ISSN 2078-9181 DOI 10.15622/sp.61

РОССИЙСКАЯ АКАДЕМИЯ НАУК
Отделение нанотехнологий и информационных технологий

САНКТ-ПЕТЕРБУРГСКИЙ
ИНСТИТУТ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ РАН

ТРУДЫ СПИИРАН

proceedings.spiiras.nw.ru

ВЫПУСК 6(61)

Санкт-Петербург
2018

18+

SPIIRAS PROCEEDINGS

Issue № 6(61), 2018

Scientific, educational, and interdisciplinary journal primarily specialized
in computer science, automation, and applied mathematics

Trudy SPIIRAN ♦ Founded in 2002 ♦ Труды СПИИРАН

Founder and Publisher

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences

Editorial Board's address

14-th line VO, 39, SPIIRAS, St. Petersburg, 199178, Russia,
e-mail: publ@iias.spb.su, web: http://www.proceedings.spiiras.nw.ru/

The journal is indexed in Scopus

ТРУДЫ СПИИРАН

Выпуск № 6(61), 2018

Научный, научно-образовательный, междисциплинарный журнал с базовой специализацией в области информатики, автоматизации и прикладной математики
Журнал основан в 2002 году

Учредитель и издатель

Федеральное государственное бюджетное учреждение науки
Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН)

Адрес редакции

199178, Санкт-Петербург, 14-я линия, д. 39,
e-mail: publ@iias.spb.su, сайт: http://www.proceedings.spiiras.nw.ru/

Журнал индексируется в международной базе данных Scopus

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук»

CONTENTS

СОДЕРЖАНИЕ

DOI 10.15622/sp.61.1

P. STEINLE, C. TINGWELL, S.A. SOLDATENKO

OBSERVATION IMPACT ASSESSMENT ON THE PREDICTION OF THE EARTH SYSTEM DYNAMICS USING THE ADJOINT-BASED METHOD

Steinle P., Tingwell C., Soldatenko S.A. **Observation Impact Assessment on the Prediction of the Earth System Dynamics Using the Adjoint-Based Method.**

Abstract. Mathematical models of the Earth system and its components represent one of the most powerful and effective instruments applied to explore the Earth system's behaviour in the past and present, and to predict its future state considering external influence. These models are critically reliant on a large number of various observations (in situ and remotely sensed) since the prediction accuracy is determined by, amongst other things, the accuracy of the initial state of the system in question, which, in turn, is defined by observational data provided by many different instrument types. The development of an observing network is very costly, hence the estimation of the effectiveness of existing observation network and the design of a prospective one, is very important. The objectives of this paper are (1) to present the adjoint-based approach that allows us to estimate the impact of various observations on the accuracy of prediction of the Earth system and its components, and (2) to illustrate the application of this approach to two coupled low-order chaotic dynamical systems and to the ACCESS (Australian Community Climate and Earth System Simulator) global model used operationally in the Australian Bureau of Meteorology. The results of numerical experiments show that by using the adjoint-based method it is possible to rank the observations by the degree of their importance and also to estimate the influence of target observations on the quality of predictions.

Keywords: variational data assimilation, adjoint model, forecast sensitivity, observation impact, Earth system.

1. Introduction. Mathematical models of the Earth system and its components such as the atmosphere, ocean, hydrosphere and biosphere, represent one of the most powerful and effective instruments applied to explore the Earth system's behaviour in the past and present, and to predict its future state considering external influence (e.g. [1-4] and references herein). These models include and parametrically describe numerous physical, chemical and biological processes and cycles such as water cycle, carbon and nitrogen cycles etc. Prediction of the Earth system dynamics under the influence of natural forcing and anthropogenic interventions represents one of the challenging issues of modern science. From the standpoint of dynamical systems theory, the Earth system consists of several interactive dynamical subsystems. Each of them covers a broad space-time spectrum of motions and a wide variety of physical and chemical processes. The Earth system components have specific physical, chemical and dynamical properties, unique structure and behaviour. They are closely related to each other via fluxes of energy, matter, water, aerosols, carbon dioxide and other chemical substances. Modern Earth system models are highly complex and resource

intensive. These models, which can range substantially in their complexity, can be a simple concept or a set of partial differential equations that can be solved numerically by high-performance computers. Formally, the Earth system (or its any component) can be considered as dynamical system generated by the following vector-valued evolutionary differential equation:

$$dx/dt = \mathcal{L}\left(x(t), \alpha\right); \tag{1}$$

$$x\big|_{t=0} = x_0, \tag{2}$$

where $\mathcal{L}$ is a nonlinear differential operator, x is a state vector, x_0 is a given vector-valued function defining the initial state of a system, and α is a vector of parameters.

Since equation (1) is solved numerically, it should be transformed to the discretised form. Equation (1) discretised on the model space-time grid can be written in the following compact from:

$$x_{k+1} = \mathcal{M}_{k,k+1}\left(x_k\right) + \varepsilon_k, \tag{3}$$

where $x_k \in \mathbb{R}^n$ is the *n*-dimensional state vector at time t_k representing the complete set of variables that determine the internal state of a system in question, $\mathcal{M}_{k,k+1} : \mathbb{R}^n \to \mathbb{R}^n$ is a discrete nonlinear operator that propagates the state variables from time t_k to time t_{k+1}, and $\varepsilon_k \in \mathbb{R}^n$ is model errors. Note that the model discrete operator indirectly includes known model parameters. It is usually assumed that the model (3) is "perfect" ($\varepsilon_k = 0$), i.e. given the initial condition x_0, equation (3) uniquely specify the path of dynamical system in its phase space.

Numerical models used in Earth system simulations are critically reliant on large amounts of Earth observation data that are required to correctly define the initial conditions through the process known as data assimilation (DA) (e.g. [5, 6]). As the practice shows, the quality of prediction is strongly affected by the observations – their volume, temporal and special distribution, and accuracy of measurements. In many applications, to simulate and predict the long-time behaviour of dynamical system (e.g. in climate studies) observation data are used to adjust a predictive model trajectory to newly obtained observations (see Figure 1; this figure was created based on the ideas discussed in [7]). To date DA remains one of the key issues in geophysical sciences. The basic goal of DA is to merge observations of any type with certain prior information which needs to be estimated in some way. For example, this prior information referred to as the background can be estimated by models used in prediction.

One of the most popular and effective DA methods is four-dimensional DA (4D-Var). In general terms, 4D-Var DA aims to define the initial state of a dynamical system in question by combining (in statistically optimal manner) the observations of state variables of a real physical system together with a background. 4D-Var procedures are mathematically formulated as an optimization problem, in which the initial condition plays the role of control vector and model equations are considered as constraints. The theoretical foundations of the study and the solution of such problems were laid in the classical works of R.E. Bellman [8], L.S. Pontryagin et al. [9], J.-L. Lions [10], G.I. Marchuk [11]. The variational approach was first used in the prediction of atmospheric processes by Sasaki [12] and then, starting from famous research papers [13-15], has been extensively explored in a vast number of publications.

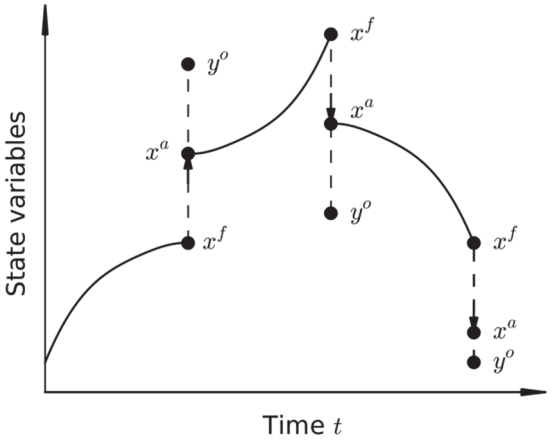


Fig. 1. The scheme of model trajectory adjustment to new observations

The ACCESS (Australian Community Climate and Earth System Simulator) at the Bureau of Meteorology [16] utilizes the 4D-Var scheme in incremental formulation developed at the UK Met Office [19]. The general idea of 4D-Var approach can be simply illustrated as follows. Suppose that at a certain initial time the background state x^b and some physical quantities y^{o} measured by instruments are known. Then [5]

$$x = x^b + \varepsilon^b, \quad y^{\mathrm{o}} = \mathcal{H}\left(x\right) + \varepsilon^{\mathrm{o}}, \tag{4}$$

where $\mathcal{H}$ is the (nonlinear) projection operator, that maps the space of model state into the space of observations, ε^b and ε^{o} are the errors of the

background and observations respectively. Within the framework of 4D-Var, the initial state x_0 is estimated via the following optimization problem [5]:

$$x_0^a = \arg\min \mathcal{J}\left(x_0\right), \tag{5}$$

$$\mathcal{J}\left(x_0\right) = \frac{1}{2}\left\|x_0 - x_0^b\right\|_{\mathrm{B}^{-1}}^2 + \frac{1}{2}\left\|\mathcal{H}\left(x\right) - y^{\mathrm{o}}\right\|_{\mathrm{R}^{-1}}^2, \tag{6}$$

where B and R are the error covariance matrices of the background and observations, respectively, $\|\cdot\|_{\mathrm{A}}$ is the inner product with respect to the A matrix metrics, i.e. $\|\mathrm{a}\|_{\mathrm{A}^{-1}}^2 = \mathrm{a}^{\mathrm{T}}\mathrm{A}^{-1}\mathrm{a}$.

The cost (objective) function (6) is interpreted as follows. The first term that is the background term represents the deviation between the model initial state x_0 and the background x_0^b and calculated in the Euclidean norm L^2 described by the background covariance matrix B. The second term, the observation term, measures the deviation between observations y^o and the "model equivalent" of observations $\mathcal{H}\left(x\right)$. This term is calculated in the L^2 norm described by the observation-error covariance matrix R and is summed over the assimilation window.

The 4D-var problem is simply a minimization problem with constraints on *x* given by the model equation (3). If the observation operator is linear, we obtain a quadratic problem whose unique solution is provided by the Best Linear Unbiased Estimator (BLUE) [5]:

$$x_0^a = x_0^b + \mathrm{Kd}, \tag{7}$$

where $\mathrm{K}=\left(\mathrm{B}^{-1} + \mathrm{H}^{\mathrm{T}}\mathrm{R}^{-1}\mathrm{H}\right)^{-1}\mathrm{H}^{\mathrm{T}}\mathrm{R}^{-1}$ is the Kalman gain matrix, H is a linearized observation operator, and $\mathrm{d} = y^{\mathrm{o}} - \mathrm{H}x^b$ is the innovation vector.

When the observation operator is nonlinear, the variational data assimilation system considers a series of state variables x_j along which the nonlinear operator $\mathcal{H}$ can be linearized. This approach known as an incremental variational data assimilation was introduced in [18]. The first state variable is taken as the background state $x_0 = x^b$, and at iteration *j* the objective function is:

$$\mathcal{J}\left(\delta x_0\right) = \frac{1}{2}\left\|\delta x_0 - \mathrm{b}_j\right\|_{\mathrm{B}^{-1}}^2 + \frac{1}{2}\left\|\mathrm{H}_j\left(\delta x\right) - \mathrm{d}_j\right\|_{\mathrm{R}^{-1}}^2, \tag{8}$$

where $\delta x_0 = x_{0,j} - x_{0,j-1}$ is the output result of the minimization, $\mathrm{b}_j = x_0^b - x_{0,j-1}$, $\mathrm{d}_j = y^{\mathrm{o}} - \mathrm{H}(x_{j-1})$, and H_j is the observation operator linearized around the state estimate x_{j-1}. To achieve the absolute minimum (not the local one), the first guess should be close enough to the truth.

The essential component of 4D-Var system is an adjoint model which, as will be shown below, plays a major role in the exploration of model forecast sensitivity to observations and in the assessment of observation impact on the accuracy of prediction of the Earth system and its components. We would like to emphasize that significant contribution to the theory of adjoint equations was made by G.I. Marchuk and his scientific school (e.g. [11, 19, 20]).

Commonly, the impact of observations on the prediction skill of Earth system models is evaluated by executing the so-called Observing System Experiment (OSE), also known as a Data Denial Experiment (DDE). In a DDE, the forecast skill of two individual runs are compared—one with all observations assimilated and the other with a given observation type (or individual instrument) withheld or added (e.g. [21]). Any change in the forecast accuracy is referred to the observations, which have been withheld. The approach can also be used to assess the impact of target or newly available observations. DDEs can be very helpful but come with disadvantages: they are computationally expensive and not suited to assess the impact of a single station in an observing network or individual measurement device. In addition, DDEs only provide information on the dataset that was withheld, and no information on the value of other subsets of observations.

Another technique, which is able to calculate the individual impact that each assimilated observation has, and is capable to continually generate and aggregate forecast impacts for all observations, was suggested in [22, 23]. This approach makes use of the adjoint models utilized within 4D-Var systems. The observation impact is measured by the reduction in the forecast error expressed as a total "moist" or "dry" energy norm. This method was subsequently implemented in several research and operational centres (e.g. [24, 25]). It is important that such a method uses the same computer code as 4D-Var systems.

This paper aims to illustrate the application of the adjoint-based approach to two coupled low-order chaotic dynamical systems and to the ACCESS global model. We emphasize that this technique is a powerful instrument that allows for not only evaluating the current observing network but also assessing the value of network components which will be used in the future, and, therefore, solve the problems of designing an observing network.

Low-order chaotic dynamical systems considered in this paper, represent computational tools which can be helpful for exploring various aspects of numerical modeling and predicting the behaviour of complex dynamical systems arising in geophysical, environmental, biological, engineering and other branches of science. For these models, the computational cost is insignificant. Consequently, they can be viewed as testing tools to mimic the behaviour of complex systems and, in particular, to explore the forecast sensitivity with respect to observations.

2. Method. As mentioned above, the simplest, but computationally expensive method to assess the impact of observations coming from various sources is the OSE. The main idea behind this method is as follows. Suppose we calculate the forecast (the future state of dynamical system in question) by integrating the model equations over a given time interval $[t_0, t^f]$, where t^f is a verification time (the time at which the forecast accuracy is assessed). Initial conditions for this experiment are determined through 4D-Var utilizing all types of observations. Assume that the forecast accuracy is verified by the use some quantitative measure E_f. Then we integrate the model equation utilizing via 4D-Var all types of observations excluding y_s^o. For this run the forecast accuracy is characterized by E_b. The difference $E_f - E_b$ quantifies the impact of observations y_s^o on the forecast accuracy. However, this approach is computationally ineffective and inconvenient to assess the impact of observations of different types and individual measurements.

Meanwhile, using the adjoint-based technique we can assess the impact of any or all available observations in a computationally efficient way. This method is very appropriate since adjoint models are embedded in 4D-Var systems. Observation impact is computed using (a) sensitivity functions which are components of the adjoint sensitivity gradient of some cost function that characterizes the forecast error, and (b) innovations $y^o - \mathcal{H}\left(x^b\right)$ [23].

Let $\mathcal{R}$ be a scalar response function which is dependent on the system state variables at verification time t^f : $\mathcal{R} = \mathcal{R}\left(x^f\right)$. From the Taylor expansion we can derive the first-order variation of $\mathcal{R}$ at time t^f:

$$\delta\mathcal{R} = \left\langle \delta x^f, \partial\mathcal{R}/\partial x^f \right\rangle = \left\langle \mathrm{M}x_0, \partial\mathcal{R}/\partial x^f \right\rangle. \quad (9)$$

Here $\langle , \rangle$ denotes the dot product, and the forecast variation is expresses via tangent linear model: $\delta x^f = \mathrm{M}\delta x_0$, where M is a linearized model operator. Let M^* be the adjoint of M such that $\langle \mathrm{M}x, y \rangle = \langle x, \mathrm{M}^* y \rangle$.

Since the adjoint of a real matrix equals to its transpose then $\mathrm{M}^{*} = \mathrm{M}^{\mathrm{T}}$ and the equation (9) takes the form [23]:

$$\delta\mathcal{R} = \left\langle \delta x_0, \mathrm{M}^{\mathrm{T}} (\partial\mathcal{R} / \partial x^f) \right\rangle; \tag{10}$$

and the sensitivity of response function to the initial state can be expressed as [23]:

$$\frac{\partial\mathcal{R}}{\partial x_0} = \mathrm{M}^{\mathrm{T}} \frac{\partial\mathcal{R}}{\partial x^f}. \tag{11}$$

Thus, running the adjoint model backward in time with the sensitivity of $\mathcal{R}$ at the verification time as input, one can calculate the sensitivity of $\mathcal{R}$ with respect to the initial conditions. Generally, any differentiable scalar function that represents the forecast accuracy can be considered as the response function (e.g. single model variable or some function of model state variables). Commonly, the forecast error relative to the "true" state x^t is measured in terms of the "total energy norm" [26]:

$$E = (x^f - x^t)^{\mathrm{T}} \mathrm{C} (x^f - x^t), \tag{12}$$

where C is a diagonal matrix of weighting coefficients. The sensitivity of E with respect to initial conditions is expressed as [27]:

$$\frac{\partial E}{\partial x_0} = 2\mathrm{M}^{\mathrm{T}} C (x^f - x^t). \tag{13}$$

At some initial time t_0, there are two state estimations: x_0^a, which is obtained using 4D-Var, and x_0^b, which is obtained via previous model run. Thus, two forecast errors, E_f and E_b, can be defined as [23]:

$$E_f = (x_a^f - x^t)^{\mathrm{T}} \mathrm{C} (x_a^f - x^t), \quad E_b = (x_b^f - x^t)^{\mathrm{T}} \mathrm{C} (x_b^f - x^t), \tag{14}$$

where x_a^f and x_b^f are the predicted states initiated from x_0^a and x_0^b.

To estimate the impact of observations on the forecast error reduction, the response function can be defined as the difference between E_f and E_b:

$$\delta\mathcal{R} \equiv \frac{1}{2}\Delta E = \frac{1}{2}\left(E_f - E_b\right). \tag{15}$$

The linear approximation of error reduction $\delta E \approx \Delta E$ is given by [23]:

$$\delta E = \left\langle \left(y^o - \mathrm{H}\left(x_0^b\right)\right),\ \frac{1}{2}\mathrm{K}^{\mathrm{T}}\left(\frac{\partial E_f}{\partial x^a} + \frac{\partial E_b}{\partial x^b}\right)\right\rangle, \qquad (16)$$

where K^{T} is a transpose of the Kalman gain matrix.

The equation (16) gives the estimate of the forecast error reduction δE produced by any or all observations. Figure 2 (adopted from [23]) shows the schematic representation of the discussed approach for evaluating the forecast sensitivity with respect to observations and assessing the impact of various observations on the forecast accuracy. To estimate the impact of all types of observations we only need to perform a single system's run.

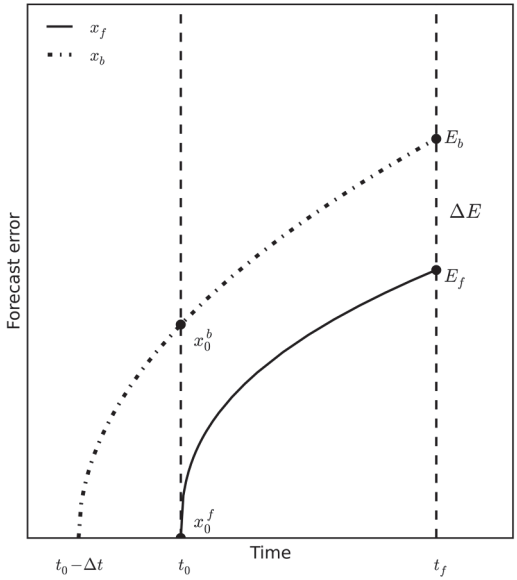


Fig. 2. Schematic representation of the adjoint-based method for observation impact assessment

It is obvious that if the assimilated observations improve the forecast accuracy at the verification time x^f, then the forecast error δE is reduced, and the value δE will be negative. However, if the assimilated observations diminish the forecast quality, the value δE will be positive.

3. Results. For illustrative purposes only we first apply the method discussed in Section 2 to estimate the observation impact on the prediction of dynamics of coupled chaotic dynamical system [27] described in the appen-

dix. This model has six state variables. The following information should be available to solve the problem: the "true" trajectory of a system x^t, the background (the first guess) trajectory x^b, and observations y^o.

In our calculations, we used synthetic data. The "true" trajectory is obtained by integrating model equations numerically with the initial conditions x_0^t taken on the system's attractor. The background (the first guess) trajectory x^b is obtained by integration of the system equations with predefined initial conditions x_0^b which is specified as $x_0^b = x_0^t + \delta^b$, where δ^b is a normally distributed random perturbation with a standard deviation of σ_b applied to all elements of the state vector.

To take into account the background errors, the assumption $\mathrm{B}_0 = \sigma_b^2 \mathrm{I}$ is used, where I is the identity matrix. We assume that the value of $\sigma_b = 0.2$ is applied to all elements of the state vector. Observations y^o are defined for every $2\Delta t$ within the assimilation window, which has a total temporal length of $50\Delta t$. The observed values are generated by adding Gaussian random noise with zero mean and specified standard deviation σ_o to the true state x^t. In calculation we assume that $\sigma_o^{(1)} = 0.05$ ("accurate" observations) and $\sigma_o^{(2)} = 0.1$ ("inaccurate" observations) for "fast" variables, and $\sigma_o^{(1)} = 0.1$ and $\sigma_o^{(2)} = 0.2$ for "slow" variables.

Since observation grid and model grid are the same, the linearized observation operator is simply an identity mapping $\mathrm{H} \equiv 1$. Under the assumption that observation errors are the same for all variables, the observation covariance matrices are defined as $\mathrm{R}_k = \mathrm{R} = \sigma_o^2 \mathrm{I}$.

To minimize the objective function, the conjugate gradient method, resulting in the analysis x_0^a, has been applied. The forecast trajectory is then obtained by integrating the model equations given initial conditions x_0^a.

To estimate the prediction accuracy and the reduction of the forecast error due to observations we use the relative error in energy norm:

$$E_r = \left[(x^t - x^f)^{\mathrm{T}} (x^t - x^f) \big/ (x^t)^{\mathrm{T}} x^t \right]^{\frac{1}{2}}. \qquad (17)$$

The impact of observations is assessed using the ensemble of trajectories generated by randomly produced initial conditions. Table 1 shows the relative error reductions averaged over 500 ensemble members for both "accurate" and "inaccurate" observations. In this table, the forecast errors are computed for

different verification time t^f . The coupled model used in these experiments is chaotic, therefore, its behaviour is highly sensitive to initial conditions. Thus, the forecast accuracy strongly depends on how accurately we can specify the initial state of the dynamical system in question. In turn, the accuracy of initial conditions depends on available observations, the model used in producing forecasts, and data assimilation system. In numerical experiments, the accuracy of observations is specified by the standard deviation σ_o (see above).

Table 1. Observation impact estimates for different verification times for "accurate" $(\delta E_r^{(1)})$ and "inaccurate" $(\delta E_r^{(2)})$ initial conditions

	Verification time			
	0.5	1.0	1.5	2.0
$(\delta E_r^{(1)})$	-0.91	-2.33	-4.32	-3.27
$(\delta E_r^{(2)})$	-0.58	-1.74	-2.53	-1.22

Table 1 illustrates that both "accurate" and "inaccurate" observations show positive impact on the forecast accuracy since the relative energy norm reductions are negative. The impact of “accurate” observations” is, however, larger than the impact of “inaccurate” observations. It is important that the observation impact estimate δE_r is valid over a limited lead time t_{lim}^f since the adjoint model used in calculation of δE_r is derived from a linear forward propagation model known as a tangent linear model. Numerical experiments shown that $t_{lim}^f \approx 2.2$ of dimensionless time units.

For coupled chaotic dynamical system developed on the bases of model [28], the prediction error reductions by observations computed for different verification times t^f and "accurate" observations are presented in Table 2.

Table 2. Observation impact estimates for different verification times for "accurate" initial conditions

	Verification time			
	1	2	3	4
δE_r	-4.53	-3.39	-2.34	-0.39

This table shows that the shorter the forecast range the larger the error reduction or, in other words, the prediction accuracy. These results were obtained by ensemble simulations with 500 ensemble members. For reference, the relative observation impacts (in percentage points) calculated for each observation variable at $t^f = 3$ are shown in Table 3. Observations of *z*-component provide the highest impact on the forecast error reduction while observations of *Y*-component the smallest impact.

Table 3. Relative observation impact (in percentage points) of each model variable for verification time of $t^f = 3$

Verification time				
x	y	z	X	Y
26.5	21.8	36.2	14.7	0.8

Let us now discuss some results obtained via ACCESS global model [16]. The model grid covers the globe with a horizontal resolution of N512 (1024×769 grid points along longitude and latitude, respectively, with average distance between grid points about 25 km), with 70 vertical levels up to ~80 km altitude. The linear perturbation forecast model (a tangent linear model with moist physics) and its adjoint used in 4D-Var and forecast-to-observations experiments has the same vertical resolution as the nonlinear model with a horizontal resolution of N215 (about 60 km).

In the Bureau of Meteorology, a total of 40 million observations are processed daily. Most of these data are satellite measurements. However, only about 10 percent of all observations (~ 4 million) are used in the assimilation system to calculate the initial conditions for the global ACCESS prediction. The following is a summary of the observation types assimilated in the ACCESS 4D-Var global system:

– Surface observations: SYNOP (synoptic network weather stations), SHIP (ship-based instruments), WINPRO (wind profilers), DRIBU (buoy-based instruments) ;

– Upper air observations: TEMP (radiosondes), PILOT (wind observations from pilot balloons and radar profilers), aircraft reports (AIREPS, AMDARS);

– Satellite winds: scatterometer surface winds (ASCAT), atmospheric vector winds (AMV);

– Microwave radiances: ATOVS (AMSU A, B and MHS);

– Infrared radiances: ATOVS (HIRS), AIRS);

– Infrared atmospheric sounding interferometer (IASI);

– Cross-track infrared sounder (CrIS), microwave humidity sounder (MHS), atmospheric infrared sounder (AITS); advanced technology microwave sounder (ATMS);

– Satellites and occultation data from various global navigation satellite systems (GNSS) such as the Global Positioning System;

– Geostationary operational environmental satellite system (GEOS).

The analysis (initial conditions for the prediction model) is generated through the 4D-Var system with a 6-h assimilation window. Observation impacts represent an estimate of the change in a 24-h forecast error as a

consequence of the assimilation of observations. Forecast error is measured in terms of a moist energy norm calculated from the surface to the 150 hPa level over the globe and the northern and southern hemispheres. The adjoint-based observation impacts were calculated from 00Z 1 January 2017 to 00Z 31 December 2017 in 6-h intervals. The experiment details are summarized in Table 4. The total energy norm used to calculate the forecast error reduction due to observations is defined as follows [24]:

$$E = \delta x^{\mathrm{T}} \mathrm{C} \delta x = \frac{1}{M_D} \iiint A r^2 \cos\varphi d\Sigma d\eta \,, \tag{18}$$

$$A = \frac{1}{2}\left(\rho u'^2 + \rho v'^2 + \frac{\rho g^2}{\theta^2 N^2}\theta'^2 + \frac{1}{\rho c^2} p'^2 + \varepsilon \frac{\rho L^2}{c_p} q'^2 \right), \tag{19}$$

where M_D is the mass of the atmosphere in the integration domain D; u, v are the zonal and meridional wind components, respectively; θ is the reference potential temperature, p is pressure; q is the specific humidity; c_p is the heat capacity at constant pressure; L is the latent heat of water vapor condensation; g is the acceleration of gravity; ρ is the air density; c is the speed of sound; N^2 is the square of the Brunt-Väisälä frequency; Σ is the horizontal domain of integration defined by the local projection matrix and η is the vertical coordinate. The primed variables denote the components of linear perturbation forecast model.

Table 4. Summary of experiment

Data period	From 00Z 1 January 2016 to 00Z 31 December 2016
Impact measure	24-hour forecast error reduction on the moist energy norm calculated from the surface to 150 hPa level over the globe and the northern and southern hemispheres.
Prediction system	Operational version of the Bureau of Meteorology prediction system with the resolution of N512 for the forecasting model and N216 for the inner loop of 4D-Var, in horizontal, and 70 levels in vertical. The adjoint of perturbation forecast model includes the moist physics.

The procedure of observation impact assessment is a post-processing routine. For each analysis time (four times per day), the adjoint forecast-to-observation system produces an ASCII output file that contains the information on all the observations including satellite and in situ, non-satellite observation types. The following information is contained in the ASCII output file:

- Sequential number of observation;
- The observation value;

– The value of innovation (the difference between observation and background values);
– Sensitivity of the forecast to observation;
– Latitude and longitude of observation;
– Pressure level of observation;
– Identifier of the instrument type (radiosonde, surface station, wind profiler etc.);
– Identifier of the observation variable type (temperature, moisture, pressure, horizontal wind components etc.);
– The time offset of the observation from the analysis time;
– Observation error variance;
– WMO (World Meteorological Organization) station identification number;
– Satellite identifier, satellite instrument and channel number.

Calculation of the observation impact is a multi-step process. At each analysis time, the ASCII output file is processed into a set of JSON (Java Script Object Notation) files, from which a set of Python-based tools aggregate the individual forecast sensitivities based on observation type and/or station and statistically analyze and visualize the results.

The calculated average observation impact per day of each type of observation on the global forecast is illustrated in Figure 3. This figure shows that all subsets of observations are beneficial, i.e. the impact measure is negative. That is, each individual observation type leads to the reduction of the forecast error of the total energy norm. The most significant observation impact is demonstrated by the Infrared Atmospheric Sounding Interferometer (IASI). Sonde (TEMP) data has the second largest impact. High impact is also provided by AMSUA (microwave sounder radiances), JPSSO-CrIS (cross track infrared sounders), Aircraft and SYNOP (surface observations at land stations). However, MTSAT (atmospheric motion retrievals), HIRS (infrared sounder radiances) and WINPRO (wind profilers) show very small impact. The small impact of WINPRO likely results from the cessation of NOAA's supply of the wind profiler data to the Global Observing System. Currently, only European (CWINDE, 29 wind profilers), Japanese (WINDAS, 31 wind profilers) and Australian (11 wind profilers) wind profiler networks serve as a source of observations. The European and Japanese profilers are operating in densely observed areas of the world, and so the low impact is not surprising. In addition, there are about 50 standalone wind profilers around the globe, which provide data via the Global Observing System.

The volume of satellite information is about 90% of the volume of meteorological information processed via the 4D-Var system. Conse-

quently, the impact of satellite data is always high. However, the analysis of impact per observations for different observation subsets shows (see Figure 4) that the most substantial contributions into the reduction of forecast error per one observation demonstrate PILOT, BUOY and satellite data such as ESA and GOES.

