number of information modules in the designed systems can vary depending on the task, therefore, in the conducted studies, it was determined that with an increase in the number of information modules and combinations of distorted symbols, the number of undetected errors in percentage compliance remains practically unchanged. It is suggested to use RNS without additional error correction for designing information collection systems, as it is possible to choose a set of modules in such a way that the maximum number of errors that will not affect the system functionality.

Keywords: Corrective codes, residual number system, Chinese remainder theorem, distributed data storage, reverse transformation methods, error correction.

DOI 10.15276/imms.v12.no4.346
УДК 004.633.2

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 346-351

РОЗРОБКА МУЛЬТИСЕРВЕРНОГО КОРПОРАТИВНОГО МЕСЕНДЖЕРА ЗІ СПЕЦІАЛЬНИМ ЗАХИСТОМ МІЖСЕРВЕРНОГО ТРАФІКУ

М.А. Лісовський, О.А. Стопакевич

Національний університет «Одеська Політехніка»,
просп. Шевченка, 1, Одеса, 65044, Україна; e-mail: mix313131@gmail.com

Запропонована реалізація месенджеру, яка передбачає, що для доступу до даних, пов'язаних з месенджером, користувачу необхідно зв'язатись з сервером локальної корпоративної мережі, в іншому випадку дані будуть знаходитись у шифрованому стані. Логіка з'єднання будується наступним чином: Клієнт підключається до Серверу, що знаходиться в межах Корпоративної Мережі. Після підтвердження версій Клієнта та Сервера, встановлюється зв'язок і Клієнт переходить у робочий режим - дані, збережені в середині Клієнту розшифровуються і Користувач отримує можливість працювати з цими даними. Коли Користувач відправляє повідомлення іншому Користувачу, Клієнт шифрує повідомлення та відправляє його на Сервер з вказанням Відправника та Одержувача. У випадку, коли Одержувач знаходиться в одній мережі з Відправником, Сервер перенаправляє повідомлення без додаткових дій. В іншому випадку - Сервер чекає на ще декілька повідомлень (від будь-яких Відправників), якщо встановлена кількість набирається - повідомлення починають пакуватись в одне велике, після чого шифруються ще раз та відправляються до іншого Серверу. Якщо після встановленого часу кількість не набирається - Сервер додає невистачаючу кількість Повідомлень у вигляді випадкового набору значень, які не матимуть Відправника та Отримувача (для того, щоб другий Сервер відкинув ці повідомлення як "шум").
Ключові слова. Мультісерверний месенджер. Захист міжсерверного трафіку. Шифрування.

В сьогоденні питання інформаційної безпеки особливо важливе, і воно розповсюджується на всі елементи інформаційної системи, у тому числі - засоби з'єднання між окремими філіями бізнесових офісів. Наразі, частіш за все для обміну інформацією використовується електронна пошта, яка має цілу низку вразливостей. Прикладами таких вразливостей можуть бути CVE-2015-0235, що є атакою на сховище електронної пошти з метою переповнення буферу і виконання коду, або CVE-2017-1274, що є атакою на рівні отримання пошти (POP3, IMAP), та ін. Всі ці атаки представляють собою велику загрозу безпеці та можуть призвести до витоку, пошкодження або знищення критичної інформації. Альтернативою використанню пошти є використання месенджерів на кшталт Telegram, WhatsApp, Viber, тощо. Але ці месенджери також становлять велику загрозу безпеці інформації. Прикладом є ймовірність втрати пристрою, на якому є активним обліковий запис месенджеру та зберігається чутлива інформація. У разі втрати, цей пристрій може потрапити у руки зловмисника і, хоча, власник аккаунту може завершити сеанс на цьому пристрої, це не дасть жодного результату, якщо пристрій не буде підключений до мережі Інтернет, бо тоді зловмисник матиме змогу скопіювати всі скачані на пристрій важливі файли, та подивитись всю збережену у хеші переписку, що є менш ніж бажаним. Однак, це не єдина загроза, притаманна месенджерам. Вони, як і будь-яке інше програмне забезпечення, мають якісь вразливості, як незначні, наприклад CVE-2021-41861, який описує баг, присутній версіям Telegram 7.5.0-7.8.0 на ОС Android, при якому фото з функцією самознищення не самовидалялось з пристрою, але відображалось як видалене для

користувачів; так і значні вразливості, як, наприклад, CVE-2019-12569, вразливість десктопної версії мессенджера Rakuten Viber до 10.7.0, при якій зловмисник міг виконати довільні команди, скориставшись небезпечними шляхами, які використовувались для URI програми. Завдяки цій вразливості зловмисник може, переконавши цільового користувача перейти за шкідливим посиланням, завантажити бібліотеки з директорії, на яку вказував URI, та отримати можливість виконувати довільні команди з привілеями цільового користувача.

Звісно, існують месенджери корпоративного призначення, але вони, загалом, працюють виключно у локальній мережі і не можуть використовуватись для з'єднання між віддаленими філіями. Саме тому існує необхідність у створенні месенджера, призначеного для безпечного з'єднання як всередині локальної корпоративної мережі, так і між кількома корпоративними мережами.

Запропонована реалізація такого месенджеру передбачає, що для доступу до даних, пов'язаних з месенджером, користувачу необхідно зв'язатись з сервером локальної корпоративної мережі, в іншому випадку дані будуть знаходитись у шифрованому стані. Логіка з'єднання будується наступним чином: Клієнт підключається до Серверу, що знаходиться в межах Корпоративної Мережі. Після підтвердження версій Клієнта та Сервера, встановлюється зв'язок і Клієнт переходить у робочий режим - дані, збережені в середині Клієнту розшифровуються і Користувач отримує можливість працювати з цими даними. Коли Користувач відправляє повідомлення іншому Користувачу, Клієнт шифрує повідомлення та відправляє його на Сервер з вказанням Відправника та Одержувача. У випадку, коли Одержувач знаходиться в одній мережі з Відправником, Сервер перенаправляє повідомлення без додаткових дій. В іншому випадку - Сервер чекає на ще декілька повідомлень (від будь-яких Відправників), якщо встановлена кількість набирається - повідомлення починають пакуватись в одне велике, після чого шифруються ще раз та відправляються до іншого Серверу. Якщо після встановленого часу кількість не набирається - Сервер додає невистачаючу кількість Повідмлень у вигляді випадкового набору значень, які не матимуть Відправника та Отримувача (для того, щоб другий Сервер відкинув ці повідомлення як "шум").

Розглянемо в деталях як саме влаштовані різні аспекти роботи месенджера. Почнемо з підключення Клієнту до Серверу. Існує декілька сценаріїв, які можуть виникнути під час підключення.

Перший сценарій - успішне підключення, як зображено на рис. 1.

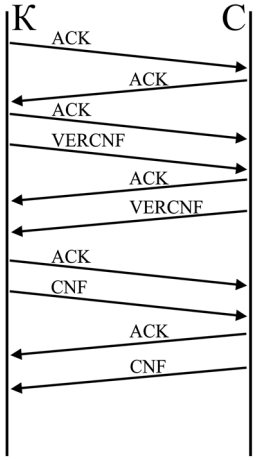


Рис 1. Процес з'єднання Клієнта та Сервера

При такому сценарії Клієнт та Сервер відправляют один одному свої версії у вигляді VERCNF (Version Confirm), та підтверджують вірність версій один

одного, відправляючи CNF (Confirmed) у вигляді відповіді, після чого виконують з'єднання. У цьому випадку наступним кроком для Клієнту буде розшифрування власного вмісту.

Другий та третій сценарії дуже схожі - це сценарії, у яких версія або Клієнту або Серверу виявилась недійсною. При такому сценарії сторона, що виявила недійсну версію відправляє CD замість CNF, що означає Connection Denied (У з'єднанні відмовлено). У випадку такої ситуації Користувач побачить повідомлення "З'єднання невдале" та буде необхідно повідомити про це до свого Системного Адміністратору.

Перед кожним повідомленням відправляється ACK (Acknowledge) від однієї з сторін, для того щоб визначити який з каналів є активним в поточній передачі.

Коли користувач надсилає повідомлення, Клієнт будує запит до сервера наступним чином: ID відправника, ID отримувача, зашифроване повідомлення (може включати до себе субструктуру текст+медіа-вкладення або одне з двох, в такому випадку поле тексту або поле медіа-вкладення буде пустим).

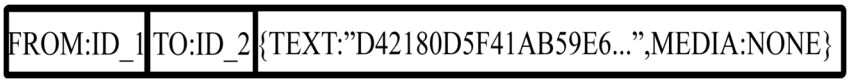


Рис 2. Структура повідомлення Клієнта до Серверу

Сам ID користувача прив'язується до відповідального Серверу, а задля спрощення орієнтації по ID, він може мати форму %server_name%_%username% (наприклад SCS1_TishchenkoG, де SCS1 - Science Company Server 1, TishchenkoG - Тіщенко Г.). Після отримання повідомлення, Сервер перевіряє чи знаходиться Одержувач всередині цієї локальної мережі чи ні, якщо ні - переходить до пакування повідомлення на другий Сервер, якщо Одержувач знаходиться в межах цієї локальної мережі, повідомлення просто пересилається до нього.

У випадку з прикріпленнями, вони посилаються у зашифрованій формі, а медіа мають при собі чек-суму, для перевірки автентичності медіа-вкладень, у випадку коли вони прикладаються до повідомлення, інакше замість медіа знаходиться None.

Коли повідомлення надходить до Серверу, він чекає n часу на надходження ще k повідомлень, після чого або генерує додаткові "мусорні" повідомлення, щоб довести загальну кількість до встановленої, або, якщо встановлена кількість була досягнута, починає процес стиснення повідомлень до одного потоку.

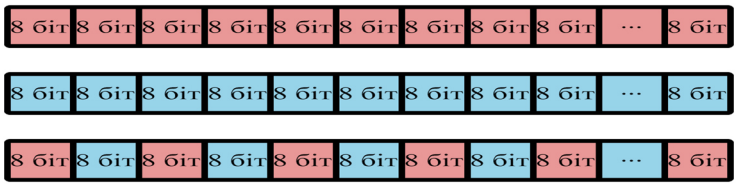


Рис 3. Процес стиснення повідомлень, де червоним кольором зображено повідомлення 1, а блакитним - повідомлення 2, повідомлення з кольорами, що чергуються - результуюче повідомлення.

На рис. 3 зображена узагальнена інтерпретація процесу стиснення, де кожне повідомлення розділяється на блоки розміром 8 біт та поєднуються в один ланцюг. Це дозволяє досягти мінімальну необхідну кількість можливих з'єднань з Сервером ззовні, що значно ускладнює типові атаки на порти з'єднання, які могли бути активними, але не використовуватись.

Після утворення нового ланцюга, він повторно шифрується, з метою приховати внутрішні ідентифікатори користувачів від аналізу потенційними

зловмисниками, після чого Сервер 1 повідомляє Сервер 2 про початок відправки повідомлень та починає пересилання.

Таким чином, навіть якщо хтось зможе продивитись трафік, в нього не буде жодного розуміння стосовно змісту повідомлень, що надсилаються від одного Серверу до іншого.

Тому загальна схема з'єднання двох Серверів буде виглядати наступним чином

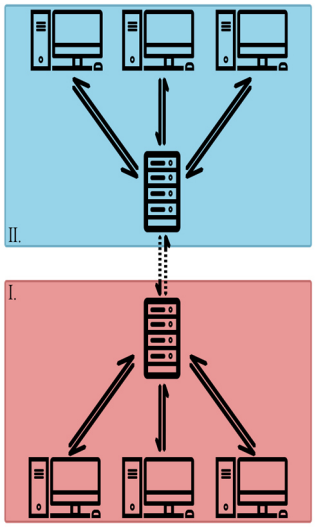


Рис 4. Загальна схема з'єднання Серверів двох філій корпорації.

На рис 4 червоним кольором позначено локальну мережу філії I, а синім - локальну мережу філії II. У кожній з мереж кожен пристрій має двостороннє з'єднання з локальним Сервером, який, в свою чергу, має двостороннє з'єднання з віддаленим Сервером іншої філії через мережу Інтернет. Ця схема передбачає відсутність у користувачів до мережі Інтернет всередині локальної мережі, а єдине, що з'єднує їх з іншою філією - їх локальний Сервер, який, в свою чергу, має доступ до мережі інтернет, але не приймає на свій зовнішній порт інші з'єднання, таким чином значно знижуючи ризики атак ззовні.

Однією з найголовніших частин будь-якого месенджеру є його інтерфейс, оскільки, незалежно від використаних технологій, популярність месенджеру буде залежати від того, наскільки комфортно ним користуватись.

Тому інтерфейс має бути простим для розуміння та ефективно використовувати простір екрану, даючи при цьому можливість змінювати зовнішній вигляд відповідно до побажань користувача.

Спираючись на ці потреби було створено макет інтерфейсу, зображений на рис 5.

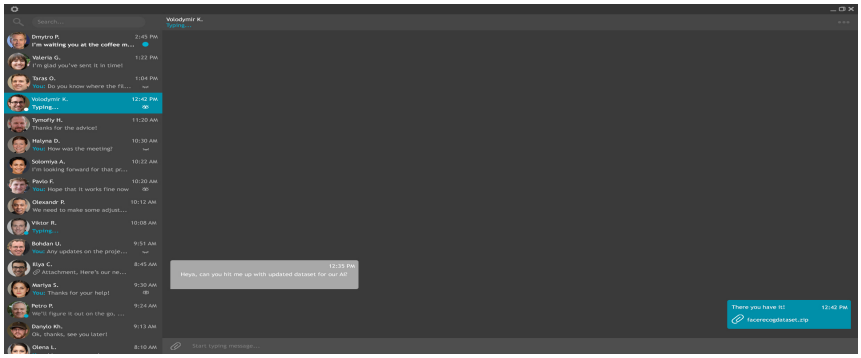


Рис 5. Макет інтерфейсу мессенджеру

У цьому макеті екран розподіллено на дві частини: список чатів та вікно активного чату. Це дозволяє користувачу мати легкий доступ до будь-якого чату навіть при листуванні з іншим користувачем.

Додатково, у лівій стрічці розташовано пошук по чатам, завдяки чому користувач матиме змогу легко та швидко знайти повідомлення та/або файли, які він отримав раніше, без необхідності перевіряти кожен з чатів вручну.

Для більш простого розуміння, кожен чат позначається додатковими символами - це зачинене та відчинене око та блакитна крапка. Перші два позначають чи прочитав інший користувач те повідомлення, яке було йому надіслане. Третій символ позначає непрочитані повідомлення, які отримав користувач, зроблено це для того, щоб візуально звернути увагу користувача на нові повідомлення.

Також у стрічці зліва відображаються: ім’я, зображення профіля, дата останнього повідомлення, стан повідомлення та передпоказ останнього повідомлення або “Typing…”, у випадку коли інший користувач набирає нове повідомлення.

У самому вікні чату користувач бачить свої повідомлення по правому боку, а повідомлення іншого користувача - по лівому. За бажанням користувач може змінити відображення повідомлень по одному боці.

Повідомлення у чаті демонструє час, коли воно було надіслано та його зміст, включаючи зображення та/або інші прикріплені файли. У вікні чату, якщо інший користувач ще не переглянув повідомлення, біля нього буде відображена блакитна крапка, яка буде відображати для користувача цей факт. А у випадку коли повідомлення все ще надсилається - буде відображатись символ годинника.

Також у користувача присутнє меню налаштувань, що зображене на Рис 6.

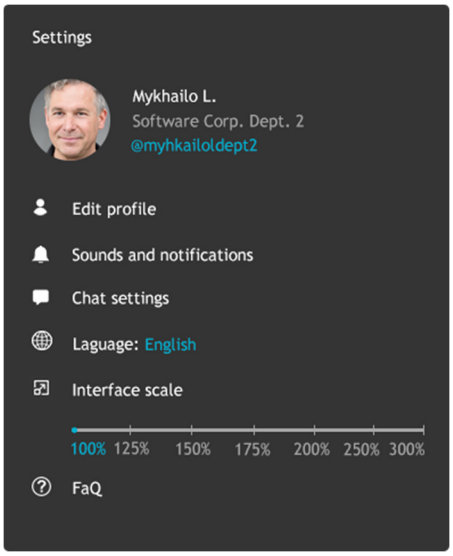


Рис 6. Меню налаштувань

Це вікно відкривається поверх вікна месенджера та має в собі наступні елементи: профіль користувача та можливість налаштувати профіль, звуки, чати, мову та розмір інтерфейсу, а також ЧаПи, які допоможуть користувачам швидше розібратись як саме влаштован месенджер.