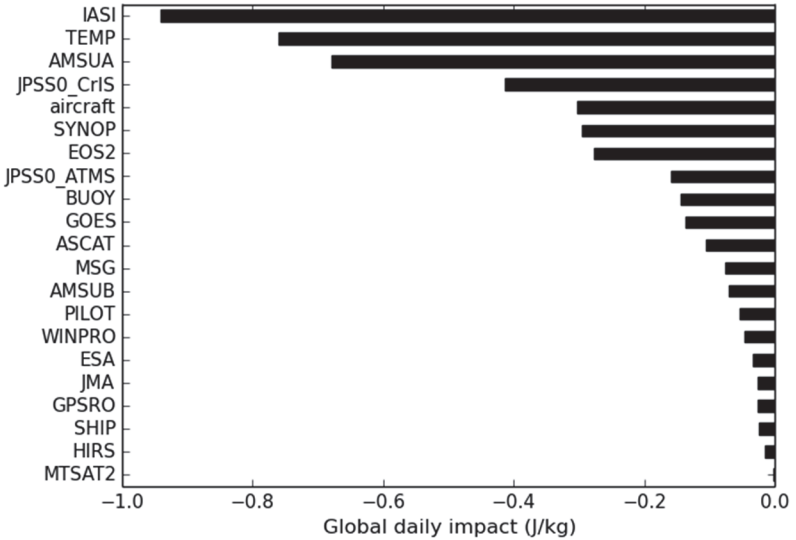


Fig. 3. Average observation impact per day (J/kg) over the globe of each type of observation

The spatial distribution of observation sources is inhomogeneous over the globe. In the northern hemisphere the ratio of land to ocean is about 1 to 1.5. In contrast, in the southern hemisphere the fraction of ocean is about 80% while the land fraction is only about 20%. Consequently, the network of synoptic and aerological stations (upper air observations) in the northern hemisphere is significantly denser than in the southern hemisphere, and the number of synoptic stations and aerological stations in the northern hemisphere significantly exceeds the corresponding number of stations in the southern hemisphere. Figures 5 and 6 illustrate the contribution of observations from both northern and southern hemispheres to the reduction of global forecast error. Sonde observations, Aircraft data, IASI and AMSUA obtained in the northern hemisphere demonstrate the most important impact on the forecast quality (skill). In the southern hemisphere, data from IASI and AMSUA contribute the greatest to the forecast error reduction. Satellite data such as MTSAT2, EOS2 and HIRS are less influential in terms of influence on the forecast accuracy.

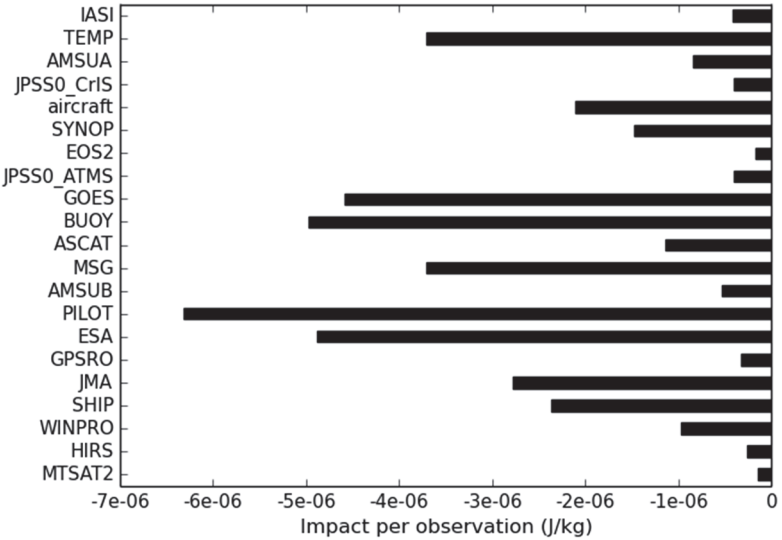


Fig. 4. Average impact per observation (J/kg) for the global forecast

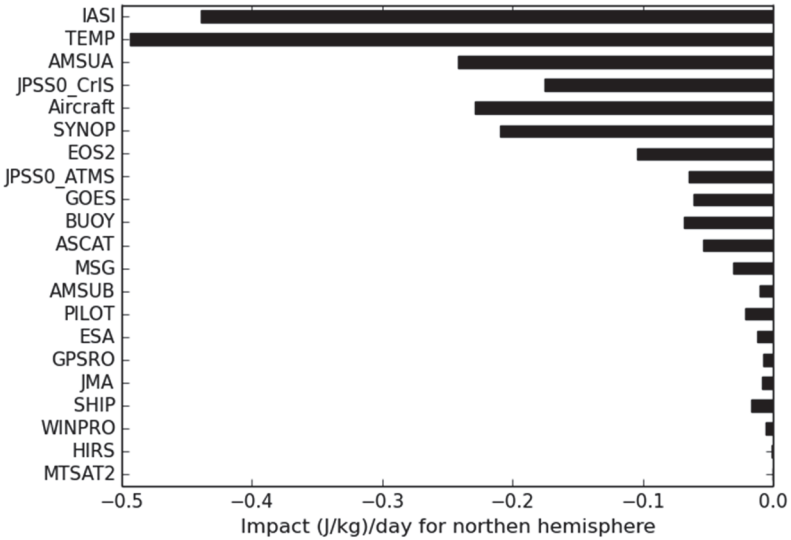


Fig. 5. Daily observation impact (J/kg) for the northern hemisphere calculated for the period January – December 2017

Results obtained show that all observation types positively influence the forecast accuracy both over the globe and northern and southern hemispheres. However, satellite data play a crucial role in weather prediction in the

southern hemisphere. We should keep in mind that the adjoint-based method used in calculations is restricted by the linearity of the algorithm (the perturbation forecast model is linear and, certainly, its adjoint is also linear), which makes it valid only to evaluate short-range forecasts (0 to 48 hours).

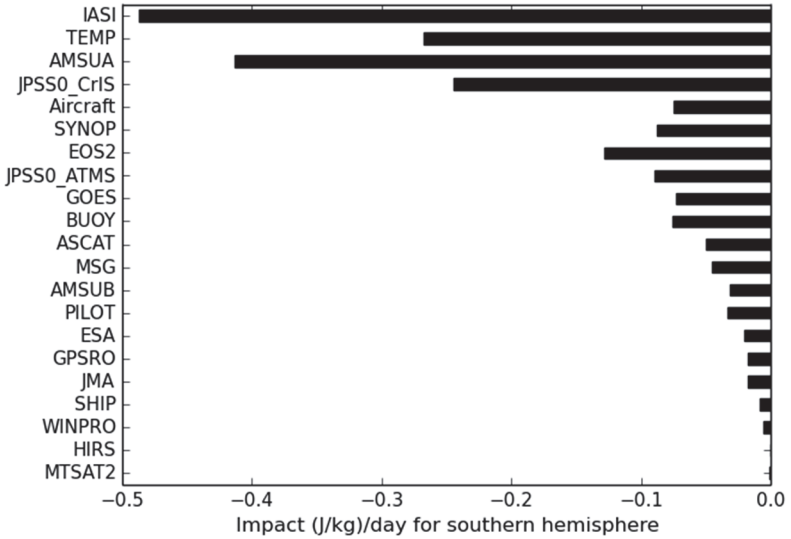


Fig. 6. Daily observation impact (J/kg) for the southern hemisphere calculated for the period January – December 2017

4. Discussion and conclusion. In this paper, we evaluated impacts of various types of observations, in situ observations (ground and ocean-based synoptic and ship observations, wind profiler information, radiosonde and wind balloon upper air observations, and aircraft data) and satellite information, on the accuracy of short-term prediction of global weather conditions using the adjoint-based approach embedded into the ACCESS and its 4D-Var. The impact is measured by the reduction in the 24-hour forecast error expressed as a moist energy norm calculated globally. Overall, all observation types have a positive impact on the forecast accuracy.

Using the adjoint-based method we can support the decision-making process regarding the evolution of the observing network. This powerful technique represents a tool for guiding the design and running of an efficient and effective observing network at the national level and internationally. Future studies are required to explore the seasonal and inter-annual variability of observation impacts.

We also presented two coupled chaotic dynamical systems developed on the bases of the original Lorenz chaotic models [27, 28] which can

imitate the essential features of natural, societal and technical dynamical systems that possess the deterministic chaos. These two coupled systems together with variational assimilation subsystems provide the computational framework for testing different numerical methods and algorithms, and estimating the observation impacts on the prediction of chaotic dynamics.

The Bureau of Meteorology has been applying an adjoint-based approach to estimate the impact of observations on forecasts for about five years. The results obtained are consistent with the available estimates calculated in different research and prediction centres around the world (e.g. [29 – 35]).

5. Appendix. In this appendix we describe two low-order coupled chaotic dynamical systems used in this study. We start from the model obtained by coupling of two ("fast" and "slow") versions of the original Lorenz system [27] (from hereon, L63) with specific time scales differing by a factor ε. We emphasize that L63 is deterministic, however, describes the phenomenon known as "deterministic chaos": over time the behaviour of a simulated system begins to resemble a random process, even though the system is defined by deterministic laws and described by deterministic equations. This phenomenon was first uncovered by Lorenz as he observed the sensitive dependence of atmospheric convection model output on initial conditions [27].

The L63 system is derived by strong spectral truncation of Saltzman's equations, which describe the Rayleigh-Benard convection, and consists of three autonomous ordinary differential equations for time-dependent variables x, y, and z: with x corresponding to the intensity of the convective motion in terms of the stream function, y to the temperature differences between rising and descending currents, and z to the departure of the vertical temperature gradient from its equilibrium magnitude. The L63 contains three positive parameters σ, r, and b, with σ being the Prandtl number, r the normalized Rayleigh number, and b a geometric parameter characterizing length scale of the convective cell.

The L63 can imitate some essential properties of the general circulation of the atmosphere and ocean since the heat flux from equator to the poles can be represented by variable z, which is proportional to meridional temperature gradient that can be represented by parameter r.

A coupled nonlinear chaotic dynamical system is represented by the following set of autonomous differential equations [36]:

a) The "fast" subsystem:

$$\dot{x} = \sigma\left(y - x\right) - c\left(aX + k\right),$$

$$\dot{y} = rx - y - xz + c(aY + k),$$

$$\dot{z} = xy - bz + c_z Z.$$

b) The “slow” subsystem:

$$\dot{X} = \varepsilon\sigma\left(Y - X\right) - c\left(x + k\right),$$

$$\dot{Y} = \varepsilon(rX - Y - aXZ) + c(y + k),$$

$$\dot{Z} = \varepsilon(aXY - bZ) - c_z z,$$

where the lower-case letters x, y and z represent the state variables of the "fast" model, upper-case letters X, Y and Z denote the state variables of the "slow" model, $\sigma > 0$, $r > 0$, and $b > 0$ are the parameters of the original L63 model, ε is a time-scale factor (e.g. if $\varepsilon = 0.1$, then the "slow" subsystem is ten times slower than the "fast" subsystem), c is a coupling strength parameter for x, X, y, and Y variables, c_z is a coupling strength parameter for z and Z variables, k is a “decentring” parameter, and a is a parameter representing the amplitude scale factor ($a = 1$ indicates that "slow" and "fast" subsystems have the same amplitude scale).The coupling strength parameters c and c_z control the interconnection between "fast" and "slow" subsystems: the smaller the parameters c and c_z, the weaker the interdependence between two subsystems.

Essential properties of this chaotic system have been considered in detail earlier [37]. It was underlined that the temporal dynamics of coupled L63 is strongly conditioned by its parameters. Standard values of the L63 parameters corresponding to chaotic behaviour are $\sigma = 10$, $b = 8/3$, and $r = 28$ [27]. These parameter values are used in this study since the motions in the atmosphere and ocean are inherently chaotic. Note that for $\sigma = 10$ and $b = 8/3$, there is a critical value for parameter r, equal to 24.74, and any r larger than 24.74 induces chaotic behaviour of the L63 system.

The time-scale factor ε, “decentring” parameter k and the amplitude scale factor a are taken to be 0.1, 0 and 1 respectively. Without loss of generality, we can assume that and $c = c_z$. The coupling strength c determines the strength of interactions between "fast" and "slow" subsystems and, therefore, the dynamics of the entire coupled system. In numerical experiments we assumed that the coupling between two subsystems is weak, therefore c=0.15 [37].

The system of model equations is numerically integrated by fourth order Runge-Kutta algorithm with a time step $\Delta t = 5 \times 10^{-3}$. To discard the initial transient period the numerical integration starts at time $t_{-\tau} = 2^{15}\Delta t$ with the initial conditions generated randomly around the point $x(-\tau) = (0.01; 0.01; 0.01; 0.02; 0.02; 0.02)^{\mathrm{T}}$ and finishes at time t=0. This guarantees that the calculated state vector $x_0 = x(0)$ is on the system’s attractor.

The state vector x_0 is then used as the initial conditions for further numerical experiments. Note that for $\Delta t = 5 \times 10^{-3}$, the numerical integration with length of 200 time steps corresponds to one dimensionless unit of time.

The next coupled chaotic dynamical system is also composed of "fast" and "slow" models. The "fast" model represents the chaotic dynamical system developed by Lorenz to study the large-scale atmospheric motions [28] (herein is referred to as L84), while the "slow" model, in the absence of coupling to the "fast" model, is a simple harmonic oscillator [38]. The following two sets of autonomous differential equations describe:

a) The "fast" model [28]:

$$\dot{x} = -y^2 - z^2 - ax + aF,$$

$$\dot{y} = xy - cy - bxz + G + aX,$$

$$\dot{z} = xz - cz + bxy + aY.$$

b) The "slow" model [38]:

$$\dot{X} = -\omega Y - \beta Y,$$

$$\dot{Y} = \omega X - \beta z,$$

where x is the intensity of the symmetric, globally averaged westerly wind current (equivalent to the meridional temperature gradient); y and z are the amplitudes of cosine and sine phases of a series of superposed large scale eddies, which transport heat poleward; F and G represent the thermal forcing terms due to the average north-south temperature contrast and the earth-sea temperature contrast, respectively, ω is the ocean oscillation frequency, X and Y are zonal asymmetries in sea surface temperature, which interact with the model atmosphere's eddy fields (y and z).

The "fast" model (the original model proposed by Lorenz) is a Galerkin truncation of the Navier-Stokes equations and gives the simplest approximation of the general circulation of the atmosphere. This model has been widely used in climatological studies and its properties have been explored extensively.

Values of the model parameters used in numerical experiments are as follows [45]: $a = 0.12$, $b = 4b$, $c = 0.5$, $F = 8$, $G = 0.25$, $\beta = \gamma = 0.1$, and $\omega = 2\pi\lambda / 4$, where $\lambda = 0.0274$.

The model equations are solved numerically by fourth order Runge-Kutta algorithm with a time step $\Delta t = 5 \times 10^{-3}$. The initial transient period is discarded, as was mentioned above for the L63 model.

References

1. Biermann F. Earth System Governance: World Politics in the Anthropocene. MIT Press. 2014. 267 p.
2. Hajima T. et al. Modeling in Earth system science up to and beyond IPCC AR5. *Progress in Earth and Planetary Science*. 2014. vol. 1. no. 1. 29 p.
3. Goose H. Climate system dynamics and modelling. Cambridge University Press. 2015. 273 p.
4. Gellelman A., Rood R.B. Demystifying climate models: a user's guide to Earth system models. Springer Nature. 2016. 274 p.
5. Kody L., Andrew S., Konstantinos Z. Data assimilation: A mathematical introduction. Springer. 2015. 242 p.
6. Fletcher S.J. Data assimilation for the geosciences: From theory to application. Elsevier. 2017. 908 p.
7. Leith C.E. Numerical models of weather and climate. *Plasma physics and controlled fusion*. 1993. vol. 35. 919 p.
8. Bellman R.E. Dreyfus S.E. Applied dynamic programming. Princeton University Press. 2015. vol. 2050. 392 p.
9. Pontryagin L.S., Boltyanskii V.G., Gamkrelidze R.V., Mishchenko E.F. The mathematical theory of optimal processes. WileyEnglish. 1962. 360 p.
10. Lions J.L. Control optimal des systemes gouvernes par des equations aux derivees Partielles. Dunod. 1968. 426 p.
11. Marchuk G.I. Numerical methods in weather prediction. Academic Press. 1974. 288 p.
12. Sasaki Y. Some basic formalism in numerical weather analysis. University of Oklahoma. 1970. vol. 98. pp. 875–883.
13. Penenko V.V., Obraztsov N.N. Variational method of adapting of meteorological fields. *Meteorology and Hydrology*. 1976. vol. 11. pp. 3–16.
14. Le Dimet F.-X., Talagrand O. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*. 1986. vol. 38. no. 2. pp. 97–110.
15. Courtier P., Talagrand O. Variational assimilation of meteorological observations with the adjoint equations. Part 2: Numerical results. *Quarterly Journal of the Royal Meteorological Society*. 1987. vol. 113. no. 478. pp. 1329–1347.
16. Puri K. et al. Implementation of the initial ACCESS numerical weather prediction system. *Australian Meteorological and Oceanographic Journal.* 2013. vol. 63. pp. 265–284.
17. Rawlins F. et al. The Met Office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society.* 2007. vol. 133. no. 623. pp. 347–362.
18. Courtier P., Thepaut J.-N., Hollingsworth A. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society.* 1994. vol. 120. no. 519. pp. 1367–1387.
19. Marchuk G.I. Adjoint equations and analysis of complex systems. Kluver Academic Publishers. 1995. 468 p.
20. Marchuk G.I., Agoshkov V.I., Shutyaev V.P. Adjoint equations and perturbation theory. CRC Press. 1996. 288 p.
21. Kelly G., Thépaut J.-N. Evaluation of the impact of the space component of the Global Observing System through Observing System Experiments. ECMWF Newsletter. 2017. vol. 113. pp. 16–28.
22. Baker N., Daley R. Observation and background adjoint sensitivity in the adaptive observation targeting problem. *Quarterly Journal of the Royal Meteorological Society.* 2000. vol. 126. pp. 1434–1454.
23. Langland R.H., Baker N.L. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A: Dynamic Meteorology and Oceanography.* 2004. vol. 56. no. 3. pp. 189–201.
24. Lorenc A.C., Marriott R.T. Forecast sensitivity to observations in the Met Office Global numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society.* 2014. vol. 140. no. 678. pp. 209–224.

25. Soldatenko S., Tingwell C., Steinle P., Kelly-Gerreyn B.A. Assessing the impact of surface and upper-air observations on the forecast skill of the ACCESS numerical weather prediction model over Australia. *Atmosphere.* 2018. vol. 9. no. 1. 23 p.
26. Errico M.R. Interpretations of an adjoint-derived observational impact measure. *Tellus A: Dynamic Meteorology and Oceanography.* 2007. vol. 59. no. 2. pp. 273–276.
27. Lorenz E.N. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences.* 1963. vol. 20. pp. 130–141.
28. Lorenz E.N. Irregularity: A fundamental property of the atmosphere. *Tellus A.* 1984. vol. 36. no. 2. pp. 98–110.
29. Lupu C., Cardinali C., McNally A.P. Adjoint-based forecast sensitivity applied to observation error variances turning. *Quarterly Journal of the Royal Meteorological Society.* 2015. vol. 141. no. 693. pp. 3157–3165.
30. Cardinali C., Healy S. Impact of GPS radio occultation measurements in the ECMWF system using adjoint-based diagnostics. *Quarterly Journal of the Royal Meteorological Society.* 2014. vol. 140. no. 684. pp. 2315–2320.
31. English S. et al. Impact of satellite data. European Centre for Medium-Range Weather Forecasts. 2013. vol. 711. 48 p.
32. Jung B.-J., Kim H.M. Adjoint-derived observation impact using WRF in the Western North Pacific. *Monthly Weather Review.* 2013. vol. 141. no. 11. pp. 4080–4097.
33. Hoover B.T., Langland R.H. Forecast and observation-impact experiments in the Navy Global Environmental Model with assimilation of ECWMF Analysis Data in the global domain. *Journal of the Meteorological Society of Japan.* 2017. vol. 95. no. 6. pp. 369–389.
34. Cioaca A., Sandu A., de Sturler E. Efficient method for computing observation impact in 4D-Var data assimilation. *Computational Geosciences*. 2013. vol. 17. no. 6. pp. 975–990.
35. Janiskoca M., Cardinali C. On the impact of the diabatic component in the forecast sensitivity observation impact diagnostics. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications.* 2017. vol. 3. pp. 483–511.
36. Sequeira L., Kirtman B. Predictability of a low-order interactive ensemble. *Nonlinear Processes in Geophysics*. 2012. vol. 19. no. 2. pp. 273–282.
37. Soldatenko S., Steinle P., Tingwell C., Chichkine D. Some aspects of sensitivity analysis in variational data assimilation for coupled dynamical systems. *Advances in Meteorology.* 2015. vol. 2015. 22 p.
38. Wittenberg A.T., Anderson J.L. Dynamical implications of prescribing part of a coupled system: results from a low-order model. *Nonlinear Processes in Geophysics*. 1998. vol. 5. no. 3. pp. 167–179.

Steinle Peter — Ph.D., leader of data assimilation team, Australian Bureau of Meteorology. Research interests: numerical modelling of processes in the Earth system, numerical prediction of weather and climate, data assimilation, optimization, applied mathematics. The number of publications — 210. peter.steinle@bom.gov.au; 700, Collins srt., Docklands, Melbourne, 3000, Victoria, Australia; office phone: +61(3)9669-4848.

Tingwell Chris — Ph.D., senior researcher of data assimilation team, Australian Bureau of Meteorology. Research interests: numerical modelling of processes in the Earth system, numerical prediction of natural phenomena, data assimilation, astrophysics. The number of publications — 200. chris.tingwell@bom.gov.au; 700, Collins srt., Docklands, Melbourne, 3000, Victoria, Australia; office phone: +61(3)9669-4239.

Soldatenko Sergei Anatolievich — Ph.D., Dr. Sci., professor, leading researcher of atmosphere-ocean interactions laboratory, State Scientific Centre of the Russian Federation the Arctic and Antarctic Research Institute (AARI). Research interests: mathematical modeling of geophysical processes, data assimilation, risk assessment and modeling. The number of publications — 190. prof.soldatenko@yandex.ru; 38 Bering str., St. Petersburg, 199397, Russia; office phone: +7(812)337-3146.

УДК 519.6 **DOI** 10.15622/sp.61.1

П. Стайнли, К. Тингвелл, С.А. Солдатенко

ОЦЕНКА ВЛИЯНИЯ НАБЛЮДЕНИЙ НА ПРОГНОЗИРОВАНИЕ ДИНАМИКИ ЗЕМНОЙ СИСТЕМЫ С ПОМОЩЬЮ МЕТОДА СОПРЯЖЕННЫХ УРАВНЕНИЙ

Стайнли П., Тингвелл К., Солдатенко С.А. **Оценка влияния наблюдений на прогнозирование динамики земной системы с помощью метода сопряженных уравнений.**

Аннотация. Математические модели земной системы служат мощным и эффективным инструментом, используемым для изучения поведения процессов, протекающих в сферических оболочках нашей планеты, в прошлом и настоящем, а также для прогнозирования их в будущем с учетом внешних воздействий. Качество моделирования и прогнозирования природных процессов с применением соответствующих математических моделей в значительной степени зависит от достоверности и объема информации, характеризующей состояние рассматриваемой физической системы (например, атмосферы) в некоторый начальный момент времени. Источниками этой информации служат различные стационарные и подвижные технические средства, интегрируемые в единую глобальную наблюдательную сеть. Поскольку развитие средств наблюдения дорогостоящее мероприятие, очень важно иметь возможность оценивать эффективность как существующей, так и планируемой наблюдательной сети. Цель настоящей работы состоит в том, чтобы, с одной стороны, рассмотреть подход, основанный на сопряженных уравнениях и позволяющий оценивать влияние различных наблюдений на точность прогнозирования эволюции основных компонентов земной системы (атмосферы и океана), и с другой стороны — проиллюстрировать применение этого подхода на примере двух хаотических малопараметрических динамических систем и глобальной модели ACCESS (моделирование австралийского климата и земной системы), используемой в Австралийском метеорологическом бюро. Результаты численных экспериментов демонстрируют высокие возможности метода сопряженных уравнений, который позволяет ранжировать измерительную информацию, получаемую с помощью различных технических средств, по степени ее важности, а также оценить влияние наблюдений на качество прогнозов.

Ключевые слова: вариационное усвоение информации, сопряженные уравнения, чувствительность прогноза к наблюдениям, земная система.

Стайнли Питер — Ph.D., руководитель группы по ассимиляции данных, Австралийское бюро метеорологии. Область научных интересов: численное моделирование процессов в земной системе, численное прогнозирование погоды и климата, ассимиляция данных, оптимизация, прикладная математика. Число научных публикаций — 210. peter.steinle@bom.gov.au; ул. Коллинз, 700, Докландз, Мельбурн, 3001, Виктория, Австралия; р.т.: +61(3)9669-4848.

Тингвелл Крис — Ph.D., старший научный сотрудник группы по ассимиляции данных, Австралийское бюро метеорологии. Область научных интересов: численное моделирование процессов в земной системе, численное прогнозирование природных явлений, ассимиляция данных, астрофизика. Число научных публикаций — 200. chris.tingwell@bom.gov.au; ул. Коллинз, 700, Докландз, Мельбурн, 3001, Виктория, Австралия; р.т.: +61(3)9669-4239.

Солдатенко Сергей Анатольевич — д-р физ.-мат. наук, профессор, ведущий научный сотрудник лаборатории процессов взаимодействия океана и атмосферы, Государствен-

ный научный центр "Арктический и антарктический научно-исследовательский институт". Область научных интересов: математическое моделирование геофизических процессов, усвоение информации, оценка и моделирование рисков. Число научных публикаций — 190. prof.soldatenko@yandex.ru; ул. Беринга, 38, Санкт-Петербург, 199397; р.т.: +7(812)337-3146.

Литература

1. *Biermann F.* Earth System Governance: World Politics in the Anthropocene // MIT Press. 2014. 267 p.
2. *Hajima T. et al.* Modeling in Earth system science up to and beyond IPCC AR5 // Progress in Earth and Planetary Science. 2014. vol. 1. no. 1. 29 p.
3. *Goose H.* Climate system dynamics and modelling // Cambridge University Press. 2015. 273 p.
4. *Gellelman A., Rood R.B.* Demystifying climate models: a user's guide to Earth system models // Springer Nature. 2016. 274 p.
5. *Kody L., Andrew S., Konstantinos Z.* Data assimilation: A mathematical introduction // Springer. 2015. 242 p.
6. *Fletcher S.J.* Data assimilation for the geosciences: From theory to application // Elsevier. 2017. 908 p.
7. *Leith C.E.* Numerical models of weather and climate // Plasma physics and controlled fusion. 1993. vol. 35. 919 p.
8. *Bellman R.E., Dreyfus S.E.* Applied dynamic programming // Princeton University Press. 2015. vol. 2050. 392 p.
9. *Pontryagin L.S., Boltyanskii V.G., Gamkrelidze R.V., Mishchenko E.F.* The mathematical theory of optimal processes // WileyEnglish. 1962. 360 p.
10. *Lions J.L.* Control optimal des systemes gouvernes par des equations aux derivees Partielles // Dunod. 1968. 426 p.
11. *Marchuk G.I.* Numerical methods in weather prediction // Academic Press. 1974. 288 p.
12. *Sasaki Y.* Some basic formalism in numerical weather analysis // University of Oklahoma. 1970. vol. 98. pp. 875–883.
13. *Penenko V.V., Obraztsov N.N.* Variational method of adapting of meteorological fields // Meteorology and Hydrology. 1976. vol. 11. pp. 3–16.
14. *Le Dimet F.-X., Talagrand O.* Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects // Tellus A. 1986. vol. 38. no. 2. pp. 97–110.
15. *Courtier P., Talagrand O.* Variational assimilation of meteorological observations with the adjoint equations. Part 2: Numerical results // Quarterly Journal of the Royal Meteorological Society. 1987. vol. 113. no. 478. pp. 1329–1347.
16. *Puri K. et al.* Implementation of the initial ACCESS numerical weather prediction system // Australian Meteorological and Oceanographic Journal. 2013. vol. 63. pp. 265–284.
17. *Rawlins F. et al.* The Met Office global four-dimensional variational data assimilation scheme // Quarterly Journal of the Royal Meteorological Society. 2007. vol. 133. no. 623. pp. 347–362.
18. *Courtier P., Thepaut J.-N., Hollingsworth A.* A strategy for operational implementation of 4D-Var, using an incremental approach // Quarterly Journal of the Royal Meteorological Society. 1994. vol. 120. no. 519. pp. 1367–1387.
19. *Marchuk G.I.* Adjoint equations and analysis of complex systems // Kluver Academic Publishers. 1995. 468 p.
20. *Marchuk G.I., Agoshkov V.I., Shutyaev V.P.* Adjoint equations and perturbation theory // CRC Press. 1996. 288 p.

21. *Kelly G., Thépaut J.-N.* Evaluation of the impact of the space component of the Global Observing System through Observing System Experiments // ECMWF Newsletter. 2017. vol. 113. pp. 16–28.
22. *Baker N., Daley R.* Observation and background adjoint sensitivity in the adaptive observation targeting problem // Quarterly Journal of the Royal Meteorological Society. 2000. vol. 126. pp. 1434–1454.
23. *Langland R.H., Baker N.L.* Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system // Tellus A: Dynamic Meteorology and Oceanography. 2004. vol. 56. no. 3. pp. 189–201.
24. *Lorenc A.C., Marriott R.T.* Forecast sensitivity to observations in the Met Office Global numerical weather prediction system // Quarterly Journal of the Royal Meteorological Society. 2014. vol. 140. no. 678. pp. 209–224.
25. *Soldatenko S., Tingwell C., Steinle P., Kelly-Gerreyn B.A.* Assessing the impact of surface and upper-air observations on the forecast skill of the ACCESS numerical weather prediction model over Australia // Atmosphere. 2018. vol. 9. no. 1. 23 p.
26. *Errico M.R.* Interpretations of an adjoint-derived observational impact measure // Tellus A: Dynamic Meteorology and Oceanography. 2007. vol. 59. no. 2. pp. 273–276.
27. *Lorenz E.N.* Deterministic nonperiodic flow // Journal of the Atmospheric Sciences. 1963. vol. 20. pp. 130–141.
28. *Lorenz E.N.* Irregularity: A fundamental property of the atmosphere // Tellus A. 1984. vol. 36. no. 2. pp. 98–110.
29. *Lupu C., Cardinali C., McNally A.P.* Adjoint-based forecast sensitivity applied to observation error variances turning // Quarterly Journal of the Royal Meteorological Society. 2015. vol. 141. no. 693. pp. 3157–3165.
30. *Cardinali C., Healy S.* Impact of GPS radio occultation measurements in the ECMWF system using adjoint-based diagnostics // Quarterly Journal of the Royal Meteorological Society. 2014. vol. 140. no. 684. pp. 2315–2320.
31. *English S. et al.* Impact of satellite data // European Centre for Medium-Range Weather Forecasts. 2013. vol. 711. 48 p.
32. *Jung B.-J., Kim H.M.* Adjoint-derived observation impact using WRF in the Western North Pacific // Monthly Weather Review. 2013. vol. 141. no. 11. pp. 4080–4097.
33. *Hoover B.T., Langland R.H.* Forecast and observation-impact experiments in the Navy Global Environmental Model with assimilation of ECWMF Analysis Data in the global domain // Journal of the Meteorological Society of Japan. 2017. vol. 95. no. 6. pp. 369–389.
34. *Cioaca A., Sandu A., de Sturler E.* Efficient method for computing observation impact in 4D-Var data assimilation // Computational Geosciences. 2013. vol. 17. no. 6. pp. 975–990.
35. *Janiskoca M., Cardinali C.* On the impact of the diabatic component in the forecast sensitivity observation impact diagnostics // Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications. 2017. vol. 3. pp. 483–511.
36. *Sequeira L., Kirtman B.* Predictability of a low-order interactive ensemble // Nonlinear Processes in Geophysics. 2012. vol. 19. no. 2. pp. 273–282.
37. *Soldatenko S., Steinle P., Tingwell C., Chichkine D.* Some aspects of sensitivity analysis in variational data assimilation for coupled dynamical systems // Advances in Meteorology. 2015. vol. 2015. 22 p.
38. *Wittenberg A.T., Anderson J.L.* Dynamical implications of prescribing part of a coupled system: results from a low-order model // Nonlinear Processes in Geophysics. 1998. vol. 5. no. 3. pp. 167–179.