У меню Edit profile користувач може змінити: зображення та ім’я профілю. Департамент користувача призначається самим сервером, до якого користувач вперше підключається.

Sounds and notifications дозволяє змінити гучність або повністю вимкнути звуки сповіщень та відображення спливаючих повідомлень, з можливістю вимкнути або повністю або на певний проміжок часу.

Chat settings дозволяє налаштувати фонове зображення чатів, кольорову схему інтерфейсу, а також розташування повідомлень користувача (по правому чи по лівому боку).

Налаштування мови дозволяє обрати мову інтерфейсу або додати власний переклад, у разі коли мова не входить в список мов за замовчуванням.

Розмір інтерфейсу дозволяє вибрати необхідний розмір, згідно з обраним користувачем монітором.

Список літератури

1. Wang X., Wu R., Ma J., Long G. Research on Vulnerability Detection Technology for WEB Mail System. *Procedia Computer Science.* 2018. Vol.131. P.124-130
2. Vulnerability found in top messaging apps let hackers eavesdrop. 2021. https://www.pandasecurity.com/en/mediacenter/security/vulnerability-messaging-apps/
3. CVE-2015-0235 - https://nvd.nist.gov/vuln/detail/CVE-2015-0235
4. CVE-2017-1274 - https://nvd.nist.gov/vuln/detail/CVE-2017-1274
5. CVE-2019-12569 - https://nvd.nist.gov/vuln/detail/CVE-2019-12569

DEVELOPMENT OF MULTISERVER CORPORATE MESSENGER WITH SPECIAL PROTECTION OF INTERSERVER TRAFFIC

M.A. Lisovsky, O.A. Stopakevych

National Odesa Polytechnic University,
ave. Shevchenko, 1, Odesa, 65044, Ukraine; e-mail: mix313131@gmail.com

The proposed implementation of the messenger, which assumes that in order to access the data related to the messenger, the user needs to contact the server of the local corporate network, otherwise the data will be in an encrypted state. The connection logic is built as follows: The Client connects to the Server located within the Corporate Network. After confirming the versions of the Client and Server, the connection is established and the Client goes into working mode - the data stored in the middle of the Client is decrypted and the User gets the opportunity to work with this data. When a User sends a message to another User, the Client encrypts the message and sends it to the Server specifying the Sender and Recipient. If the Recipient is in the same network as the Sender, the Server forwards the message without additional actions. Otherwise - the Server waits for several more messages (from any Senders), if the set number is reached - the messages begin to be packed into one big one, after which they are encrypted again and sent to another Server. If after the set time the number is not dialed - the Server adds an insufficient number of Messages in the form of a random set of values that will not have a Sender and a Recipient (so that the second Server rejects these messages as "noise").

Keywords. Multiserver messenger. Protection of inter-server traffic. Encrypts

DOI 10.15276/imms.v12.no4.352
УДК 004.633.2

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 352-357

ОРГАНІЗАЦІЯ ЕЛЕКТРОННОГО ДОКУМЕНТООБІГУ З ВИКОРИСТАННЯМ ЧАТ-БОТУ

В.В. Подуфалов, Н.І. Кушніренко, Н.Г. Козаченко

Національний університет «Одеська Політехніка», просп. Шевченка, 1, Одеса, 65044, Україна; e-mail: infsec2011@gmail.com

Через події останніх років, такі як вірус COVID-19, а пізніше російсько-українська війна, було унеможливлено очне навчання в багатьох навчальних закладах в Україні. Через загрозу здоров'ю та життю студентів, викладачів та інших робітників, навчальні заклади були змушені перейти на дистанційне навчання або роботу. Тому в цей час доволі гостро постало питання організації електронного документообігу, адже працювати із усіма файлами вручну доволі не доцільно, особливо коли кількість оброблювальних в день файлів може становити декілька сотень. Через це було прийняте рішення створити програмний продукт, який зміг би спростити дистанційну роботу із документами студентам та викладачам ВНЗ або невеликим компаніям. Метою роботи є автоматизація процесу управління документами за допомогою чат-боту та програмного додатка. У роботі проведено аналіз месенджерів, на основі яких може бути розроблений чат-бот, а також аналіз мов програмування, які дадуть можливість взаємодіяти із Telegram та розробити програмний додаток максимально ефективно. У результаті виконання роботи був розроблений чат-бот на основі Telegram, який призначений для відправки студентами або викладачами файлів та програмний додаток для упорядкування та завантаження документів викладачем або студентом відповідно. Результати даної роботи можуть бути використані на персональних комп'ютерах студентів, викладачів або на кафедрі університету.

Ключові слова: чат-бот, упорядкування електронних документів, автоматизація електронного документообігу.

Вступ

За останні декілька років сталися події, які зробили складним, а інколи взагалі неможливим очне навчання в навчальних закладах в Україні: спочатку це був вірус COVID-19, а тепер російсько-українська війна [1-2]. Загроза життю та здоров'ю людей змусила навчальні заклади і більшість працівників перейти на дистанційне навчання або роботу, адже продовжувати функціонувати в очному режимі навчальні заклади не могли: станом на 1 вересня і з початку повномасштабного вторгнення російські війська пошкодили 2,4 тисячі навчальних закладів в Україні, з них 270 зруйновані повністю. І навіть на даний момент майже 50% навчальних закладів в Україні продовжують дистанційне навчання [3]. Відповідно зараз як ніколи гостро постало питання організації електронного документообігу, адже відправляти всі файли вручну досить трудомістко, особливо коли кількість відправлених в день файлів становить декілька сотень.

На основі аналізу існуючих на ринку рішень, таких як Сервіс «Вчасно», Сервіс «Мій Арт-Офіс», Alfresco, Сервіс «Аврора», Microsoft SharePoint, Directum та NauDoc було зроблено висновок, що у більшості з цих систем документообігу є обмеження за часом безкоштовного використання або за кількістю користувачів у системі, інші ж мають забагато зайвого функціоналу, наприклад, цифровий підпис або юридична цінність документів. Таким чином, не було знайдено програмних рішень для сфери освіти, які б задовольняли усім потребам і водночас не були б перенавантажені зайвими у обраній сфері можливостями. Тому було вирішено створити програмний продукт, який зможе полегшити дистанційну роботу викладачам ВУЗів, буде максимально простим і водночас матиме можливість

виконувати усі основні функції, необхідні студентам та викладачам.

Мета статті та постановка завдань

Метою розробки програмного продукту є автоматизація процесу управління документами за допомогою програмного додатку та чат-боту на базі Telegram.

Для досягнення поставленої мети необхідно розв'язати наступні *задачі*:

а) виконати аналіз предметної області та існуючих програмних рішень;

б) виконати аналіз найбільш гнучких технологій для роботи із Telegram та документами;

в) розробити чат-бота через Telegram;

г) розробити програму-додаток для упорядкування та завантаження документів.

Програмний продукт, що розроблюється, буде розглядатися на прикладі взаємодії студента та викладача і повинен бути здатний виконувати найпростіші функції, а саме:

- мати можливість приймати, зберігати та відправляти файли;
- упорядковувати файли за ім'ям, прізвищем та групою навчання відправника;
- використовувати базу даних для зберігання файлів;
- надавати можливість вибору директорії зберігання файлів;
- розрізняти користувачів чат-боту між собою;
- приймати файли будь-якого формату, розширення та розміру.

Основна частина

У відкритому доступі не було знайдено безкоштовних програмних продуктів, які б могли ефективно вирішувати задачі документообігу, які повсякденно виникають в професійній діяльності студентів та викладачів і не були б ускладнені непотрібним функціоналом. Більшість подібних програм розраховані на великі компанії, в яких в процесі документообігу задіяно сотні людей та офіційні папери, які необхідно підписувати та завіряти, тощо.

Взявши до уваги те, що в процесі обміну документами будуть задіяні як викладачі так і студенти, було вирішено побудувати додаток на основі взаємодії з одним з популярних месенджерів з використанням чат-боту. Це надало б всім користувачам можливість працювати із простим та знайомим інтерфейсом та мати постійний доступ до чат-боту на будь-якому приладі – телефон, персональний комп'ютер, ноутбук.

У якості месенджерів для реалізації чат-боту було розглянуто декілька популярних варіантів, таких як Telegram, Viber та WhatsApp. Проаналізувавши перелічені месенджери, їх слабкі та сильні сторони, було вирішено використовувати Telegram через ті можливості, які він надає користувачу та програмісту, який буде з ним працювати.

Великим плюсом використання чат-боту саме на основі Telegram є його захищеність [4]. У сервісі використовують одразу декілька систем захисту інформації, застосовуючи такі алгоритми шифрування, як AES-256, SHA-256, RSA-2048 та DH-2048 при використанні Telegram на телефоні або персональному комп'ютері, а також HTTPS-протокол, якщо користувач використовує web-версію Telegram, а ключи шифрування, звісно, не передаються третім особам [5].

Також однією з переваг месенджеру Telegram є відкритий API – сервіс операційної системи, який дозволяє користуватись усіма функціями Telegram через програмний код [6]. Наприклад, сервер деякого сайту замість того, щоб робити запит для створення події з встановленими параметрами напряму до серверів Telegram, який має обробити цей запит та передати відповідь на нього назад на сайт, може створити запит до API сервера Telegram, видаляючи із цього ланцюга

запит на сайт.

Для написання програмного коду для керування командами чат-боту потрібно було обрати мову програмування, яка використовуючи найменш можливу кількість ресурсів, дасть можливість виконувати усі потрібні функції. У якості мов програмування було розглянуто декілька: C++, C#, Python та Java [7]. Маючи певне уявлення про структуру кожної з розглянутих мов програмування, було зроблено висновок, що найбільш доцільно розробляти обране програмне забезпечення на Python.

До переваг цієї мови програмування можна віднести більш простий синтаксис ніж у деяких аналогів, велику кількість додаткових пакетів для вирішення майже будь-якої задачі, високу затребуваність на ринку праці, широке застосування у компаніях будь-яких масштабів, тощо [8-9].

На основі обраних технологій розроблено програмний продукт для автоматизації документообігу. Схема роботи програмного продукту наведена на рисунку (рис. 1).

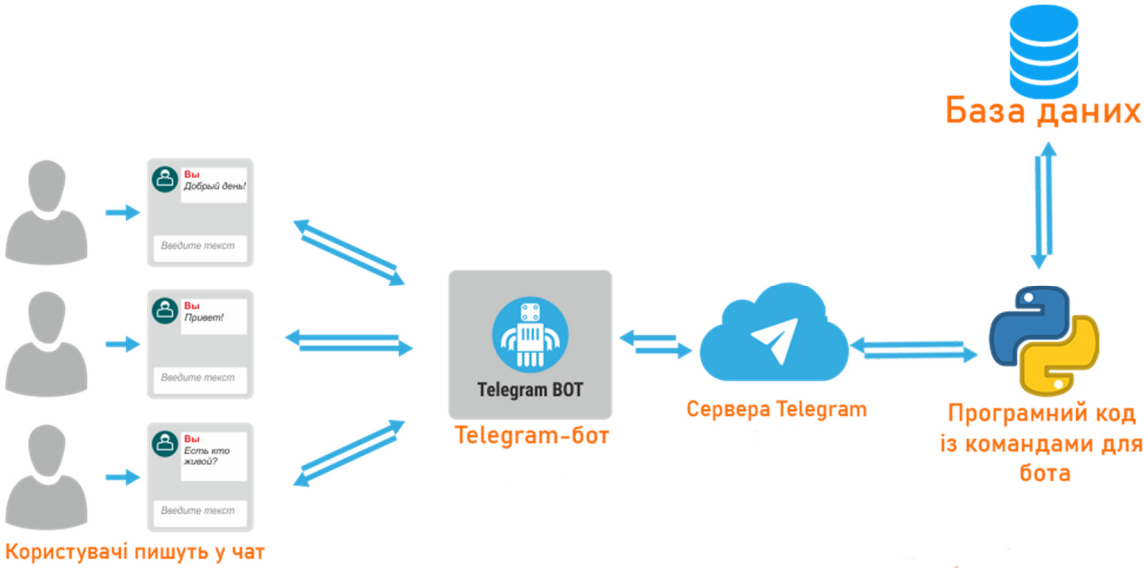


Рис. 1. Схема програмного продукту

Алгоритм функціонування програмного продукту складається з наступних кроків.

1. Студент або викладач, який бажає відправити документи (лабораторні або модульні роботи, курсові роботи, тощо), повинен зайти у Telegram, знайти бота для зберігання та упорядкування файлів за допомогою пошукового запиту «@SendFilesONPUBot».
2. Почати діалог з ботом із стандартної команди «/start».
3. Слідувати крокам у повідомленнях бота, а саме: ввести команду«/filesend»,
4. Обрати, чи є користувач студентом або викладачем.
5. Якщо було обрано “викладач” - треба ввести пароль, відомий тільки викладачам.
6. Ввести ім’я, прізвище, групу та предмет. (Якщо відправляє студент - ввести групу навчання, якщо викладач - групу, для якої буде відправлятися файл)
7. Після цього бот буде готовий прийняти файли від користувача. Водночас може бути відправлено від 1 до 100 файлів будь-якого розширення.
8. Після отримання документів бот відправить користувачу повідомлення про те, що усі файли було успішно збережено, на чому користувач завершає роботу із чат-ботом.

На четвертому та п’ятому кроках визначається, у яку таблицю бази даних буде записана інформація, передана в наступних кроках. Якщо користувач обирає варіант “Викладач”, це означає, що файли, до яких йому буде надано доступ у програмному додатку - це виконані роботи студентів, а тому, щоб студенти не могли отримати доступ до завершених робіт інших студентів, потрібен метод захисту бази даних.

На шостому кроці відбувається ідентифікація користувача: після того, як користувач обирає студент він або викладач, бот зберігає унікальний ідентифікаційний номер користувача, а коли користувач відправляє документ – бот порівнює ідентифікаційний номер відправника команди із ідентифікаційним номером відправника файлу. Якщо вони відрізняються – бот видає помилку визначення користувача. Це зроблено для того, щоб ніхто не зміг втрутитися у процес відправки файлів іншої людини.

Усі файли, що були отримані ботом зберігаються у заздалегідь створеній таблиці у базі даних SQLite у бінарному форматі разом із ім’ям користувача, його прізвищем, групою навчання, назвою предмета та назвою кожного відправленого файлу. Коли викладачеві знадобиться зберегти роботи студентів для перевірки, або студент захоче завантажити бланк для виконання модульної роботи, відправлений викладачем, він відкриє додаток, який дозволяє знайти необхідні файли та зберегти їх на персональний комп’ютер.

Після відкриття програми користувач повинен обрати свою роль - Викладач чи Студент. Після цього користувач отримає доступ до програми, що дозволить йому переглядати доступні дані, упорядковувати їх за групою та предметом і завантажувати необхідні файли на персональний комп’ютері (рис. 2).

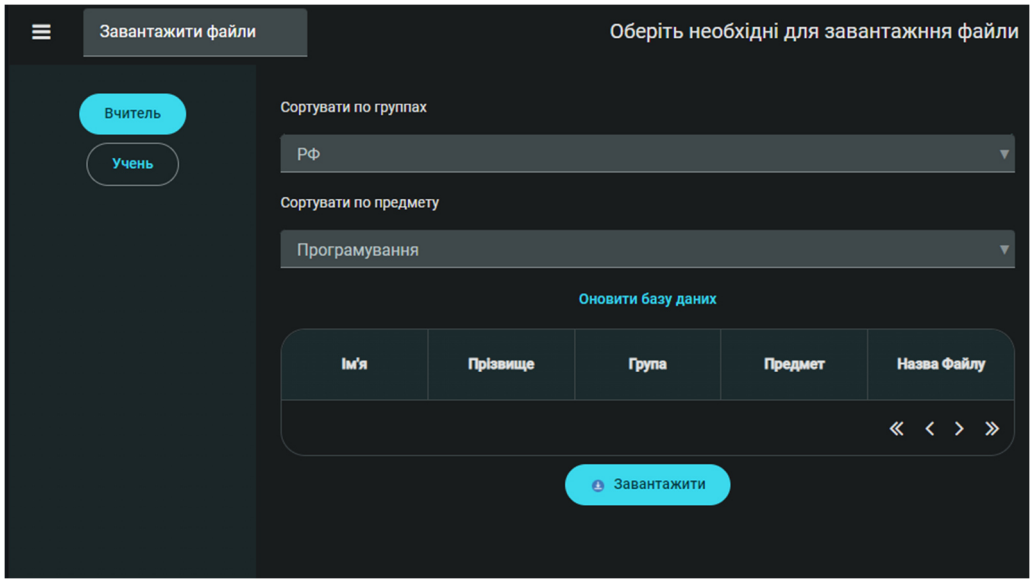


Рис. 2. Інтерфейс програми для завантажування файлів

Після натискання на кнопку «Завантажити» користувач повинен обрати директорію, куди будуть збережені обрані файли.