УДК 621.390 **DOI** 10.15622/sp.61.2

Я.А. Скороходов

МОДЕЛИРОВАНИЕ ФУНКЦИОНИРОВАНИЯ КОСМИЧЕСКИХ И НАЗЕМНЫХ СИСТЕМ НАБЛЮДЕНИЯ ЗА ВОЗДУШНЫМ ДВИЖЕНИЕМ

Скороходов Я.А. **Моделирование функционирования космических и наземных систем наблюдения за воздушным движением.**

Аннотация. В настоящее время создаются и постепенно наращиваются орбитальные группировки космических аппаратов с возможностью приема, обработки и ретрансляции сигналов системы ADS-B (от англ. «Automatic Dependent Surveillance — Broadcast» — автоматическое зависимое наблюдение — широковещание), обеспечивающие глобальность и непрерывность наблюдения за воздушным движением. В соответствии с концепцией использования технологии ADS-B каждый участник воздушного движения передает в широковещательном режиме свои идентификационные данные, местоположение и параметры состояния. Так как при разработке системы не предполагалось принимать сигналы на борту космического аппарата, существуют определенные проблемы, связанные с их энергетической доступностью, наличием коллизий сообщений от разных источников, влиянием эффекта Доплера и другими факторами. Разработана имитационная модель системы наблюдения за воздушным движением на основе приема сигналов, содержащих идентификационную и навигационную информацию и передающихся по радиоканалу в широковещательном режиме. Программно реализованные алгоритмы имитационного моделирования позволяют задавать различные ограничения и допущения (используя различные модели распределения источников излучений, пунктов приема сигналов авиационных систем связи, канала передачи информации, распределения частоты и длительности сигналов) и получать оценки целевых показателей функционирования космических и наземных систем обеспечения безопасности движения воздушных судов с учетом различных пространственных и энергетических факторов и условий распространения радиосигналов, а также реального размещения контролируемых объектов и динамики их движения в мировом воздушном пространстве. Приведены методики и примеры использования имитационной модели для расчета целевых показателей функционирования космических и наземных систем авиационного наблюдения.

Ключевые слова: авиационное наблюдение, ADS-B, автоматическое зависимое наблюдение – вещание, математическое моделирование, космические системы, обработка информации.

1. Введение. В последние годы в целях повышения безопасности воздушного движения активно развиваются системы авиационного наблюдения. Наряду с традиционными системами наземного базирования в практику управления воздушным движением внедряются перспективные средства космического базирования [1-5]. Для выработки обоснованных требований или решений при проектировании космических и наземных систем авиационного наблюдения на основе приема сигналов, содержащих идентификационную и навигационную информацию (например, сигналов системы автоматического зависимого наблюдения — радиовещания), а также оптимального планирования

применения уже существующих систем, необходимо выполнение ряда научно-исследовательских и опытно-конструкторских работ, чтобы оценить энергетическую доступность сигналов и влияние на безошибочный прием сообщений других факторов, в частности возможных интерференций сигналов от разных источников [6, 7].

Во всех случаях как для космических, так и для наземных систем проведение натурных экспериментов является дорогостоящей и не всегда выполнимой операцией. В первом случае высокая финансовая ресурсоемкость экспериментов обусловлена необходимостью оснащения отделяемых элементов ракет-носителей (последних ступеней), разгонных блоков или космических аппаратов (КА) дополнительным оборудованием приема и обработки сигналов, излучаемых на частоте 1090 МГц [8-10]. Для наземной сетевой системы экспериментальные исследования затруднены большой территориальной разнесенностью пунктов наблюдения.

В ряде случаев для принятия решений достаточно использовать результаты математического моделирования. В настоящее время разработано большое количеств универсальных и специализированных средств имитационного моделирования, в частности «AnyLogic», «GPSS» [11, 12]. Однако их использование для решения указанной в статье задачи не представляется возможным в силу следующих факторов:

– большое количество источников излучений сигналов (десятки и сотни тысяч);

– неравномерное и неоднородное распределение источников излучений в пространстве, динамически изменяющаяся радиоэлектронная обстановка, что особенно справедливо для воздушных объектов, учитывая высокую скорость их перемещения;

– сложные протоколы передачи информации, в частности, для воздушных судов количество передаваемых сигналов (типов MODE A/C, MODE S) за анализируемых интервал времени обусловлено количеством радиоэлектронных станцией, в радиолокационном поле которых движется объект;

– большое количество факторов, влияющих на ослабление сигналов спутниковых радиолиний, ряд из которых плохо формализованы (например, влияние гидрометеоров на распространение электромагнитных волн);

– сложные баллистические модели.

Проблемами моделирования систем управления воздушным движением и планирования использования воздушного пространства в системах организации воздушного движения занимается ряд организа-

ций, в том числе и в России — ФГУП «ГосНИИАС» [13]. В то же время тема моделирования космических систем контроля движения воздушных судов на основе приема сигналов автоматического зависимого наблюдения в опубликованных работах не рассматривалась.

Цель настоящего исследования — разработка имитационно-аналитической модели функционирования систем авиационного наблюдения, которая может быть использована для решения следующих задач:

– оценивания энергетической доступности сигналов для заданной структуры орбитальной группировки (ОГ) или наземной сети и других условий (с учетом различных допущений и ограничений при моделировании) [14];

– оценивания целевых показателей обнаружения судов для заданной структуры космической системы (наземной сети);

– выбора структуры ОГ (параметров размещения сети) для заданных показателей обнаружения воздушных судов.

Правильность результатов моделирования зависит от качества описания реальной действительности в используемых моделях. С целью наиболее адекватного представления моделируемых систем произведен анализ распределения источников излучения сигналов на частоте 1090 МГц и маршрутов их движения, а также факторов, влияющих на распространение сигналов систем авиационного наблюдения.

2. Анализ характеристик воздушного движения. Для оценивания характеристик воздушного движения и разработки статической и динамической моделей распределения источников излучений сигналов на частоте 1090 МГц использовались форматированные данные, включающие координаты воздушных судов в сферической геодезической системе координат, их ICAO (от англ. «International Civil Aviation Organization» — международная организация гражданской авиации) коды и уникальные идентификаторы рейсов, регистрация которых происходила в течение 24 часов каждые 15 минут с веб-сервисов, позволяющих следить за перемещением воздушных судов в режиме реального времени, например интернет-ресурсов Flightradar24, Planefinder, ADS-B Exchange. Результаты регистрации с указанием времени получения информации заносились в базу данных. В простейшем случае база данных может состоять из двух таблиц, содержащих статические и динамические характеристики источников излучений сигналов. Визуализация выборочных данных в геодезической системе координат (СК) представлена на рисунке 1.

Анализ рисунка 1 позволяет выделить регионы с повышенной плотностью воздушного трафика, регионы, где отсутствуют средства авиационного наблюдения или не предоставляется общий доступ к полетным данным, а также основные междугородние и трансконтинентальные маршруты движения воздушных судов. При более детальном анализе можно выделить кластеры сосредоточения воздушных судов, соотнесенные с региональными центрами соответствующих стран.

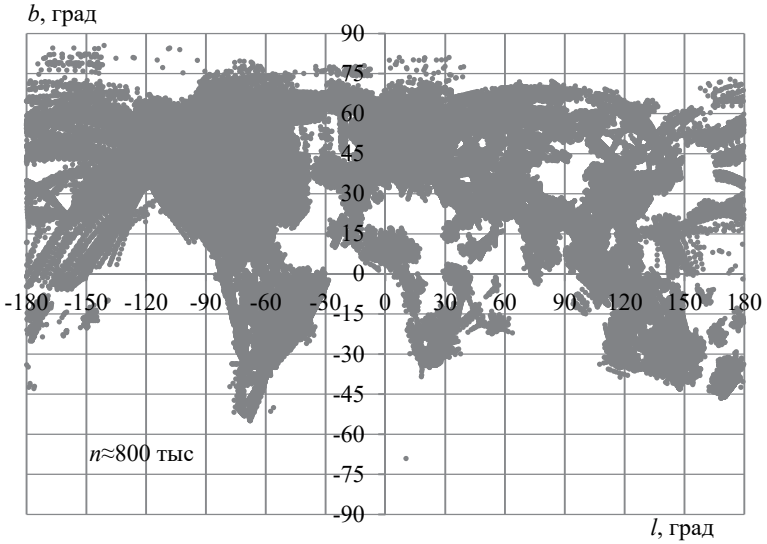
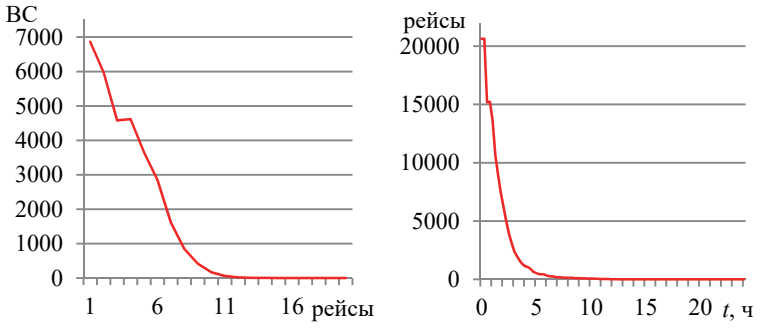


Рис. 1. Визуализация выборочных данных в геодезической СК

Обработка и систематизация полного объема информации позволила оценить обобщенные статистические данные о воздушном трафике. Размер выборки (общее количество полученных сообщений) за указанный интервал времени составил 769855, количество зафиксированных воздушных судов (без учета воздушных судов, не имеющих ICAO кода) — 31666, количество рейсов — 125558. Полигоны частот количества рейсов, совершенных одним воздушным судном за анализируемый интервал времени, и продолжительности рейсов представлены на рисунке 2.

График на рисунке 2а не учитывает 14662 рейсов воздушных судов, не имеющих ICAO кода, что связано с особенностью работы самого интернет ресурса и не позволяет отнести рейс к какому-либо воздушному судну.



а) *б)*

Рис. 2. Полигоны частот: а) количество рейсов, совершенных одним ВС за анализируемый интервал времени; б) продолжительность рейсов

Числовые оценки характеристик воздушного движения за сутки представлены в таблице 1.

Таблица 1. Характеристики воздушного движения

№ п/п	Характеристика	Значение		
		Мин.	Макс.	Среднее
1	Количество источников излучений сигналов на частоте 1090 МГц	7022	9320	8019
2	Количество рейсов, совершенных одним ВС за сутки	1	20	3,5
3	Продолжительность полета ВС, часов	0,25	24	1,2

Минимальное и максимальное значения продолжительности полетов, указанные в таблице 1, в данном случае обусловлены минимальным интервалом регистрации информации и общей длительностью проведения эксперимента.

График изменения значений количества источников излучений сигналов на частоте 1090 МГц в период с 6 января 2018 года по 9 января 2018 года (продолжительность регистрации 84 часа) изображен на рисунке 3. Начало регистрации соответствует 19.00 в московском часовом поясе. Анализ графика показывает, что изменение количества источников излучений имеет периодический характер и зависит от времени суток.

Учитывая периодичность изменения представленных характеристик, можно сделать вывод, что данные суточного мониторинга воздушного движения являются репрезентативными и позволяют на их основе

создать как статическую, так и динамическую модели распределения источников излучений сигналов авиационных бортовых систем связи.

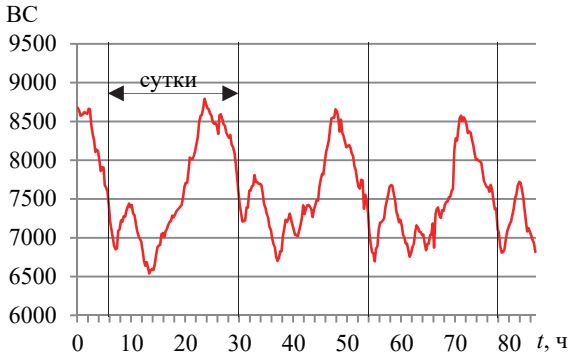


Рис. 3. Количество $N_S(t)$ источников излучений

Кроме того, используя данные, полученные из публичных интернет-ресурсов, можно создать более сложные модели и алгоритмы прогнозирования траектории движения воздушных судов с целью исключения их столкновений и оптимизации управления воздушным движением [15-18].

3. Анализ факторов, влияющих на ослабление сигналов авиационных систем обеспечения безопасности воздушного движения. На ослабление сигналов авиационных систем обеспечения безопасности воздушного движения в реальных условиях влияют различные факторы, к основным из которых следует отнести дальность передачи, атмосферные явления и поляризационные эффекты.

Затухание энергии сигнала в свободном пространстве. Расчет потерь $L_1(d, f)$ передачи при распространении сигнала в свободном пространстве осуществляется по следующей зависимости [19]:

$$L_1(d, f) = 32{,}4 + 20\log f + 20\log d, \qquad (1)$$

где f — частота передачи сообщения, МГц; d — расстояние от источника s_i до приемника r_j, км.

Атмосферные потери. Потери мощности сигнала в атмосфере в основном обусловлены тропосферными кислородом и парами воды, а также дождем и прочими осадками. При моделировании канала связи между воздушным судном и КА атмосферные потери не учитываются, поскольку большая часть полета воздушного судна проходит над слоем атмосферы, содержащим основную часть имеющихся в ней водяно-

го пара и кислорода. Выражение для вычисления ослабления $L_2(f, \beta)$ мощности сигнала в газах атмосферы имеет вид [20]:

$$L_2(f,\beta) = \gamma_{H_2O}(f) l_{H_2O}(\beta) + \gamma_{O_2}(f) l_{O_2}(\beta), \quad (2)$$

где β — угол места; $\gamma_{H_2O}(f)$, $\gamma_{O_2}(f)$ — погонное ослабление для водяного пара и кислорода соответственно, дБ/км; $l_{H_2O}(\beta)$, $l_{O_2}(\beta)$ — эффективные длины трасс в атмосфере, содержащей водяные пары и кислород, км.

Аналитические аппроксимации для определения коэффициентов $\gamma_{H_2O}(f)$, $\gamma_{O_2}(f)$ погонного ослабления, зависящие от частоты f сигнала, приведены в [7, 19], выражения эффективных длин трасс $l_{H_2O}(\beta)$, $l_{O_2}(\beta)$ имеются в работе [20].

Угол места β рассчитывается в топоцентрической системе координат, начало которой определяется геодезической широтой $b_П$ и долготой $l_П$ измерительного пункта на поверхности земного эллипсоида, согласно следующему выражению:

$$\beta = \operatorname{asin}(y_T / d), \quad (3)$$

где y_T — проекция координат объекта на ось OY топоцентрической системе координат (СК).

Математическая формализация распределения гидрометеоров на Земле в целом и динамики атмосферы характеризуется высокой сложностью и является отдельно научной задачей, поэтому влияние атмосферных осадков на распространение электромагнитных волн и ослабление мощности сигналов авиационных систем связи учитывается только в специально оговоренных случаях, например, с целью определения электромагнитной доступности сигналов в экстремальных условиях.

Алгоритм расчета ослабления $L_3(f, \beta, h_0, J)$ сигналов в гидрометеорах (дожде) достаточно подробно приведен в работе [21]. Исходными данными для расчета ослабления $L_3(f, \beta, h_0, J)$ сигнала в дожде являются: угол места β, частота f сигнала, высота h_0 изотермы 0°C над средним уровнем моря, интенсивность дождя J, частотно зависимые коэффициенты для прогнозирования ослабления сигнала в дожде с заданной интенсивностью J. Значения высоты h_0 изотермы 0°C и интенсивности J дождя для региона размещения измерительного пункта

можно взять из справочных данных, в том числе в книге [21] частотно зависимые коэффициенты приводятся в рекомендациях Международного союза электросвязи.

Поляризационные потери суммируются из потерь связанных с расхождением плоскостей поляризаций приемной и передающей антенн, эффектом Фарадея, потерь из-за деполяризации радиоволн в осадках.

Выражение для определения потерь $L_4(\varphi)$ из-за расхождения плоскостей поляризации передающей и приемной антенн имеет вид:

$$L_4(\varphi) = -20\log[\cos(\varphi)], \tag{4}$$

где φ — угол между фокальными осями антенн судна и КА, при φ = π/2 связи не будет.

Угол ψ поворота плоскости поляризации вследствие фазовой дисперсии сигналов, обусловленной эффектом Фарадея, можно приближенно определить по формуле [19]:

$$\psi = 2{,}32 \cdot 10^{19} \big/ f^2 \,[1 - 0{,}91 \cdot \cos(\beta)]^{0{,}5}, \tag{5}$$

а потери сигнала $L_5(\psi)$:

$$L_5(\psi) = -20\log[\cos(\psi)]. \tag{6}$$

Деполяризация волн в осадках обусловлена несферичностью формы и траекторией падения частиц гидрометеоров, прежде всего капель дождя. Поглощение радиоволн в гидрометеорах вследствие деполяризации увеличивается прямопропорционально частоте сигнала и является весьма значительным для сигналов с частотой выше 5 ГГц. Поскольку формализация эффекта деполяризации в осадках является достаточно сложной задачей и анализируемая частота существенно меньше 5 ГГц, при моделировании ослабление сигнала из-за деполяризации радиоволн не учитывается.

4. Имитационная модель функционирования систем авиационного наблюдения. Модель функционирования систем авиационного наблюдения составляют частные модели, а именно модель распределения источников излучений сигналов на частоте 1090 МГц, модель распределенных пунктов приема сигналов, баллистические модели, модель канала передачи информации, модель распределения частоты и длительностей сигналов (рисунок 4).

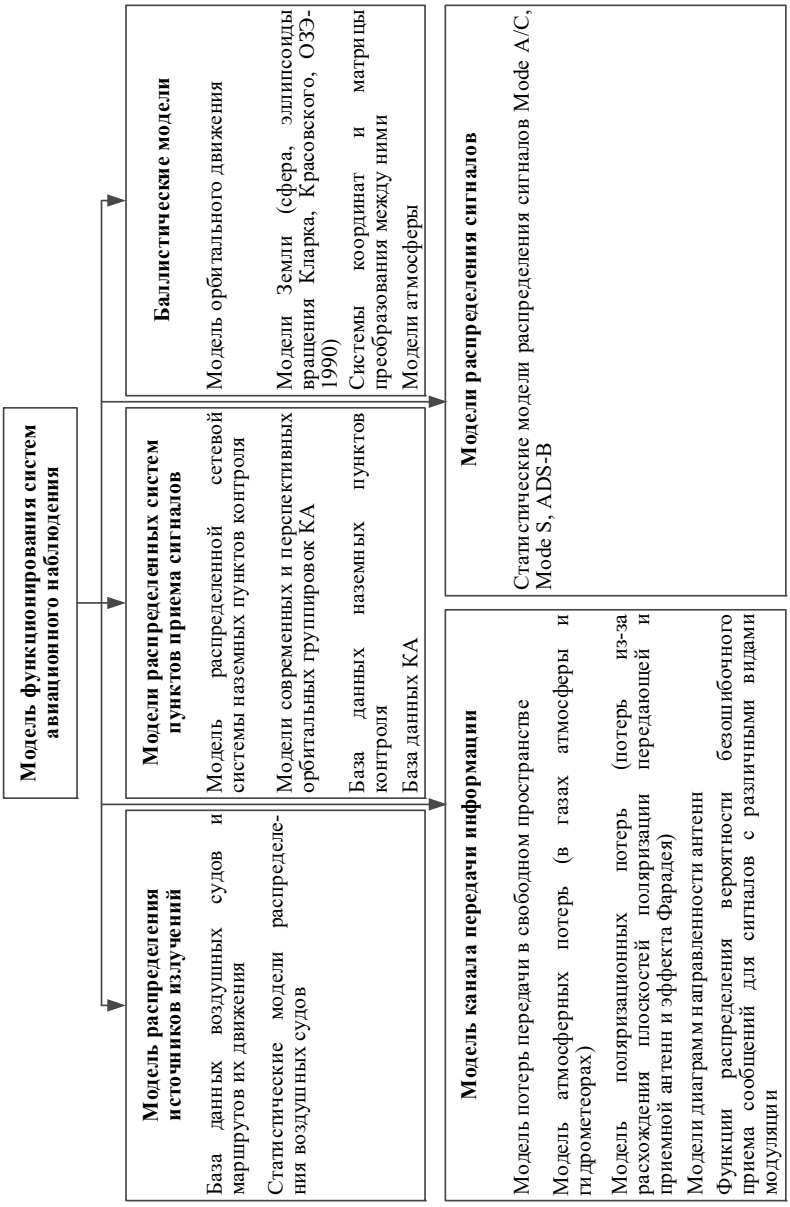


Рис. 4. Состав имитационной модели систем авиационного наблюдения

Модель распределения источников излучений сигналов на частоте 1090 МГц представлена в статическом и динамическом вариантах. Достоинством статической модели распределения источников излучений является простота программной реализации, в тоже время результаты моделирования при условии статической модели, учитывая высокую скорость перемещения воздушных судов, будут адекватными лишь на коротких интервалах времени.

Статическая модель распределения источников излучений представляет собой множество $S = \{s_i, i = 1,..., N_S\}$ элементов — источников излучений сигналов на частоте 1090 МГц, каждый из которых характеризуется своим местоположением C_i, заданным в геодезической сферической системе координат. Местоположение воздушных судов может быть задано двумя способами, в первом случае координаты C_i местоположений объектов наблюдения определяются статистически по заданному закону распределения, как правило, равномерному, во втором случае используется «снимок» реального местоположения объектов для заданного момента времени *t*.

В случае динамической модели распределения источников излучений каждому из элементов $s_i \in S$, $i = 1,..., N_S$ множества контролируемых объектов ставится в соответствие множество полетов $F_i = \{f_{ij}, i = 1,..., N_S, j = 1,..., n_i\}$ где $n_i = \text{card} < F_i >$. В свою очередь полет f_{ij} составляют координаты $C_i(t_k)$ объекта в геодезической сферической системе координат, соотнесенные с моментом времени $t_k \in [t_0, t_0 + T]$. Маршруты движения воздушных судов определяются исходя из анализа воздушного трафика за заданный промежуток времени.

Выбор типа модели (статической или динамической) распределения источников излучений определяется, прежде всего, целью моделирования, первый тип может применяться для сравнения результатов имитационного моделирования и аналитических вычислений, предложенных другими авторами [6], второй тип — для получения обоснованных результатов применения систем авиационного наблюдения с учетом неоднородности распределения судов в мировом воздушном пространстве, маршрутов и динамики их движения, а также других факторов.

Модель распределенных систем пунктов приема сигналов включает модель ОГ КА и модель сети наземных пунктов контроля воздушного движения. Модель ОГ КА представлена множеством элементов $r_j \in R$, $j = 1,..., N_R$, структура ОГ задается следующими параметрами: количество плоскостей *n*, параметры орбит в каждой плоскости (долгота восходящего узла Ω, наклонение *i*, апогей r_A, перигей $r_П$, широта перицентра ω), количество КА в каждой плоскости, углы, на которые разнесены КА друг относительно друга в одной плоскости.

Модель сети наземных пунктов контроля движения воздушных судов представлена множеством элементов $q_k \in Q$, $k = 1,..., N_Q$, каждый из которых характеризуется своим местоположением C_k на земной поверхности.

К *баллистическим моделям* относятся: модель орбитального движения, системы координат и матрицы преобразования между ними, модели Земли. Модель орбитального движения задана отображением $\eta : C_j(t_k) \to C_j(t_{k+1})$, где $C_j(t_k)$, $C_j(t_{k+1})$ — положение КА r_j в моменты t_k и t_{k+1} соответственно. Методы расчета орбитального движения КА являются общеизвестными и приведены в различной научно-методической литературе, например в [22].

Для проверки корректности работы моделей ОГ КА и орбитального движения в разработанном программном комплексе имитационного моделирования реализован компонент трехмерной визуализации функционирования КС. Визуальное представление функционирования КС на примере ОГ КА «Iridium NEXT» изображено на рисунке 5. В рассматриваемой ОГ 75 КА расположены в 6 плоскостях по 11 в каждой, 9 КА из которых являются резервными [23, 24].

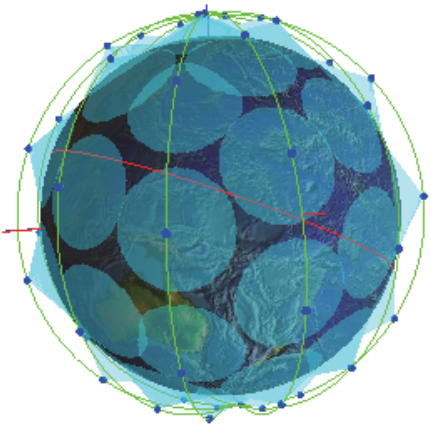


Рис. 5. Трехмерная визуализация ОГ КА «Iridium NEXT»

Модель Земли задается общим земным эллипсоидом с основными геометрическими параметрами: большая полуось a = 6378,136 км и коэффициент сжатия α = 1/298,258.

Модель канала передачи информации учитывает расстояние $d(s_i, r_j)$ от источника излучения s_i до радиоприемного устройства r_j, временные задержки Δt на распространение сигналов, ослабление

$L_1(d_{ij}, f)$ энергии сигнала в свободном пространстве, где $d_{ij} = d(s_i, r_j)$, f — частота передачи, потери $L_2(f, \beta)$ в газах атмосферы, потери $L_3(\beta, f, h_0, J)$ в гидрометеорах (в специально оговоренных случаях), потери $L_4(\psi)$ из-за расхождения плоскостей поляризации передающей и приемной антенн, потери $L_5(\varphi)$ из-за эффекта Фарадея, характеристики приемопередающих устройств, влияющие на мощность излучаемых сигналов и вероятность безошибочного декодирования сообщений, а также эффект Доплера.

Таким образом, мощность e' сигнала на входе радиоприемного устройства (РПУ) определяется исходя из выражения:

$$e' = e + L_1(d, f) + L_2(f, \beta) + L_3(f, \beta, h_0, J) + \\ + L_4(\varphi) + L_5(\psi) + G_{\text{прд}} + G_{\text{прм}}, \tag{7}$$

где $G_{\text{прд}}$, $G_{\text{прм}}$ — коэффициенты направленного действия передающей и приемной антенн соответственно.

Угол β определяется на каждом шаге моделирования и соответствует углу места КА в СК $X_{\text{Св}}Y_{\text{Св}}Z_{\text{Св}}$, начало которой совпадает с центром масс источника сообщений s_i, определяемой геодезической широтой B_i и долготой L_i и высотой H_i над поверхностью земли. Ось $X_{\text{Св}}$ направлена в сторону Северного полюса Земли по касательной к меридиану корабля s_j; ось $Y_{\text{Св}}$ — по внешней нормали к земному эллипсоиду, а ось $Z_{\text{Св}}$ дополняет систему до правой [25].

Коэффициент $G_{\text{прд}}$ направленного действия определяется для вычисленного на предыдущем шаге угла места β и диаграммы направленности $\mu = \beta \to G_{\text{прд}}$ типовой авиационной антенны, приведенной в [26].

Диаграмма направленности спутниковой антенны аппроксимируется зависимостью:

$$G_{\text{прм}}(\theta) = G_{\text{МВ}} - 12(\theta/\theta_{-3\text{dB}})^2, \tag{8}$$

где $G_{\text{прм}}(\theta)$ — коэффициент усиления спутниковой антенны (дБи) при угле отклонения от оси θ (градусов); $G_{\text{МВ}}$ — коэффициент усиления в главном луче спутниковой антенны (дБи); $\theta_{\text{-3dB}}$ — ширина луча спутниковой антенны по уровню –3 дБ (градусов).

Зависимость вероятности $p_e(m)$ ошибочного декодирования сообщений, состоящих из m бит, от коэффициента SNR сигнал/шум задана отображением $\varphi = SNR \to p_e(m)$, например, для идеального когерентного и некогерентного приема сигналов с амплитудной модуляцией соответствующие зависимости приведены в книге [27].

Модель распределения сигналов основана на данных, опубликованных Международным союзом электросвязи (комиссией по радиосвязи) в работе [26]. Модель включает сигналы, передаваемые воздушными судами на частоте 1090 МГц (Mode A/C, Mode S, Mode S ADS-B) (таблица 2).

Принимая во внимание две авиационные антенны, расположенные в верхней и нижней частях фюзеляжа, и пренебрегая отраженными от земли сигналами, переданными нижней антенной, следует считать, что излучаются только 40% от указанного в таблице количества сигналов типов MODE A/C, MODE S и 3 сигнала ADS-B в секунду [26, 28].

Таблица 2. Распределение частоты и длительности сигналов, передаваемые воздушными судами на частоте 1090 МГц

Тип воздушного судна	Характеристика		
	Тип сигналов	Максимальная длительность (мкс)	Количество сообщений в секунду
Отвечающие стандартам ИКАО	Режим A/C	20,3	0−120
	Все вызовы в режиме S	64	0−60
	Короткое сообщение в режиме S	64	6−40
	Длинное сообщение в режиме S	120	6−20
	ADS-B/1090 ES	120	6
Не отвечающие стандартам ИКАО	Короткий импульс ВОРЛ	3,5	6−40
	Длинный импульс ВОРЛ	35	6−20

Модель функционирования систем контроля воздушной обстановки реализована в виде комплекса алгоритмов, взаимосвязь которых схематически изображена на рисунке 6.

Комплекс состоит из следующих алгоритмов:

1. Алгоритм статистического моделирования источников излучений.

2. Алгоритм моделирования источников излучений с использованием данных интернет-ресурсов наблюдения за полетами воздушных судов.

3. Алгоритм детектирования сигналов (MODE A/C, MODE S, ADS-B) в секундном интервале t_k.

4. Алгоритм моделирования приема сообщений КА (наземным пунктом контроля) в секундном интервале $t_k \in [t_0, t_0 + T]$.

5. Алгоритм моделирования работы космической системы (сети наземных пунктов) контроля воздушной обстановки.