Саме на цьому етапі відбувається упорядкування файлів: введені після команди /filesend дані використовуються для створення відповідних каталогів. У обраному користувачем шляху файлової системи буде створено директорію «Групи», всередині якої створюються папки із назвами предметів а пізніше груп , які існують у базі даних на даний момент. У створених папках із назвами груп буде створено нові директорії, назви яких мають вигляд «Ім’я_Прізвище», де і будуть збережені файли відповідного студента (рис. 3).

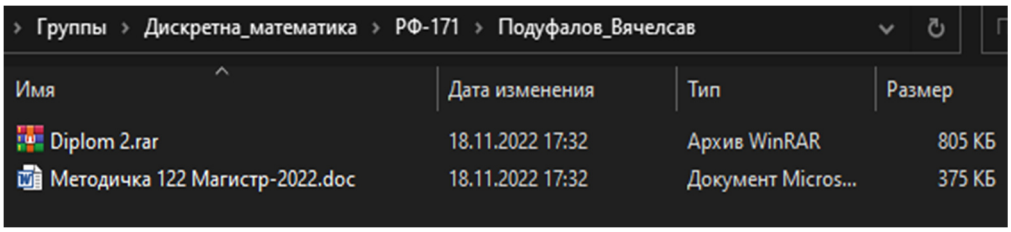


Рис. 3. Шлях зберігання файлів

Висновки

В роботі виконана розробка програмного продукту для автоматизації упорядкування документів студентів та викладачів, відправлених за допомогою чат-боту месенджера Telegram. Для досягнення поставленої мети було розв'язано наступні задачі:

- виконано аналіз предметної області та існуючих програмних рішень. Не було знайдено програмних рішень для сфери освіти, які б задовольняли усім потребам і водночас не були б перенавантажені зайвими у обраній сфері можливостями;
- виконано аналіз найбільш гнучких технологій для роботи із Telegram та документами. Обрано мову Python, яка має найбільш компактний програмний код та надає безліч можливостей завдяки стороннім бібліотекам. У якості системи управління базами даних було обрано SQLite через легкість використання та наявність усього необхідного функціоналу. Месенджер Telegram для реалізації чат-боту було обрано через його популярність серед студентів і викладачів а також через відкритий API, що дозволяє керувати ботом за допомогою програмного коду;
- розроблено чат-бота через Telegram, який має наступні можливості:

1) приймати файли будь-якого розміру та розширення;
2) приймати, зберігати та упорядковувати файли;
3) розрізняти користувачів чат-боту між собою;
4) перевіряти, яку роль має відправник файлів: Студент або Викладач;

- розроблено додаток для упорядкування та завантаження документів для викладачів та студентів, який має наступний функціонал:

1) упорядкування документів за ім'ям, прізвищем, групою та предметом;
2) використання бази даних для зберігання файлів;
3) надання можливості вибору директорії зберігання файлів;
4) можливість пошуку результатів по базі даних.

У підсумку можемо зазначити, що розроблений програмний продукт з використанням чат-боту в Telegram може значно полегшити процес прийому файлів викладачем від студентів і навпаки тому, що користувачам більше не потрібно буде зберігати кожен файл окремо та вручну і упорядковувати їх по необхідним папкам самостійно, усе це за нього тепер може робити розроблений додаток, що дасть викладачам більше часу на перевірку прийнятих робіт або на наукову та методичну діяльність, а студентам - більше часу для виконання інших завдань.

Список літератури

1. Pfizer. Коронавірус: загальна інформація. URL: https://www.pfizermed.com.ua/public/covid19-first
2. Російсько-українська війна URL: https://www.bbc.com/ukrainian/topics/czp6w66edqpt
3. Zmina.info. З початку вересня навчання в Україні розпочали майже 4 мільйони учнів. URL: https://zmina.info/news/z-pochatku-veresnya-navchannya-v-ukrayiny-rozpochaly-majzhe-4-miljony-uchniv-shhe-blyzko-piv-miljona-dosi-lyshayutsya-za-kordonom/

4. Security Guidelines for Client Developers. URL: https://core.telegram.org/mtproto/security_guidelines
5. Шнайер Б. Прикладная криптография: протоколы, алгоритмы и исходный код на С. Київ: Диалектика, 2018.
6. Telegram Bot API Documentation. URL: https://core.telegram.org/bots/api
7. Nicolas M. Building Telegram Bots: Develop Bots in 12 Programming Languages using the Telegram Bot API. Apress, 2019.
8. Matthes E. Python Crash Course, 2nd Edition: A Hands-On, Project-Based Introduction To Programming. No Starch Press, Inc., 2019.

DEVELOPMENT OF A CHATBOT FOR ORGANIZING ELECTRONIC DOCUMENTS

V. Podufalov, N. Kushnirenko, N. Kozachenko

National Odessa Polytechnic University,
1, Shevchenko Ave., Odesa, 65044, Ukraine; e-mail: infsec2011@gmail.com

Due to the events of the last years, such as the COVID-19 virus and later the Russian-Ukrainian war, it was impossible to study face-to-face in many educational institutions in Ukraine. Due to the threat to the health and lives of students, teachers and other employees, educational institutions were forced to switch to distance learning or work. Therefore, at this time, the issue of organizing electronic document management became quite acute, because it is not advisable to work with all files manually, especially when the number of files processed per day can be several hundred. Therefore, it was decided to create a software product that could simplify remote work with documents for students and teachers of universities or small companies. The aim of the work is to automate the document management process using a chatbot and a software application. The work includes the analysis of messengers, on the basis of which a chatbot can be developed, as well as the analysis of programming languages that will make it possible to interact with Telegram and develop a software application as efficiently as possible. As a result of the work, a chatbot based on Telegram was developed, which is designed to send files by students or teachers and a software application for organizing and downloading documents by a teacher or student, respectively. The results of this work can be used on personal computers of students, teachers or at the university department.
Keywords: chatbot, organization of electronic documents, automation of electronic document management.

DOI 10.15276/imms.v12.no4.358
УДК 621.377: 681.518

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 358-366

МАТЕМАТИТЧНЕ МОДЕЛЮВАННЯ НАДІЙНОСТІ ТЕНЗОМЕТРИЧНИХ СИСТЕМ НА ОСНОВІ ЕВРИСТИЧНИХ МОДЕЛЕЙ ВИМІРЮВАЛЬНИХ ПРОЦЕДУР

С.А. Положаєнко, Ф.Г. Гаращенко, Л.Л. Прокоф’єва

Національний університет «Одеська політехніка»
пр.-т Шевченка, 1, Одеса, 65022, Україна; e-mail: sanp277@gmail.com

Тенденція зростання складності апаратних засобів систем вимірювання залишається сталою у зв’язку з масовим застосуванням засобів обчислювальної техніки в процесах вимірювання. Надлишкова складність нових засобів, висока вартість комплектуючих та програм і достатньо низький рівень якості виробництва не дозволяють виключити можливість виникнення похибок, які спричиняють порушення працездатності засобів в цілому, а також зниження їх продуктивності. Термін «надійність засобів», аналогічно терміну «надійність апаратури» в задачах діагностики тензометричної апаратури, означає, що «відмови», в даному випадку (мається на увазі наявність в складі систем вимірювань крім суто апаратних засобів, ще і програмних) , як результат появи похибок, має якісно відмінну фізичну природу, ніж відмови суто апаратури. Але це свідчить про можливість використання певних термінів та показників надійності технічних засобів при дослідженні якості. Зокрема, це виправдовується необхідністю розв’язування задачі розподілу ресурсів (або витрат) власно між апаратними та програмними засобами при забезпеченні заданого показника надійності систем вимірювань. Перевірка правильності функціонування апаратних та програмних засобів, що входять до складу систем вимірювань, здійснюється на етапі налаштування та тестування. Як правило, основним фактором налаштування є витрачений на нього час. Тому в низці моделей оцінювання надійності систем вимірювань, поряд з необхідним часом їх функціонування в штатних режимах (власно реалізації процесу вимірювання), необхідно розглядати ще і інший часовий фактор — час налаштування апаратних та програмних засобів для використання їх за призначенням. Дієвим шляхом визначення надійності тензометричних систем, а особливо на етапі проєктування, є застосування математичного моделювання, результати якого визначаються коректністю покладених в його основу моделей вимірювальних процедур.

Ключові слова: тензометрична система, вимірювальна процедура, надійність апаратних засобів, математичне моделювання, евристична модель.

Вступ

Для зручності аналізу показників надійності складних СВ доцільно представити їх у вигляді сукупності менш складних складових, які, за звичай [1 — 3], називаються модулями. Ці модулі, в свою чергу, можуть бути поділеними на більш дрібні частини, тощо. Таким чином, апаратні (АМ) або програмні (ПМ) модулі являють собою розрахункові елементи у випадку визначенні надійності систем вимірювань (СВ).

При дослідженні надійності функціонування СВ, за звичай, необхідно розв’язувати дві задачі. Перша — за заданою структурою СВ, яка утворена певною сукупністю АМ або ПМ, що мають власні показники надійності, віднайти показник надійності СВ загалом. Цю задачу будемо називати прямою. Друга задача — досягнення максимального (можливого) значення показника надійності при обмеженнях на ресурси, в якості яких виступають: час, вартість, масо-габаритні характеристики, тощо. Або може розв’язуватися задача мінімізації

величини обмеження при досягненні заданого значення показника надійності. Будь-яку з цих задач будемо називати зворотною.

Огляд досліджень

Наявні літературні джерела, присвячені організації та виконанню залізничних перевезень, в питаннях щодо прогнозування стану вантажів при здійсненні технологічних операцій транспортування, зосереджують увагу на захисті [1 — 5] та забезпеченні бажаних умов зберігання вантажів [6 — 9]. Контролю поточного стану вантажів за допомогою датчиків, розташованих у вагонах, присвячено, зокрема, роботи [10, 11]. Наразі не виявлено робіт, в яких піднімаються питання математичного моделювання стану вантажів та нетяглового РС, в яких вони перевозяться

Мета роботи

Мета статті полягає у формалізації підходів щодо визначення надійності тензометричних (вимірювальних) систем на основі застосування моделей вимірювальних процедур, що враховують експертні оцінки (евристичних моделей). **Постановка задачі**. Нехай розглядається певна СВ, яка складається з М окремих модулів, що поєднані між собою. По структурі СВ формується стохастичний граф, що має М+2 вершини. Вершина 0 означає «виток», а вершина М+1 — «стік» графу. Кожний АМ або ПМ викликається на розв'язок з заданою вірогідністю, виходячи з мети функціонування або значень вихідних даних.

Розглянемо, для конкретності подальших розмірковувань, випадки, коли пряма задача полягає у вірогідності безпохибкового відшукання розв'язку задачі вимірювання апаратними та програмними засобами СВ, якщо відомі вірогідності безпохибкових розв'язків задач всіх апаратних (АЗ) та програмних (ПЗ) засобів. Крім того, зворотна задача полягає у відшуканні максимуму вірогідності безпохибкового розв'язку задачі вимірювання при обмеженнях на загальний час налаштування всіх модулів, а також у визначенні мінімального часу налаштування всієї СВ при заданій вірогідності її безпохибкового функціонування.

Пряма задача. Для визначення вірогідності безпохибкового розв'язку задачі вимірювання скористаємося методом розрахунку ймовірносно-часових характеристик перебування заявки в мережі масового обслуговування (МО) [4], яку використаємо в якості моделі СВ. Одначе, безпосередньо в тому вигляді, як описано зазначений метод в [4], його використання не є коректним, оскільки в мережі МО час перебування заявок в модулях підсумовується. В задачі вимірювання повинен виконуватися **принцип слабкої ланки**, який властивий основному поєднанню елементів в теорії надійності. Тому невірно застосовувати перетворення Лапласа щільностей розподілу часу до похибок у модулях. Необхідно замість них поставити вірогідності правильної роботи відповідних модулів, які здійснюють обчислення на заданих часових інтервалах. При цьому будемо брати до уваги, що модулі статистично незалежні.

$$\mathbf{G} = \mathbf{G}(t), \quad t = (t_0, t_1, ..., t_{M+1}),$$

елементами якої є добутки

$$p_{ij}\ P_i(t_i), \quad i, j = (0, 1, ..., M+1),$$

де p_{ij} — вірогідність переходу від i-го АМ (або ПМ) до j-го АМ (або ПМ); $P_i(t_i)$ — вірогідність безпохибкового функціонування i-го АМ (або ПМ) на протязі часу t_i.

Оскільки як 0-а та М+1 вершини графу фіктивні, то приймаємо, що час перебування в них дорівнює нулю, а вірогідність безпохибкової роботи — одиниці.

Введемо поняття «кроку», маючи на увазі під ним перехід від одного АМ (або ПМ) до іншого. Щоб віднайти вірогідність безпохибкової роботи за два кроки, необхідно підсумувати, з відповідними ймовірностями, добутки ймовірностей по всіх шляхах, які вміщують дві вершини (одна з них — нульова). Це досягається підведенням матриці $\mathbf{G}$ у квадрат. При підведенні $\mathbf{G}$ у куб отримуємо вірогідності безпохибкового функціонування за три кроки і так далі.

Побудуємо матрицю

$$\mathbf{T} = \mathbf{I} + \mathbf{G}(t) + \mathbf{G}^2(t) + ... = \mathbf{I}(\mathbf{I} - \mathbf{G}(t))^{-1}, \quad (1)$$

де $\mathbf{I}$ — одинична матриця.

Елемент матриці $\mathbf{T}$ з номером $(0, M+1)$ являє собою вираз для вірогідності безпохибкової роботи всієї СВ з урахуванням всіх можливих послідовностей дії АМ та ПМ. У відповідності до правил обчислення елементів зворотної матриці (наприклад, [5]), вираз для вірогідності безпохибкової роботи СВ можна представити у вигляді

$$\mathbf{Y}(t) = \mathbf{Q}(t)/\mathbf{R}(t), \quad (2)$$

де $\mathbf{Q}(t)$ — алгебраїчне доповнення елемента з номером (M+1, 0) матриці $(\mathbf{I} - \mathbf{G}(t))$; $\mathbf{R}(t)$ — головний визначник матриці $(\mathbf{I} - \mathbf{G}(t))$.

Виконавши вказані перетворення, можна отримати шуканий вираз для вірогідності безпохибкового функціонування СВ з урахуванням задіяння всіх можливих маршрутів обчислень.

Зворотна задача. Визначимо мінімальний час налаштування СВ при заданій вірогідності її безпохибкового функціонування $P_{\text{зад}}$. Процес налаштування *i*-го АМ (або ПМ) визначається часом його налаштування τ_i. У відомих аналітичних та емпіричних моделях оцінювання надійності АМ та ПМ (зокрема, в [2]) параметр τ_i, як і параметр t_i — заданий час роботи *i*-го АМ (або ПМ), входять у відповідний вираз для показника надійності СВ. При цьому приймається, що оцінки шуканих показників є детермінованими відомими виразами, що визначаються по результатах випробовувань. Подібних моделей існує певна множина, а тому розглядати їх всі недоцільно. В якості прикладу, не порушуючи узагальнення підходу, наведемо лише одну з самих початкових моделей — модель Муси [3]. Вірогідність безпохибкової роботи i-го модуля, відповідно до даної моделі, можна представити наступної формулою

$$P_i(t_i, \tau_i) = \exp(-\lambda_i t_i), \quad (3)$$

де λ_i — інтенсивність появи похибки; t_i та τ_i — відповідно, час обчислень та налаштування i-го модуля. За звичай $\lambda_i = 1/T_i$, T_i — початковий середній час безпохибкової роботи модуля; $\tau_i = K_i/(N_{\text{пох}_i} T_i)$, K_i — коефіцієнт стискання часу налаштування (тестування) у порівнянні з часом вимірювань; $N_{\text{пох}_i}$ — первинне (уявне) число похибок в модулі.

Знайдемо мінімальний час налаштування СВ

$$\tau = \sum_{i=0}^{M+1} \tau_i,$$

при якому

$$P(t, \tau) \geq P_{\text{зад}},$$

беручи до уваги, що для АМ (або ПМ) виконується (3). Розв'язок отримаємо на основі методу невизначених множників Лагранжа [6]. Запишемо функцію Лагранжа

$$F(t,\tau,\gamma)=\sum_{i=0}^{M+1}\tau_i+\gamma\left[P(t,\tau)-P_{\text{зад}}\right], \tag{4}$$

де γ — множник Лагранжа.

Тоді, диференціюючи (4) по аргументах τ_i та γ і, прирівнюючи отримані вирази до нуля, отримаємо систему рівнянь

$$\begin{cases}\dfrac{\partial F(t,\tau,\gamma)}{\partial\tau_i}=0,\\ P(t,\tau)-P_{\text{зад}}=0.\end{cases} \tag{5}$$

Розв'язуючи (5) відносно $\tau_i=\tau_i^0$, отримаємо

$$\tau^0=\sum_{i=0}^{M+1}\tau_i^0\ .$$

Визначимо максимальне значення вірогідності безпохибкового функціонування СВ при заданому часі його налаштування. Будемо шукати максимум функції $P(t,\tau)$ при заданому часі налаштування $\tau_{\text{зад}}$. Функція Лагранжа в цьому випадку має вигляд

$$F(t,\tau,\gamma)=P(t,\tau)+\gamma\left(\sum_{i=0}^{M+1}\tau_i-\tau_{\text{зад}}\right). \tag{6}$$

Диференціюючи (6) по аргументах τ_i та γ і, прирівнюючи до нуля, отримаємо систему рівнянь

$$\begin{cases}\dfrac{\partial F(t,\tau,\gamma)}{\partial\tau_i}=0,\\ \sum\limits_{i=0}^{M+1}\tau_i-\tau_{\text{зад}}=0.\end{cases} \tag{7}$$

Розв'язуючи (7), знаходимо шукані значення τ_i^0, а підставляючи їх у вираз для $P(t,\tau)$, віднайдемо максимальне значення вірогідності безпохибкового функціонування СВ.

Отримані теоретичні результати проілюструємо декількома прикладами.

Приклад 1. Необхідно знайти значення вірогідності безпохибкового функціонування СВ, що складається з трьох модулів, для якої задано наступні характеристики

$$K_i=1;\ t_1 1c; t_2=7c;\ t_3=10c; N_{\text{пох}_1}=10; N_{\text{пох}_2}=5; N_{\text{пох}_3}=3;$$
$$\lambda_1=0{,}01\,1/c;\lambda_2=0{,}02\,1/c;\lambda_1=0{,}03\,1/c; p_{01}=1; p_{12}=0{,}7; p_{13}=0{,}3;$$
$$p_{23}=0{,}6; p_{24}=0{,}4; p_{31}=0{,}8; p_{32}=0; p_{34}=0{,}2\cdot P_i(t_i,\tau_i).$$

Стохастичний граф наведеної СВ має 5 вершин, з яких нульова та четверта — фіктивні, а три інші — попарно зв'язані між собою. Тобто, в даному випадку М=3. Матриці $\mathbf{G}$ та $(\mathbf{I}-\mathbf{G})$ мають вигляд

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & p_{12}P_1 & p_{13}P_1 & 0 \\ 0 & 0 & 0 & p_{23}P_2 & p_{24}P_2 \\ 0 & p_{31}P_3 & 0 & 0 & p_{34}P_3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$(\mathbf{I}-\mathbf{G}) = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -p_{12}P_1 & -p_{13}P_1 & 0 \\ 0 & 0 & 1 & -p_{23}P_{21} & -p_{24}P_2 \\ 0 & -p_{31}P_3 & 0 & 1 & -p_{34}P_3 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Елемент матриці $(\mathbf{I}-\mathbf{G})^{-1}$ з номером (0, 4) $\mathbf{Y}=\mathbf{Q}/\mathbf{R}$, де $\mathbf{Q}$ — алгебраїчне доповнення елемента (4, 0) матриці $(I-G)$. Розкриваючи вказані визначники, отримаємо

$$Y = P(t,\tau) = \frac{p_{12}P_1 \cdot p_{23}P_2 \cdot p_{34}P_3 + p_{12}P_1 \cdot p_{24}P_2 + p_{13}P_1 \cdot p_{34}P_3}{1 - p_{13}P_1 \cdot p_{31}P_3 - p_{12}P_1 \cdot p_{23}P_2 \cdot p_{31}P_3}.$$

(8)

В (7) аргументи t_i та τ_i для кратності опущені. Підставляючи вихідні дані в (8) та поклавши $\tau_{\text{ш}} = 0$, отримаємо $P(t,0) = 0{,}563$.

Приклад 2. Нехай задано значення вірогідності для СВ $P_{\text{зад}} \geq 0{,}9$. Необхідно віднайти час налаштувань кожного модуля (АМ та ПМ), а також сумарний мінімальний час налаштування СВ.

Вчиняючи, як було описано вище для зворотної задачі, віднайдемо

$$\tau_1^0 = 3460\,c;\ \tau_2^0 = 1871\,c;\ \tau_3^0 = 847\,c;\ \tau^0 = 6178\,c.$$

Приклад 3. Нехай задано час налаштування СВ $\tau_{\text{зад}} = 6000\,c$. Необхідно віднайти час налаштування кожного модуля τ_i^0 та максимальне значення вірогідності його безпохибкового функціонування при вихідних даних, що наведено в прикладі 1.

В даному випадку отримаємо

$$\tau_1 = 3327\,c;\ \tau_2 = 1738\,c;\ \tau_3 = 835\,c;\ \max P(t,\tau) = 0{,}98.$$

Наведені числові приклади ілюструють можливість використання розглянутого методу при розв'язуванні прикладних задач, які пов'язано з аналізом надійності апаратних та програмних модулів систем вимірювання, а також забезпечення необхідних показників їх надійності при налаштуванні.

В практиці визначення надійності (або, що те саме — діагностування технічного стану), зокрема, тензометричних систем, не завжди можна застосувати точні моделі, які було розглянуто вище. Це пов'язано з відсутністю повної вихідної інформації щодо виміру параметрів, зокрема, АЗСВ, часу та умов експлуатації, якості проведених технічних оглядів та ремонтів. Тому в прикладних дослідженнях цілком допустимо використання **евристичних моделей** вимірювальних процедур, Особливості застосування останніх полягає у наступному.

Діагностування являє собою одну з найбільш інтелектуальних процедур в процесі експлуатації складних технічних об'єктів (СТО). Одначе, при створенні об'єктів діагностування (ОД) знання щодо них, які вкладаються в алгоритми, програми АЗ діагностування, часто виявляються недостатніми для забезпечення необхідного рівня готовності ОД в процесі його експлуатації. Це відноситься до таких ОД, як, наприклад, автономні вимірювальні системи. Такі ОД являють

собою складні динамічні системи зі значною структурою, часовою та функціональною надлишковістю, та такі, що складаються з великої кількості елементів з різноманітними принципами дії, режимами роботи, процедурами обслуговування і умовами експлуатації. Процеси деградації в елементах таких СТО мають різноманітні закономірності і часто недостатньо вивчені. Досить проблематичною є установка необхідних датчиків щодо забезпечення задачі діагностування (навіть, якщо вони існують) на ряді елементів та організація інтерфейсу для передачі діагностичної інформації. Зазначене вище зумовлює обмеженість вихідної бази знань (БЗ) системи діагностування та призводить до зниження рівня достовірності рішень, що приймаються, про актуальні та прогнозні технічні стани ОД. Очевидно, що для таких ОД досить проблематично створити точні моделі, але, з достатньою для практики точністю, можна застосувати евристичні моделі, які відбивають найважливіші особливості відповідних ОД (в тому числі тензометричних систем, які розглядаються у чинній роботі). Розглянемо отримання евристичних моделей для діагностики тензометричних (вимірювальних) систем.

В загальній теорії вимірювань [7] під **вимірювальною процедурою** (ВП) розуміється операція порівняння об'єктів за деякими ознаками, що вміщує в собі визначення відношень між об'єктами та спосіб їх порівняння. При цьому під об'єктом мається на увазі рівень зумовленості (інтенсивності) тої або іншої властивості (якості).

Нехай є множина S рівнів зумовленості певної діагностичної ознаки, і на цьому рівні існує множина відносин V, наприклад, відносин домінування. Введемо множину L певних елементів (термінів, символів, найменувань) та множину W відношень на ньому, наприклад, відношень порядку,, а також однозначне відображення g елементів множини S на множину L. Сукупність процедур формування вказаних множин, а також відображення g і являють собою процедуру вимірювання, а кортеж $\{S, V, L, W, g\}$ при цьому виступає в якості «шкали».

ВП, яка нами розглядається, характерна тим, що множини S, V, L, W формуються на підставі експертної оцінки (тобто евристично), на підставі якої (оцінки) реалізується відображення g. Всі ці операції, завдяки вказаній особливості процедури, є по суті евристичними, а тому ВП, що розглядаються, слід вважати *евристичними ВП* (ЕВП).

Як і інші ВП, ЕВП можуть здійснюватися в різних «шкалах». Вибір «шкали» в кожному конкретному випадку зумовлено конкретною процедурою. Відмінності між «шкалами» визначаються припустимим перетворенням g, яке встановлює зв'язок між всіма парами $\{S, V\}$, які вибрано для опису пар $\{L, W\}$. Потенційно ЕВП можуть здійснюватися у будь-якій з трьох відомих в теорії вимірювань шкал: номінальній, порядковій, інтервальній.

Нехай S — множина виділених значень певного діагностичного показника (ДП), а V — множина відносин домінування в ньому. Нехай також L — множина певних термінів, що змістовно визначають значення ДП в прийнятій мові, а W — множина відношень еквівалентності на множині L. Задача вимірювання складається у приписуванні кожному $s_i \in S, i = \overline{1, n}$ певного терміналу $l_j \in L,\ j = \overline{1, m}$. Така процедура являє собою процедуру вимірювання (порівняння) в номінальній шкалі. Її особливість, в нашій уяві, має ту відмінність, що вона реалізується за експертною оцінкою, в ході визначення надійності ОД на

підставі визначення працездатності останнього. В даному випадку задача полягає в отриманні певної *агрегативної оцінки* значень ДП, яка, в певному сенсі, найкращим чином узгоджується з експертною оцінкою.

Будемо вважати, що термін $l_j \in L,\ j = \overline{1,m}$ є чинним для значення ДП $s_i \in S, i = \overline{1,n}$, і, що s_i та l_j пов'язані відношенням еквівалентності g_{ij}. Оцінка g_{ijk}, яку отримано при k-му дослідженні працездатності ОД, з певною вірогідністю p_k визначає істинне відношення g_{ij} між рівнем ДП s_i, що спостерігається, та відповідним йому терміном (позначенням) — l_j. Внаслідок можливих похибок експертних оцінок отримані значення g_{ijk} можуть, у загальному випадку, не співпадати з g_{ij}. Необхідно побудувати таку процедуру агрегатування оцінок g_{ijk}, яка дозволяє мінімізувати неспівпадіння агрегатованої оцінки $\overline{g}_{ij}$ з g_{ij}.

Отримання оцінок g_{ijk} здійснюється наступним чином. Будемо вважати, що $g_{ijk} = 1$, якщо k-та оцінка підтверджує значення ДП, яке ця оцінка визначає, терміном l, та $g_{ijk} = 0$ — у відмінному випадку. Агрегативну оцінку $\overline{g}_{ij}$ значення ДП сформуємо відповідно до правила: $\overline{g}_{ij} = \alpha\, g_{ijk}$, а прийняття рішення про те, що виміряне значення ДП оцінюється терміном l, здійснимо у відповідності з умовою $\max \overline{g}_{ij} \to z$, де z — значення порогу, яке варіюється. Можна показати, що максимальна вірогідність співпадіння l_j з істинним значенням ДП досягається при $z = 0{,}5$.

Можливість похибок в експертних оцінках викликає необхідність аналізу їх узгодженості, яка, в свою чергу, є мірою достовірності, яку отримано в результаті проведення ЕВП інформації. Оскільки процедура, що розглядається, являє собою різновид процедури *ранжування*, то для оцінки узгодженості тут доцільно застосовувати відомий в теорії порядкових статистик [8] математичний апарат.

Розглянемо іншу організацію тієї ж процедури, яка відрізняється тим, що кожний її учасник упорядковує використані терміни множини L у відповідності до того, наскільки адекватно вони характеризують значення ДП. Формальна відмінність цієї процедури полягає у порядковій шкалі та у тому, що між термінами множини L можливі відношення еквівалентності. Це означає, що значення ДП, яке визначається k-ою експертною оцінкою, може бути в одному акті вимірювання визначене в різних термінах з L. При цьому можуть бути використані технології безпосереднього упорядкування та парних порівнянь.

Область застосування ЕВП в номінальній та порядковій шкалах — це, в основному, безпосереднє визначення типу та місця локалізації дефекту (постановка діагнозу), хоча вони застосовуються, очевидно, і при вимірюванні інтенсивності прояву того або іншого ДП. Одначе часто, для інтегрування у відповідну агрегативну систему підтримки прийняття рішень інформацію, що надходить в результаті ЕВП, бажано отримувати в інтервальній шкалі. Це пов'язано з тим, що саме в таких шкалах представляється інформація , яка отримується від вимірювальних датчиків (приладів).

Розробка ЕВП в інтервальних шкалах, як і в попередніх випадках, потребує формалізації операцій отримання первісної інформації у формі експертних

(евристичних) оцінок. Для підвищення достовірності і точності цієї інформації експертні оцінки повинні формуватися в реальних умовах експлуатації ОД.

Основний принцип, що реалізується у запропонованих далі процедурах, полягає у наступному. Значення ДП, що оцінюються, розглядаються як випадкові величини, вичерпною характеристикою яких є закони їх розподілу. Конкретний вигляд закону визначається конкретними ж фізичними особливостями процесів деградації, а параметри, за звичай, оцінюються на основі статистичних даних. Недостатню кількість статистичних даних можна компенсувати за допомогою введення вірогідносних показників, наприклад,

$$p = \sum_{k=1}^{K} p_k, \quad p_k = \frac{n + \left(1 - g_{ijk}\right)}{nm\left(n+1\right)}. \tag{8}$$

Найбільш часто значення ДП описуються нормальним розподілом. Для оцінки параметрів такого розподілу можна застосувати два підходи. Перший засновано на упорядкуванні членами ГС інтервалів можливих значень ДП по вірогідності потрапляння в них значення, яке спостерігається, з використанням процедур ранжування (впорядкування). Вірогідність потрапляння значень ДП в v-й інтервал визначається величиною

$$p_v = \frac{\sum_{k=1}^{K}\left(1 - g_{ijk}\right)}{r_v},$$

де r_v — ранг v-го інтервалу значень ДП в упорядкуванні k-ї експертної оцінки, K — загальне число висунутих експертних оцінок.

Отримане значення можна розглядати як частотність потрапляння ДП в v-й інтервал, тобто як ординату гістограми випадкової величини. Цю гістограму може бути згладжено відповідними неперервними розподілами, для яких стандартними способами отримано оцінки параметрів [9].

Другий підхід щодо оцінки параметрів розподілу засновано на завданні квантилів цього розподілу. Кожна експертна оцінка вказує довірчий інтервал $[a, b]$, в якому перебувають значення параметра ДП, що вимірюється, а також величину α, що характеризує ступінь «підтвердження» цього даною експертною оцінкою. Вважаючи a та b квантилями, які дорівнюють, відповідно

$$a = \frac{1+\alpha}{2} \cdot 100\%; \quad b = \frac{1-\alpha}{2} \cdot 100\%, \tag{9}$$

параметри законів розподілу можна тоді визначити за допомогою відомих [3] методів математичної статистики.

Отримана розглянутим способом інформація може розглядатися як важлива складова бази знань системи підтримки прийняття рішень при визначенні технічного стану складних тензометричних (вимірювальних) систем в процесі їх експлуатації.

Висновки

Запропонований метод моделювання надійності тензометричних систем дозволяє виконувати оцінку надійності складних СВ при відомих показниках надійності складових модулів та їх ймовірнісному зв'язку у стохастичному графі.