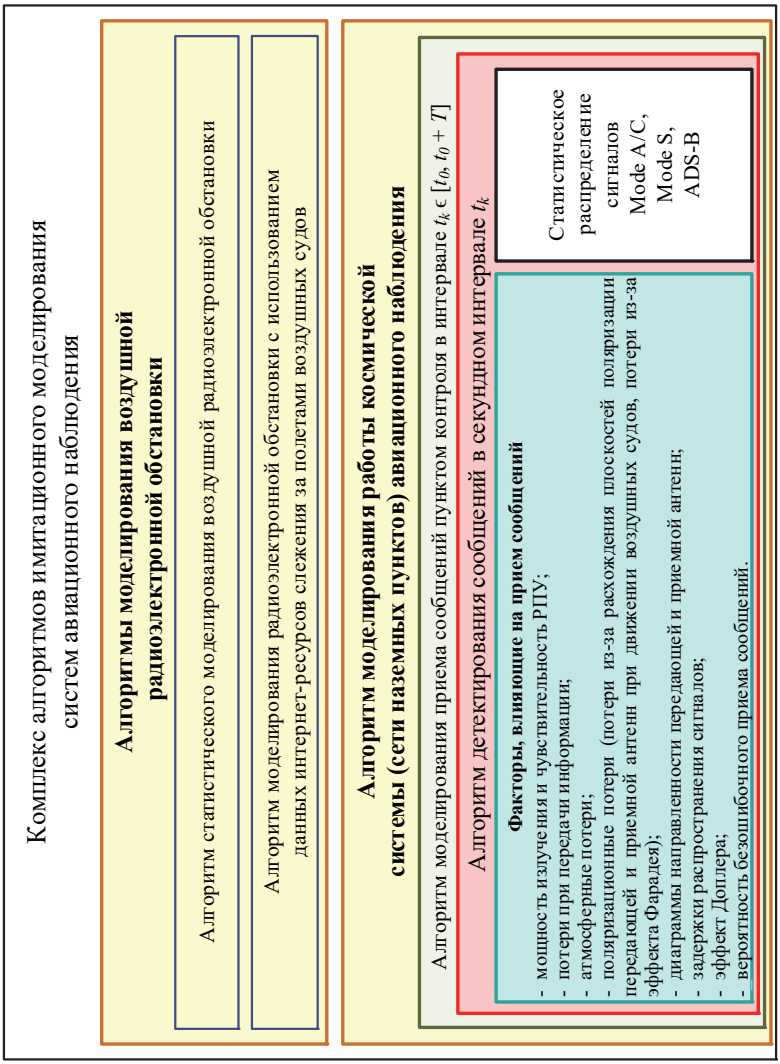


Рис. 6. Комплекс алгоритмов имитационного моделирования функционирования систем авиационного наблюдения

Разработанный комплекс алгоритмов программно реализован. Исходными данными для моделирования космической системы (рассмотрен более сложный случай по сравнению с сетью наземных пунктов) являются время t_0 начала и интервал T моделирования, количество КА в ОГ $R = \{r_j, j = 1,..., N_R\}$ и параметры их орбит, множество $S = \{s_i, i = 1,..., N_S\}$ воздушных судов и маршруты их движения. Перед началом моделирования каждому из воздушных судов назначается его тип (MODE A/C, MODE S или ADS-B) и мощность e излучения сигналов в зависимости от заданной модели распределения различных типов ретрансляторов по парку воздушных судов в мире [26]. В соответствии с рекомендациями международного союза электросвязи мощность e сигнала, излучаемого авиационным транспондером, принимает значения из множества { 21 дБВт, 24 дБВт, 27 дБВт, 29 дБВт }.

В процессе работы программного комплекса для каждого момента времени $t_k \in [t_0, t_0 + T]$ выполняется следующая последовательность шагов:

1) Рассчитываются текущие положения КА $r_j \in R$, $j = 1,..., N_R$, входящих в состав ОГ, по заданным начальным значениями оскулирующих элементов орбиты и координаты воздушных судов;

2) Для КА $r_j \in R$, $j = 1,..., N_R$ определяется энергетическая доступность $\nu_r(s_i, r_j)$ источников излучений $s_i \in S$, $i = 1,..., N_S$, исходя из условий превышения минимально допустимого угла места β (s_i, r_j) КА r_j в системе координат, связанной с местоположением анализируемого источника излучений s_i, и требуемого потока мощности сигналов авиационных систем связи на входе приемника r_j с учетом заданной системы ограничений и допущений (моделей потерь, диаграмм направленности антенн и т.д.);

3) Для всех объектов наблюдения $s_i \mid \nu_r(s_i, r_j) = 1$, находящихся в зоне покрытия КА r_j, и момента t, используя статистическую модель распределения частоты излучения и длительности сигналов, передаваемых воздушными судами на частоте 1090 МГц (таблица 2), формируется список $\mathbf{M}_j = \{M_{ji} : s_i \mid \nu_r(s_i, r_j) = 1\}$ сообщений, содержащих идентификационную (все типы сообщений MODE A/C, MODE S, ADS-B) и навигационную (только ADS-B) информацию с привязкой к источнику излучения s_i. Моменты начала передачи сообщений генерируется датчиком случайных чисел из определенного диапазона;

4) Для каждого из сообщений в списке $\mathbf{M}_j$ проверяется условие наличия коллизии с сообщениями других судов $s_k \in S$, $s_i \neq s_k$, учитывая задержки распространения сигналов от различных источников, потери мощности и так далее. Вычисляется вероятность безошибочного прие-

ма сообщения с амплитудной модуляцией для отношения сигнал/шум, где энергия шума при наличии интерференции сигналов соответствует максимальной энергии одного из «мешающих» сообщений;

5) В случае правильного детектирования сообщения $m_z \in M_{ji}$ (любого из типов MODE A/C, MODE S, ADS-B) источник s_i, его передавший, считается идентифицированным, кроме того, если сообщение имеет тип ADS-B — локализованным;

6) Формируется информация как интегрированная за интервал моделирования $t \in [t_0, t_k]$, так и текущая, относящаяся к моменту времени t_k.

Результатом работы комплекса алгоритмов являются целевые показатели функционирования систем авиационного наблюдения, к которым следует отнести:

– зависимость количества $n_1(t, S)$ переданных сообщений судами, находившимися в зоне покрытия хотя бы одного КА (зоне радиовидимости наземного пункта) от времени моделирования t, то есть максимально возможное (потенциальное) количество принятых сообщений;

– зависимость количества $n_2(t, S)$ переданных сообщений, содержащих координаты местоположения, судами, находившимися в зоне покрытия хотя бы одного КА (зоне радиовидимости наземного пункта) от времени моделирования t;

– зависимость количества $n_3(t, S)$ детектированных сообщений от времени моделирования t;

– зависимость количества $n_4(t, S)$ детектированных сообщений, содержащих опознавательный код и координаты местоположения, то есть сообщений типа ADS-B, от времени моделирования t;

– зависимость количества $n_5(t, S)$ воздушных судов, находившихся в зоне покрытия одного из КА (наземных пунктов авиационного наблюдения) и передавших хотя бы одно сообщение, от времени моделирования t, то есть максимально возможное количество идентифицированных воздушных судов;

– зависимость количества $n_6(t, S)$ воздушных судов, находившихся в зоне покрытия одного из КА (наземных пунктов авиационного наблюдения) и передавших хотя бы одно сообщение, содержащее координаты местоположения, от времени моделирования t, то есть максимально возможное количество локализованных воздушных судов;

– зависимость количества $n_7(t, S)$ идентифицированных судов от времени моделирования t;

– зависимость количества $n_8(t, S)$ обнаруженных судов, то есть судов с локализованным местоположением, от времени моделирования t;

– среднее количество $N_1(T, S)$ переданных сообщений за интервал моделирования T;

– среднее количество $N_2(T, S)$ переданных сообщений, содержащих координаты местоположения, то есть сообщений типа ADS-B, за интервал моделирования T;

– среднее количество $N_3(T, S)$ безошибочно принятых сообщений за интервал моделирования T;

– среднее количество $N_4(T, S)$ безошибочно принятых сообщений ADS-B, содержащих опознавательный код и координаты местоположения, за интервал моделирования T;

– среднее количество $N_5(T, S)$ судов, передавших хотя бы одно сообщение, за интервал моделирования T;

– среднее количество $N_6(T, S)$ судов, передавших хотя бы одно сообщение типа ADS-B, за интервал моделирования T;

– среднее количество $N_7(T, S)$ идентифицированных судов за интервал моделирования T;

– среднее количество $N_8(T, S)$ обнаруженных судов за интервал моделирования T;

– среднее количество $N_8(T, S, z)$ обнаруженных судов в зависимости от региона z наблюдения за интервал моделирования T.

Все перечисленные зависимости $n_i(t, S)$, $i = 1,\ldots, 8$ также могут быть вычислены в зависимости от витка l (для КА) или другого заданного интервала времени. На практике функциональные зависимости $n_1(t, S)$, $n_2(t, S)$, $n_5(t, S)$, $n_6(t, S)$ и значения $N_1(T, S)$, $N_2(T, S)$, $N_5(T, S)$, $N_6(T, S)$ неизвестны, однако при моделировании систем сбора сведений о движении воздушных судов отношения $n_3(t,S)/n_1(t,S)$, $n_4(t,S)/n_2(t,S)$, $n_7(t,S)/n_5(t,S)$, $n_8(t,S)/n_6(t,S)$ могут говорить о конструктивном и программном превосходстве одной системы по сравнению с другой или выбора наилучшей структуры и параметров.

Имея оценки перечисленных базовых характеристик, можно определить производные характеристики, например требуемые пропускную способность канала связи или объем ЗУ бортовой аппаратуры КА.

Следует отметить, что достоинством разработанной модели является универсальность. Если задать в качестве частных моделей радиоэлектронной обстановки и распределения сигналов модель распределения источников излучения сигналов АИС и модель планирования слотов для передачи сообщений соответственно, представленный модельно-алгоритмический комплекс можно использовать для оценивания целевых показателей функционирования систем автоматической идентификации морских судов [29, 30].

5. Результаты моделирования функционирования космической системы авиационного наблюдения. С целью *оценивания ста-*

тистических характеристик энергетических параметров сигналов в зоне обзора КА используется статическая модель функционирования КА (без учета орбитального движения) и предполагается равномерное распределение воздушных судов на заданной высоте h в области прямой видимости КА, что позволяет оценить экстремальные (минимальное и максимальное) значения мощности сигнала на входе детектора при различных значениях мощности передатчика и запас по мощности с учетом различных значений чувствительности радиоприемного устройства.

Статистические характеристики энергетических параметров сигналов для КА с высотой орбиты h_r = 500 км и источников излучений, размещенных на высоте h_s = 10 км над поверхностью земли, представлены в таблице 3. Максимальное усиление $G_{\text{МВ}}$ спутниковой антенны принималось равным 6 дБ, ширина диаграммы направленности по уровню половинной мощности $\theta_{\text{-3dB}}$ = 20º, угол направления максимального усиления относительно оси симметрии 40º.

В таблице в графе «Запас по мощности» через черту указаны значения для приемников с чувствительностями –134 и –117 дБ соответственно [31]. Из анализа таблицы видно, что на границе зоны видимости КА сигналы являются энергетически недоступными как для приемника с чувствительностью –117 дБ, так и с –134 дБ. В лучшем случае для приемников с чувствительностью –134 дБ имеется гарантированный запас мощности для любого класса передающих устройств (вероятность битовой ошибки менее 10^{-3}), для приемников с чувствительностью –117 дБ сигналы являются практически недоступными, для передатчиков с максимальной мощностью отношение сигнал/шум составляет всего 4,60 дБ, что соответствует низкой вероятности безошибочного приема сообщений.

Таблица 3. Статистические характеристики энергетических параметров сигналов систем авиационной наблюдения для линии связи между воздушным судном и КА

Характеристика	Максимум суммарных потерь	Минимум суммарных потерь
Суммарные потери, дБ	–186,17	–141,40
Потери $L_1(d_{ij}, f)$ мощности на распространение, дБ	–162,44	–149,50
Потери $L_4(\varphi_{ij})$ из-за расхождения плоскостей поляризации приемной и передающей антенн, дБ	–0,86	–0,02
Угол ψ поворота плоскости поляризации, град.	64,62	32,05
Потери $L_5(f, \beta)$ из-за эффекта Фарадея, дБ	–7,36	–1,44
Коэффициент $G_{\text{прд}}$ усиления передающей антенны, дБи	2,08	3,56

Продолжение таблицы 3

Характеристика		Максимум суммарных потерь	Минимум суммарных потерь
Коэффициент $G_{прм}$ усиления спутниковой антенны, дБи		–17,59	6,00
Запас по мощности, дБ	21	–31,17/–48,17	14,60/–3,40
	24	–28,17/–45,17	17,60/–0,40
	27	–25,17/–42,17	20,60/2,60
	29	–23,17/–40,17	22,60/4,60
Характеристики, поясняющие пространственное положение ИРИ относительно пункта наблюдения			
Расстояние d (s_i, r_j) от источника s_i до приемника r_j, км		2915,38	657,20
Угол β возвышения над горизонтом, град		–3,04	46,28
Угол φ (s_i, r_j) между плоскостями поляризаций приемной и передающей антенн, град.		25,04	3,79
Угол θ отклонения луча от оси спутниковой антенны, град.		68,04	39,93

Оценивание *количества* $v_r(S, t)$ *источников излучения в зоне покрытия КА* проводилось с использованием статической и динамической моделей воздушной радиоэлектронной обстановки:

1. Моделирование движения КА — приемника сигналов авиационных систем связи с параметрами орбиты: долгота восходящего узла $\Omega = 0$ град.; наклонение $i = 98$ град.; апогей $r_A = 500$ км.; перигей $r_П = 500$ км.; широта перицентра $\omega = 0$ град. Интервал моделирования составляет сутки ($T = 86400$ с.), время t_0 начала моделирования соответствует положению КА на орбите с истинной аномалией $\theta_0 = 0$ град. Модель воздушной радиоэлектронной обстановки включает более 8000 тыс. судов $s_i \in S$, $i = 1,...,N_S$, местоположения которых не изменяются за время моделирования [14].

На рисунке 7 изображены графики количества судов $\nu(S, t)$, находящихся в зоне обзора КА r_j в момент времени t, и количества судов $\nu_r(S, t)$, находящихся в зоне обзора КА с учетом их энергетической доступности.

Для расчета электромагнитной доступности модель распределения типов авиационных ретрансляторов по их мощности была заимствована из работы [26], пороговое значение для обнаружения принималось равным –134 дБВт, максимальное усиление спутниковой антенны $G_{МВ} = 6$ дБ, ширина диаграммы направленности по уровню половинной мощности $\theta_{-3dB} = 20$ град.

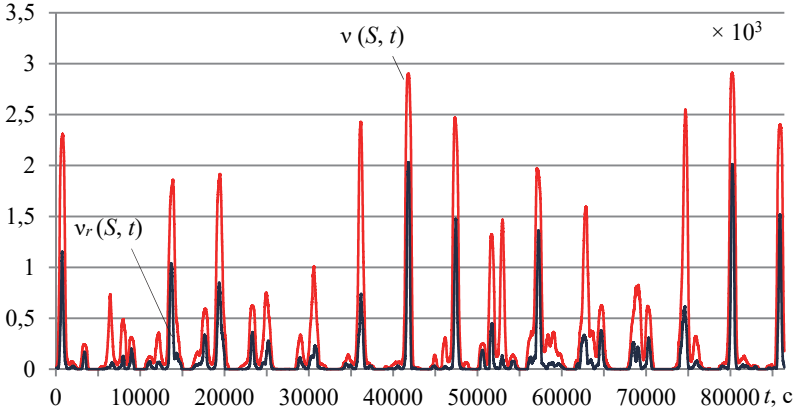


Рис. 7. Количество судов ν (*S*, *t*), попавших в зону обзора КА, и судов $\nu_r(S, t)$ в зоне обзора с учетом энергетической доступности

С целью сокращения объема излагаемого материала расчет остальных целевых показателей производится только для динамической модели воздушной РЭО.

2. Основным отличием условий моделирования по сравнению с предыдущим пунктом является применение динамической модели воздушной РЭО, включающей данные о полетах более чем 30 тысяч воздушных судов $s_i \in S$, $i = 1,...,N_S$ и более чем 120 тысяч рейсах $f_{ij} \in F_i$, $j = 1,...,n_i$, $n_i = \mathrm{card} < F_i >$.

Динамическая модель воздушной РЭО реализуется посредством смены местоположений воздушных судов $s_i \mid a_i(t_k) = 1$, где $a_i(t_k)$ — индикатор активности источника s_i в момент времени t_k, через равные интервалы обновления T_u, то есть при выполнении простейшего условия t_k % $T_u \equiv 0$ (% — операция деления по модулю). На рисунке 8 представлен график изменения текущего количества $a(S, t)$ активных источников, определяемого выражением:

$$a(S,t_k) = \sum_{s_i \in S} a_i(t_k), \; t_k \in [t_0, t_0 + T]. \tag{9}$$

Графики изменений количества $\nu_r(S, t)$ воздушных судов в зоне покрытия КА и коэффициента энергетической доступности $k = \nu_r(S,t)/\nu(S,t)$ от текущего положения КА на орбите изображены на рисунке 9.

Визуальное сравнение графиков на рисунках 7 и 9а показывает существенное различие в характере их поведения, что говорит о воз-

можности использования статической модели РЭО только на коротких интервалах моделирования.

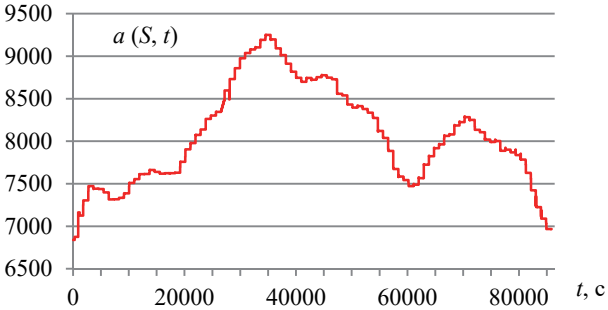


Рис. 8. Количество $a\ (S, t)$ активных источников излучений сигналов на частоте 1090 МГц

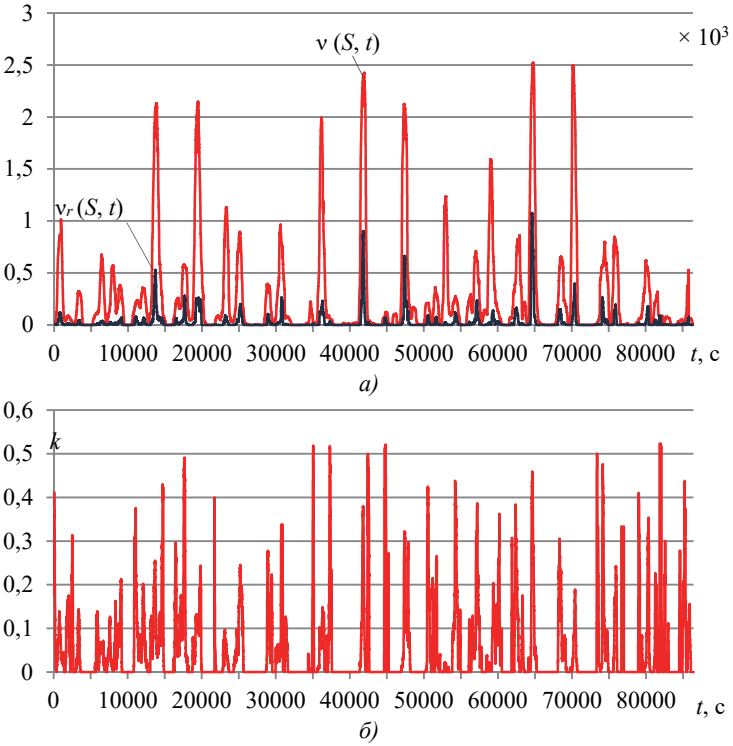
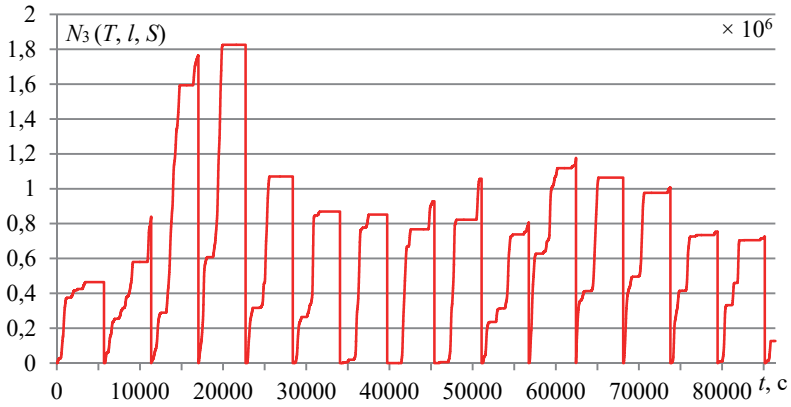


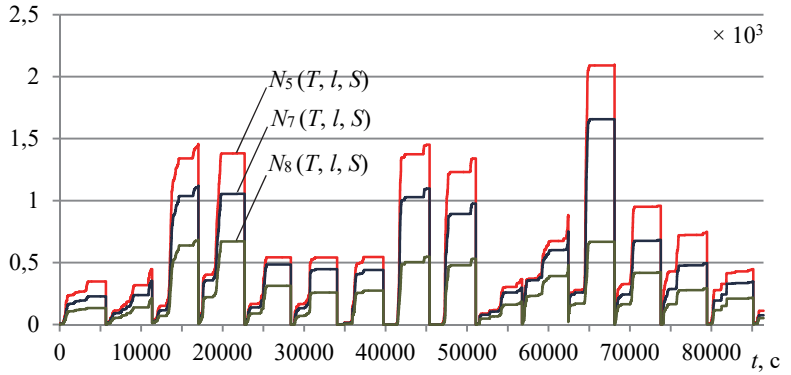
Рис. 9. Анализ радиовидимости: а) количество судов $\nu\ (S, t)$, попавших в зону обзора КА, и судов $\nu_r\ (S, t)$ в зоне обзора с учетом энергетической доступности; б) коэффициент энергетической доступности

Из анализа графика на рисунке 9б видно, что коэффициент k энергетической доступности не превосходит 50%. В то время как количество судов в зоне обзора КА достигает 3000 единиц, доступными (с энергетической точки зрения) будет не более 1500 тыс.

Результаты моделирования представлены на рисунке 10.



а)



б)

Рис. 10. Результаты моделирования: а) количество $N_3(t, l, S)$ детектированных сообщений всех типов; б) количество $N_5(t, l, S)$ потенциальных, количество $N_7(t, l, S)$ идентифицированных и количество $N_8(t, l, S)$ локализованных ВС

Общее количество моделируемых объектов составляет $N_S = 31666$, максимально возможное (потенциальное) количество

идентифицируемых воздушных судов составляет $N_5(T, S) = 13146$. Отличие примерно в два раза значений N_S и $N_5(T, S)$ обусловлено в основном кратковременностью излучений сигналов или времени полета воздушных судов (от 15 минут), что подтверждается на рисунке 2а. Минимальное значение количества $N_3(T, l, S)$ безошибочно принятых сообщений всех типов за один виток составляет 464907 ($l = 1$), максимальное — 1825714 ($l = 4$), среднее — 1029203, минимальное количество $N_7(T, l, S)$ идентифицированных судов составляет 229 ($l = 1$), максимальное — 1677 ($l = 12$), среднее — 803 штук, минимальное количество $N_8(T, l, S)$ локализованных судов — 114 ($l = 1$), максимальное — 650 ($l = 4$), среднее — 377. Общее количество детектированных сообщений за время моделирования ($T = 86400$ с) составляет $N_3(T, S) = 15205831$, идентифицированных судов $N_7(T, S) = 9198$, локализованных $N_8(T, S) = 4919$.

Результаты моделирования показывают, что вследствие интерференций сигналов существенно снижается количество безошибочно принятых сообщений и идентифицированных воздушных судов [6]. Для разрешения этой проблемы требуются дополнительные схемно-технические решения, например, пространственно-разнесенного приема.

Анализ полученных результатов позволяет определить зависимость среднего количества обнаруженных судов от региона и времени наблюдения T. Очевидно, что наиболее трудными с точки зрения обнаружения воздушных судов, как и ожидалось, являются европейский, китайский и североамериканский регионы. Вероятности обнаружения при указанных выше условиях моделирования за сутки равны 0,41, 0,34 и 0,18 (значения получены с учетом общего количества моделируемых объектов равного 31666) соответственно.

6. Результаты моделирования функционирования наземной системы авиационного наблюдения. *Статистические характеристики энергетических параметров сигналов* систем авиационного наблюдения для воздушных судов, размещенных на высоте $h = 10$ км, представлены в таблице 4.

Значения получены путем статистического анализа энергетических характеристик переданных сигналов объектами, равномерно распределенными в зоне обзора пункта контроля. В вычислениях были приняты следующие допущения: вертикальная поляризация приемной антенны, максимальное усиление $G_{\text{МВ}} = 3$ дБ, ширина диаграммы направленности в вертикальной плоскости $\theta_{-3\text{dB}} = 160°$, интенсивность осадков $J = 59$ мм/ч, высота изотермы 0° над уровнем моря $h_0 = 4{,}5$ км.

Таблица 4. Статистические характеристики энергетических параметров сигналов системы авиационного наблюдения для линии связи между ВС и НП

Характеристика		Без учета осадков		С учетом осадков	
		Максимум суммарных потерь	Минимум суммарных потерь	Максимум суммарных потерь	Минимум суммарных потерь
Суммарные потери, дБ		-152,40	-110,95	-152,91	-110,95
Потери $L_1(d_{ij}, f)$ мощности на распространение, дБ		-144,08	-114,33	-144,08	-114,33
Потери $L_2(f, \beta)$ в газах атмосферы, дБ		-1,31	-0,02	-1,31	-0,02
Потери $L_3(\beta, f, h_0, J)$ в дожде, дБ		0,00	-0,00	-0,51	-0,003
Потери $L_4(\varphi_{ij})$ из-за расхождения плоскостей поляризации приемной и передающей антенн, дБ		-0,01	0,00	-0,01	0,00
Коэффициент $G_{\text{прм}}$ усиления приемной антенны, дБи		3,00	-1,56	3,00	-1,56
Коэффициент $G_{\text{прд}}$ усиления авиационной антенны, дБи		-10,00	4,96	-10,00	4,96
Запас по мощности, дБ	21	2,60/-14,40	44,05/27,05	2,09/-14,91	44,05/27,05
	24	5,60/-11,40	47,05/30,05	5,09/-11,91	47,05/30,05
	27	8,60/-8,40	50,05/33,05	8,09/-8,91	50,05/33,05
	29	10,60/-6,40	52,05/35,05	10,09/-6,91	52,05/35,05
Характеристики, поясняющие пространственное положение ИРИ относительно пункта наблюдения					
Расстояние d (s_i, r_j) от источника s_i до приемника r_j, км		351,83	114,51	351,83	114,51
Угол β возвышения над горизонтом, град		0,00	60,81	0,00	60,81
Угол θ отклонения луча от оси главного лепестка авиационной антенны, град		90,00	29,14	90,00	29,14

Из анализа таблицы видно, что на границе зоны прямой видимости (угол $\beta = 0°$) для РПУ с чувствительностью –117 дБВт сигналы

являются практически недоступными, для РПУ с чувствительностью –134 дБВт и излучений мощностью 29 дБВт без учета осадков чистый запас 10,60 дБ обеспечивает безошибочный прием 112-битового сообщения (MODE S ES) с амплитудной модуляцией с вероятностью 0,95 (при когерентной демодуляции), с учетом осадков запас 10,09 дБ — с вероятностью 0,92, для излучений мощностью 27 и менее вероятность правильного приема составляет менее 0,66. В лучшем случае (взаимном расположении источника и приемника сообщений) имеется гарантированный запас мощности при использовании любого класса передающих и приемных устройств для декодирования сообщений с вероятностью ошибки $p_e < 10^{-3}$.

Оценивание *количества* $v_r(S, t)$ *источников излучения в зоне радиовидимости сети наземных пунктов* $r_j \in R$, $j = 1,..., N_R$. Зоны радиовидимостидля наземных пунктов контроля определялись следующим образом: сеть источников s_i излучений сигналов мощностью e (мощность e принимается равной 21, 24, 27 или 29 дБВт [26]) располагалась на высоте h над интересующим регионом с заданными приращениями по широте ΔB и долготе ΔL. Объект s_i считался наблюдаемым $\rho_i = 1$ для заданного пункта r_j контроля, если угол β (s_i, r_j) возвышения над горизонтом объекта s_i для пункта r_j контроля больше нуля и мощность переданного сигнала на входе радиоприемного устройства больше порогового значения чувствительности.

Визуализация зон прямой видимости сети наземных пунктов $r_j \in R$, $j = 1,..., N_R$, соотнесённых с местоположением 15 наиболее загруженных аэропортов в европейском регионе, с помощью ГИС OpenStreetMap представлена на рисунке 11.

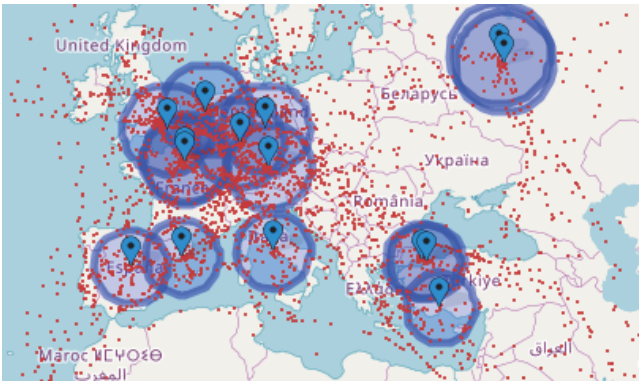


Рис. 11. Визуализация зон видимости наземных пунктов наблюдения для ВС на высоте h = 10 км на картографической основе

На рисунке 12 показаны графики изменения количества ν (*S*, *t*) и ν_r (*S*, *t*). В моделировании чувствительность РПУ принималась равной –117 дБ, осадки не учитывались. Общее количество обнаруженных объектов за время моделирования составляет 4391.

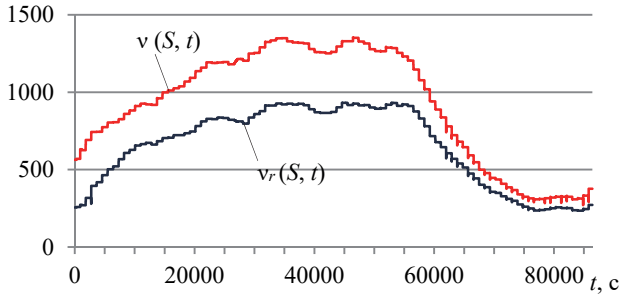


Рис. 12. Количество воздушных судов в зоне видимости сети *R*

На рисунке 13 представлены числовые показатели функционирования наземной сети, смоделированные по данным открытых источников информации, например, полученных путем web-скреппинга (парсинга интернет-сайтов Flightradar24, Planefinder, ADS-B Exchange и т.д.).

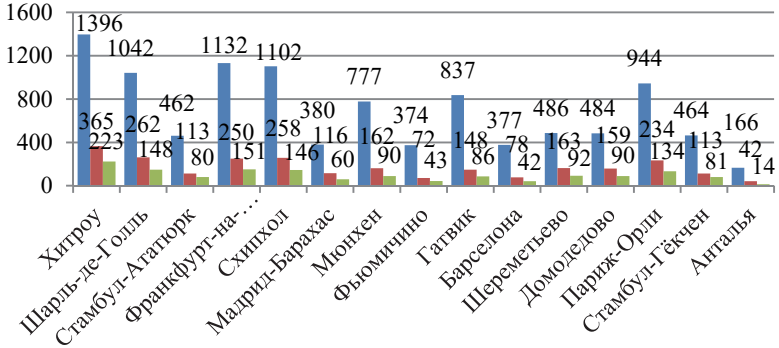


Рис. 13. Показатели N_8 (*T*, *S*), max [ν (*t*, *S*)], E [ν (*t*, *S*)] функционирования сети наземных пунктов r_j, *j* = 1,…, 15

На диаграмме, представленной на рисунке 13, каждому моделируемому пункту наблюдения $r_j \in R$, $j = 1,...,N_R$ соответствует три характеристики, а именно, начиная с левого столбца, общее количество N_8 (*T*, *S*) обнаруженных объектов за время моделирования *T* =

86400 с, максимальное количество max [ν (*t*, *S*)] одновременно наблюдаемых объектов пунктом r_j, среднее количество E [ν (*t*, *S*)] объектов в зоне радиовидимости пункта r_j. С уменьшением чувствительности РПУ и появлением дополнительных факторов, влияющих на ослабление мощности сигналов, например осадков, перечисленные характеристики снижаются. Результаты моделирования совпадают с данными анализа воздушного движения, предоставленными провайдером аэронавигационных услуг «National Air Traffic Services» [32].

Верификация имитационной модели функционирования космической системы контроля движения воздушных судов проводилась с использованием макета бортовой специальной аппаратуры приема и обработки сигналов системы АЗН–В, размещенного на борту летно-подъемного средства, путем сравнения полученных экспериментальных значений, в частности количества переданных сообщений судами в заданном территориальном районе, количества безошибочно принятых сообщений, количества обнаруженных объектов за заданный интервал времени с результатами имитационного моделирования.