Існує можливість отримувати часові оцінки для вибору оптимальних величин часу налаштування складових модулів СВ в цілому. При розв'язанні задачі забезпечення заданої надійності СВ нескладно врахувати вплив на результуючу надійність відповідні показники складових модулів СВ за допомогою коефіцієнтів K_i.

Метод допускає застосування будь-яких моделей «росту надійності», які можуть бути отримані теоретично або на основі експериментальних даних.

Крім того, застосування евристичних моделей при визначенні надійності тензометричних систем дозволяє зняти проблему їх «складності», в сенсі деталізації моделей, та обрати лише найбільш вагомі показники, що визначають результуючу надійність ОД.

Список літератури

1. Липаев В.В. Качество программного обеспечения. М.: Финансы и статистика, 1983. 261 с.
2. Карповский Е.Я., Чижов С.А. Надежность программной продукции. К.: Техніка, 1990. 160 с.
3. Баглюк С.И., Мальцев М. Г., Смагин В. А., Филимонихин Г. В. Надежность функционирования программного обеспечения. СПб.: Судостроение, 1991. 278 с.
4. Смагин В.А., Бубнов В.П., Филимонихин Г.В. Расчет вероятностно-ременных характеристик пребывания задач в сетевой модели массового обслуживания. *Изв. ВУЗов*. 1989. Т. XXXII, № 2. С. 47 – 59.
5. Гантмахер Ф. Р. Теория матриц. М.: Наука, 1988. 548 с.
6. Смирнов В. И. Курс высшей математики. М.: Высшая школа, 2001. 327 с.
7. Берка К. Измерения – понятия, теория, проблемы. М: Прогресс, 1987. 263 с.
8. Вирьянский З.Я., Калявин В.П. Использование информации операторов в системах диагностирования транспортных средств. *Труды академии транспорта*, 1994. Вып. 1. С. 46 – 53.
9. Корн Г., Корн Т. Справочник по математике. М.: Наука, 1977. 831 с.

MATHEMATICAL MODELING OF THE RELIABILITY OF TENSOMETRIC SYSTEMS BASED ON HEURESTIC MODELS OF MEASUREMENT PROCEDURES

S.A. Polozhaenko, F.G. Garaschenko, L.L.Prokofieva

National Odesa Polytechnic University
Shevchenko ave., 1, Odesa, Ukraine, e-mail: sanp277@gmail.com

The trend of increasing the complexity and hardware of measurement systems remains constant in connection with the massive use of computer technology in measurement processes. The excessive complexity of the newly created AZSV, the high cost of components and software, and the sufficiently low level of production quality do not allow us to rule out the possibility of errors, which cause a violation of the AZSV's performance as a whole, as well as a decrease in their productivity. The term "reliability of the AZSV, similar to the term "reliability of the equipment" in the tasks of diagnostics of strain gauge equipment, means that "failures", in this case (it means the presence in the composition of the SV, in addition to purely hardware, as well as software), as a result of the appearance of errors, has a qualitatively different physical nature than purely AZ failures. This indicates the possibility of using certain terms and indicators of the reliability of technical means in the study of the quality of AZSV. In particular, this is justified by the need to solve the problem of resource (or cost) distribution between the AZ and the software (software) while ensuring the given reliability indicator of the JI. Checking the correct functioning of AZ and software, which are part of the JI, is carried out at the stage of configuration and testing. As a rule, the main factor in the adjustment is the time spent on it. Therefore, in a number of models for assessing the reliability of JI, along with the necessary time of their operation in regular modes (the actual implementation of the measurement process), it is necessary to consider another time factor - the time of setting up AZ and PZ in relation to the use of these means as intended. An effective way to determine the reliability of strain gauge systems, and especially at the design stage, is the use of mathematical modeling, the results of which are determined by the correctness of the models of measurement procedures based on it.

Keywords: strain gauge system, measurement procedure, hardware reliability, mathematical modeling, heuristic model.

DOI 10.15276/imms.v12.no4.367
УДК 681.51

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 367-381

МЕТОД СИНТЕЗУ ГІБРИДНИХ СИСТЕМ АВТОМАТИЧНОГО КЕРУВАННЯ ДЛЯ БАГАТОВИМІРНИХ ТЕХНОЛОГІЧНИХ ОБ'ЄКТІВ

А. О. Стопакевич, О.А. Стопакевич

Державний університет інтелектуальних технологій та зв'язку,
1, Кузнечна, Одеса, 65029, stopakevich@gmail.com
Національний університет «Одеська політехніка»,
1, пр. Шевченка, Одеса, 65044, stopakevich@opu.ua

Стаття присвячена розробці методу синтезу багатовимірних гібридних систем автоматичного керування, які включають паралельне функціонування системи децентралізованих одновимірних регуляторів та багатовимірного регулятора виходу, керуючих багатовимірним об'єктом. Проводиться дослідження отриманої системи керування за прямими показниками якості, робастністю та крихкістю. Найбільш істотним результатом роботи є розробка методу настройки лінійно-квадратичного регулятора, який має працювати додатково до децентралізованої системи керування, настройки якої отримані за методом BLT (Biggest Log Modulus). BLT - це відомий метод детюнінгу настройок ПІ-регуляторів, отриманих за правилами Циглера-Нікольсона з метою забезпечення робастності багатовимірної систем керування. Розроблений метод синтезу системи керування не є складним й при його застосуванні не втрачається можливість корекції настройок децентралізованих регуляторів. З застосуванням розробленого методу був проведений синтез нової гібридної системи автоматичного керування ректифікаційною установкою. Аналіз динамічних властивостей розробленої системи керування показав, що застосування паралельно включеного додаткового лінійного-квадратичного регулятора дозволяє забезпечити невелике перерегулювання при керуванні за завданням та високу точність підтримки величин керованих змінних при максимальних збуренням. При цьому не зменшується робастність та система керування залишається некрихкою, тобто є можливість зміни настройок регуляторів без істотної втрати робастності.

Ключові слова: система керування, гібридна, прямі показники, лінійно-квадратичний регулятор, робастність, BLT, ПІ-регулятор, ректифікаційна колона, ректифікаційна установка.

Вступ

Системи автоматичного керування (САК) багатовимірними технологічними об'єктами керування як правило будуються як децентралізовані або централізовані.

Ідея побудови децентралізованої системи керування полягає в тому, що необхідно обрати найменш пов'язані канали керування й за моделями каналів об'єкту керування (ОК) синтезувати одновимірні регулятори [1-4]. Розробка такої системи керування з першого погляду є простою, однак виникають проблеми з забезпеченням якості перехідних процесів в розробленій системі й запасом її стійкості. Джерелом проблеми є те, що в реальній системі канали не є автономними й впливають один на одного, тому фактично кожен децентралізований регулятор керує не процесом, який описується моделлю каналу. Як мінімум це призводить до зниження якості САК. Однак чим більша кількість каналів керування задіяна в ОК, тим ця проблема стає більш гострою, що відображується вже не на зниженні якості, а на втраті стійкості САК. Децентралізовані системи керування хоч й не забезпечують навіть теоретично високої якості керування, тим не менш є зручними для промислового застосування, особливо якщо використовуються стандартні регулятори ПІД-типу. По-перше, визначити настройки регуляторів можна за експериментальним розгінними характеристиками чи апроксимативними моделям

окремих каналів за інженерними правилами, що не вимагає застосування комп'ютерного моделювання й математичних САПР. По-друге, децентралізована система керування не має структурно пов'язаних між собою блоків, тому допускає вимкнення окремих регуляторів якщо вони розбалансовують ОК, а регулятори ПІД-типу дозволяють змінювати параметри чисто емпірично для досягнення задовільного перехідного процесу за каналом керування. Науковий підхід до синтезу децентралізованих систем вимагає наявності математичної моделі динаміки усього ОК. Перший крок – це аналіз ступеню зв'язності [5-6] каналів керування з метою вибору каналів керування без домінуючих перехресних впливів. Другий крок – корекція настройок регуляторів для врахування невідповідності динаміки окремих каналів й динаміки каналів в зв'язаній системі. Класичним методом корекції настройок є метод BLT [7], який орієнтується для застосування в децентралізованих САК з типовими ПІ-регуляторами. Це метод досліджується й модифікується й зараз [8-11].

Централізовані регулятори синтезуються за цілою моделлю ОК, яка включає всі канали керування. В задачах синтезу таких регуляторів немає проблем, які виникають в децентралізованих системах керування. Тим не менше, виникають деякі інші. Якісне проведення процедури синтезу таких систем вимагає істотно більше часу, більшого професіоналізму інженерів, а часто й коштовних досліджень динамічних властивостей об'єкту керування. В результаті отримується негнучка САК, яка не може бути декомпозована й параметри регулятора не можуть бути змінені емпірично в випадку виникнення незадовільних перехідних процесів в САК. Централізовані регулятори також вимагають значної уваги до точності цифрової реалізації й якості програмного забезпечення, особливо це відноситься до модельно-прогнозуючих регуляторів, які мають виконувати оптимізаційну задачу в режимі он-лайн.

Класичним «проміжним» варіантом побудови САК є варіант «розв'язки», в якому в децентралізовану САК додаються нові блоки, задачею яких є компенсація перехресних впливів. Фактично розрахувати додатково до децентралізованої САК компенсатор й сподіватись на покращення якості керування можна тільки в випадках ОК без істотних перехресних впливів й з достатньо лінійними динамічними характеристиками. В іншому випадку такий компенсатор скоріше за все зменшить запас стійкості САК й дуже ймовірно погіршить якість перехідних процесів. Задача розв'язується оптимізаційним шляхом, в якому необхідно визначити настройки регулятора для нового «розв'язаного ОК» з блоками компенсації, а краще визначити компенсатор й настройки регуляторів шляхом розв'язку оптимізаційної задачі за критерієм мінімізації перехресних впливів.

В роботі [12] запропонована ще одна «проміжна» стратегія, яка перетворює багатовимірний оптимальний регулятор в субоптимальний регулятор, який складається з блоків, додавання кожного з яких вносить відоме покращення в сенсі інтегрального квадратичного критерію. Це дуже спрощує впровадження й відлагодження САК: послідовно вмикаючи блоки можна визначити який з них є проблемним. Однак ця стратегія вимагає відмови від регуляторів ПІД-типу й розв'язку нескладної оптимізаційної задачі.

В цій роботі пропонується ще одна «проміжна» стратегія, яку назвемо гібридною. Ідея цієї стратегії – застосовувати декілька паралельно функціонуючих регуляторів принципово різної структури. При цьому гнучкість САК має бути забезпечена, тобто окремі елементи САК можуть бути відімкнені без втрати стійкості усієї САК.

Мета роботи

Мета роботи – розробити новий метод синтезу САК гібридної структури, в якому базова децентралізована система з ПІ-регуляторами буде вдосконалюватись

шляхом додавання паралельного централізованого регулятора. Базовим методом синтезу децентралізованої системи керування обрано метод BLT.

Задачі роботи

В роботі будуть розв'язані дві задачі: 1) розробка методу синтезу САК гібридної структури; 2) синтез та дослідження показників якості синтезованої САК ректифікаційної установки.

Синтез децентралізованих систем керування за методом BLT

Перший крок – необхідно обрати найбільш автономні канали керування. В більшості випадків для цього достатньо застосовувати матрицю Брістоля (RGA) [5,6], яка розраховується за матрицею статичних коефіцієнтів *M* ОК. В подальшому будемо розглядати випадок, коли всі канали керування є статичними. Метод RGA дозволяє відкинути неробочі конфігурації та вказує на такі конфігурації, які наближаються до автономних (з мінімумом перехрестних впливів). Все це робиться за порівнянням коефіцієнтів результативної матриці з 1 (ідеальний варіант). Правило розрахунку RGA:

$$RGA(M) \; \square \; M \otimes M^{-T},$$

де $\otimes$ – поелементний добуток матриць.

В результаті буде отримана матриця, елементи якої $\lambda_{i,j}$ відображають міру зв'язності (i,j) каналу. Правила вибору наступні:

- бажано обирати канали, у яких $\lambda_{i,j}$ близькі до 1;
- не бажано обирати канали з $\lambda_{i,j}$ близькі до 0, це означає, що не вдається уникнути взаємозв'язків каналів в системі керування;
- заборонено використовувати канали з $\lambda_{i,j}$<0, це означає, що перехресні впливи є домінантними й скоріше за все в номінальному режимі САК буде нестійкою.

За результатами аналізу необхідно сформувати матрицю передавальних функцій (МПФ) ОК *P*, у якій по діагоналі будуть канали керування, які відповідають правилам вибору.

Другий крок – необхідно визначити настройки ПІ-регуляторів за моделями каналів діагоналі МПФ *P* за методом Циглера-Нікольсона[13]. Оригінальний метод передбачає переведення процесу в автоколивальний режим за допомогою П-регулятора й отримання настройок. Однак, якщо модель каналу TF в вигляді передавальної функції (ПФ) відома, то незалежно від порядку настройки ПІ-регулятора можна отримати за допомогою наступного MATLAB-коду.

```
[Gm,Pm,Wcg]=margin(TF);     %Визначення меж стійкості
if Wcg <=0, Wcg=0.01; end   %Щоб не ділити на 0
ku=Gm;                      %Критичний коефіцієнт
pu=2*pi/Wcg;                %інтервал коливань
Kp = ku/2.2; Ti = pu/1.2;
C = pidstd(Kp,Ti);
```

Якщо модель каналу має вид

$$P_{j,j}(s) = \frac{k_o}{T_o \cdot s + 1} e^{-\tau_o \cdot s},$$

то для ПІ регулятора стандартної форми настройки можна отримати за правилом

$$k_{ZNr} \approx \frac{0.9 \cdot T_o}{k_o \cdot \tau}, T_{ZNi} \approx \frac{\tau_o}{0.3}.$$

Третій крок – необхідно визначити фактор *F* й скоректувати за ним отримані настройки всіх ПІ-регуляторів. Для цього формуємо цикл по *F* від 1 до 4.5 з малим кроком (наприклад, 0.05). В циклі обчислюємо настройки всіх регуляторів C_f(s) за формулою

$$k_r = k_{ZNr} / F, T_i = T_{ZNi} \cdot F$$

й далі розраховуємо величину $Lc_{\max}$ (на прикладі ОК 2x2)

$$W(i\omega) = 1 + \det\left[I + P(i\omega) \cdot C(i\omega)\right]$$

$$P(s) = \begin{bmatrix} P_{11}(s) & P_{12}(s) \\ P_{21}(s) & P_{22}(s) \end{bmatrix}, C(s) = \begin{bmatrix} \dfrac{k_{r1}\left(T_{i,1}s+1\right)}{T_{i,1}s} & 0 \\ 0 & \dfrac{k_{r2}\left(T_{i,2}s+1\right)}{T_{i,2}s} \end{bmatrix}$$

Якщо $Lc_{max} \approx 2 \cdot N$(дБ), де N–розмірність ОК.

В MATLAB це реалізується за допомогою наступного коду

```
PC=P*C; PCsym=tf2sym(PC); W=1+det(1+PCsym);
Ssym=W/(1+W); S=sym2tf(Ssym);
[mag,~]=bode(S); mag=mag(1,:);
dB=20*log10(mag); Lcmax=max(dB);
```

Наприклад, для відомої моделі ректифікаційної колони (РК) Wood & Berry [14]

$$P(s) = \begin{bmatrix} \dfrac{12.8e^{-s}}{16.7 \cdot 7s+1} & \dfrac{-18.9e^{-3s}}{21 \cdot s+1} \\ \dfrac{6.6e^{-7s}}{10.9 \cdot s+1} & \dfrac{-19.4e^{-3s}}{14.4 \cdot s+1} \end{bmatrix}, \quad RGA(P(0)) = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Розраховані настройки $C_1(s)$ наступні: $k_{ZNr} = 0.96,\ T_{ZNi} = 3.25$.

Розраховані настройки $C_2(s)$ наступні: $k_{ZNr} = -0.19,\ T_{ZNi} = 9.20$.

Побудований графік залежності $Lc_{\max}$ від F показаний на рис. 1.