В целом моделирование систем авиационного наблюдения позволяет сформировать технические решения при их проектировании, обосновать принципы планирования применения орбитальной группировки космических аппаратов, определить такие характеристики системы, как частоты обновления информации о местоположении судов, интервала времени от момента получения сообщения до доставки его потребителю, требования к объему бортового запоминающего устройства и скорости передачи информации по каналу связи.

7. Заключение. Разработана и программно реализована имитационная модель, позволяющая получать оценки целевых показателей функционирования космических и наземных систем идентификации и определения местоположения воздушных судов с учетом различных пространственных и энергетических факторов и условий распространения радиосигналов, а также реального размещения контролируемых объектов в мировом воздушном пространстве.

Имитационная модель функционирования может быть использована для подготовки исходных данных и обоснования тактико-технических требований для проектируемых опытных образцов космической и наземной техники, в частности, задаваясь местоположением наземных пунктов приема и обработки информации, определить требования к объему бортового запоминающего устройства и скорости передачи информации по спутниковому каналу связи.

С целью уменьшения количества интерференций сообщений от различных источников в зоне обзора КА и повышения полноты сведений о движении воздушных судов на определенных типах КА, в част-

ности типа «Iridium NEXT», устанавливаются активные фазированные антенные решетки, поэтому моделирование приема сигналов авиационных систем связи с учетом диаграмм направленности спутниковых антенн такого типа является одним из направлений дальнейшей работы.

В случае больших объемов данных, моделирование для ОГ, состоящих из 15 и более КА, может занимать достаточно продолжительное время, несмотря на распараллеливание вычислений для каждого КА отдельно с учетом ресурсов процессора (количества ядер). По этой причине в дальнейшем планируется разработать алгоритмы моделирования с использованием технологии массивно-параллельных вычислений.

Литература

1. *Barson J.V.* Automatic Dependent Surveillance-Broadcast – The First Step in the FAA's Next-Generation Air Transportation System // Aviation, Space and Environmental Medicine. 2009. vol. 80. no. 4. pp. 422–423.
2. *Zhang J., Wei L., Yanbo Z.* Study of ADS-B Data Evaluation // Chinese Journal of Aeronautics. 2011. vol. 24. no. 4. pp. 461–466.
3. *Strohmeier M., Schäfer M., Lenders V., Martinovic I.* Realities and Challenges of NextGen Air Traffic Management: The Case of ADS-B // IEEE Communications Magazine. 2014. vol. 52. no. 5. pp. 111–118.
4. *Schäfer M. et al.* OpenSky Report 2016: Facts and Figures on SSR Mode S and ADS-B Usage // 35th IEEE/AIAA Digital Avionics Systems Conference (DASC). 2016. pp. 1–9.
5. *Carandente M., Rinaldi C.* Aireon surveillance of the globe via satellite // Tyrrhenian International Workshop on Digital Communications – Enhanced Surveillance of Aircraft and Vehicles (TIWDC/ESAV). 2014. pp. 53–55.
6. *Van Der Pryt R., Vincent R.* A Simulation of Signal Collisions over the North Atlantic for a Spaceborne ADS-B Receiver Using Aloha Protocol // Positioning. 2015. vol. 6. no. 03. pp. 23–31.
7. *Van Der Pryt R., Vincent R.* A Simulation of the Reception of Automatic Dependent Surveillance-Broadcast Signals in Low Earth Orbit // International Journal of Navigation and Observation. 2015. 11 p.
8. *Van Der Pryt R., Vincent R.* The CanX-7 Nanosatellite ADS-B Mission: A Preliminary Assessment // Positioning. 2017. vol. 08. pp. 1–11.
9. *Brodsky Y., Rieber R., Nordheim T.* Balloon-borne air traffic management (ATM) as a precursor to space-based ATM // Acta Astronautica. 2012. vol. 70. pp. 112–121.
10. *Knudsen B.G. et al.* ADS-B in space: Decoder implementation and first results from the GATOSS mission // 14th Biennial Baltic Electronic Conference (BEC). 2014. pp. 57–60.
11. Сайт компании «AnyLogic». URL: https://www.anylogic.ru (дата обращения: 19.05.2018).
12. Сайт компании «Элина компьютер». URL: http://elina-computer.ru/static/gpss-world.html (дата обращения: 19.05.2018).
13. Сайт ФГУП «ГосНИИАС». URL: https://www.gosniias.ru/ (дата обращения: 19.05.2018).
14. *Скороходов Я.А., Малышев Д.В.* Анализ энергетической доступности сигналов системы АЗН-В для низкоорбитальных космических аппаратов с использованием статистического моделирования // Информация и космос. 2017. № 4. С. 137–141.
15. *Baek K., Bang H.* ADS-B based Trajectory Prediction and Conflict Detection for Air Traffic Management // International Journal of Aeronautical and Space Sciences. 2012. vol. 13. no. 3. pp. 377–385.

16. *Xi L., Jun Z., Yanbo Z., Wei L.* Simulation Study of Algorithms for Aircraft Trajectory Prediction Based on ADS-B Technology // 7th International Conference on System Simulation and Scientific Computing. 2008. pp. 322–327.
17. *Zhang K., Qiao Y., Zhang C.* Trajectory Tracking Using Auto-adaptive Multi-model Filtering Method in ADS-B System // Second International Conference on Robot, Vision and Signal Processing. 2013. pp. 93–97.
18. *Thipphavong D.P., Schultz C.A., Lee A.G., Chan S.H.* Adaptive Algorithm to Improve Trajectory Prediction Accuracy of Climbing Aircraft // Journal of Guidance, Control, and Dynamics. 2013. vol. 36. no. 1. pp. 15–24.
19. *Allnutt J.E.* Satellite to Ground Radiowave Propagation: 2nd Edition // The Institution Of Engineering And Technology. 2010. 704 p.
20. *Сомов А.М., Корнев С.Ф.* Спутниковые системы связи // М.: Горячая линия-Телеком. 2012. 244 с.
21. *Кантор Л.Я., Ноздрин В.В.* Электромагнитная совместимость систем спутниковой связи // М.: НИИР. 2009. 280 с.
22. *Roy A.E.* Orbital Motion // CRC Press. 2004. 544 p.
23. *Gupta O.P.* Global Augmentation of ADS-B Using Iridium NEXT Hosted Payloads // Integrated Communications, Navigation and Surveillance Conference (ICNS). 2011. pp. 1–15.
24. *Garcia M., Dolan J., Hoag A.* Aireon's initial on-orbit performance analysis of space-based ADS-B // Integrated Communications, Navigation and Surveillance Conference (ICNS). 2017. pp. 4A1-1–4A1-8.
25. *Нариманов Г.С.* Основы теории полета космических аппаратов // М.: Машиностроение. 1972. 608 с.
26. Динамические статистические исследования по спутниковому приему сигналов ADS-B для глобального слежения за рейсами гражданской авиации // Всемирная конференция радиосвязи. 2015. 19 с. URL: https://www.itu.int/md/dologin_md.asp?lang=es&id=R15-WRC15-C-0100!!MSW-R (дата обращения: 17.05.2016).
27. *Sklar B.* Digital Communication. Fundamentals and Application: 2nd Edition // Prentice Hall. 2017. 1104 p.
28. *Van Der Pryt R., Vincent R.* A Simulation of Reflected ADS-B Signals over the North Atlantic for a Spaceborne Receiver // Positioning. 2016. vol. 7. pp. 51–62.
29. *Скороходов Я.А., Андреев А.М.* Моделирование функционирования космического сегмента системы автоматической идентификации морских судов // Информационно-управляющие системы. 2018. Т. 93. № 2. С. 36–48.
30. *Скороходов Я.А., Махров К.Б., Малышев Д.В.* Имитационная модель функционирования космической системы контроля движения морских судов // Труды Военно-космической академии имени А.Ф. Можайского. 2017. С. 23–33.
31. *Werner K., Bredemeyer J., Delovski T.* ADS-B over satellite: Global air traffic surveillance from space // Tyrrhenian International Workshop on Digital Communications – Enhanced Surveillance of Aircraft and Vehicles (TIWDC/ESAV). 2014. pp. 47–52.
32. Сайт издательского дома «Авиатранспортное обозрение». URL: http://www.ato.ru/blogs/ blog-alekseya-sinickogo/30-tysyach-aviareysov-v-sutki-nad-evropoy (дата обращения: 19.01.2016).

Скороходов Ярослав Анатольевич — к-т техн. наук, начальник лаборатории военного института (научно-исследовательского), Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: моделирование процессов и систем специального назначения, теория статистических решений, методы обработки и анализа измерительной информации, теория выбросов траекторий случайных процессов, системы искусственного интеллекта. Число научных публикаций — 30. yaroslavskor@gmail.com; ул. Красного курсанта, 18, Санкт-Петербург, 197198; р.т.: +7 (812) 347-9759.

DOI 10.15622/sp.61.2

Ya.A. Skorokhodov

SIMULATION OF SPACE AND GROUND-BASED AVIATION SURVEILLANCE SYSTEMS FUNCTIONING

Skorokhodov Ya. A. **Simulation of Space and Ground-Based Aviation Surveillance Systems Functioning.**

Abstract. At present, the orbital constellations of satellites with the possibility of receiving, processing and relaying the ADS-B («Automatic Dependent Surveillance — Broadcast») system signals that ensure globality and continuity of the air traffic monitoring are being created and gradually enhanced. In accordance with the concept of ADS-B technology usage, each air traffic participant broadcasts its identity, location, and status parameters in broadcast mode. Due to the fact that the system was not designed to receive signals onboard the satellite, there are certain problems related to their energy availability, the presence of collisions of messages from different sources, the effect of Doppler effect and other factors. Developed simulation model of aviation surveillance systems based on the reception of signals containing identification and navigation information and transmitted over the air in the broadcasting mode. Software-implemented simulation algorithms allow to set various constraints and assumptions (radiation sources distribution, aviation communication systems signals receiving points, information transmitting channel, signals frequency and duration distribution models) and obtain the target indicators estimates of the space and ground-based aviation surveillance systems functioning, taking into account various spatial and energy factors and conditions for the radio signals propagation, and controlled objects actual placement and their movement dynamics in the world airspace. Simulation model use methods and examples for calculating the space and ground aviation surveillance systems functioning target indicators are presented.

Keywords: mathematical modeling, aviation surveillance systems, automatic dependent surveillance - broadcasting, space systems, information processing.

Skorokhodov Yaroslav Anatolevich — Ph.D., head of the laboratory (research) of the Military Institute (research), Mozhaisky Military Space Academy. Research interests: processes and special purposes systems modeling, statistical solutions theory, measurement information processing and analysis methods, random processes trajectories emission theory, artificial intelligence system. The number of publications — 30. yaroslavskor@gmail.com; 18, Krasnogo Kursanta str., St. Petersburg, 197198, Russia; office phone: +7 (812) 347-9759.

References

1. Barson J.V. Automatic Dependent Surveillance-Broadcast – The First Step in the FAA's Next-Generation Air Transportation System. *Aviation, Space and Environmental Medicine*. 2009. vol. 80. no. 4. pp. 422–423.
2. Zhang J., Wei L., Yanbo Z. Study of ADS-B Data Evaluation. *Chinese Journal of Aeronautics.* 2011. vol. 24. pp. 461–466.
3. Strohmeier M., Schäfer M., Lenders V., Martinovic I. Realities and Challenges of NextGen Air Traffic Management: The Case of ADS-B. *IEEE Communications Magazine.* 2014. vol. 52. no. 5. pp. 111–118.
4. Schäfer M. et al. OpenSky Report 2016: Facts and Figures on SSR Mode S and ADS-B Usage. 35th IEEE/AIAA Digital Avionics Systems Conference (DASC). 2016. pp. 1–9.

5. Carandente M., Rinaldi C. Aireon surveillance of the globe via satellite. Tyrrhenian International Workshop on Digital Communications – Enhanced Surveillance of Aircraft and Vehicles (TIWDC/ESAV). 2014. pp. 53–55.
6. Van Der Pryt R., Vincent R. A Simulation of Signal Collisions over the North Atlantic for a Spaceborne ADS-B Receiver Using Aloha Protocol. *Positioning.* 2015. vol. 6. no. 03. pp. 23–31.
7. Van Der Pryt R., Vincent R. A Simulation of the Reception of Automatic Dependent Surveillance-Broadcast Signals in Low Earth Orbit. *International Journal of Navigation and Observation*. 2015. 11 p.
8. Van Der Pryt R., Vincent R. The CanX-7 Nanosatellite ADS-B Mission: A Preliminary Assessment. *Positioning.* 2017. vol. 08. pp. 1–11.
9. Brodsky Y., Rieber R., Nordheim T. Balloon-borne air traffic management (ATM) as a precursor to space-based ATM. *Acta Astronautica*. 2012. vol. 70. pp. 112–121.
10. Knudsen B.G. et al. ADS-B in space: Decoder implementation and first results from the GATOSS mission. 14th Biennial Baltic Electronic Conference (BEC). 2014. pp. 57–60.
11. Sajt kompanii «AnyLogic» [The site of the company «AnyLogic»]. Available at: https://www.anylogic.ru (accessed: 19.05.2018). (In Russ.).
12. Sajt kompanii «Elina kompiuter» [The site of the company «Elina computer»]. Available at: http://elina-computer.ru/static/gpss-world.html (accessed: 19.05.2018). (In Russ.).
13. Sajt FGUP «GosNIIAS» [The site of the company «State Recearch Instintute of Aviation Systems State Scientific Center of Russian Federation»]. Available at: https://www.gosniias.ru/ (accessed: 19.05.2018). (In Russ.).
14. Skorokhodov Ya., A. Malyshev D. V. [The Analysis of ADS-B Signals Energy Availability to Low-Space Satellites with Use of Statistical Modeling]. *Informatsiia i kosmos – Information and space*. 2017. vol. 4. pp. 143–147. (In Russ.).
15. Baek K., Bang H. ADS-B based Trajectory Prediction and Conflict Detection for Air Traffic Management. *International Journal of Aeronautical and Space Sciences.* 2012. vol. 13. no. 3. pp. 377–385.
16. Xi L., Jun Z., Yanbo Z., Wei L. Simulation Study of Algorithms for Aircraft Trajectory Prediction Based on ADS-B Technology. 7th International Conference on System Simulation and Scientific Computing. 2008. pp. 322–327.
17. Zhang K., Qiao Y., Zhang C. Trajectory Tracking Using Auto-adaptive Multi-model Filtering Method in ADS-B System. Second International Conference on Robot, Vision and Signal Processing. 2013. pp. 93–97.
18. Thipphavong D.P., Schultz C.A., Lee A.G., Chan S.H. Adaptive Algorithm to Improve Trajectory Prediction Accuracy of Climbing Aircraft. *Journal of Guidance, Control and Dynamics.* 2013. vol. 36. no. 1. pp. 15–24.
19. Allnutt J.E. Satellite to Ground Radiowave Propagation: 2nd Edition. The Institution of Engineering and Technology. 2010. 704 p.
20. Somov A.M., Kornev S.F. *Sputnikovye sistemy svjazi* [Satellite Communication Systems]. M.: Gorjachaja linija-Telekom. 2012. 244 p. (In Russ.).
21. Kantor L.Ya., Nozdrin V.V. *Jelektromagnitnaja sovmestimost' sistem sputnikovoj svjazi* [Satellite Communication Systems Electromagnetic Compatibility]. M.: NIIR. 2009. 280 p. (In Russ.).
22. Roy A.E. Orbital Motion. CRC Press. 2004. 544 p.
23. Gupta O.P. Global Augmentation of ADS-B Using Iridium NEXT Hosted Payloads. Integrated Communications, Navigation and Surveillance Conference (ICNS). 2011. pp. 1–15.
24. Garcia M., Dolan J., Hoag A. Aireon's initial on-orbit performance analysis of space-based ADS-B. Integrated Communications, Navigation and Surveillance Conference (ICNS). 2017. pp. 4A1-1–4A1-8.

25. Narimanov G.S. *Osnovy teorii poleta kosmicheskikh apparatov* [Space Vehicles Flight Theory Fundamentals]. Moscow: Mashinostroenie Publ. 1972. 608 p. (In Russ.).
26. Dynamic Statistical Studies on Satellite Reception of the ADS B Signal for Global Flight Tracking for Civil Aviation. World Radiocommunication Conference (WRC-15). 2015. 19 p. Available at: https://www.itu.int/md/R15-WRC15-C-0100/en (accessed: 17.05.2016).
27. Sklar B. Digital Communication. Fundamentals and Application: 2nd Edition. Prentice Hall. 2017. 1104 p.
28. Van Der Pryt R., Vincent R. A Simulation of Reflected ADS-B Signals over the North Atlantic for a Spaceborne Receiver. *Positioning.* 2016. vol. 7. pp. 51–62.
29. Skorokhodov Ya.A., Andreev A.M. [Modeling the Vessels Automatic Identification System Space Segment Functioning]. *Informatsionno-upravliaiushchie sistemy – Information and Control Systems*. 2018. Issue 93. vol. 2. pp. 36–48. (In Russ.).
30. Skorokhodov Ya.A., Mahrov K.B., Malyshev D.V. [Imitation model of marine ships traffic control space system functioning]. *Trudy Voenno-kosmicheskoj akademii imeni A.F. Mozhajskogo – Proceedings of the Mozhaisky Military Aerospace Academy*. 2017. vol. 657. pp. 23–33. (In Russ.).
31. Werner K., Bredemeyer J., Delovski T. ADS-B over Satellite. Global Air Traffic Surveillance from Space. Tyrrhenian International Workshop on Digital communications – enhanced surveillance of aircraft and vehicles. 2014. pp. 47–52.
32. Sajt izdatel'skogo doma «Aviatransportnoe obozrenie» [The site of the publishing house «Air Transport Review»]. Available at: http://www.ato.ru/blogs/ blog-alekseya-sinickogo/30-tysyach-aviareysov-v-sutki-nad-evropoy (accessed: 19.01.2016). (In Russ.).

УДК 62-50 **DOI** 10.15622/sp.61.3

В.И. Воротников, А.В. Вохмянина

МЕТОД ЛИНЕАРИЗУЮЩЕЙ ОБРАТНОЙ СВЯЗИ В ЗАДАЧЕ УПРАВЛЕНИЯ ПО ЧАСТИ ПЕРЕМЕННЫХ ПРИ НЕКОНТРОЛИРУЕМЫХ ПОМЕХАХ

Воротников В.И., Вохмянина А.В. **Метод линеаризующей обратной связи в задаче управления по части переменных при неконтролируемых помехах.**

Аннотация. Рассматривается задача гарантированного перевода за конечное время подверженной неконтролируемым помехам нелинейной динамической системы в положение, где заданная часть фазовых переменных равна нулю. Эта задача относится к задачам частичного (по отношению к части переменных) управления. Помехи не имеют каких-либо статистических описаний. Управления формируются по принципу обратной связи и удовлетворяют заданным «геометрическим» ограничениям.

Для решения указанной задачи используется метод линеаризующей обратной связи, позволяющий свести решение рассматриваемой нелинейной задачи управления к решению соответствующих линейных игровых антагонистических задач (с нефиксированным временем окончания). Приводятся конструктивно проверяемые достаточные условия, обеспечивающие гарантированное решение рассматриваемой задачи для заданной области начальных значений фазовых переменных. В отличие от ранее выполненных работ, посредством обратной связи линеаризуется более общий класс нелинейных управляемых систем, для которого допускаются оценки некоторой части переменных, и управление может осуществляться по отношению к большей части переменных.

В качестве примера изучается случай, когда рассматриваемая нелинейная управляемая система описывает пространственный разворот асимметричного твердого тела при управлении посредством моментов внутренних сил, создаваемых двигателями-маховиками. В этом случае система включает динамические уравнения Эйлера и кинематические уравнения в переменных Родрига — Гамильтона, описывающие вращательное движение основного тела, а также уравнения вращения маховиков. Рассматриваются две задачи гарантированного пространственного разворота тела при неконтролируемых внешних помехах, где цели управления определяются по части фазовых переменных указанной системы: задача переориентации тела, а также задача «прохождения» (с произвольной скоростью) телом заданного углового положения в пространстве.

Показано, что предложенный в статье подход позволяет с единых позиций получить и дополнить как некоторые уже известные решения этих задач, так и предложить новое решение задачи переориентации посредством более простых управляющих моментов, включающее оценку (завышенную) соотношения допустимых уровней управляющих моментов и неконтролируемых помех. Приводятся результаты численных расчетов, показывающие эффективность применяемых управляющих моментов.

Ключевые слова: управление по части переменных, неконтролируемые помехи, линеаризующая обратная связь, переориентация гиростата.

1. Введение. Часто в приложениях управление динамической системой достаточно осуществить не по всем фазовым переменным, а только по отношению к их некоторой части. Это касается, например, следующих задач: «жесткой» встречи (по отношению к координатам, но не по скоростям) двух объектов [1]; управления движением по от-

ношению к скоростям [2]; «прохождения» (с произвольной скоростью) твердым телом заданного углового положения в пространстве [3].

Более общими являются задачи управления на многообразиях (или в конфигурационном пространстве) [4], а также по отношению к заданной функции фазовых переменных (по выходу) [5]. Некоторое обсуждение задач частичного управления (совместно с проблемами частичной устойчивости и стабилизации) можно найти в работах [5, 6].

Отметим, что в задачах частичного (по части переменных) управления, в отличие от задач частичной (по части переменных) стабилизации [3, 6], речь идет об управлении за конечное, а часто и минимальное, время. Кроме того, в задачах частичного управления само конечное положение рассматриваемой замкнутой системы может не быть ее положением равновесия (система «проходит» данное положение, не останавливаясь в нем). Термин «стабилизация» в этом случае просто теряет смысл. В то же время значительно возрос интерес и к задачам стабилизации по части переменных на конечном промежутке времени (finite-time partial stabilization) [7-16], для решения которых используется метод функций Ляпунова [17] в соответствующей модификации [18, 19].

В настоящей статье рассматривается задача гарантированного перевода за конечное время подверженной помехам (возмущениям) нелинейной динамической системы в положение, где заданная часть фазовых переменных равна нулю. Помехи не имеют каких-либо статистических описаний. Управления формируются по принципу обратной связи и удовлетворяют заданным прямым «геометрическим» ограничениям. В процессе управления текущая информация о всех фазовых переменных системы считается известной.

Предложена модификация метода линеаризующей обратной связи («эквивалентной линеаризации») [3], позволяющая получить решение указанной нелинейной задачи на основе решения соответствующих игровых антагонистических задач управления с нефиксированным временем окончания для линейных конфликтно-управляемых систем дифференциальных уравнений простейшего вида. В сравнении с работой [3] линеаризуется более общий класс нелинейных управляемых систем, и управление может осуществляться по отношению к большей части переменных.

В качестве примера изучается представляющий самостоятельный теоретический и прикладной интерес случай, когда рассматриваемая нелинейная управляемая система описывает пространственный разворот асимметричного твердого тела (космического аппарата) при действующих на него неконтролируемых внешних помехах, не имею-

щих статистического описания. Управление осуществляется посредством моментов внутренних сил, создаваемых двигателями-маховиками. Рассматриваются две задачи гарантированного пространственного разворота тела, где цели управления определяются по части фазовых переменных указанной системы: задача переориентации тела, а также задача «прохождения» (с произвольной скоростью) телом заданного углового положения в пространстве.

Показано, что предложенный в статье подход позволяет с единых позиций получить и дополнить уже известные решения [20-22] этих задач. Также предложено новое решение задачи переориентации посредством более простых управляющих моментов. Расчеты показывают эффективность применяемых управляющих моментов.

2. Постановка задачи. Рассмотрим нелинейную управляемую систему обыкновенных дифференциальных уравнений вида:

$$\begin{gathered}\dot{y}_i = Y_i^{(0)}(\mathbf{x}) + \sum_{k=1}^{r} Y_{ik}(\mathbf{x}) u_k + Y_i^{(1)}(\mathbf{x}) v_i; \\ \dot{z}_j = Z_j(\mathbf{x}) \quad (i = \overline{1, m};\ j = \overline{1, p}),\end{gathered} \tag{1}$$

в которой $\mathbf{x} = (\mathbf{y}, \mathbf{z})$ — вектор фазовых переменных y_i, z_j; $\mathbf{u}$, $\mathbf{v}$ — векторы управляющих воздействий (управлений) u_k и неконтролируемых возмущений (помех) v_i.

Функции $Y_i^{(0)}$, Y_{ik}, $Y_i^{(1)}$, Z_j в системе (1) определены и непрерывны вместе со своими частными производными по y_i, z_j в области $S = \{\|\mathbf{x}\| \le H\}$, $H = \text{const} > 0$; $\|.\|$ — евклидова норма. Считаем, что $Y_i^{(0)}(\mathbf{0}) = 0$, $Z_j(\mathbf{0}) = 0$. Полагаем также, что $p = r \le m$.

Управления u_k формируются по принципу обратной связи $u_k = u_k(t, \mathbf{x})$, и в процессе управления известна информация о текущих значениях всех координат фазового вектора $\mathbf{x}$ системы (1). Реализации $u_k[t]$ управлений (здесь $u_k[t] = u_k(t, \mathbf{x}[t])$, $\mathbf{x}[t]$ — решения системы (1) при $u_k = u_k(t, \mathbf{x})$) являются измеримыми функциями, удовлетворяющими заданным «геометрическим» ограничениям:

$$|u_k| \le \alpha_k = \text{const} > 0. \tag{2}$$

Помехи не имеют каких-либо статистических описаний и могут реализовываться в виде любых измеримых функций в рамках ограничений:

$$|v_i| \le \beta_i = \text{const} > 0. \tag{3}$$

Отметим, что для любой допустимой реализации неконтролируемых помех v_i решения замкнутой системы дифференциальных уравнений (1) с формально-математической точки зрения понимаются [23] как абсолютно непрерывные функции времени $\mathbf{x}[t]$, удовлетворяющие этой системе (или соответствующей системе дифференциальных включений) при почти всех значениях $t \in [t_0, t_1]$.

Задача 1. Найти управления u_k при любых допустимых реализациях неконтролируемых помех v_i переводящие систему (1) из заданной области $S_0 = \{\|\mathbf{x}_0\| \le H_0 < H\}$ начальных значений $\mathbf{x}_0 = \mathbf{x}[t_0]$ фазового вектора за конечное время $\tau = t_1 - t_0$ в положение, где:

$$z_k[t_1] = 0 \quad (k = \overline{1,r}). \tag{4}$$

Значения остальных фазовых переменных системы (1) в момент времени t_1 могут быть произвольными. Момент времени t_1 не фиксируется.

Замечание 1. Задача 1 является задачей частичного (по части переменных) управления [3, 6] для системы (1). Положение (4) не является, вообще говоря, положением равновесия не только автономной системы (1) (при $\mathbf{u} = \mathbf{0}$, $\mathbf{v} = \mathbf{0}$), но и замкнутой системы (1) (при $\mathbf{v} = \mathbf{0}$). В этом случае система (1) «проходит» положение (4) в конечный момент времени t_1.

Замечание 2. Неравенства (2) и (3) определяют покомпонентную форму учета ограничений на управления u_k и помехи v_i. Связь между значениями α_k и β_i уровней управлений u_k и помех v_i, а также размерами области S_0 начальных значений $\mathbf{x}_0$ фазового вектора нелинейной системы (1), априори неизвестна и устанавливается в процессе решения задачи 1.

3. Вспомогательная линейная управляемая система. Рассмотрим матричную функцию

$$F(\mathbf{x}) = \left(\partial \Phi_s(\mathbf{x},\mathbf{u}) \Big/ \partial u_k \right) \quad (s,k = \overline{1,r}),$$

$$\Phi_s(\mathbf{x},\mathbf{u}) = \sum_{i=1}^{m} \left\{ \partial Z_s(\mathbf{x}) \Big/ \partial y_i \left[\sum_{k=1}^{r} Y_{ik}^{(1)}(\mathbf{x}) u_k \right] \right\},$$

и допустим, что в области $S_1 = \{\|\mathbf{x}\| \le H_1,\ H_0 < H_1 < H\}$ выполняется условие:

$$\operatorname{rank} F(\mathbf{x}) = r. \tag{5}$$

В этом случае в области S_1 система линейных алгебраических относительно u_k уравнений:

$$\Phi_s(\mathbf{x},\mathbf{u}) + \sum_{i=1}^{m} \partial Z_s(\mathbf{x}) \Big/ \partial y_i \; Y_i^{(0)}(\mathbf{x}) + \sum_{k=1}^{r} \partial Z_s(\mathbf{x}) \Big/ \partial z_k \; Z_k(\mathbf{x}) = u_s^* \quad (s = \overline{1,r})$$

(где u_s^* — вспомогательные управляющие воздействия, которыми распорядимся далее) имеет единственное решение:

$$u_k = f_k(\mathbf{x},\mathbf{u}^*), \quad f_k(\mathbf{0},\mathbf{0}) = 0 \quad (k = \overline{1,r}), \tag{6}$$

и функции f_k непрерывны в области S_1; $\mathbf{u}^* = (u_1^*, \ldots, u_r^*)$.

Также введем обозначения:

$$v_k^* = v_k^*(\mathbf{x},\mathbf{v}) = \sum_{i=1}^{m} \partial Z_k(\mathbf{x}) \Big/ \partial y_i \left[Y_i^{(1)}(\mathbf{x}) v_i \right] \quad (k = \overline{1,r}),$$

и будем интерпретировать v_k^* как «вспомогательные помехи».

Учитывая непосредственно проверяемые соотношения:

$$\ddot{z}_k = \sum_{i=1}^{m} \partial Z_k(\mathbf{x}) \Big/ \partial y_i \; \dot{y}_i + \sum_{s=1}^{r} \partial Z_k(\mathbf{x}) \Big/ \partial z_s \; \dot{z}_s =$$

$$= \Phi_k(\mathbf{x},\mathbf{u}) + \sum_{i=1}^{m} \partial Z_k(\mathbf{x}) \Big/ \partial y_i \; Y_i^{(0)}(\mathbf{x}) +$$

$$+ \sum_{s=1}^{r} \partial Z_k(\mathbf{x}) \Big/ \partial z_s \; Z_s(\mathbf{x}) + \sum_{i=1}^{m} \partial Z_k(\mathbf{x}) \Big/ \partial y_i \left[Y_i^{(1)}(\mathbf{x}) v_i \right] \quad (k = \overline{1,r}),$$

заключаем, что из замкнутой системы (1), (6) в области S_1 можно выделить линейную управляемую систему дифференциальных уравнений:

$$\ddot{z}_k = u_k^* + v_k^* \quad (k = \overline{1,r}). \tag{7}$$

На основе решения соответствующих игровых антагонистических задач управления для построенной вспомогательной линейной управляемой системы дифференциальных уравнений (7) далее будет строиться решение поставленной задачи 1 для исходной нелинейной управляемой системы (1). В этом смысле равенства (6) предопределяют общую структурную форму управлений u_k в задаче 1.

Для реализации указанной цели считаем возможным в области S_1 проведение оценки уровней «вспомогательных помех» v_k^* (при учете заданных ограничений (3) на помехи v_i):

$$\left|v_k^*\right| \le \beta_k^*, \quad \beta_k^* = \sup_{\mathbf{x} \subseteq S_1} \left\lfloor v_k^*(\mathbf{x}, \mathbf{v}) \right\rfloor \quad (k = \overline{1,r}). \tag{8}$$

Естественно, оценка уровней v_k^* может быть затруднена и зависит от конкретного структурного вида нелинейной управляемой системы (1). Например, в случае, когда система (1) описывает пространственный разворот твердого тела (см. раздел 7), такая оценка не вызывает затруднений.

4. Вспомогательная линейная игровая задача. Для линейной управляемой системы (7) решим задачу о быстрейшем «прохождении» положения (4). Управление осуществляется посредством u_k^* при любых допустимых реализациях v_k^*, удовлетворяющих неравенствам (8). Конечные значения $\dot{z}_k$ $(k = \overline{1,r})$ не фиксируются (произвольны).

Данную задачу трактуем как игровую антагонистическую (с нефиксированным временем окончания). В этой задаче один из игроков распоряжается вспомогательными управлениями u_k^* и стремится уменьшить время $\tau = t_1 - t_0$ достижения системой (7) положения (4). Второй игрок (противник) стремится увеличить τ или вообще избежать позиции (4), и имеет в распоряжении «вспомогательные помехи» v_k^*. При этом построенная линейная система (7) трактуется как конфликтно-управляемая.

Для решения данной задачи ограничение на u_k^* примем в виде:

$$\left\|\mathbf{u}^*\right\| = \left[\sum_{k=1}^{r} u_k^{*2}\right]^{1/2} \le \alpha^* = \text{const} > 0. \tag{9}$$

Ограничение на v_k^* также примем в виде:

$$\left\|\mathbf{v}^*\right\| = \left[\sum_{k=1}^{r} v_k^{*2}\right]^{1/2} \le \beta^* = \left[\sum_{k=1}^{r} \beta_k^{*2}\right]^{1/2}, \tag{10}$$

где значения β_k^* определяются согласно соотношениям (8).

В случае $\alpha^* > \beta^*$, когда уровень α^* управлений u_k^* выше уровня β^* помех v_k^*, решение указанной линейной игровой задачи при ограничениях (9), (10) дает тормозящая экстремальная стратегия Н. Н. Красовского [1].

Управления u_k^*, диктуемые этой стратегией, имеют вид:

$$u_k^*(t, \mathbf{z}, \dot{\mathbf{z}}, \theta) = \begin{cases} -\alpha^* L_k \left[\sum_{s=1}^{r} L_s^2 \right]^{-1/2}, & \omega > 0 \\ 0, & \omega \le 0; \end{cases}$$
$$\omega = \left[\sum_{s=1}^{r} L_s^2 \right]^{1/2} - \tfrac{1}{2}(\alpha^* - \beta^*)(\theta - t)^2; \qquad (11)$$
$$L_k = z_k + (\theta - t)\dot{z}_k \quad (k = \overline{1,r}),$$

где переменная θ находится как наименьший положительный корень уравнения $\omega = 0$.

Минимальное гарантированное время τ достижения положения (4) в рассматриваемой линейной игровой задаче определяется [1] как наименьший положительный корень уравнения:

$$\left[\sum_{s=1}^{r} L_s^{*2} \right]^{1/2} - \tfrac{1}{2}\left(\alpha^* - \beta^*\right)\tau^2 = 0, \;\; L_s^* = z_{s0} + \tau \dot{z}_{s0}.$$

Это значение τ является гарантированным временем управления в исходной нелинейной задаче 1, и соответствует случаю $v_k^* = -(\beta^*/\alpha^*)u_k^*$ «наихудших» помех v_k^*, — оптимальных управлений «противника». Если же v_k^* отличаются от «наихудших», то встреча с положением (4) произойдет быстрее, чем за время τ.

Управления (11), а также соответствующие им решения (движения) $\mathbf{z}[t], \dot{\mathbf{z}}[t]$ системы дифференциальных уравнений (7), (11) могут быть построены конструктивно. Такую возможность дает аппроксимационная схема [1] коррекции управлений u_k^* вида (11) в дискретные моменты времени. При этом управления (11) и решения (движения) системы (7), (11) можно рассматривать как предельный переход от соответствующих управлений и решений (движений), порождаемых указанной дискретной схемой. Получаемые в результате указанного предельного перехода решения системы дифференциальных уравнений (7), (11) являются абсолютно непрерывными функциями времени, удовлетворяющими этой системе почти всюду на рассматриваемом отрезке времени.

5. Условия разрешимости задачи 1. При решении задачи 1 управления по части переменных (по $z_1, \ldots, z_r$) для нелинейной си-

стемы (1) на основе решения соответствующей вспомогательной задачи управления для линейной системы (7) возникает следующая ситуация. Поведение не только переменных $z_1, \dots, z_r$, но и поведение функций $Z_1(\mathbf{x}), \dots, Z_r(\mathbf{x})$ нелинейной системы (1), будет определяться поведением фазовых переменных, входящих в фазовые векторы $\mathbf{z}, \dot{\mathbf{z}}$ линейной системы (7).

При определенном условии данное обстоятельство можно использовать для оценки всего фазового вектора $\mathbf{x}[t]$ нелинейной системы (1) и в конечном счете для оценки уровней (2) управлений u_k.

Пусть, например, в области S имеют место тождества:

$$Z_s(\mathbf{x}) \equiv \sum_{k=1}^{r} y_k Z_{sk}^{(1)}\left(y_{r+1},\dots,y_m,\mathbf{z}\right) + Z_s^{(2)}\left(y_{r+1},\dots,y_m,\mathbf{z}\right),$$
$$Z_s^{(2)}\left(0,\dots,0,\mathbf{0}\right) \equiv 0 \quad (s=\overline{1,r}). \tag{12}$$

Рассмотрим матричную функцию:

$$Q(\mathbf{x}) = \left(\partial Z_s(\mathbf{x}) \Big/ \partial y_k \right) \quad (s,k=\overline{1,r})$$

и допустим, что в области $S_2 = \{\|\mathbf{x}\| \le H_2,\ H_0 < H_2 < H\}$ выполняется условие:

$$\operatorname{rank} Q(\mathbf{x}) = r. \tag{13}$$

При выполнении условий (12), (13) система уравнений:

$$\dot{z}_s = Z_s(\mathbf{x}) \quad (s=\overline{1,r}),$$

рассматриваемая как линейная алгебраическая система относительно y_k, имеет единственное решение:

$$y_k = g_k(y_{r+1},\dots,y_m,\mathbf{z},\dot{\mathbf{z}}),$$
$$g_k(0,\dots,0,\mathbf{0},\mathbf{0}) = 0 \quad (k=\overline{1,r}). \tag{14}$$

Функции g_k непрерывны в области S_2; $\mathbf{g} = (g_1, \dots, g_k)$.

Соотношения (14) означают, что поведение фазового вектора $\mathbf{x}$ нелинейной системы (1) будет определяться поведением фазовых переменных линейной вспомогательной системы (7), а также поведением переменных $y_{r+1}, \dots, y_m$ исходной системы (1).

Поэтому фазовый вектор $\mathbf{x}[t]$ системы (1) можно представить следующим образом:

$$\mathbf{x}[t] \triangleq \mathbf{w}[t], \quad \mathbf{w} = [\mathbf{g}(y_{r+1},...,y_m,\mathbf{z},\dot{\mathbf{z}}), y_{r+1},...,y_m,\mathbf{z}].$$

Если переменные y_{r+1}, …, y_m исходной системы (1) можно оценить тем или иным образом (например, зная первые интегралы этой системы), то указанное обстоятельство можно использовать для конструктивной проверки заданных ограничений (2) на управления u_k.

Пусть оценки (8) получены. Достаточные условия разрешимости задачи 1 можно сформулировать следующим образом.

Теорема 1. Пусть выполняются следующие условия:

1) условие (5) в области S_1;

2) условия (12) и (13) в области S_2;

3) для любого $\mathbf{x}_0 \in S_0$ найдется число α^* ($\alpha^* > \beta^*$) такое, что:

$$\begin{aligned} &\left| f_k\left(\mathbf{w}[t], \mathbf{u}^*[t]\right)\right| \le \alpha_k \ (k = \overline{1,r}), \\ &\left\|\mathbf{w}[t]\right\| \le \min(H_1, H_2), \ \ t \in [t_0, t_0 + \tau]. \end{aligned} \tag{15}$$

Тогда для любого $\mathbf{x}_0$ из области S_0 управления (6), (11) обеспечивают гарантированное «прохождение» системой (1) положения (4) в конечный момент времени $t_1 = t_0 + \tau$ при любых допустимых реализациях помех v_i.

Доказательство. При выполнении условия (5) в области S_1 из замкнутой нелинейной системы (1), (6) можно выделить линейную конфликтно-управляемую систему (7). При этом поведение переменных z_1, … , z_r нелинейной системы (1), (6), (11) определяется поведением этих же переменных замкнутой линейной системы (7), (11).

Поэтому для всех значений $\mathbf{x}_0$, таких, что фазовый вектор $\mathbf{x}[t]$ в процессе управления остается в области S_1, управления (6), (11) обеспечивают гарантированное «прохождение» системой (1) положения (4) в конечный момент времени $t_1 = t_0 + \tau$. Однако при этом не гарантируется выполнение заданных ограничений (2) на управления (6), (11).

Если в области S_2 выполнены условия (12), (13), а для любого $\mathbf{x}_0 \in S_0$ выполнены условия (15), то для всех $\mathbf{x}_0 \in S_0$ управления (6), (11) не только обеспечивают гарантированное «прохождение» системой (1) положения (4) в конечный момент времени $t_1 = t_0 + \tau$, но и удовлетворяют заданным ограничениям (2). Теорема доказана.

6. Дополнительные возможности предложенного подхода. Рассмотрим два варианта модификации, позволяющие расширить возможности предложенного подхода.

6.1. Управление по большей части переменных (по отношению к y_k,z_k). Допустим, что в области S_2 имеют место соотношения (14), причем дополнительно имеют место тождества:

$$g_k(y_{r+1},...,y_m,\mathbf{0},\mathbf{0}) \equiv 0 \quad (k=\overline{1,r}). \tag{16}$$

Покажем, что в данном случае при соответствующем изменении вспомогательной игровой задачи для линейной конфликтно-управляемой системы (7) и при выполнении условий теоремы 1 управления (6) будут гарантированно переводить систему (1) из заданной области S_0 начальных значений $\mathbf{x}_0$ за конечное время $\tau = t_1 - t_0$ в положение, где:

$$y_k[t_1] = z_k[t_1] = 0 \quad (k=\overline{1,r}). \tag{17}$$

Значения остальных фазовых переменных системы (1) в момент времени t_1 могут быть произвольными. Момент времени t_1 не фиксируется.

Для этого, в отличие от раздела 4, для линейной системы (7) решим задачу о быстрейшем гарантированном приведении в положение:

$$z_k[t_1] = \dot{z}_k[t_1] = 0 \quad (k=\overline{1,r}). \tag{18}$$

Управление также осуществляется посредством u_k^* при любых допустимых реализациях v_k^*, удовлетворяющих неравенствам (8). Данную задачу трактуем как игровую антагонистическую, в которой один из игроков распоряжается управлениями u_k^* и стремится уменьшить время $\tau = t_1 - t_0$ приведения системы (7) в положение (18). Второй игрок (противник) стремится увеличить значение τ или вообще избежать позиции (18), и имеет в распоряжении «помехи» v_k^*.

Решение этой задачи при ограничениях $|u_k^*| \le \alpha_k^*$, $|v_k^*| \le \beta_k^* = \rho_k\alpha_k^*$, $0 < \rho_k < 1$ дают законы управления [1, 24]:

$$u_k^*(z_k,\dot{z}_k) = \begin{cases} \alpha_k^* \operatorname{sgn}\psi_k^\rho(z_k,\dot{z}_k), & \psi_k^\rho \neq 0 \\ \alpha_k^* \operatorname{sgn} z_k = -\alpha_i^* \operatorname{sgn}\dot{z}_k, & \psi_k^\rho = 0 \end{cases}$$
$$\psi_k^\rho = -z_k - \left[2(1-\rho_k)\alpha_k^*\right]^{-1}|z_k||\dot{z}_k| \tag{19}$$
$$(k=\overline{1,r}),$$

где ψ_k^ρ — функции переключений.

При $v_k^* \neq -\rho_k u_k^*$ движения системы (7), (19) на фазовых плоскостях переменных $z_k, \dot{z}_k$ будут сначала происходить (до достижения кривых переключений $\psi_k^{\rho} = 0$) между дуг парабол, являющихся траекториями систем $\ddot{z}_k = (1 \pm \rho_k) u_k^*$. Далее движения будут происходить вдоль кривых переключений в скользящем режиме до достижения требуемых конечных значений (18); на этих участках движения управления u_k^* принимают значения $\pm\alpha_k^*$ с бесконечно частыми сменами знака.

Указанные решения (движения) системы (7), (19) с формально-математической точки зрения трактуем как абсолютно непрерывные функции времени, удовлетворяющие почти всюду соответствующим дифференциальным включениям.

Величина $\tau = \max(\tau_k)$, где τ_k — минимальное гарантированное время для каждой подсистемы системы (7), определяет минимальное гарантированное время в рассматриваемой задаче. Значения τ_k соответствуют случаям $v_k^* = -\rho_k u_k^*$ «наихудших» помех v_k^*, — оптимальных управлений «противника». Если же v_k^* отличаются от «наихудших», то встреча с положением (18) произойдет быстрее. Отметим, что те подсистемы системы (18), которые придут в требуемое положение раньше, чем последняя из них, остаются в этом положении, и соответствующие управления u_k^* в этих подсистемах парируют помехи v_k^*.

Теорема 2. Пусть выполняются следующие условия:

1) условие (5) в области S_1;

2) условия (12), (13) и (16) в области S_2;

3) для любого $\mathbf{x}_0 \in S_0$ найдутся числа α_k^* ($\alpha_k^* > \beta_k^*$) такие, что выполнены соотношения (15).

Тогда для любого $\mathbf{x}_0$ из области S_0 управления (6), (19) обеспечивают гарантированный перевод системы (1) в положение (4) за конечное время τ при любых допустимых реализациях помех v_i.

Доказательство. Если в области S_2 выполнены условия (12), (13), (16), то приведение переменных $\mathbf{z}, \dot{\mathbf{z}}$ линейной системы (7) посредством управлений (19) в положение (18) будет означать приведение нелинейной системы (1) посредством управлений (6), (19) в положение (17).

Поэтому для всех значений $\mathbf{x}_0$ таких, что фазовый вектор $\mathbf{x}[t]$ в процессе управления остается в области $S_1 \bigcap S_2$, управления (6), (19) обеспечивают гарантированное приведение нелинейной системы (1) в положение (17) за конечное время τ. Однако при этом не гарантируется выполнение заданных ограничений (2) на управления (6), (19).

При выполнении условий (15) для всех $\mathbf{x}_0 \in S_0$ фазовый вектор $\mathbf{x}[t]$ остается в области $S_1 \bigcap S_2$, и для всех $\mathbf{x}_0 \in S_0$ управления (6), (19) не только обеспечивают гарантированное приведение системы (1) в положение (17), но и удовлетворяют ограничениям (2). Теорема доказана.

Замечание 3. Линеаризация обратной связью рассматривается для более общего (в сравнении с [3], где $m = p = r$) класса нелинейных управляемых систем (1), поскольку в данном случае $m > p = r$. Кроме того, управление осуществляется по большей части (по $2r$) переменным, в то время как в [3] управление осуществляется только по r переменным. Дополнительное ограничение на правые части системы (1) в виде условия (12) не связано с линеаризацией, и существенно облегчает проведение оценок фазового вектора $\mathbf{x}[t]$ в процессе управления.

6.2. Упрощение конструкции управлений (6). Пусть $\Phi_s^{(1)}(\mathbf{x})$, $\Phi_s^{(2)}(\mathbf{x})$, $\Phi_s^{(1)}(\mathbf{0}) = \Phi_s^{(2)}(\mathbf{0}) = 0$ — две совокупности функций такие, что в области S имеют место тождества:

$$\sum_{i=1}^{m} \partial Z_s(\mathbf{x}) \Big/ \partial y_i \; Y_i^{(0)}(\mathbf{x}) + \sum_{k=1}^{r} \partial Z_s(\mathbf{x}) \Big/ \partial z_k \; Z_k(\mathbf{x}) \equiv \Phi_s^{(1)}(\mathbf{x}) + \Phi_s^{(2)}(\mathbf{x}) \quad (s = \overline{1,r}).$$

Если выполняется условие (5), то в области S_1 система линейных относительно u_k алгебраических уравнений:

$$\Phi_s(\mathbf{x}, \mathbf{u}) + \Phi_s^{(1)}(\mathbf{x}) = u_s^*$$

имеет единственное (при фиксированном выборе $\Phi_s^{(1)}$) решение:

$$u_k = f_k^*(\mathbf{x}, \mathbf{u}^*), \quad f_k^*(\mathbf{0}, \mathbf{0}) = 0 \quad (k = \overline{1,r}), \tag{20}$$

и функции f_k^* непрерывны в области S_1.

Введем обозначения:

$$v_k^{**} = v_k^{**}(\mathbf{x}, \mathbf{v}) = \sum_{i=1}^{m} \left\{ \partial Z_k(\mathbf{x}) \Big/ \partial y_i \; [Y_i^{(2)}(\mathbf{x}) v_i] \right\} + \Phi_k^{(2)}(\mathbf{x}) \quad (k = \overline{1,r}),$$

и будем интерпретировать v_k^{**} как «вспомогательные помехи».

В результате из замкнутой системы (1), (20) в области S_1 можно выделить линейную управляемую систему дифференциальных уравнений:

$$\ddot{z}_k = u_k^* + v_k^{**} \quad (k = \overline{1,r}). \tag{21}$$

Решение поставленной задачи 1 для исходной нелинейной управляемой системы (1) можно также найти на основе решения соответствующей игровой задачи управления для построенной вспомогательной линейной управляемой системы (21).

Равенства (20) определяют вторую (более простую) структурную форму управлений u_k в задаче 1, причем упрощение структурной формы (6) достигается за счет усложнения «вспомогательных помех» в образующихся вспомогательных линейных конфликтно-управляемых системах. Имеющийся произвол в выборе функций $\Phi_s^{(1)}$ и $\Phi_s^{(2)}$ может использоваться для поиска наиболее приемлемого решения.

Для реализации указанной цели считаем возможным в области S_1 проведение оценки уровней «вспомогательных помех» v_k^{**} (при учете заданных ограничений (3)):

$$\left|v_k^{**}\right| \le \beta_k^{**}, \quad \beta_k^{**} = \sup_{\mathbf{x} \subseteq S_1} \left[v_k^{**}(\mathbf{x}, \mathbf{v}) \right] \quad (k = \overline{1,r}).$$

Оценка уровней v_k^{**} (как и оценка уровней v_k^{*}) может быть затруднена, и зависит от конкретной структуры нелинейной управляемой системы (1). В случае, когда система (1) описывает пространственный разворот твердого тела (см. раздел 7), такая оценка может быть получена на основе принципа «назначения и последующего подтверждения» уровней v_k^{**}.

Теорема 3. Пусть выполняются условия (2), (3) теоремы 1, и для любого $\mathbf{x}_0 \in S_0$ найдется число α^* ($\alpha^* > \beta^*$) такое, что выполнены соотношения:

$$\left| f_k^* \left(\mathbf{w}[t], \mathbf{u}^*[t] \right) \right| \le \alpha_k \ (k = \overline{1,r}),$$

$$\left\| \mathbf{w}[t] \right\| \le \min(H_1, H_2), \ t \in [t_0, t_0 + \tau].$$

Тогда для любого $\mathbf{x}_0$ из области S_0 управления (20), (11) обеспечивают гарантированное «прохождение» системой (1) положения (4) в конечный момент времени $t_1 = t_0 + \tau$ при любых допустимых реализациях v_i.

Доказательство теоремы 3 аналогично доказательству теорем 1, 2.

7. Приложение к задачам пространственного разворота асимметричного твердого тела. Допустим, что вдоль главных центральных осей инерции асимметричного твердого тела (космического аппарата) закреплены оси вращения однородных симметричных маховиков. Поскольку геометрия масс данной механической системы

«несущее основное тело — маховики» не меняется при вращениях маховиков, то эта система является гиростатом, и ее вращательное движение вокруг центра масс описывается системой обыкновенных дифференциальных уравнений [3, 25]:

$$\begin{aligned}
&(A_1 - J_1)\dot{\omega}_1 = (A_2 - A_3)\omega_2\omega_3 + J_2\omega_3\dot{\varphi}_2 - J_3\omega_2\dot{\varphi}_3 - u_1 + v_1,\\
&(A_2 - J_2)\dot{\omega}_2 = (A_3 - A_1)\omega_1\omega_3 + J_3\omega_1\dot{\varphi}_3 - J_1\omega_3\dot{\varphi}_1 - u_2 + v_2,\\
&(A_3 - J_3)\dot{\omega}_3 = (A_1 - A_2)\omega_1\omega_2 + J_1\omega_2\dot{\varphi}_1 - J_2\omega_1\dot{\varphi}_2 - u_3 + v_3,\\
&J_k(\ddot{\varphi}_k + \dot{\omega}_k) = u_k \quad (k = \overline{1,3}).
\end{aligned} \tag{22}$$

В системе (22): ω_k — проекции вектора угловой скорости основного тела на главные центральные оси $\mathbf{i}_k$ эллипсоида инерции гиростата; A_k, J_k — главные центральные моменты инерции гиростата и осевые моменты инерции маховиков; φ_k — углы поворота маховиков, оси вращения которых закреплены вдоль осей $\mathbf{i}_k$.

Управляющие моменты u_k (моменты внутренних сил) приложены к маховикам и создаются специальными двигателями. Моменты v_k характеризуют внешние неконтролируемые возмущения (в том числе и некоторые внешние силы), действующие на основное тело. Управляющие моменты u_k и помехи v_k удовлетворяют соответственно ограничениям (2) и (3).

Ориентацию основного тела рассматриваемого трехроторного гиростата будем определять уравнениями для кватернионов (кинематическими уравнениями в переменных Родрига — Гамильтона [25]):

$$\begin{aligned}
&2\dot{\eta}_1 = \eta_4\omega_1 + \eta_2\omega_3 - \eta_3\omega_2, \quad 2\dot{\eta}_2 = \eta_4\omega_2 + \eta_3\omega_1 - \eta_1\omega_3,\\
&2\dot{\eta}_3 = \eta_4\omega_3 + \eta_1\omega_2 - \eta_2\omega_1, \quad \eta_1^2 + \eta_2^2 + \eta_3^2 + \eta_4^2 = 1.
\end{aligned} \tag{23}$$

Требуется найти управляющие моменты u_k, осуществляющие гарантированный разворот основного тела за конечное время $\tau = t_1 - t_0$ в заданное угловое положение. Не нарушая общности, считаем:

$$\eta_k[t_1] = 0,\ \eta_4[t_1] = 1, \tag{24}$$

когда в момент времени $t_1 > t_0$ происходит совмещение связанной с телом и заданной систем координат.

Начальные угловые скорости основного тела и маховиков считаем нулевыми. В момент времени t_1 угловая скорость основного тела или также нулевая или не фиксируется («прохождение» заданного положения), а угловая скорость маховиков не фиксируется.

В данном случае систему (22), (23), записанную в переменных, соответствующих структуре системы (1), можно представить в виде (здесь и далее под знаком суммы суммирование по j от 1 до 3):

$$\begin{aligned}
\dot{y}_1 &= (A_1 - J_1)^{-1}\left[(A_2 - A_3)y_2y_3 + J_2y_3y_5 - J_3y_2y_6 - u_1 + v_1\right],\\
\dot{y}_2 &= (A_2 - J_2)^{-1}\left[(A_3 - A_1)y_1y_3 + J_3y_1y_6 - J_1y_3y_4 - u_2 + v_2\right],\\
\dot{y}_3 &= (A_3 - J_3)^{-1}\left[(A_1 - A_2)y_1y_2 + J_1y_2y_4 - J_2y_1y_5 - u_3 + v_3\right],\\
\dot{y}_{3+k} &= -\dot{y}_k + J_l^{-1}u_k \quad (k = \overline{1,3}),\\
\dot{z}_1 &= \tfrac{1}{2}(y_1\sqrt{1 - \sum z_j^2} - y_2z_3 + y_3z_2),
\end{aligned} \tag{25}$$

$\dot{z}_2 = \tfrac{1}{2}(y_2\sqrt{1 - \sum z_j^2} + y_1z_3 - y_3z_1)$, $\dot{z}_3 = \tfrac{1}{2}(y_3\sqrt{1 - \sum z_j^2} - y_1z_2 + y_2z_1)$,
причем, как будет показано, достаточно рассматривать ее в области:

$$\sum z_j^2 < 1. \tag{26}$$

7.1. Задачи управления по части переменных. Уточним рассматриваемые задачи пространственного разворота тела применительно к нелинейной управляемой системе (25).

Задача 2 (*переориентации твердого тела*). Найти управляющие моменты u_k, гарантированно переводящие систему (25) из области:

$$S_0 = \{\mathbf{x}_0 = (\mathbf{y}_0, \mathbf{z}_0) = (\mathbf{0}, z_{10}, z_{20}, z_{30}),\ \sum z_{j0}^2 < 1\} \tag{27}$$

за конечное время $\tau = t_1 - t_0$ в положение:

$$y_k[t_1] = z_k[t_1] = 0 \quad (k = \overline{1,3}). \tag{28}$$

Значения $y_{3+k}[t_1]$, а также момент времени t_1, не фиксируются.

Задача 3 (*«прохождения» твердым телом заданного положения*). Найти управляющие моменты u_k, гарантированно переводящие систему (25) из области (27) за конечное время $\tau = t_1 - t_0$ в положение:

$$z_k[t_1] = 0 \quad (k = \overline{1,3}). \tag{29}$$

Значения $y_k[t_1]$ и $y_{3+k}[t_1]$, а также момент времени t_1, не фиксируются.

Покажем, что предложенный в разделах 2-5 подход к решению задачи 1, а также его модификации, указанные в разделе 6, позволяют:

1) Охватить более сложный (в сравнении с [3], где задачи 2, 3 решаются посредством моментов внешних сил, создаваемых реактивными двигателями) класс задач пространственного разворота асимметричного твердого тела в неопределенной внешней среде посредством моментов внутренних сил, создаваемых двигателями-маховиками.

2) С единых позиций получить и дополнить ранее найденные [20-22] решения задач 2, 3, а также дать новое решение в задаче 2 посредством более простых управлений.

7.2. Выполнимость условий теорем 1, 2. Имеем $m = 6$, $p = r = 3$ (поэтому используемые в статье индексы k, s и j меняются далее от 1 до 3) и для системы (25) выполнено (при $Z_s^{(2)} \equiv 0$) условие (12).

Компоненты f_{ks}, q_{ks} матриц F, Q определяются следующим образом:

$$f_{kk} = -q_{kk}(A_k - J_k)^{-1}, \quad q_{kk} = \tfrac{1}{2}\sqrt{1 - \sum z_j^2},$$

$$f_{12} = q_{21}(A_2 - J_2)^{-1}, \quad f_{13} = q_{23}(A_3 - J_3)^{-1}, \quad f_{21} = q_{12}(A_1 - J_1)^{-1},$$

$$f_{23} = q_{32}(A_3 - J_3)^{-1}, \quad f_{31} = q_{13}(A_1 - J_1)^{-1}, \quad f_{32} = q_{23}(A_2 - J_2)^{-1},$$

$$q_{12} = -q_{21} = -\tfrac{1}{2}z_3, \quad q_{13} = -q_{31} = \tfrac{1}{2}z_2, \quad q_{23} = -q_{32} = -\tfrac{1}{2}z_1$$

и, следовательно, в области (26) имеем rank F = rank Q = 3. Поэтому в области (26) выполнены условия 1, 2 теоремы 1 и условие 1 теоремы 2.

Управляющие моменты u_k типа (6) имеют вид (выписано только выражение для u_1; выражения для u_2 и u_3 получаются из u_1 циклической перестановкой индексов $1 \to 2 \to 3$, $4 \to 5 \to 6$; здесь и далее суммирование по l, как и по j, от 1 до 3):

$$\begin{aligned} u_1 = -\frac{2(A_1 - J_1)}{\sqrt{1 - \sum z_j^2}}\Big[&u_1^*(1 - z_2^2 + z_3^2) + u_2^*\Big(z_1 z_2 + z_3\sqrt{1 - \sum z_j^2}\Big) + \\ &+ u_3^*\Big(z_1 z_3 - z_2\sqrt{1 - \sum z_j^2}\Big) + \tfrac{1}{4}z_1\sum y_l^2\Big] + \\ &+ (A_2 y_2 + J_2 y_5)y_3 - (A_3 y_3 + J_3 y_6)y_2. \end{aligned} \tag{30}$$

Из замкнутой нелинейной системы (25), (30) можно выделить вспомогательную линейную конфликтно-управляемую систему типа (7):

$$\begin{gathered}\ddot{z}_k = u_k^* + v_k^* \quad (k = \overline{1,3}),\\ v_1^* = \tfrac{1}{2}[\sqrt{1-\sum z_j^2}\, v_1 (A_1 - J_1)^{-1} + z_2 v_3 (A_3 - J_3)^{-1} - \\ - z_3 v_2 (A_2 - J_2)^{-1}]\\ (1 \to 2 \to 3).\end{gathered} \tag{31}$$

Используя неравенства Коши — Буняковского, имеем:

$$\left|v_k^*\right| \le \beta_k^* = \beta^* = \frac{1}{2}\sqrt{\sum \frac{\beta_l^2}{\left(A_l - J_l\right)^2}}. \tag{32}$$

В силу вида области S_0 (см. выражения (27)) выполнены условия $\dot{z}_{k0} = 0\ (k = \overline{1,3})$. Поэтому при решении задач 2, 3 в линейных управляемых системах (31), (11) и (31), (19) имеют место соотношения:

$$\sum z_j^2[t] \le \sum z_{j0}^2 < 1, \quad \left|z_k[t]\right| \le \left|z_{k0}\right|$$

и, следовательно, систему (25) достаточно рассматривать в области (26).

Соотношения (14) сводятся к уравнениям:

$$\begin{gathered}y_1 = \frac{2}{\sqrt{1-\sum z_j^2}}[\dot{z}_1(1 - z_2^2 + z_3^2) + \dot{z}_2(z_1 z_2 + z_3\sqrt{1-\sum z_j^2}) + \\ + \dot{z}_3(z_1 z_3 - z_2\sqrt{1-\sum z_j^2})] \quad (1 \to 2 \to 3).\end{gathered} \tag{33}$$

Поскольку $y_k = 0$ при $z_k = \dot{z}_k = 0\ \ (k = \overline{1,3})$, то имеют место соотношения (16) и, следовательно, выполнено условие 2 теоремы 2.

Переменные y_4, y_5, y_6, используя первый интеграл системы (25):

$$\sum \left[A_l y_l + J_l y_{3+l}\right]^2 = \text{const},$$

можно оценить следующим образом:

$$\left|A_k y_k[t] + J_k y_{3+k}[t]\right| \le t\left[\sum \beta_l^2\right]^{1/2} \quad (k = \overline{1,3}). \tag{34}$$

Управляющие моменты (30), (11) и (30), (19) обеспечивают гарантированное приведение тела соответственно в положения (29) и (28) за конечное время при любых $\alpha_k^* > \beta_k^*$ (в задаче 2) и любом $\alpha^* > \beta^*$ (в задаче 3). Но заданные ограничения (2) могут при этом не выполняться.

Условие 3 теорем 1, 2 применительно к задачам 2, 3 можно проверить по-разному. Например, в работах [20-22] получены прямые оценки допустимых уровней β_k помех v_k, определяющие возможности решения задач 2, 3 посредством управляющих моментов (30), (19) и (30), (11) при ограничениях (2). Эти оценки используют, в частности, неравенства (32).