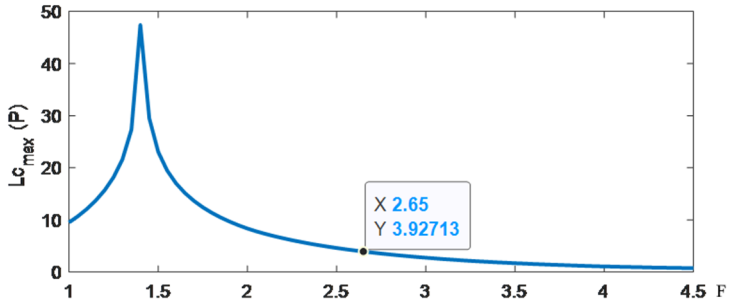


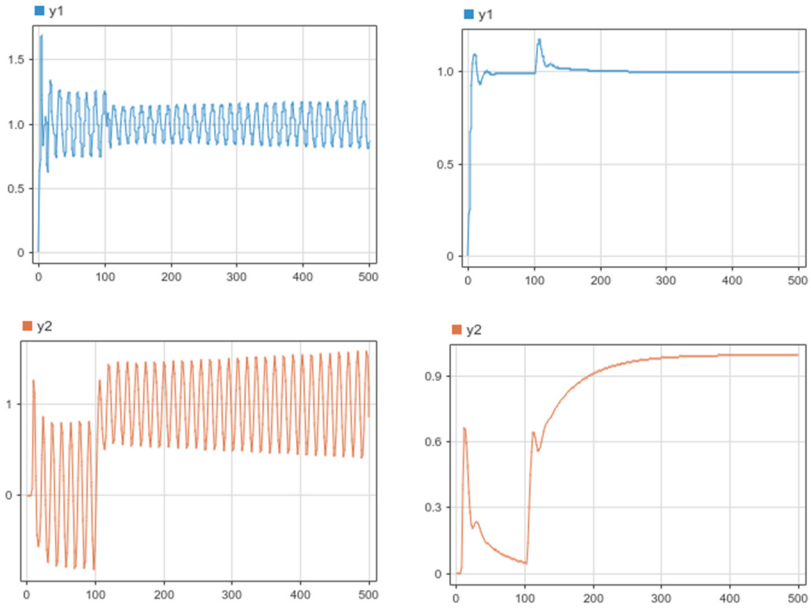
Рис.1. Залежність Lc_{max} від коефіцієнту F

Бачимо, що величина F=2.65, яка відповідає 2·N=4 дБ. Перевіримо, чи дійсно якість керування покращується. Для цього промелюємо САК з номінальними настройками за методом Циглера-Нікольсона та з відкоректованими за коефіцієнтом F.

В моделі використовуються дискретні ПІ-регулятори з кроком 0.1. Спочатку подається завдання в одиницю на перший регулятор, через 100 одиниць часу – на другий.

Бачимо з графіку, що при номінальних настройках (F=1) перехідні процеси незадовільні оскільки дуже коливальні, а й за другою керованою змінною перехідний процес є слабко розбіжним. При корекції настройок коефіцієнтом F перехідні процеси істотно покращуються, хоча й присутня велика амплітуда в другому каналі при зміні завдання першому регулятору, однак в принципі він є задовільним.

Таким чином, використовуючи метод BLT отримується працездатна САК з певним запасом стійкості. Однак прямі показники якості в таких САК далекі від оптимальних, що може приводити до надмірних відхилень керованих змінних й тривальних перехідних процесів. Ідея цієї роботи полягає в тому як покращити прямі показники САК шляхом модернізації, а не заміни вже працюючої в системі автоматизації системи керування, настройки якої отримані за методом BLT.



а) F=1 б) F=2.65

Рис. 2. Перехідні процеси в САК з ПІ регулятором і ОК P

Метод синтезу гібридної САК

Метод полягає в синтезі гібридної багатовимірної (MIMO) САК з цифровим MIMO децентралізованим (САК$_1$) та MIMO цифровим централізованим (САК$_2$) регуляторами. Спочатку синтезується САК$_1$, а потім, якість її функціонування покращується за допомогою синтезу САК$_2$, де об'єктом є САК$_1$. Під покращенням будемо розуміти покращення прямих показників якості САК при максимальних збуреннях і додержання робастності.

Формується два представлення моделі ОК. Перше представленням є аналогова матриця передавальних функцій (МПФ), яку позначимо P_a. Другим представлення є цифрова модель, яку позначимо $P_ц$. Модель $P_ц$ формується в три етапи: 1) запізнення в каналах МПФ РК апроксимується ланкою Паде; 2) МПФ РК перетворюється в простір станів стандартним чином; 3) визначається оптимальний крок дискретності та модель дискретизується. Матриці моделі $P_ц$ позначимо як $\{A_P, B_P, C_P, D_P\}$.

Для синтезу аналогових ПІ регуляторів використовується метод BLT, приклад застосування якого наведений вище. САК$_1$ включає ПІ регулятори, які спочатку синтезуються поканально. Коли в САК$_1$ використано N одновимірних регуляторів $C = diag(C_1, ..., C_N)$ і об'єктом P_a, критерій якості має вигляд $L_c^{\max} = \max_{\omega}\{20 \cdot \log(W/(W+1))\}$, де $W = -1 + \det(I + C \cdot P_a)$. Бажана величина $L_c^{\max} = 2$ дБ$\cdot N$. Для досягнення робастності вводиться коефіцієнт F, який зазвичай становить від 2 до 5 і на який ділиться отриманий за методом ZN коефіцієнт передачі кожного ПІ-регулятора й множиться час ізодрому. Цей коефіцієнт встановлюється емпірично, орієнтуючись на бажану величину $L_c^{\max}$. Вважається,

що метод BLT дає процеси, близькі до оптимальних за критерієм IAE (інтеграл модуля похибки керування).

Для подальшого синтезу MIMO модель децентралізованого аналогового регулятора дискретизується звичайним чином та отримується MIMO система цифрових ПІ регуляторів S_0 $u_i = C_{PI} \cdot w_i + D_{PI} \cdot (r - y_i)$, $w_{i+1} = A_{PI} \cdot w_i + B_{PI}(r - y_i)$, де r – вектор завдань.

Для САК з централізованим регулятором виберемо цифровий лінійно-квадратичний регулятор (ЛКР) з наглядачем стану [15]. Наш досвід показав, що застосування ЛКР з інтегральною складовою для керування САК з ПІ-регуляторами призводить до проблем зі зниженням запасу стійкості систем керування, оскільки в динаміці САК фактично виникає подвійний інтегратор. Тому включати інтегратор в структуру ЛКР не будемо.

Керуючий вплив ЛКР регулятора в гібридній системі додається до керуючого впливу ПІ-регуляторів. Тому матриці об’єкта для ЛКР регулятора мають вигляд

$$Az_2 = \begin{bmatrix} A_{PK} - B_{PK} \cdot D_{PI} \cdot C_{PK} & B_{PK} \cdot C_{PI} \\ -B_{PI} \cdot C_{PK} & A_{PI} \end{bmatrix}, Bz_2 = \begin{bmatrix} B_{PK} \\ 0 \end{bmatrix}, Cz_2 = \begin{bmatrix} C_{PK} & 0 \end{bmatrix}.$$

Матриці цифрового ЛКР регулятора з включеним наглядачем стану мають вигляд

$$Ar_2 = Az_2 - Bz_2 \cdot K_2 - L_2 \cdot Cz_2, \quad Br_2 = L_2, \quad Cr_2 = K_2, Dr_2 = 0,$$

де матриці K_2 і L_2 розраховуються за допомогою Matlab

$$K_2 = dlqr(Az_2, Bz_2, Q_{k2}, R_{k2}), \quad L_2 = dqlr(Az_2', Cz_2', Q_{L2}, R_{L2})'.$$

Вагові матриці визначаються за правилом $Q_{k2} = Q_{z2} \cdot Cz_2' \cdot Cz_2$, $R_{k2} = R_{z2} \cdot I$, де Q_{z2}, R_{z2} зазвичай одиничні, але можуть бути при необхідності змінені. Вагові матриці наглядача повного порядку обираються одиничними $Q_{L2} = I, R_{L2} = I$.

Якість гібридної системи оцінюється за комплексом таких критеріїв: 1) прямі показники якості при керуванні за завданням; 2) прямі показники якості при максимальних збуреннях; 3) величина зміни прямих показників якості при зміні параметрів моделі об’єкта (робастність САК); 4) величина зміни прямих показників якості при зміні настройок ПІ регуляторів (крихкість САК).

Приклад синтезу гібридної САК за запропонованим методом

Розробимо нову гібридну САК ректифікаційної установки (РУ) хімічної промисловості, яка включає дві ректифікаційні колони (РК) і розділяє три компоненти: бензол, толуол, м-ксілен, модель РУ та задача її керування розглянуті в роботі [16]. В цій РУ перша РК виступає в ролі префракціонатора, а друга РК має три продуктові потоки. Колони пов’язані між собою значною кількістю рециркуляційних потоків та допоміжних теплообмінників. Технологічна схема РУ показана на рис. 3.

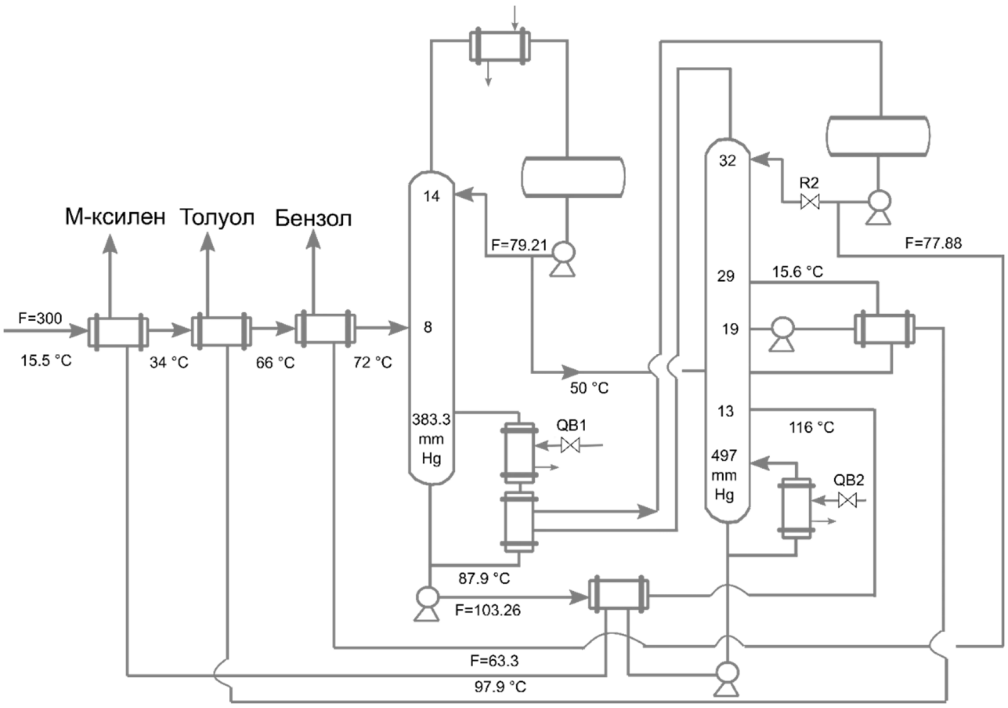


Рис. 3. Технологічна схема розглянутої РУ [16]

Входи моделі: витрати пари в ребойлер РК №1 (QB1), витрата флегми в РК №2 (R2), витрата бокового потоку РК №2 (S), витрата пари в ребойлер РК №2 (QB2).

Виходи моделі: молярна концентрація бензолу в кубі РК №1 XB1, молярна концентрація бензолу в дистиляті РК №2 XD2, молярна концентрація ксилену боковому продукті XS2, молярна концентрація ксилену в кубовому продукті РК №3 XB2.

Математична модель РУ в вигляді МПФ має наступний вигляд [16]:

$$P_a = \begin{bmatrix} \frac{-7.39\cdot e^{-s}}{(11\cdot s+)\cdot(s+1)} & 0 & 0 & 0 \\ \frac{-0.11\cdot(200\cdot s+1)\cdot e^{-5\cdot s}}{(20\cdot s+1)^3} & \frac{10.1\cdot e^{-s}}{(28\cdot s+1)\cdot(4\cdot s+1)} & \frac{1.18\cdot e^{-11\cdot s}}{(31\cdot s+1)\cdot(6\cdot s+1)} & \frac{-18.3\cdot e^{-s}}{(28\cdot s+1)\cdot(5\cdot s+1)} \\ \frac{1.9\cdot e^{-2\cdot s}}{(4\cdot s+1)^3} & \frac{1.7\cdot(200\cdot s+1)\cdot e^{-1.4\cdot s}}{(108\cdot s+1)\cdot(s+1)^2} & \frac{-3.15\cdot e^{-s}}{(3.5\cdot s+)\cdot(0.3\cdot s+1)} & \frac{-1.27\cdot(188\cdot s+1)\cdot e^{-s}}{(68\cdot s+1)\cdot(s+1)} \\ \frac{4.9\cdot e^{-16\cdot s}}{(40\cdot s+1)\cdot(3\cdot s+1)} & \frac{-8.21\cdot e^{-2.5\cdot s}}{(24\cdot s+1)\cdot(3\cdot s+1)} & \frac{12\cdot e^{-1.5\cdot s}}{(29\cdot s+1)\cdot(3\cdot s+1)} & \frac{19.4\cdot e^{-s}}{(26\cdot s+1)\cdot(3\cdot s+1)} \end{bmatrix}$$

Коефіцієнти передачі безрозмірні і розраховані виходячи з 50% номіналу регулюючих органів, 0.2 мольного відсотку діапазону вимірювання давача кубу РК №1 й 1 мольного відсотку діапазону вимірювання для інших давачів.

Хоча більшість каналів моделі доволі точно зводяться до інерційної ланки першого порядку з запізненням (FOPDT), три канали є винятком. Криві розгону в них характеризуються різким стрибком до приблизно у 2 рази більшого коефіцієнту передачі, а потім повільно зменшуються до значення цього коефіцієнту. Два канали з цих трьох мають також домінуюче запізнення.

Математична модель як і попередня передбачає 4 типи збурень: ZF1: бензол – 20%, толуол – 50%, ксилен – 30% (% мольні); ZF2: бензол – 30%. толуол – 50%, ксилен – 20% (% мольні); ZF3: бензол – 30%, толуол – 40%, ксилен – 30% (% мольні); ZF4: бензол – 20%, толуол – 60%, ксилен – 20% (% мольні)

Моделі збурень мають вигляд [16]:

$$P_D = \begin{bmatrix} - \\ \dfrac{2.42 \cdot e^{-5 \cdot s}}{(3 \cdot s + 1) \cdot (26 \cdot s + 1)^2} \\ \dfrac{0.592 \cdot e^{-5 \cdot s}}{(7 \cdot s + 1)^2} \\ \dfrac{-1.51 \cdot e^{-19 \cdot s}}{(45 \cdot s + 1) \cdot (5 \cdot s + 1)^2} \end{bmatrix} \cdot ZF_{1,2} + \begin{bmatrix} - \\ \dfrac{-2.47 \cdot e^{-5 \cdot s}}{(3 \cdot s + 1) \cdot (22 \cdot s + 1)^3} \\ \dfrac{1.83 \cdot e^{-6 \cdot s}}{(25 \cdot s + 1) \cdot (2 \cdot s + 1)} \\ \dfrac{-4.52 \cdot e^{-8 \cdot s}}{(50 \cdot s + 1) \cdot (7 \cdot s + 1)^3} \end{bmatrix} \cdot ZF_{3,4}$$

Синтезуємо САК$_1$ методом BLT. Структурна схема САК$_1$ зі знайденими настройками регуляторів показана на рис. 4

Модель 4 цифрових ПІ регуляторів (Δt=0.4 хв) має вигляд

$$A_{PI} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, B_{PI} = \begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.05 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.05 \end{bmatrix}, C_{PI} = \begin{bmatrix} -0.2143 & 0 & 0 & 0 \\ 0 & 0.2436 & 0 & 0 \\ 0 & 0 & -0.1523 & 0 \\ 0 & 0 & 0 & 0.1305 \end{bmatrix}, D_{PI} = \begin{bmatrix} -0.6 & 0 & 0 & 0 \\ 0 & 0.679 & 0 & 0 \\ 0 & 0 & -0.31 & 0 \\ 0 & 0 & 0 & 0.323 \end{bmatrix}.$$

Синтезуємо нову гібридну САК. Структурна схема моделі гібридної САК з централізованим ЛРК регулятором показана на рис. 5.