Покажем, что в задаче 2 эти неравенства, а также сами оценки [20, 21], можно несколько улучшить, если использовать принцип «назначения и последующего подтверждения» уровней v_k^*.

Утверждение 1. Если выполняются неравенства:

$$\sqrt{3}\left(A_k - J_k\right)\gamma_k \sqrt{\sum \frac{\beta_l^2}{\left(A_l - J_l\right)^2}} + 4\sqrt{2}\lambda_k \sqrt{\sum z_{j0}^2}\sqrt{\sum \beta_l^2} < \alpha_k;$$

$$\gamma_k = \sqrt{\left(1 + \frac{z_{k0}^2}{1 - \sum z_{j0}^2}\right)\left(1 - z_{k0}^2\right)}, \quad \lambda_1 = \sqrt{2 + \frac{z_{20}^2 + z_{30}^2}{1 - \sum z_{j0}^2}} \quad (1 \to 2 \to 3), \tag{35}$$

то задача 2 может быть решена посредством управляющих моментов (30), (19), удовлетворяющих заданным ограничениям (2).

Доказательство. «Назначим» уровни β_k^* «вспомогательных помех» v_k^* в системе (31), полагая:

$$\beta_k^* = \tfrac{1}{2}\left(1 - z_{k0}^2\right)^{1/2}\left[\sum \beta_l^2 \left(A_l - J_l\right)^{-2}\right]^{1/2}, \tag{36}$$

и рассмотрим управления (19), в которых $\alpha_k^* = \beta_k^* + \varepsilon$, где $\varepsilon > 0$ — достаточно малое положительное число.

На множестве состояний системы (31), (19) имеем [20, 21]:

$$z_k^2[t] \le z_{k0}^2, \quad \left|\dot{z}_k^2[t]\right| \le \left|z_{k0}\right|\left(\alpha_k^*\right)^{-1}\left[\left(\alpha_k^*\right)^2 - \left(\beta_k^*\right)^2\right]. \tag{37}$$

Учитывая оценку (34), а также равенства:

$$\tau = \max(\tau_k), \quad \tau_k = 2\sqrt{\left|z_{k0}\right|\left(\alpha_k^* - \beta_k^*\right)^{-1}},$$

определяющие минимальное гарантированное время управления в системе (31), с учетом фазового портрета системы (31), (19), используя

неравенство Коши — Буняковского и неравенства (37), получаем соотношения:

$$\left|\{A_2 y_2[t] + J_2 y_5[t]\} y_3[t]\right| \le 2\left(1 - \sum z_j^2[t]\right)^{-1/2} \left(\sum \beta_l^2\right)^{1/2} \times$$

$$\times \left| \tau_3 \dot{z}_3[t] \{1 - z_1^2[t] - z_2^2[t]\} + \tau_1 \dot{z}_1[t] \{z_3[t] z_1[t] + z_2[t] \sqrt{1 - \sum z_j^2[t]}\} + \right.$$

$$\left. + \tau_2 \dot{z}_2[t] \{z_3[t] z_2[t] - z_1[t] \sqrt{1 - \sum z_j^2[t]}\} \right| \le$$

$$\le 2\left\{1 + z_3^2[t]\left(1 - \sum z_j^2[t]\right)^{-1}\right\}^{1/2} \left(\sum \beta_l^2\right)^{1/2} \left[\sum \{\tau_k \dot{z}_k[t]\}^2\right]^{1/2} \le$$

$$\le 4\left\{1 + z_{30}^2 \left(1 - \sum z_{j0}^2\right)^{-1}\right\}^{1/2} \times$$

$$\times \left[\left(\alpha_3^* + \beta_3^*\right) z_{30}^2/\alpha_3^* + \left(\alpha_1^* + \beta_1^*\right) z_{10}^2/\alpha_1^* + \left(\alpha_2^* + \beta_2^*\right) z_{20}^2/\alpha_2^*\right]^{1/2} \left(\sum \beta_l^2\right)^{1/2} \le$$

$$\le 4\sqrt{2}\left\{1 + z_{30}^2 \left(1 - \sum z_{j0}^2\right)^{-1}\right\}^{1/2} \left(\sum z_{j0}^2\right)^{1/2} \left(\sum \beta_l^2\right)^{1/2}$$

$$(1 \to 2 \to 3, \quad 4 \to 5 \to 6).$$

Это значит, что полученные верхние оценки выражений типа $\{A_2 y_2[t] + J_2 y_5[t]\} y_3[t]$, входящих в «вспомогательные помехи» v_k^{**} в системе (31), (19), не зависят от «назначенных» значений β_k^*, а также от выбранных значений α_k^*. (Поэтому данные оценки совпадают с полученными ранее в работах [20, 21].)

Кроме того, в силу равенства:

$$\tfrac{1}{4} \sum y_l^2[t] = \left(1 - \sum z_j^2[t]\right)^{-1} \sum \dot{z}_k^2[t],$$

на основании неравенств (37) заключаем, что при достаточно малом значении ε выражения:

$$\left(A_k - J_k\right) z_k[t] \left(1 - \sum z_j^2[t]\right)^{-3/2} \sum y_l^2[t],$$

входящие в u_k, также являются достаточно малыми.

В результате при достаточно малом $\varepsilon > 0$ на множестве состояний линейной конфликтно-управляемой системы (31), (19) имеем соотношения:

$$\left|u_k[t]\right| \le 2\left(A_k - J_k\right)\left\{1 + z_{k0}^2 \left(1 - \sum z_{j0}^2\right)^{-1}\right\}^{1/2} \left[\sum \left(\alpha_l^*\right)^2\right]^{1/2} +$$

$$+ 4\sqrt{2}\lambda_k \left(\sum z_{j0}^2\right)^{1/2} \left(\sum \beta_l^2\right)^{1/2} \quad (k = \overline{1,3}).$$

На основании неравенства Коши — Буняковского справедливы более точные (в сравнении с (32)) оценки:

$$\left|v_1^*[t]\right| = \tfrac{1}{2}\,|\,[\sqrt{1-\sum z_j^2[t]}\,v_1(A_1-J_1)^{-1} +$$
$$+\,z_2[t]v_3(A_3-J_3)^{-1} - z_3[t]v_2(A_2-J_2)^{-1}]\,| \le$$
$$\le \tfrac{1}{2}\left(1-z_{10}^2\right)^{1/2}\left[\sum \beta_l^2\left(A_l-J_l\right)^{-2}\right]^{1/2} \quad (1\to 2\to 3).$$

В результате получаем оценки $|v_k^*| \le \beta_k^*$, что подтверждает «назначенные» уровни «вспомогательных помех» v_k^*.

Поэтому при достаточно малом значении ε заданные ограничения (2) на управляющие моменты (30), (19) выполнены при выполнении неравенств (35). Утверждение доказано.

Утверждение 2 [22]. Если выполняются неравенства:

$$\sqrt{3}\left(A_k-J_k\right)\gamma\sqrt{\sum\frac{\beta_l^2}{\left(A_l-J_l\right)^2}} + 8\lambda\sqrt{\left(\sum\beta_l^2\right)} < \alpha_k,$$
$$\gamma = \frac{1+3\sum z_{j0}^2}{\sqrt{\left(1-\sum z_{j0}^2\right)^3}}, \quad \lambda = \sqrt{\frac{\sum z_{j0}^2}{1-\sum z_{j0}^2}}, \tag{38}$$

то задача 3 может быть решена посредством управляющих моментов (30), (11), удовлетворяющих заданным ограничениям (2).

Полученные оценки (35), (38) связывают максимально допустимые уровни управляющих моментов (30), (19) и (30), (11), неконтролируемых возмущений v_k, а также область допустимых начальных значений z_{k0} в задачах 2, 3 соответственно. В «предельном» случае $z_{k0} = 0$ (или когда область допустимых начальных значений z_{k0} является малой), неравенства (35) и (38) совпадают и записываются в значительно более простой форме:

$$\sqrt{3}\left(A_k-J_k\right)\sqrt{\sum\frac{\beta_l^2}{\left(A_l-J_l\right)^2}} < \alpha_k. \tag{39}$$

Отметим, что при $J_l = 0$ оценка (39) совпадает с «предельной» оценкой области допустимых уровней помех в задаче переориентации твердого тела посредством моментов внешних сил [3].

Оценки (35), (38) являются завышенными, поскольку для их вывода используется ряд неравенств, а также делаются предположения

о «наихудшем» поведении вспомогательной системы. Кроме того, для простоты используется единая для всего промежутка управления оценка уровней управляющих моментов, хотя реализации построенных управляющих моментов являются функциями, значения которых меняются в процессе управления в достаточно широких пределах.

Такая ситуация является характерной и, по-видимому, неизбежной при стремлении получить достаточно простые и наглядные соотношения допустимых уровней управлений и помех в сложных нелинейных управляемых системах. Реальные соотношения уровней построенных в рамках предложенного подхода управляющих моментов и неконтролируемых помех значительно ниже, причем в каждый конкретный момент времени эти соотношения меняются в достаточно широких пределах.

Выделим два этапа использования неравенств, приводящих к завышенности оценок (35), (38). На первом этапе используется неравенство Коши — Буняковского для оценки уровней β_k^* «вспомогательных помех» v_k^*, приводящее к относительно небольшому завышению их истинных уровней. При этом величины β_k^* входят в u_k^* и, следовательно, входят в управляющие моменты (30), (19) и (30), (11), что также приводит к некоторому (относительно небольшому) завышению их уровней.

На втором этапе используется цепочка неравенств для оценки самих управляющих моментов (30), (19) и (30), (11), но полученные завышенные величины (в отличии от величин β_k^*) не входят в управляющие моменты. Поэтому управляющие моменты (30), (19) и (30), (11) работоспособны и эффективны и за пределами полученных оценок, что подтверждают результаты численных расчетов.

7.3. Результаты численных расчетов. Для трехроторного гиростата с $A_1 = 4\times10^4$, $A_2 = 8\times10^4$, $A_3 = 5\times10^4$ (кгм^2); $J_1 = 4\times10^3$, $J_2 = 8\times10^3$, $J_3 = 5\times10^3$ (кгм^2) рассмотрим задачу 3 при начальных значениях $z_{10} = 0.353$, $z_{20} = 0.434$, $z_{30} = 0.432$ (рад).

Пусть уровни β_k неконтролируемых помех v_k определяются равенством $\beta_k^* = 10^{-3}$ (рад/с^2). В случае $\beta_k = \beta$ имеем $\beta = 55.99$ (Нм).

Допустим, что гарантированное время переориентации $\tau = 70$ (с). Значения τ, β_k^* предопределяют значение уровня α^* вспомогательных управлений u_k^* в управляющих моментах (30), (11):

$$\alpha^* = \sqrt{3}\times10^{-3} + \frac{\sqrt{\sum z_{j0}^2}}{\tau^2} = 1.87\times10^{-3}\ (\text{рад/с}^2).$$

Оценим уровни управляющих моментов (30), (11). Расчет показывает, что при «наихудших» помехах v_k^* уровни управляющих

моментов (30), (11) следующие (законы изменения u_k в этом случае даны на рисунке 1):

$$\alpha_1 = 122.48,\ \alpha_2 = 252.56,\ \alpha_3 = 221.19\ (\text{Нм}).$$

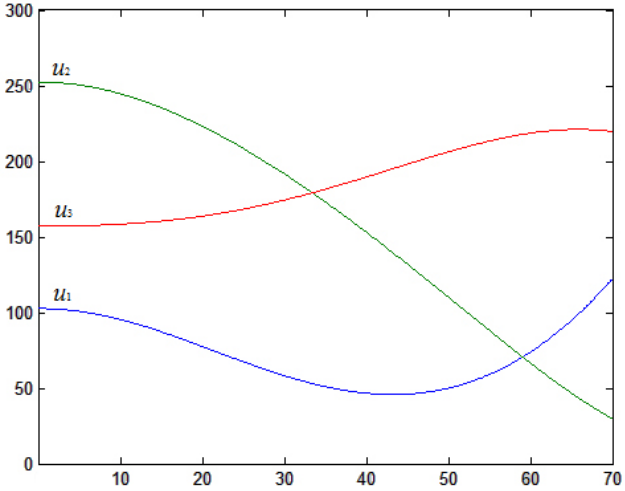


Рис. 1. Реализации управляющих моментов u_k вида (30), (11) в случае «наихудших помех» v_k^*

Сравнивая с уровнями $\beta = 55.99$ (Нм) помех, в данном случае имеем:

$$\beta < \min\,(0{,}4571\alpha_1;\ 0{,}2217\alpha_2;\ 0{,}2531\alpha_3), \qquad (40)$$

что на порядок отличается от оценки (38), которая принимает вид:

$$\beta < \min\,(0{,}0338\alpha_1;\ 0{,}0226\alpha_2;\ 0{,}0304\alpha_3).$$

Указанная существенная разница, как уже отмечалось, объясняется тем, что при получении неравенств (38) использованы завышенные оценки, но управляющие моменты (30), (11) и их уровни не зависят от этих завышенных оценок. Кроме того, реализации управляющих моментов (30), (11) не являются релейными функциями времени, и меняются в процессе управления в достаточно широких пределах. Поэтому оценка (38) не полностью характеризует реальное соотношение допустимых уровней управляющих моментов (30), (11) и действующих неконтролируемых помех, и на отдельных промежутках времени при развороте основного тела гиростата некоторые значения β/α_k превышают 1.

Расчет также показывает, что при отсутствии помех v_k (в этом случае имеют место тождества $v_k^* \equiv 0$) уровни управляющих моментов (30), (11) определяются значениями:

$$\alpha_1 = 102.71,\ \alpha_2 = 252.56,\ \alpha_3 = 157.12\ (\text{Нм}).$$

Сравнивая с уровнями $\beta = 55.99$ (Нм) помех, в данном случае имеем:

$$\beta < \min\,(0{,}5451\alpha_1;\ 0{,}2217\alpha_2;\ 0{,}3564\alpha_3),$$

а на отдельных промежутках времени некоторые значения β/α_k также превышают 1.

В целом моделирование показывает, что при одних и тех же ограничениях на управляющие моменты u_k, гарантированное время разворота основного тела гиростата в задаче 3 меньше, чем в задаче 2.

Как уже отмечалось, полученные оценки (35), (38) существенно зависят от начальных значений z_{k0}. Например, в «предельном» случае $z_{k0} = 0$ (или при малых значениях z_{k0}) области (35) и (38) совпадают и имеют вид:

$$\beta < \min\,(0{,}4201\alpha_1;\ 0{,}2102\alpha_2;\ 0{,}3360\alpha_3).$$

С другой стороны, случай $\sum z_{j0}^2 = 3/4$ также является «предельным» в том плане, что для расширения допустимой области помех на первом этапе следует перейти к управляющим моментам, получающимся из (30), (19) или (30), (11) соответствующей перестановкой индексов. Поэтому для выбранных значений параметров гиростата область допустимых значений β при использовании управляющих моментов (30), (19) или (30), (11) является промежуточной по отношению к указанным двум «предельным» областям. Например, при $z_{k0} = 0.1$ (рад) область (38) принимает вид:

$$\beta < \min\,(0{,}1894\alpha_1;\ 0{,}0947\alpha_2;\ 0{,}1669\alpha_3).$$

Отметим также, что в рамках рассматриваемого в статье подхода неуправляемые переменные y_i связаны с управляемыми переменными $z_k, \dot{z}_k$ соотношениями (33), (34). Это исключает недопустимые в процессе управления «особенности» поведения этих переменных (типа «ухода в бесконечность» за конечное время). Поэтому поведение неуправляемых переменных замкнутой системы (25) является регулярным.

Для примера, при «наихудших» помехах v_k^* законы изменения переменных $y_k[t]$ (угловых скоростей тела) и переменных $y_{3+k}[t]$ (угловых скоростей маховиков) даны соответственно на рисунках 2 и 3.

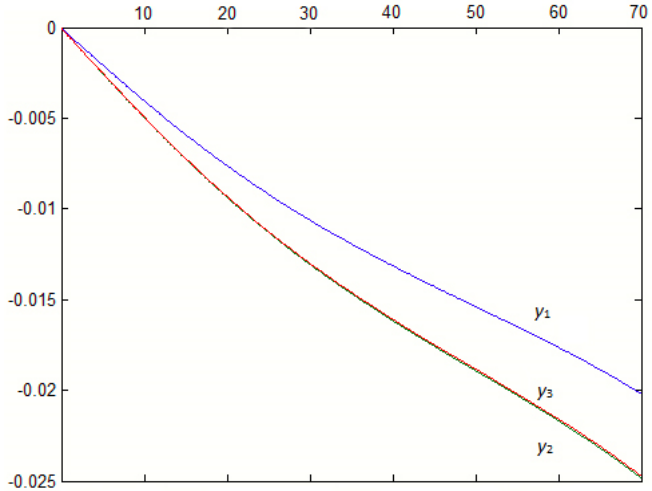


Рис. 2. Законы изменения переменных $y_k[t]$ замкнутой системы (25), (30), (11) в случае «наихудших помех» v_k^*

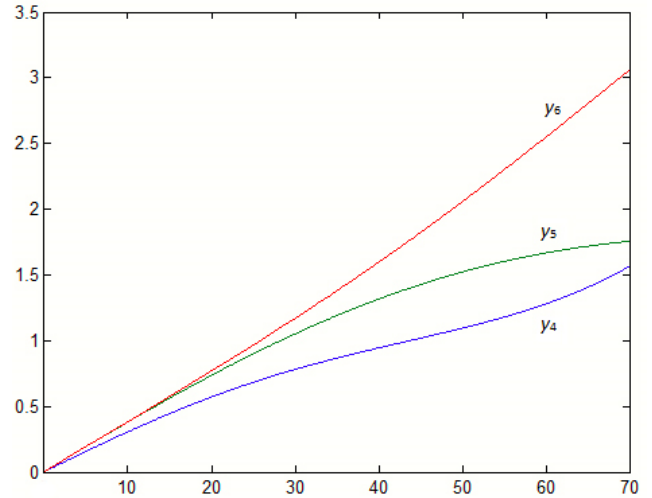


Рис. 3. Законы изменения переменных $y_{3+k}[t]$ замкнутой системы (25), (30), (11) в случае «наихудших помех» v_k^*

7.4. Упрощение управлений в задаче 2. Покажем, что оценка (35) определяет возможность решения задачи 2 не только посредством управляющих моментов (30), (19), но и посредством более простых управляющих моментов типа (20):

$$u_1 = -\frac{2(A_1 - J_1)}{\sqrt{1-\sum z_j^2}}[u_1^*(1-z_2^2-z_3^2) + u_2^*(z_1 z_2 + z_3\sqrt{1-\sum z_j^2}) + u_3^*(z_1 z_3 - z_2\sqrt{1-\sum z_j^2})] + \\ + (A_2 y_2 + J_2 y_5) y_3 - (A_3 y_3 + J_3 y_6) y_2 \quad (1 \to 2 \to 3). \tag{41}$$

В данном случае из нелинейной системы (25), (41) можно выделить линейную конфликтно-управляемую систему типа (21):

$$\ddot{z}_k = u_k^* + v_k^{**} \quad (k = \overline{1,3}), \\ v_1^{**} = \tfrac{1}{2}[\sqrt{1-\sum z_j^2}\,(v_1 + V_1)(A_1 - J_1)^{-1} + \\ + z_2(v_3 + V_2)(A_3 - J_3)^{-1} - z_3(v_2 + V_3)(A_2 - J_2)^{-1}], \\ V_1 = -\tfrac{1}{2}\frac{(A_1 - J_1)}{\sqrt{1-\sum z_j^2}} z_1 \sum y_l^2 \quad (1 \to 2 \to 3). \tag{42}$$

Утверждение 3. Если выполняются неравенства (35), то задача 2 может быть решена посредством управляющих моментов (41), (19), удовлетворяющих заданным ограничениям (2).

Доказательство. «Назначим» уровни β_k^{**} «вспомогательных помех» v_k^{**} в вспомогательной линейной системе (42), полагая $\beta_k^{**} = \beta_k^* + \varepsilon_1$ (значения β_k^* определяются соотношениями (36)). Рассмотрим также управления (19), в которых $\alpha_k^* = \beta_k^* + \varepsilon_2$. Здесь $\varepsilon_2 > \varepsilon_1 > 0$ — достаточно малые положительные числа.

Аналогично доказательству утверждения 1 можно показать, что достаточно малые положительные числа $\varepsilon_2 > \varepsilon_1 > 0$ можно выбрать таким образом, чтобы заданные ограничения (2) на управляющие моменты (41), (19) были выполнены при выполнении неравенств (35).

Кроме того, в силу соотношений (37) имеем:

$$|V_k[t]| \le 2\left(1-\sum z_{j0}^2\right)^{3/2}\left(\sum z_{j0}^2\right)^{1/2}\sum\left\{|z_{l0}|\left(\alpha_l^*\right)^{-1}\left[\left(\alpha_l^*\right)^2 - \left(\beta_l^{**}\right)^2\right]\right\} \le \\ \le 4\left(\varepsilon_2 - \varepsilon_1\right)\lambda^*, \\ \lambda^* = \left(1-\sum z_{j0}^2\right)^{3/2}\left(\sum z_{j0}^2\right)^{1/2}\sum|z_{j0}|,$$

и на основании неравенства Коши — Буняковского получаем оценки:

$$\left|v_k^{**}[t]\right| \le \beta_k^* + 2\left(\varepsilon_2 - \varepsilon_1\right)\lambda^{**},$$

$$\lambda^{**} = \lambda^* \left[\sum \left(A_l - J_l\right)^{-2}\right]^{1/2}.$$

Полагая

$$\varepsilon_2 = \left[1 + \tfrac{1}{2}\left(\lambda^{**}\right)^{-1}\right]\varepsilon_1,$$

при достаточно малом ε_1 получаем оценки $|v_k^{**}| \le \beta_k^* + \varepsilon_1$, что подтверждает «назначенные» уровни «вспомогательных помех» v_k^{**}. Утверждение доказано.

Как и при использовании управляющих моментов (30), (19), условия (35) гарантируют решение задачи 2 посредством управляющих моментов (41), (19) также только при достаточно большом (хотя и конечном) значении τ. Однако учитывая, что оценки (35) завышенные, а управляющие моменты (41), (19) работоспособны и за пределами этих оценок, нахождение гарантированного времени переориентации τ можно осуществить без учета оценок (35) следующим образом.

Итерационный алгоритм (нахождения τ в задаче 2 при управляющих моментах (41), (19)).

Шаг 1. Выбирается значение $\Delta > 0$ и «назначаются» уровни $\beta_k^{**} = \beta_k^* + \Delta$ «вспомогательных помех» v_k^{**} в системе (42) (значения β_k^* определяются соотношениями (36)). Выбираются также «пробные» значения $\alpha_k^* > \beta_k^{**}$, предопределяющие соответствующее значение τ.

Шаг 2. Подтверждается выполнимость неравенств $|v_k^{**}| \le \beta_k^{**}$ на множестве состояний линейной системы (42), (19), что подтверждает назначенные уровни v_k^{**}.

Поскольку верхние оценки выражений V_k монотонно зависят от значений $\gamma_k = \alpha_k^* - \beta_k^{**}$, причем эти оценки тем меньше, чем меньше значения γ_k, то для выбранного значения $\Delta > 0$ найдется диапазон «пробных» значений α_k^* ($\alpha_k^* > \beta_k^{**}$) и соответствующий диапазон значений τ, в котором «назначенные» уровни v_k^{**} подтверждаются.

Шаг 3. Проверка заданных ограничений (2) на управляющие моменты (41), (19). При этом в управляющие моменты (41), (19) входят значения α_k^* из того диапазона «пробных» значений, в котором «назначенный» уровень v_k^{**} подтверждается.

Если на шаге 3 заданные ограничения (2) на управляющие моменты u_k не выполняются, выбор «пробных» $\alpha_k^* > \beta_k^{**}$ (подбор подходящего значения τ из диапазона значений, в котором «назначенные» уровни

v_k^{**} подтверждается) продолжается. В противном случае гарантированное время переориентации определяется выбранным значением τ.

8. Дополнительное замечание. Для управляемых систем дифференциальных уравнений (1) и (25), линеаризуемых обратной связью, рассмотренные задачи решаются также (см., например, [5]) путем построения траектории движения в заданную точку с последующим обеспечением стабилизации (при помехах) относительно заданной траектории.

В рамках предложенного в статье подхода, в сущности, так и происходит. При использовании управлений (6), (19) и управляющих моментов (30), (19) линейные системы (7) и (31) по каждой переменной выводятся на траектории, соответствующие кривым переключений управлений (19), с последующим движением в скользящем режиме вдоль этих траекторий. Отличие в том, что указанные траектории строятся в результате решения соответствующей игровой задачи оптимального управления.

Отметим, что при отсутствии помех такой способ позволяет получить (как показывают расчеты в работе [3]) субоптимальные по быстродействию законы управления в задаче 2.

При использовании управлений (6), (11) и управляющих моментов (30), (11) движение к цели происходит по траектории, определяемой тормозящей экстремальной стратегией, и при отсутствии помех законы управления также субоптимальны по быстродействию в задаче 3.

9. Заключение. Рассмотрена задача гарантированного перевода нелинейной управляемой системы, подверженной неконтролируемым помехам (возмущениям), за конечное время в положение, где заданная часть переменных равна нулю. Данная задача является задачей частичного (по отношению к части фазовых переменных) управления. Управляемая система является аффинной (линейной по управлениям). Помехи не имеют каких-либо статистических описаний. Управления формируются по принципу обратной связи и удовлетворяют заданным «геометрическим» ограничениям. Реализации управлений и помех допускаются в классе измеримых функций времени. Предполагается, что в процессе управления известна информация о текущих значениях всех координат фазового вектора изучаемой системы.

Предложена модификация метода линеаризующей обратной связи [3], позволяющая получить решение указанной нелинейной задачи управления по части переменных на основе решения соответствующих игровых антагонистических задач управления (с нефиксированным временем окончания) для линейных конфликтно-управляемых систем дифференциальных уравнений простейшего

вида. Линеаризация посредством обратной связи рассматривается для более общего (в сравнении с [3], где $m = p = r$) класса нелинейных управляемых систем. Кроме того, управление может осуществляться (в случае использования конструкции (6), (19)) по большей части переменных.

В качестве примера изучается случай, когда рассматриваемая нелинейная управляемая система описывает пространственный разворот асимметричного твердого тела (космического аппарата) в неопределенной внешней среде, генерирующей помехи с неизвестным статистическим описанием. Управление осуществляется посредством ограниченных по величине моментов внутренних сил, создаваемых двигателями-маховиками, что имеет ряд преимуществ при управлении пространственным движением космических аппаратов.

Рассматриваются две типичные задачи гарантированного пространственного разворота тела, где цели управления определяются по части фазовых переменных указанной системы. Наряду с классической задачей переориентации, также рассматривается задача «прохождения» телом заданного углового положения в пространстве, небезынтересная в случаях необходимости быстрейшей переориентации космического аппарата для совершения кратковременных операций в момент достижения заданного положения (таких, например, как фотографирование, поражение цели, передача информации).

Показано, что предложенный в статье подход позволяет с единых позиций получить и дополнить уже известные решения [20-22] этих задач. Также предложено новое решение задачи гарантированной переориентации твердого тела посредством более простых управляющих моментов. Указанные решения получены в классе пространственных разворотов без каких-либо дополнительных ограничений на характер результирующего движения тела (типа «плоского разворота» вокруг оси Эйлера). Приводятся результаты численных расчетов, показывающие эффективность применяемых управляющих моментов.

Отметим, что имеются также и другие, не использующие линеаризацию обратной связью, подходы к анализу нелинейных задач переориентации асимметричного твердого тела при внешних помехах, основанные на методах конфликтного управления с целевым функционалом интегрального типа, адаптивного управления, а также на методе функций Ляпунова; см., например, работы [26-31]. Данные задачи важны для приложений в космической технике, в том числе для вновь возникающих приложений (связанных, например, с использованием группы космических аппаратов-перехватчиков для борьбы с астероидной опасностью [32]).

Литература

1. *Красовский Н.Н.* Игровые задачи о встрече движений // М.: Наука. 1970. 420 с.
2. *Акуленко Л.Д.* Асимптотические методы оптимального управления // М.: Наука. 1987. 365 с.
3. *Vorotnikov V.I.* Partial Stability and Control // Boston: Birkhauser. 1998. 448 p.
4. *Понтрягин Л.С., Болтянский В.Г., Гамкрелидзе Р.В., Мищенко Е.Ф.* Математическая теория оптимальных процессов // М.: Физматлит. 1961. 392 с.
5. *Fradkov A.L., Miroshnik I.V., Nikiforov V.O.* Nonlinear and Adaptive Control of Complex Systems // Kluwer Academic Publisher. 1999. 528 p.
6. *Воротников В.И.* Частичная устойчивость и управление: состояние проблемы и перспективы развития // Автоматика и телемеханика. 2005. № 4. С. 3−58.
7. *Jammazi C.* Finite-Time Partial Stabilizability of Chained Systems // Comptes Rendus Mathematique. 2008. vol. 346. no. 17-18. pp. 975−980.
8. *Jammazi C.* A Discussion on the Holder and Robust Finite-Time Partial Stabilizability of Brockett's Integrator // Control, Optimization and Calculus of Variations. 2012. vol. 18. no. 2. pp. 360−382.
9. *Jammazi C.* Continuous and Discontinuous Homogeneous Feedbacks Finite-Time Partially Stabilizing Controllable Multichained Systems // Journal of Control and Optimization. 2014. vol. 52. no. 1. pp. 520–544.
10. *Chen H., Li B.Y., Zhang B.W., Zhang L.* Global Finite-Time Partial Stabilization for a Class of Nonholonomic Mobile Robots Subject to Input Saturation // International Journal of Advanced Robotic Systems. 2015. vol. 12. no. 11. 159 p.
11. *Haddad W.M., L'Afflitto A.* Finite-Time Partial Stability and Stabilization, and Optimal Feedback Control // Journal of the Franklin Institute. 2015. vol. 352. no. 6. pp. 2329–2357.
12. *Binazadeh T., Shafiei M.H., Bazregarzadeh E.* New Approach in Guidance Law Design Based on Finite-Time Partial Stability Theorem // Journal of Space Science and Technology. 2015. vol. 8. pp. 1–7.
13. *Golestani M., Mohammadzaman I., Yazdanpanah M. J.* Robust Finite-Time Stabilization of Uncertain Nononlinear Systems Based on Partial Stability // Nonlinear Dynamics. 2016. vol. 85. no. 1. pp. 87–96.
14. *L'Afflitto A.* Differential Games, Finite-Time Partial-State Stabilization of Nonlinear Dynamical Systems, and Optimal Robust Control // International Journal of Control. 2017. vol. 90. no. 9. pp. 1861–1878.
15. *Rajpurohit T., Haddad W.M.* Stochastic Finite-Time Partial Stability, Partial-State Stabilization, and Finite-Time Optimal Feedback Control // Mathematics of Control, Signals, and Systems. 2017. vol. 29. no. 2. 10 p.
16. *Jammazi C., Abichou A.* Controllability of Linearized Systems Implies Local Finite-Time Stabilizability: Applications to Finite-Time Attitude Control // Journal of Mathematical Control and Information. 2018. vol. 35. no. 1. pp. 249–277.
17. *Ляпунов А.М.* Общая задача об устойчивости движения // Изд-во Харьковского математического общества. 1892. 251 с.
18. *Bhat S.P., Bernstein D.S.* Finite-Time Stability of Continuous Autonomous Systems // Journal on Control and Optimization. 2000. vol. 38. no. 3. pp. 751–766.
19. *Bhat S.P., Bernstein D.S.* Geometric Homogeneity with Applications to Finite-Time Stability // Mathematics of Control, Signals and Systems. 2005. vol. 17. no. 2. pp. 101–127.
20. *Воротников В.И., Мартышенко Ю.Г.* К нелинейной задаче трехосной переориентации трехроторного гиростата при игровой модели помех // Космические исследования. 2013. Т. 51. № 5. С. 412.
21. *Воротников В.И., Мартышенко Ю.Г.* К задаче переориентации трехроторного гиростата при неконтролируемых внешних помехах // Мехатроника. Автоматизация. Управление. 2016. Т. 17. № 6. С. 414–419.