Перехідні процеси при одночасній зміні завдань на 1 всім регуляторам показані на рис. 6. В табл. 1 узагальнені результати при зміні як всіх завдань, так і кожного окремо. Використаємо такі критерії для дослідження: σ – перерегулювання (%), ST – час досягнення 95% усталення процесів, максимум за модулем керованої змінною max|y| і керуючого впливу max|u|, IAE.

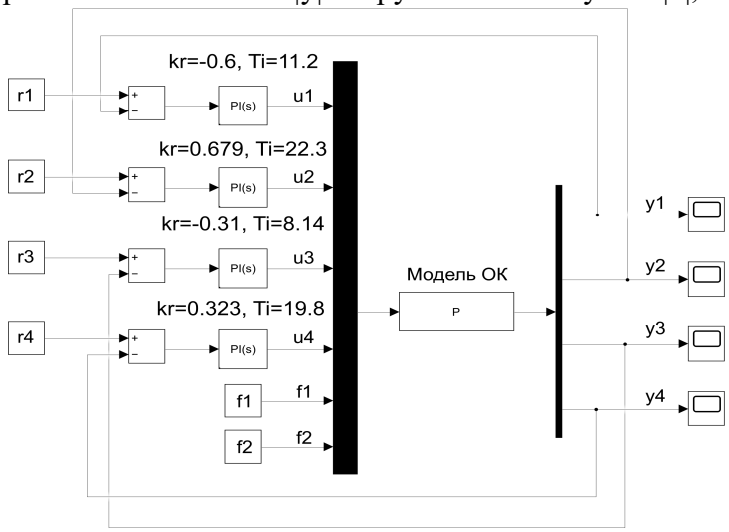


Рис. 4. Структурна схема САК$_1$

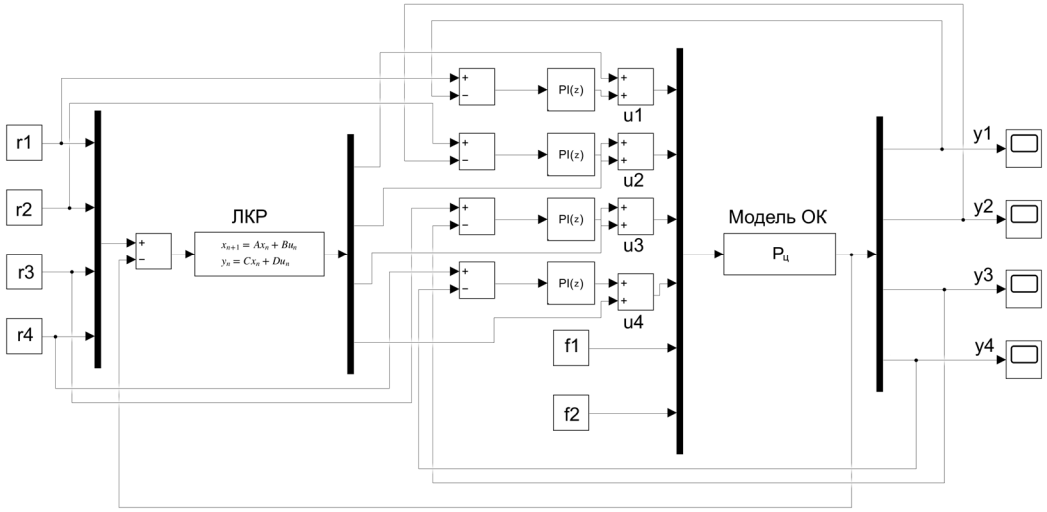


Рис. 5. Структурна схема САК$_2$ з централізованим ЛКР регулятором

Таблиця 1

Показники якості процесів в САК за завданням

Вихід	**Система**	**σ**	**ST**	**max|y|**	**IAE**	**Керування**	**max|u|**
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Зміна завдання усім регуляторам							
y_1	САК$_1$	20.999	10.081	1.21	4.1642	u_1	0.65902
	САК$_2$	9.4839	6.8498	1.0948	3.8588		1.0001
y_2	САК$_1$	14.787	111.15	1.138	41.284	u_2	1.2248
	САК$_2$	0	180.47	0.98634	35.285		1.1875
y_3	САК$_1$	46.506	75.659	1.4636	18.418	u_3	0.49627
	САК$_2$	6.9398	123.44	1.0737	21.064		0.61398
y_4	САК$_1$	14.871	105.84	1.1391	40.442	u_4	0.5991
	САК$_2$	3.0123	119.19	1.0427	53.852		0.57684
Зміна завдання першому регулятору							
y_1	САК$_1$	20.741	10.078	1.2074	4.1436	u_1	0.65148
	САК$_2$	8.1612	6.8025	1.0816	3.4274		0.9721
y_2	САК$_1$	–	119.89	0.38726	12.658	u_2	0.39037
	САК$_2$	–	286.85	0.060512	7.8982		0.2502
y_3	САК$_1$	–	118.87	0.30091	11.205	u_3	0.16149
	САК$_2$	–	198.24	0.13099	5.2903		0.063354
y_4	САК$_1$	–	115.85	0.36546	12.756	u_4	0.19154
	САК$_2$	–	132.3	0.24	12.906		0.1392
Зміна завдання другому регулятору							
y_1	САК$_1$	–	0	0	0	u_1	0
	САК$_2$	–	16.065	0.045435	0.34765		0.071124
y_2	САК$_1$	19.009	74.164	1.189	14.386	u_2	0.72073
	САК$_2$	12.937	29.246	1.1276	9.4575		0.6895
y_3	САК$_1$	–	41.402	1.5483	14.761	u_3	0.59744
	САК$_2$	–	13.94	0.76293	3.8817		0.86396
y_4	САК$_1$	–	135.04	0.24528	8.1704	u_4	0.089625
	САК$_2$	–	172.93	0.054327	1.6425		0.50249
Зміна завдання третьому регулятору							
y_1	САК$_1$	–	0	0	0	u_1	0
	САК$_2$	–	154.77	0.026237	0.73376		0.030363
y_2	САК$_1$	–	114.84	0.72437	26.168	u_2	0.8058
	САК$_2$	–	289.38	0.14569	20.012		0.60198
y_3	САК$_1$	41.843	101.96	1.4177	22.049	u_3	0.39901
	САК$_2$	4.5404	38.505	1.0476	15.792		0.4217
y_4	САК$_1$	–	111.05	0.6631	26.371	u_4	0.40594
	САК$_2$	–	127.64	0.42189	29.505		0.32537
Зміна завдання четвертому регулятору							
y_1	САК$_1$	–	0	0	0	u_1	0
	САК$_2$	–	30.389	0.011901	0.11927		0.020823
y_2	САК$_1$	–	92.265	0.67541	19.121	u_2	0.59341
	САК$_2$	–	97.251	0.16713	4.7113		0.75669
y_3	САК$_1$	–	88.405	0.95131	19.789	u_3	0.35257
	САК$_2$	–	15.736	0.74222	4.3682		0.19746
y_4	САК$_1$	27.356	84.979	1.2725	18.434	u_4	0.34473
	САК$_2$	1.1502	18.432	1.0131	10.278		0.45516

Аналіз результатів, приведених в табл. 1, показує, що в САК$_2$ з гібридним регулятором істотно знижується величина перерегулювання й максимальна амплітуда відхилень. Відомо, що метод BLT звичайно забезпечує перехідні процеси, близькі до оптимуму за IAE критерієм. Дійсно, за цим критерієм при зміні завдань усім регуляторам децентралізованої САК$_1$ показує у цілому кращі результати, однак якщо розглядати керування за завданням по кожному каналу окремо, то ситуація протилежна. Відносно часу регулювання складно виявити яка

з двох типів САК – децентралізована чи гібридна – виявляється кращою, у середньому час керування приблизно однаковий.

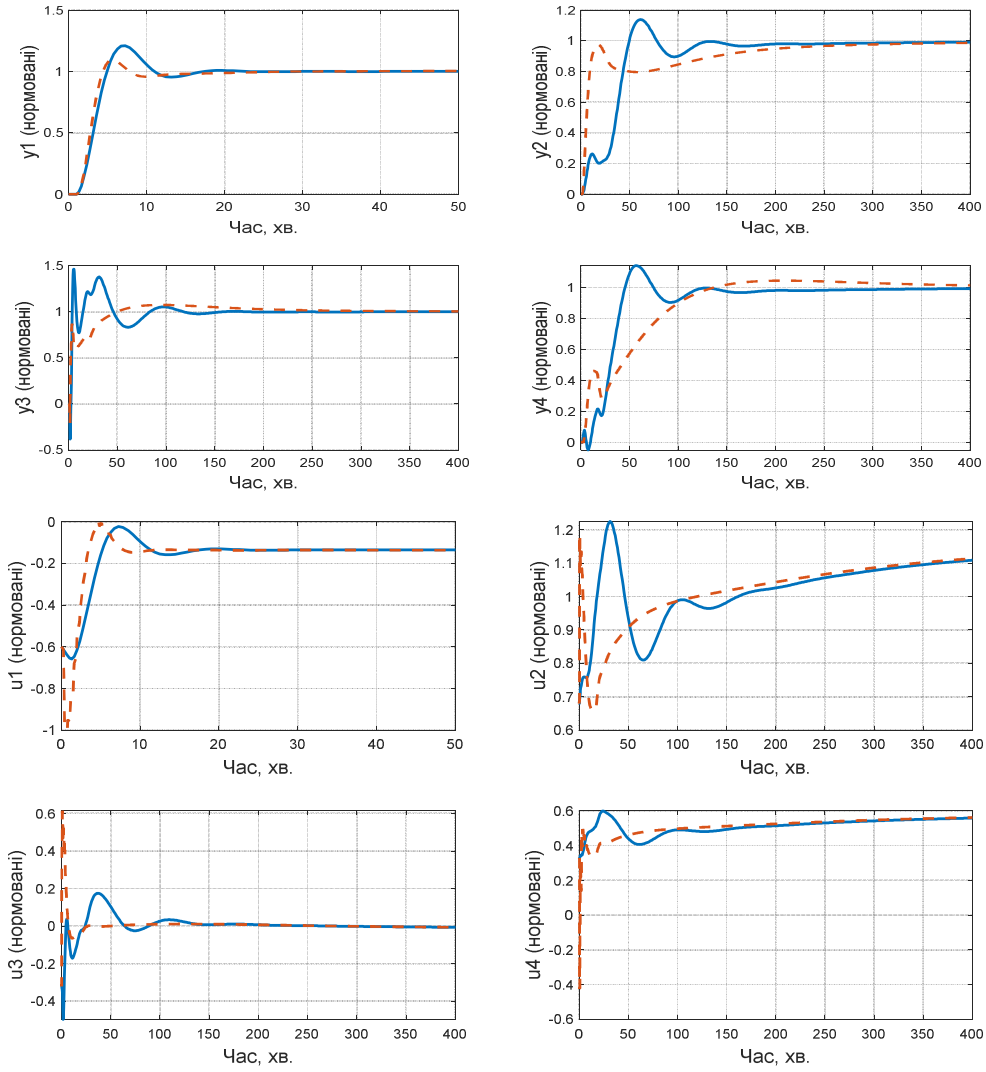
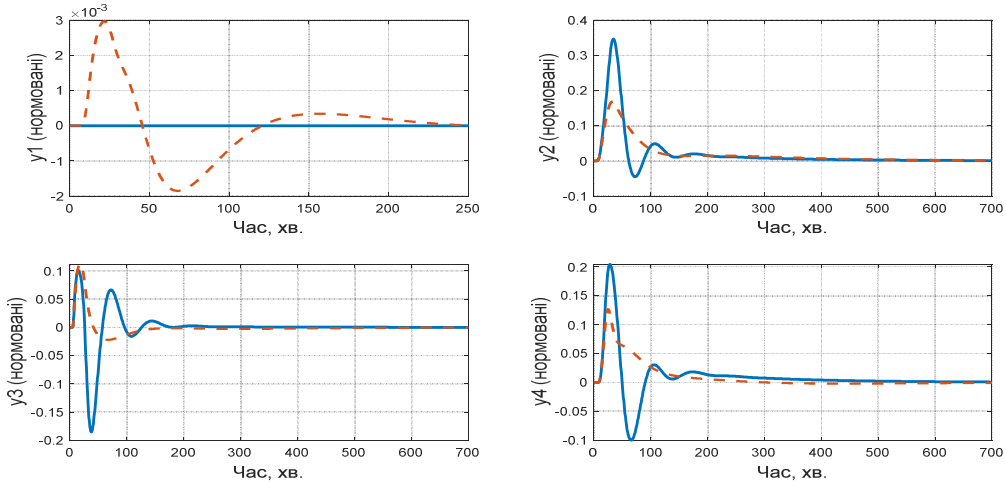


Рис. 6. Перехідні процеси при керуванні за завданням при одночасній зміні завдань на 1 всім регуляторам

Моделювання перехідних процесів при збуренні f_1=1 показані на рис. 7. В табл. 2 узагальнені результати при збуреннях f_1=1 і f_2=1 окремо.



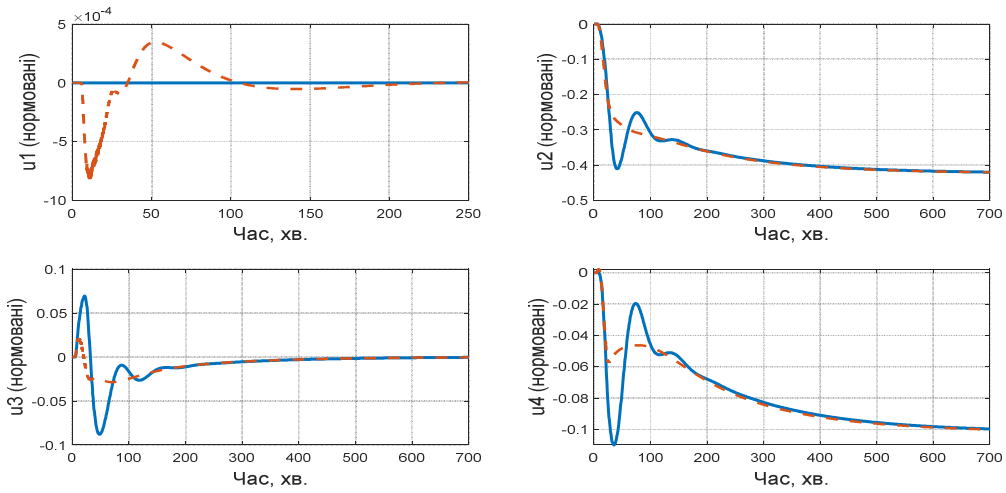


Рис. 7. Перехідні процеси при збуренні f_1=1

Результати, приведені в табл. 2. показують, що гібридна САК$_2$ у цілому забезпечує меншу амплітуду відхилень керованих змінних, особливо при збуренні f_1=1. Щодо інших критеріїв ситуація неоднозначна. При збуренні f_1=1 гібридна САК$_2$ є кращою за критерієм IAE та часом регулювання, при збуренні f_2=1 – гіршою.

Таблиця 2

Показники якості процесів в САК за збуренням

Змінна	**Система**	**σ**	**ST**	**max|y|**	**IAE**	**Змінна**	**max|u|**
Збурення f_1=1							
y_1	САК$_1$	-	0	0	0	u_1	
	САК$_2$	-	207.61	0.0029633	0.17453		0.00081919
y_2	САК$_1$	-	188.96	0.34635	15.276	u_2	0.42003
	САК$_2$	-	345.64	0.16859	13.68		0.4206
y_3	САК$_1$	-	152.81	0.18518	7.289	u_3	0.0879
	САК$_2$	-	123.45	0.11169	4.364		0.028601
y_4	САК$_1$	-	245.85	0.20418	11.068	u_4	0.10982
	САК$_2$	-	201.8	0.12671	7.4884		0.099993
Збурення f_2=1							
y_1	САК$_1$	-	0	0	0	u_1	0
	САК$_2$	-	265.43	0.01769	1.5176		0.0031928
y_2	САК$_1$	-	203.09	0.1574	8.4389	u_2	0.22112
	САК$_2$	-	282.33	0.14779	17.253		0.20241
y_3	САК$_1$	-	110.66	0.3988	14.944	u_3	0.56918
	САК$_2$	-	209.64	0.31918	20.611		0.56918
y_4	САК$_1$	-	167.29	0.24102	13.888	u_4	0.21418
	САК$_2$	-	286.18	0.13555	14.414		0.21583

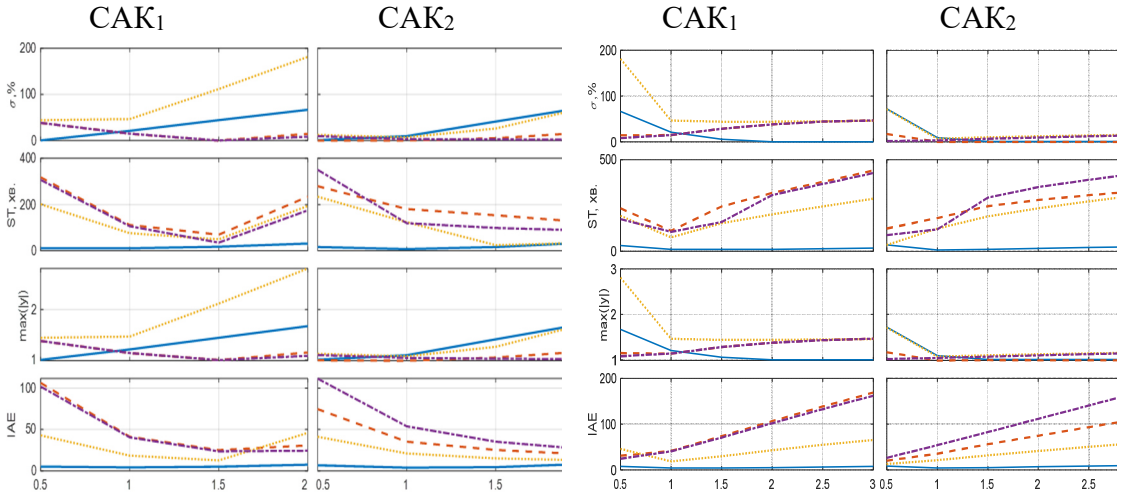
Наступний крок – перевірка поведінки обох САК при зміні параметрів моделі P_a. Як збурення виберемо зміну однакового завдання усім регуляторам. Перевірки будемо проводити для наступних випадків: зміна коефіцієнтів передачі, зміна постійних часу та зміна запізнень усіх каналів на певний множник.

Узагальнені результати експериментів представлені на рис. 8. Аналіз приведених результатів показує наступне: 1) гібридна САК$_2$ не є більш робастною до параметричної невизначеності ніж децентралізована, але й не менш; 2) САК$_1$ та САК$_2$ не витримують зміну коефіцієнтів передачі в каналах моделі P_u більше ніж в два рази; 3) критерій IAE у обох САК змінюється за близьким законом; 4) САК$_1$ при зміні параметрів моделі реагує надмірною величиною перерегулювання й перехідні процеси стають різко коливальними; 5) САК$_2$ істотно менш чутлива до зміни параметрів моделі у порівнянні з САК$_1$, в ній витримується істотне

зменшення амплітуди відхилень в порівнянні з САК$_1$, а коливальність процесів, яка особливо становиться великою в САК$_1$ при збільшенні часу запізнення, в САК$_2$ істотно зменшується.

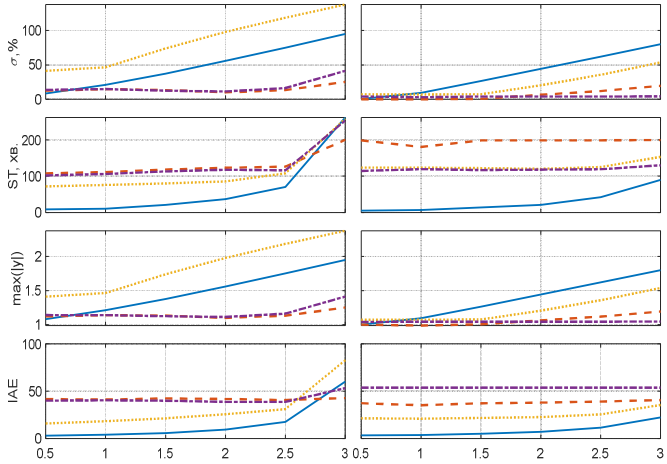
Останній експеримент – перевірка гібридної САК$_2$ на зміну параметрів ПІ регуляторів, що може знадобитись в процесі експлуатації САК$_2$. Попередній аналіз показує, що величина перерегулювання є достатнім критерієм зниження якості системи керування. Проведемо 16 експериментів щодо відхиленні *Kp* і *Ti* регуляторів в обох САК при зміні усіх завдань регуляторам й зафіксуємо величини перерегулювання. Результати експериментів зведемо в табл. 3.

Аналіз результатів, приведених в табл. 3 показує, що 1) гібридна САК$_2$ допускає зміну настройок ПІ регуляторів без втрати стійкості; 2) при зміні настройок одного з регуляторів відношення перерегулювань децентралізованої САК$_1$ та гібридної САК$_2$ в інших каналах залишається близьким до постійного. Додаткові експерименти показали, що гібридна САК$_2$ й децентралізована САК$_1$ мають однакові межі крихкості відносно діапазону змін настройок регуляторів.



Кратна зміна всіх **к** об’єкта від номіналу

Кратна зміна всіх **Т** об’єкта від номіналу



Кратна зміна всіх запізнень об’єкта від номіналу

Позначення: y_1 - , y_2 - , y_3 – , y_4 –

Рис. 8. Показники якості при кратній зміни параметрів об’єкта

Таблиця 3

Перерегулювання в гібридній та децентралізованій САК при зміні настройок ПІ-регуляторів

Зміна	**Система**	$б_{y1}$	$б_{y2}$	$б_{y3}$	$б_{y4}$
Стандартні настройки	$САК_1$	20.999	14.787	46.506	14.871
	$САК_2$	9.4839	0	6.9398	3.0123
$к_{р1}\cdot 0.5$	$САК_1$	0.29019	11.282	49.336	11.503
	$САК_2$	1.6147	0	6.9008	3.9933
$к_{р1}\cdot 1.5$	$САК_1$	44.2	14.677	44.155	14.754
	$САК_2$	39.512	0	7.4013	4.0879
$Ti_1\cdot 0.5$	$САК_1$	43.878	15.685	45.898	15.764
	$САК_2$	26.563	0	7.5148	4.1145
$Ti_1\cdot 1.5$	$САК_1$	13.075	12.744	46.553	12.826
	$САК_2$	4.0213	0	6.9121	3.9795
$к_{р2}\cdot 0.5$	$САК_1$	21.065	50.678	51.559	23.394
	$САК_2$	8.9056	0	6.7672	3.8724
$к_{р2}\cdot 1.5$	$САК_1$	20.734	5.2879	116.81	9.369
	$САК_2$	10.045	11.343	50.597	4.3861
$Ti_2\cdot 0.5$	$САК_1$	20.882	21.465	63.609	11.794
	$САК_2$	9.5139	15.948	8.0305	5.0572
$Ti_2\cdot 1.5$	$САК_1$	20.988	0.24739	40.703	13.752
	$САК_2$	9.4654	0	6.274	3.2371
$к_{р3}\cdot 0.5$	$САК_1$	21.099	17.783	45.665	18.76
	$САК_2$	9.1575	7.2011	11.313	3.3136
$к_{р3}\cdot 1.5$	$САК_1$	20.986	10.709	63.344	8.5617
	$САК_2$	9.7773	0	5.5806	4.5366
$Ti_3\cdot 0.5$	$САК_1$	20.878	22.716	62.696	18.669
	$САК_2$	9.5361	0	5.0721	4.9012
$Ti_3\cdot 1.5$	$САК_1$	20.966	10.123	41.07	11.052
	$САК_2$	9.4573	0.34308	8.633	3.2274
$к_{р4}\cdot 0.5$	$САК_1$	21.022	9.4436	74.449	20.733
	$САК_2$	10.243	40.084	26.296	35.126
$к_{р4}\cdot 1.5$	$САК_1$	21.033	13.996	39.771	8.4743
	$САК_2$	8.6982	0	7.2674	1.226
$Ti_4\cdot 0.5$	$САК_1$	20.77	37.753	46.782	35.596
	$САК_2$	9.3874	0	10.015	0.78673
$Ti_4\cdot 1.5$	$САК_1$	21.038	5.7437	50.924	0
	$САК_2$	9.5069	7.0781	6.1914	10.584

Висновки

Таким чином, розроблено новий метод синтезу САК для складних технологічних ОК з істотними перехресними впливами.

Структура системи керування включає паралельно працюючі цифровий централізований ЛКР та систему децентралізованих ПІ-регуляторів. Мета синтезу – покращення перехідних процесів децентралізованої $САК_1$ з ПІ-регуляторами, настройки якої отримані за допомогою відомого в галузі автоматизації процесів ректифікації методу BLT. Проведені експерименти показують, що перехідні процеси в гібридній $САК_2$ покращуються: істотно зменшується максимальна амплітуда відхилень керованих змінних та перерегулювань при керуванні за завданням (величина якої в деяких експериментах для $САК_1$ перевищує 100%). $САК_2$ є робастною при істотній зміні параметрів моделі об’єкта. Гібридна $САК_2$ має такий же запас крихкості, як й $САК_1$, тому гібридна $САК_2$ дозволяє змінювати настройки ПІ-регуляторів, при цьому ступінь покращення показників якості залишається відносно постійною. Розроблений метод синтезу гібридної $САК_2$ має істотні переваги відносно такого відомого способу покращення якості перехідних процесів в децентралізованих САК, як динамічна розв’язка. На підставі розробленого методу синтезована нова система керування складною

енергоефективною РУ для розділення суміші рідин в хімічній промисловості. Розроблена гібридна САК$_2$ має такі властивості: 1) при керуванні за завданням перехідні процеси мають невелике перерегулювання; 2) при керуванні за збуренням досягається висока точність підтримки регламентних змінних при максимальних збуреннях; 3) система є робастною, тому при зміні параметрів моделі об'єкту від 0.5 до 2-3 номіналів прямі показники якості погіршуються не істотно і залишаються в рамках регламентних відхилень; 4) система не є крихкою, тому витримує зміну налаштувань ПІ-регуляторів.

Список літератури

1. Albertos P., Sala A. Multivariable Control Systems: An Engineering Approach. Springer-Verlag, 2004. ISBN: 978-1-85233-738-4
2. Skogestad S., Postlethwaite I. Multivariable feedback control. Analysis and design. John Wiley & Sons, 2005. ISBN: 978-0470011683
3. Stopakevych A. Analysis of the major approaches to the design of multivariable decentralized control systems for distillation columns. *Eurasian scientific discussions:* Proceedings of the 6th International scientific and practical conference. Barcelona, Spain: Barca Academy Publishing, 2022. P. 80-84.
4. Stopakevych A., Stopakevych O. Design of Robust Decentralized Control Systems for Distillation Columns. *Problemele Energeticii Regionale*. 2022. №2 (54). P. 38-52. https://doi.org/10.52254/1857-0070.2022.2-54.04
5. Стопакевич А.О., Стопакевич О.А. Аналітичний огляд методів розв'язування задачі синтезу децентралізованих систем керування ректифікаційними колонами. *Інформатика та математичні методи в моделюванні*. 2021. Т. 11. №. 4. С. 343–356. https://doi.org/10.15276/imms.v11.no4.343
6. Sedigh A. K., Moaveni B. Control Configuration Selection for Multivariable Plants. Berlin/Heidelberg : Springer-Verlag, 2009. ISBN : 978-3642031922
7. Luyben W. L. Simple method for tuning SISO controllers in multivariable systems. *Industrial & Engineering Chemistry Process Design and Development*. 1986. Vol. 25, № 3. P. 654–660. https://doi.org/10.1021/i200034a010
8. Ahmad A., Wahid A. Application of model predictive control (MPC): tuning strategy in multivariable control of distillation column. Reaktor. 2007. Vol. 11, № 2. P. 66-70. https://doi.org/10.14710/reaktor.11.2.66-70
9. Anbu S. Multiloop Control of Continuous Stirred Tank Reactor Using Biggest Log Modulus Method. Asian Journal of Electrical Sciences. 2016. Vol. 5, № 2. P. 54–61.
10. Yapur S. F., Adam E. J. A Comparison of MIMO Tuning Controller Techniques Applied to Steam Generator. Advances in Science, Technology and Engineering Systems Journal. 2018. Vol. 3, № 3. P. 7–14.
11. Huang H. P., Jeng J. C., Chiang C. H., Pan W. A direct method for multi-loop PI/PID controller design. Journal of Process Control. 2003. Vol. 13, № 8. P. 769–786. https://doi.org/10.1016/s0959-1524(03)00009-x
12. Стопакевич А.О., Стопакевич О.А. Синтез субоптимальної системи автоматичного керування мінімальної складності вакуумною ректифікаційною колоною спиртового виробництва. *Інформатика та математичні методи в моделюванні*. 2022. Т. 12. №. 1-2. С. 84–93. https://doi.org/10.15276/imms.v12.no1-2.84
13. O'Dwyer A. Handbook of PI and PID controller tuning rules. Third edition. London : Imperial Colledge Press, 2009.
14. Wood R. K., Berry M. W. Terminal composition control of a binary distillation column. *Chemical Engineering Science*. 1973. Vol. 28, № 9. P. 1707–1717. https://doi.org/10.1016/0009-2509(73)80025-9
15. Стопакевич А.А. Системный анализ и теория сложных систем. Одесса: Астропринт, 2013. 350 с. ISBN 978-966-190-760-6.

16. Ding S. S., Luyben W. L. Control of a Heat-Integrated Complex Distillation Configuration. *IFAC Proceedings Volumes*. 1989. Vol. 22, № 8. P. 69–76. https://doi.org/10.1016/S1474-6670(17)53339-X

DEVELOPMENT OF A NEW METHOD FOR HYBRID CONTROL SYSTEMS DESIGN FOR MIMO TECHNOLOGICAL PLANTS

Andrii Stopakevych, Oleksii Stopakevych

National University of Intellectual Technologies and Communications,
1, Kuznechna street, Odesa, 65029, Ukraine, stopakevich@gmail.com
Odesa National Polytechnic University,
1, Shevchenko Ave., Odesa, 65044, Ukraine, stopakevich@op.edu.ua

The paper aims to develop a new method for the design of hybrid multivariable automatic control systems. We determine a hybrid control system as a system that includes two different types of parallel operating controllers. The paper's aim is achieved by solving the following tasks. The first task is to develop a method for the design of hybrid control systems. Investigated systems include PI controllers and a linear-quadratic controller. The second task is to investigate the resulting control system. The investigation is based on direct quality indicators, robustness, and fragility analyses. The most significant result of the paper is the development of a method for the tuning of a linear-quadratic controller, which should work with the decentralized control system in parallel. BLT (Biggest Log Modulus) method is used for decentralized control system tuning. BLT is a well-known method of PI controller detuning. The idea of BLT is to ensure the robustness of multivariable control systems tuned by Ziegler Nichols's rules. The proposed method for the design of the control system is not complicated. The possibility of correcting the tunings of decentralized controllers remains. Design of the new hybrid automatic control system for the distillation plant made using the developed method. The analysis of the dynamic properties of the designed control system showed that the use of a parallel linear-quadratic controller provides a small percentage of overshoot in the control by the reference and high accuracy of stabilizing the values of controlled variables at the largest disturbances. At the same time, the robustness does not decrease and the control system remains fragile, i.e., it is possible to change the settings of the controllers without significant loss of robustness.

Keywords. Control system, hybrid, direct indicators, linear quadratic, robustness, BLT, PI controller, distillation column, distillation unit.

ІНФОРМАТИКА ТА МАТЕМАТИЧНІ МЕТОДИ В МОДЕЛЮВАННІ

Том 12, номер 4, 2022. Одеса – 382 с., іл.

ИНФОРМАТИКА И МАТЕМАТИЧЕСКИЕ МЕТОДЫ В МОДЕЛИРОВАНИИ

Том 12, номер 4, 2022. Одесса – 382 с., ил.

INFORMATICS AND MATHEMATICAL METHODS IN SIMULATION

Volume 12, No. 4, 2022. Odesa – 382 p.

Засновник: Національний університет «Одеська політехніка»

Зареєстровано Міністерством юстиції України 04.04.2011р.

Свідоцтво: серія КВ № 17610 - 6460Р

Друкується за рішенням Вченої ради Одеського національного політехнічного університету (протокол № 4 від 25.10.2022)

Адреса редакції: Національний університет «Одеська політехніка»,

проспект Шевченка, 1, Одеса 65044 Україна

Web: www.immm.op.edu.ua (immm.opu.ua)

E-mail: immm.ukraine@gmail.com