22. *Воротников В.И., Вохмянина А.В.* К нелинейной задаче «прохождения» трехроторным гиростатом заданного углового положения в пространстве при неконтролируемых внешних помехах // Космические исследования. 2018. Т. 56. № 5. С. 382–387.
23. *Филиппов А.Ф.* Дифференциальные уравнения с разрывной правой частью // М.: Наука. 1985. 216 с.
24. *Черноусько Ф.Л., Ананьевский И.М., Решмин С.А.* Методы управления нелинейными механическими системами // М.: Физматлит. 2006. 328 с.
25. *Лурье А.И.* Аналитическая механика // М.: Физматлит. 1961. 824 с.
26. *Park Y.* Robust and Optimal Attitude Stabilization of Spacecraft with External Disturbances // Aerospace Science and Technology. 2005. vol. 9. no. 3. pp. 253–259.
27. *Ding S.H., Li S.H.* Stabilization of the Attitude of a Rigid Spacecraft with External Disturbances using Finite-Time Control Techniques // Aerospace Science and Technology. 2009. vol. 13. no. 4-5. pp. 256–265.
28. *Xia Y.Q., Zhu Z., Fu M.Y., Wang S.* Attitude Tracking of Rigid Spacecraft with Bounded Disturbances // IEEE Transactions on Industrial Electronics. 2011. vol. 58. no. 2. pp. 647–659.
29. *Hu Q., Niu G.* Attitude Output Feedback Control for Rigid Spacecraft with Finite-Time Convergence // ISA Transactions. 2017. vol. 70. pp. 173–186.
30. *Song Z., Duan C., Su H., Hu J.* Full-Order Sliding Mode Control for Finite-Time Attitude Tracking of Rigid Spacecraft // Control Theory & Applications. 2018. vol. 12. no. 8. pp. 1086–1094.
31. *Ran D., Chen X., de Ruiter A., Xiao B.* Adaptive Extended-State Observer-Based Fault Tolerant Attitude Control for Spacecraft with Reaction Wheels // Acta Astronautica. 2018. vol. 145. pp. 501–514.
32. *Минаков Е.П., Соколов Б.В., Шалдаев С.Е.* Исследование характеристик и вариантов применения окололунной системы поражения астероидов // Труды СПИИРАН. 2017. Вып. 5(54). С. 106–129.

Воротников Владимир Ильич — д-р физ.-мат. наук, профессор, профессор кафедры информационных технологий, Уральский федеральный университет имени первого Президента России Б. Н. Ельцина (УрФУ). Область научных интересов: устойчивость динамических систем, частичная устойчивость и стабилизация, теория управления, динамика управляемого твердого тела (космического аппарата). Число научных публикаций — 190. vorotnikov-vi@rambler.ru; ул. Мира, 19, Екатеринбург, 620002; р.т.: +7(3435) 256722.

Вохмянина Анастасия Владимировна — аспирант кафедры информационных технологий, Уральский федеральный университет имени первого Президента России Б. Н. Ельцина (УрФУ). Область научных интересов: теория управления, динамика управляемого твердого тела (космического аппарата). Число научных публикаций — 5. vokhmyanina.av@gmail.com; ул. Мира, 19, Екатеринбург, 620002; р.т.: +7(3435) 256722.

DOI 10.15622/sp.61.3

V.I. VOROTNIKOV, A.V. VOKHMYANINA

FEEDBACK LINIARIZATION METHOD FOR PROBLEM OF CONTROL OF A PART OF VARIABLES IN UNCONTROLLED DISTURBANCES

Vorotnikov V.I., Vokhmyanina A.V. **Feedback Liniarization Method for Problem of Control of a Part of Variables in Uncontrolled Disturbances.**

Abstract. The paper studies a problem of guaranteed transfer within a finite amount of time of a nonlinear dynamical system subjected to uncontrolled disturbances to a state where a given part of the variables equals zero. The bounded controls are offered to be generated by means of a feedback in form of nonlinear functions of phase variables of a given nonlinear controlled system of differential equations. The method of exact feedback linearization of the nonlinear system is used. As a result, the solution of the original nonlinear problem is narrowed down to solve the linear game-theoretic antagonistic control problem. Sufficient conditions are obtained with ensure that the problem has a guaranteed solution for the given domain of initial conditions.

As an example, problem of the space turn of an asymmetric rigid bode (spacecraft) is considered within the framework of the method. Three reaction wheels are employed to produce necessary torque in the axes of the spacecraft. External uncontrolled disturbances, that have no statistical description, are taken into consideration in the process of reorientation. In this case the initial nonlinear controlled systems consists of dynamic Euler equations and Rodriges – Hamilton kinematic equations based on the quaternion parameterization of attitude kinematics. Two problems of the space turn of the spacecraft are considered. 1) The rest - to - rest reorientation problem. 2) The space turn from a stationary state to a given angular position; it is not assumed that the turn takes the spacecraft to a stationary state. The proposed approach allows common positions to give some already well-known solutions of these problems. A new solution of the reorientation problem is also given. For this new solution an estimation of the admissible domain of uncontrolled disturbances is found. Results of a numerical calculations are considered.

Keywords: control of a part of variables, uncontrolled disturbances, linearization feedback, three-rotor gyrostat reorientation.

Vorotnikov Vladimir Il'ich — Ph.D., Dr. Sci., professor, professor of information technology department, Ural Federal University named after the First President of Russia B. N. Yeltsin. Research interests: stability of dynamical systems, partial stability and stabilization, control theory, dynamics of controlled solid (spacecraft). The number of publications — 190. vorotnikov-vi@rambler.ru; 19, Mira str., Ekaterinburg, 620002; office phone: +7(3435) 256722.

Vokhmyanina Anastasiya Vladimirovna — Ph.D. student of information technology department, Ural Federal University named after the First President of Russia B. N. Yeltsin. Research interests: control theory, dynamics of controlled solid (spacecraft). The number of publications — 5. vokhmyanina.av@gmail.com; 19, Mira str., Ekaterinburg, 620002; office phone: +7(3435) 256722.

References

1. Krasovskii N.N. *Igrovye zadachi o vstreche dvizhenij* [Rendeznous Game Problems]. M.: Nauka. 1971. 365 p. (In Russ.).
2. Akulenko L.D. *Asimptoticheskie metody optimal'nogo upravlenija* [Perturbation Methods in Optimal Control Problems]. M.: Nauka. 1987. 365 p. (In Russ.).

3. Vorotnikov V.I. Partial Stability and Control. Boston: Birkhauser. 1998. 448 p.
4. Pontryagin L.S. Boltyanskii V.G., Gamkrelidze R.V., Mishenko E.F. *Matematicheskaja teorija optimal'nyh processov* [The Mathematical Theory of Optimal Processes]. M.: Fizmatlit. 1961. 392 p. (In Russ.).
5. Fradkov A.L., Miroshnik I.V., Nikiforov V.O. Nonlinear and Adaptive Control of Complex Systems. Kluwer Academic Publisher. 1999. 528 p.
6. Vorotnikov V.I. **[**Partial Stability and Control: the State of the Art and Developing Prospects]. *Avtomatika i telemehanika – Automation and Remote Control.* 2005. vol. 4. pp. 3–58. (In Russ.).
7. Jammazi C. Finite-Time Partial Stabilizability of Chained Systems. *Comptes Rendus Mathematique.* 2008. vol. 346. no. 17-18. pp. 975−980.
8. Jammazi C.A Discussion on the Holder and Robust Finite-Time Partial Stabilizability of Brockett's Integrator. *Control, Optimization and Calculus of Variations.* 2012. vol. 18. no. 2. pp. 360−382.
9. Jammazi C. Continuous and Discontinuous Homogeneous Feedbacks Finite-Time Partially Stabilizing Controllable Multichained Systems. *Journal of Control and Optimization.* 2014. vol. 52. no. 1. pp. 520–544.
10. Chen H., Li B.Y., Zhang B.W., Zhang L. Global Finite-Time Partial Stabilization for a Class of Nonholonomic Mobile Robots Subject to Input Saturation*. International Journal of Advanced Robotic Systems.* 2015. vol. 12. no. 11. 159 p.
11. Haddad W.M., L'Afflitto A. Finite-Time Partial Stability and Stabilization, and Optimal Feedback Control. *Journal of the Franklin Institute.* 2015. vol. 352. no. 6. pp. 2329–2357.
12. Binazadeh T., Shafiei M.H., Bazregarzadeh E. New Approach in Guidance Law Design Based on Finite-Time Partial Stability Theorem. *Journal of Space Science and Technology.* 2015. vol. 8. pp. 1–7.
13. Golestani M., Mohammadzaman I., Yazdanpanah M.J. Robust Finite-Time Stabilization of Uncertain Nonlinear Systems Based on Partial Stability. *Nonlinear Dynamics*. 2016. vol. 85. no. 1. pp. 87–96.
14. L'Afflitto A. Differential Games, Finite-Time Partial-State Stabilization of Nonlinear Dynamical Systems, and Optimal Robust Control. *International Journal of Control.* 2017. vol. 90. no. 9. pp. 1861–1878.
15. Rajpurohit T., Haddad W.M. Stochastic Finite-Time Partial Stability, Partial-State Stabilization, and Finite-Time Optimal Feedback Control. *Mathematics of Control, Signals, and Systems*. 2017. vol. 29. no. 2. 10 p.
16. Jammazi C., Abichou A. Controllability of Linearized Systems Implies Local Finite-Time Stabilizability: Applications to Finite-Time Attitude Control. *Journal of Mathematical Control and Information.* 2018. vol. 35. no. 1. pp. 249–277.
17. Lyapunov A.M. *Obshhaja zadacha ob ustojchivosti dvizhenija* [The General Problem of the Stability of Motion]. Kharkov Mathematical Society. 1892. 251 p. (In Russ.).
18. Bhat S.P., Bernstein D.S. Finite-Time Stability of Continuous Autonomous Systems. *Journal on Control and Optimization.* 2000. vol. 38. no. 3. pp. 751–766.
19. Bhat S. P., Bernstein D. S. Geometric Homogeneity with Applications to Finite-Time Stability. *Mathematics of Control, Signals and Systems.* 2005. vol. 17. no. 2. pp. 101–127.
20. Vorotnikov V.I., Martyshenko Yu.G. **[**On the Nonlinear Problem of Three-Axis Reorientation of a Three-Rotor Gyrostat in the Game Noise Model]. *Kosmicheskie issledovanija – Cosmic Research*. 2013. Issue. 51. vol. 5. pp. 412. (In Russ.).
21. Vorotnikov V.I., Martyshenko Yu.G. [To Problem of Three-Rotor Gyrostat Reorientation under Uncontrolled External Disturbances]. *Mekhatronika, Avtomatizatsija, Upravlenie – Mechatronics, Automation, Control*. 2016. Issue. 17. vol. 6. pp. 414–419. (In Russ.).
22. Vorotnikov V.I., Vokhmyanina A.V. [Revisiting the Nonlinear Problem of the Passage of a Three-Rotor Gyrostat through a Given Angular Position in Space under Uncon-

trollable External Disturbances]. *Kosmicheskie issledovanija – Cosmic Research*. 2018. vol. 56. no. 5. pp. 382–387. (In Russ.).

23. Filippov A.F. *Differencial'nye uravnenija s razryvnoj pravoj chast'ju* [Differential equations with discontinuous righthand sides: control systems]. M.: Nauka. 1985. 216 p. (In Russ.).
24. Chernousko F.L., Ananievski I.M., Reshmin S.A. *Metody upravlenija nelinejnymi mehanicheskimi sistemami* [Control of Nonlinear Dynamical Systems: Methods and Applications]. M.: Fizmatlit. 2006. 328 p. (In Russ.).
25. Lurie A.I. *Analiticheskaja mehanika* [Analytical Mechanics]. M.: Fizmatlit. 1961. 824 p. (In Russ.).
26. Park Y. Robust and Optimal Attitude Stabilization of Spacecraft with External Disturbances. *Aerospace Science and Technology*. 2005. vol. 9. no. 3. pp. 253–259.
27. Ding S.H., Li S.H. Stabilization of the Attitude of a Rigid Spacecraft with External Disturbances using Finite-Time Control Techniques. *Aerospace Science and Technology.* 2009. vol. 13. no. 4-5. pp. 256–265.
28. Xia Y.Q., Zhu Z., Fu M.Y., Wang S. Attitude Tracking of Rigid Spacecraft with Bounded Disturbances. *IEEE Transactions on Industrial Electronics.* 2011. vol. 58. no. 2. pp. 647–659.
29. Hu Q., Niu G. Attitude Output Feedback Control for Rigid Spacecraft with Finite-Time Convergence. *ISA Transactions.* 2017. vol. 70. pp. 173–186.
30. Song Z., Duan C., Su H., Hu J. Full-Order Sliding Mode Control for Finite-Time Attitude Tracking of Rigid Spacecraft. *Control Theory & Applications*. 2018. vol. 12. no. 8. pp. 1086–1094.
31. Ran D., Chen X., de Ruiter A., Xiao B. Adaptive Extended-State Observer-Based Fault Tolerant Attitude Control for Spacecraft with Reaction Wheels. *Acta Astronautica*. 2018. vol. 145. pp. 501–514.
32. Minakov E.P., Sokolov B.V., Shaldaev S.E. [Investigation of the Characteristics of the Near-Moon System for Hitting Asteroids]. *Trudy SPIIRAN — SPIIRAS Proceedings*. 2017. vol. 5(54). pp. 106–129. (In Russ.).

УДК 519.677 **DOI** 10.15622/sp.61.4

В.В. Печенкин, М.С. Королёв, Л.В. Димитров

ПРИКЛАДНЫЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ АЛГОРИТМОВ РАНЖИРОВАНИЯ ДЛЯ ОРИЕНТИРОВАННЫХ ВЗВЕШЕННЫХ ГРАФОВ (НА ПРИМЕРЕ ГРАФОВ СОЦИАЛЬНЫХ СЕТЕЙ)

Печенкин В.В., Королёв М.С., Димитров Л.В. **Прикладные аспекты использования алгоритмов ранжирования для ориентированных взвешенных графов (на примере графов социальных сетей).**

Аннотация. Рассматриваются прикладные аспекты использования предварительного ранжирования вершин ориентированного взвешенного графа. Особое внимание уделяется широкому использованию такого приема в разработке эвристических алгоритмов дискретной оптимизации. Задача ранжирования имеет непосредственное отношение к проблеме определения центральности в социальных сетях, обработке больших массивов данных реального мира, но, как показано в статье, явно или косвенно используется при разработке алгоритмов решения прикладных задач в качестве начального этапа построения решения. Приводятся примеры использования предварительного ранжирования, в которых продемонстрировано повышение эффективности решения некоторых прикладных задач, имеющих широкое применение в математических методах оптимизации. Дано описание структуры первой фазы вычислительного эксперимента, которая связана с получением тестовых наборов данных. Полученные данные представлены взвешенными графами, которые соответствуют нескольким группам социальной сети ВКонтакте с числом вершин в диапазоне от 9000 до 24 тысяч участников. Показано, что структурные характеристики полученных графов по числу компонент связности существенно различаются. Продемонстрированы некоторые характеристики центральности (распределения степенных последовательностей), которые имеют экспоненциальный характер. Основное внимание уделяется анализу трех алгоритмов построения иерархии ранжирования вершин графов, предлагаются новые подходы к вычислению рангов вершин с использованием информации об активности пользователей в социальных сетях. Проводится сравнение распределений полученных совокупностей рангов. Вводится понятие сходимости алгоритмов ранжирования вершин графов, а также обсуждаются различия их использования при рассмотрении данных большой размерности и необходимости построения решения в случае учета только локальных изменений.

Ключевые слова: ранжирование, ориентированный граф, взвешенный граф, инкрементальный алгоритм, локальный алгоритм.

1. Введение. Рассматривается задача ранжирования вершин ориентированного взвешенного графа с точки зрения использования предварительного ранжирования при решении некоторых оптимизационных задач. Отметим, что в работе в качестве исходных эмпирических данных используются графы социальных сетей, что несколько сужает прикладное значение полученных результатов. Тем не менее ниже продемонстрировано, что алгоритм предварительного ранжирования является универсальным подходом во многих оптимизационных задачах, но в каждом конкретном случае требуется дополнительный анализ.

Постановка задачи имеет достаточно длинную историю и разнообразные подходы к ее решению. Алгоритмы ранжирования предложены в большом количестве, и имеется анализ их использования и реализации. В настоящее время в связи с тем, что появляются структурированные данные большой размерности — графы с десятками, сотнями тысяч вершин, эта задача вновь становится актуальной. Такого рода структурированные данные необходимо обрабатывать и анализировать их структуру. При этом важной задачей становится снижение временной сложности алгоритмов обработки за счет распараллеливания и обработки только локальной окрестности той части графа, в которой произошли изменения. То есть в этом случае решение задачи для всего графа можно получить из имеющегося глобального решения за счет его локального изменения на ограниченной окрестности вершин. Такая постановка задачи в современных исследованиях связана с понятием инкрементального алгоритма поиска решения, который для коррекции глобального решения использует обработку вершин, находящихся в локальной окрестности части графа, подвергшейся изменению [1, 2].

Далее будут приведены примеры использования предварительного ранжирования вершин графа, что приводит к повышению эффективности решения уже конкретной задачи. Другой рассматриваемой задачей статьи является анализ результатов конкретных алгоритмов ранжирования на предмет типа распределения полученной совокупности рангов.

2. Формальная постановка задачи. Мы рассматриваем задачу построения такого отображения множества вершин ориентированного взвешенного графа в множество вещественных чисел, которое отражает доминирование одной вершины по отношению к другой. Алгоритмы такого рода используются при ранжировании сайтов в сети Интернет и имеют приложения к разнообразным задачам. В качестве определения ориентированного взвешенного графа мы используем стандартное понятие графа $G=(V, E, w)$, где V — множество вершин графа, E — множество дуг (ориентированных ребер), $w: V \rightarrow R$ — функция веса дуг. Смысловым источником такого рода задач является классическая задача топологической сортировки, которая имеет формальную постановку, опирающуюся на следующие определения. В том случае, если функция веса действует тривиальным способом, задавая вес любой дуги равным 1, мы будем опускать ее при определении графа.

Определение 1. Дан граф $G=(V, E)$, говорим, что вершина v достижима из вершины u, если $v = u$, или $(u, v) \in E$, или существует путь $p = v_1, v_2, \ldots, v_n$ в графе G такой, что $v_1 = u$ и $v_n = v$. Факт

достижимости в этом случае обозначаем как $u \sim> v$. Для понятия «взвешенный граф», которое предполагает наличие функции веса для ориентированных дуг, используется такое же определение.

Близкой по своей природе к задаче ранжирования является задача топологической сортировки (ТС), которая может быть сформулирована следующим образом: топологической сортировкой графа является отображение $f: V \rightarrow R$ такое, что справедлива формула (1):

$$\forall v,u \in V(v \sim> u) \rightarrow (f(v) < f(u)). \quad (1)$$

Классическая постановка задачи о ТС накладывает серьезные ограничения на граф — он должен быть ациклическим. Заметим, что в общем случае такие графы являются достаточно редкими, и необходимы процедуры ранжирования, пригодные для произвольных графов. Задача ТС может быть решена в ходе обхода графа в ширину и имеет сложность $O(n+m)$, где n — количество вершин графа, m — количество дуг.

Под формальным определением задачи ранжирования вершин ориентированного взвешенного графа мы будем понимать задачу, описанную следующим образом. Под ранжированием понимается определение функции f, как это сделано при определении топологической сортировки, определенной выше, но имеющей следующие существенные отличия:

1. Допускается существование таких пар вершин графа $v, u \in V$, что $f(v) = f(u)$.
2. Отношение ранжирования строится на основе композиции отношений ранжирования для локального окружения каждой пары вершин.
3. Снимается ограничение об отсутствии циклов в графе.

Анализируемые далее алгоритмы могут быть разбиты в соответствии с нашими установками на два класса — локальные и глобальные. *Локальные* алгоритмы при вычислении ранга вершины учитывают структурные характеристики локальной окрестности вершины в графе. При вычислении ранга вершины в этом случае учитывается только направленность и вес дуг, инцидентных данной вершине. *Глобальные* алгоритмы получают решение в результате последовательности итераций, учитывающих изменения текущего решения для всех вершин графа. Использованные далее способы вычисления рангов были адаптированы для нашей задачи и нормализованы к диапазону сравнимых значений.

3. Алгоритмы ранжирования в решении прикладных задач. Многие приложения (анализ структуры указателей при организации

блоков памяти операционной системы, инкрементальная компиляция) сводятся к реализации ТС ациклических ориентированных графов в условиях их динамического изменения [3]. Вычисление различных характеристик динамически изменяющихся графов представляет определенный интерес, и в этом плане наибольшее значение имеют так называемые инкрементальные алгоритмы, которые позволяют сконструировать глобальное решение для всего графа, учитывая только локальные изменения и имеющееся решение задачи до момента изменения. Так для решения задачи построения ТС вершин графа разработаны эффективные алгоритмы динамической сортировки, которые учитывают операции добавления/удаления ребер графа и позволяют корректировать имеющееся решение сразу после изменений [4, 5].

Опишем ряд результатов, которые показывают, что задача ранжирования вершин ориентированного графа имеет большое значение для повышения эффективности оптимизационных алгоритмов. Предварительное ранжирование и индексирование данных является традиционным способом повышения скорости выполнения запросов к реляционным базам данных. Аналогичная ситуация имеет место и при работе с NoSQL базами данных (Neo4j, DEX), которые содержат информацию о динамических структурах с большим количеством отношений различного типа [6, 7].

Повышение скорости работы алгоритмов на основе ранжирования вершин графа предлагается и для решения классических оптимизационных задач. В работе [8] описан муравьиный алгоритм решения обобщенной оптимизационной задачи (рассматривается задача коммивояжёра) на основе набора эвристик. На основе инкрементального определения концентрации искусственного феромона определяются вероятности выбора следующей вершины графа при выборе продолжения промежуточного решения. В процессе работы алгоритма все вершины графа ранжируются и вероятность выбора следующей вершины зависит от ее ранга.

Работа [9] содержит описание двух алгоритмов ранжирования вершин ориентированных и неориентированных графов, которые используют матрицу Лапласа. Особое внимание к этой проблеме связано с задачей обучения интеллектуальных систем принятия решений в условиях динамического изменения данных различной природы. В качестве примеров, используемых для демонстрации эффективности алгоритмов, рассматриваются два набора данных. Первый имеет отношение к компьютерной биологии и содержит иерархическую классификацию протеинов. Второй содержит разбиение текстовых новостных статей на группы по связанным с ними тегам. Результаты вычислительного эксперимента показывают

уменьшение суммарной ошибки ранжирования при использовании динамической процедуры. К особенностям рассматриваемой задачи относится то, что на структуру графов накладываются серьезные ограничения — они должны быть κ-дольными, а в качестве используемых массивов для вычислительного эксперимента выбраны данные именно такой природы.

В работе [10] рассматривается оптимизационная задача поиска кратчайшего пути на местности, описываемой географической картой. Особенностью предложенного в работе подхода является использование математического аппарата нечетких графов, у которых функция принадлежности ребра имеет вид:

$$\mu : V^2 \to [0,1].$$

Решаемая задача связана с нечеткостью определения на географической карте объектов (дорог и полигонов) и, соответственно, расстояний между ними. Предлагается переход от карты к нечеткому графу, учитывая параметры сетей дорог и полигонов, подлежащие оптимизации. Длина пути в графе вычисляется подобно обычному алгоритму для взвешенного графа (сумма весов дуг, входящих в путь, должна быть минимальной), но используются операции сложения и нахождения экстремумов для нечетких чисел. Ранжирование вершин используется при вычислении кратчайших путей в графах, соответствующих картам.

Похожий подход используется в работе [11], в которой задается граф с нечеткими весами дуг и трапециевидной функцией принадлежности для решения задачи повышения скорости передачи данных в оптических сетях. Кратчайшие пути в этом случае ранжируются предлагаемым авторами алгоритмом. Результаты описанного вычислительного эксперимента показывают повышение производительности маршрутизации, которое связано с уменьшением времени на поиск оптимального пути и увеличением пропускной способности сети от 10 до 20%.

Ранжирование как один из этапов оптимизационного алгоритма часто используется в математическом моделировании систем видеонаблюдения. В работе [12] решается задача повторной идентификации человека в видеопотоке, который получается с нескольких камер видеонаблюдения. Полученное множество оценок видеоизображений людей преобразуется в графовую модель, для которой решается задача ранжирования, упрощающая построение консенсусной оценки. В [13] на основе ранжирования решается задача поиска скрытых объектов при визуальном отслеживании некоторой сцены.

4. Исходные данные и сравнение результатов ранжирования. Для проведения вычислительного эксперимента используются наборы

данных, которые были получены при анализе структурных характеристик виртуальных сообществ социальной сети ВКонтакте. Схема проведенного эксперимента представлена на рисунке 1.

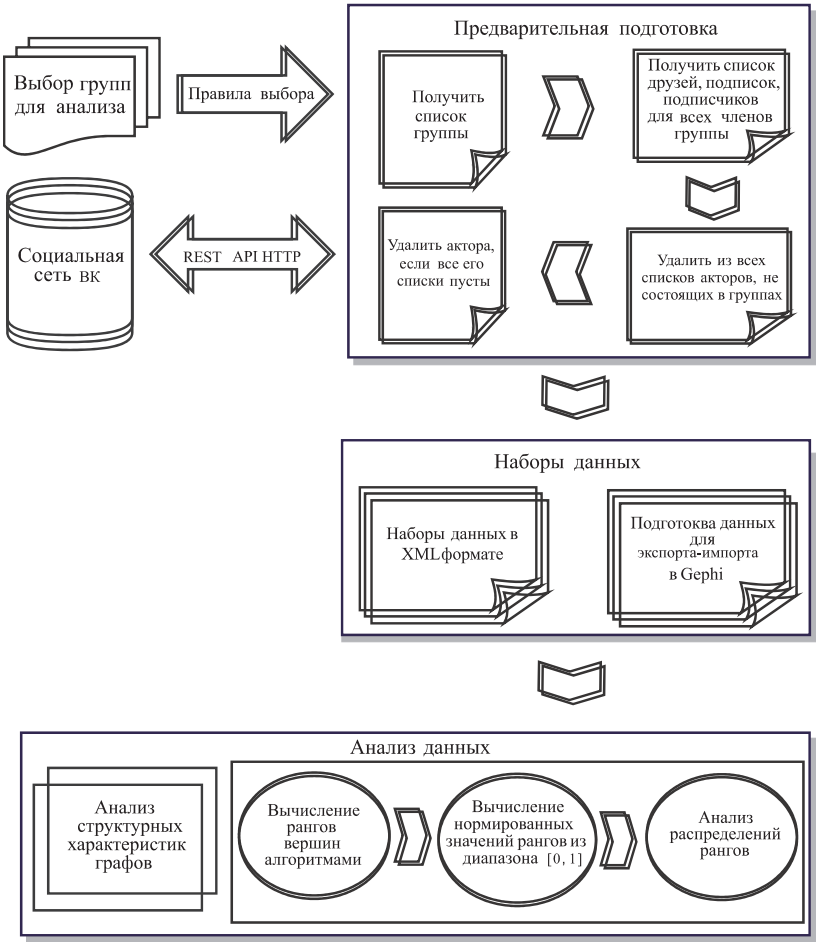


Рис. 1. Структура вычислительного эксперимента

Специально разработанное программное обеспечение было использовано для извлечения данных о пользователях групп, постах, комментариях и лайках к ним. Приложение вызывает методы HTTP интерфейса (REST HTTP API), предоставляемого социальной сетью. Разработанная программа позволяет получить данные обо всех

интересующих нас интеракциях пользователей в рамках одной группы. Для преобразования данных из объектной модели в XML файлы и обратно используется свободно распространяемая библиотека JAXB. В качестве целевых выбирались группы численностью порядка 10-30 тысяч человек.

Далее приводятся результаты анализа графов, которые получены для конкретных групп с условными названиями «Волонтеры», «Шесть рукопожатий», «Письма добра», имеющие от 10 до 24 тысяч участников. Описание процедуры получения данных и некоторые структурные характеристики графов для соответствующих групп приведены в работе [14].

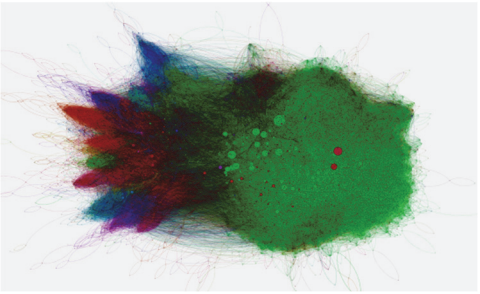
Последним шагом этапа предварительной подготовки данных является удаление всех изолированных вершин из полученных графов. Этот шаг позволяет «очистить» данные от вершин, которым могут быть присвоены минимальные ранги, так как они не связаны ни с одной вершиной графа. Такая процедура выглядит естественной и обоснованной.

Второе замечание касается представленных далее визуальных образов графов. В случае, когда размеры графов исчисляются десятками тысяч вершин, их визуальное представление позволяет увидеть структуру графа и характерные для них структурные различия. Здесь нужно отметить, что различные алгоритмы укладки графов могут сформировать принципиально различные визуализации, которые подчеркивают разные особенности их структурных характеристик. Для получения визуализации был использован алгоритм ForceAtlas 2, реализованный в открытой платформе для визуализации графов Gephi (https://gephi.org/), который объединяет несколько теоретических подходов к укладке. Этот алгоритм основан на минимизации энергии (вершины притягиваются или отталкиваются друг от друга в зависимости от их взаимного расположения и наличия связей), которая приписывается графу в целом, его вершинам и ребрам. Алгоритм хорошо подходит для построения изображений, подчеркивающих структуру графа, а также для визуализации подмножеств вершин с высокой степенью взаимодействия. Визуальное представление графов приведено на рисунке 2. Для каждого графа также указывается число вершин и компонент связности. Эти величины лишний раз подчеркивают различия в структурных характеристиках графов. Графы на рисунке 2 визуализируют отношение «подписка» (односторонняя дуга) и «дружба» (двусторонняя дуга), то есть между двумя вершинами может быть либо однонаправленная стрелка, либо двунаправленная.

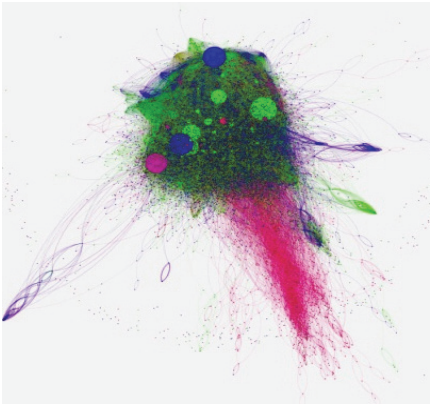
Графики на рисунке 3 дают представление о виде распределения полустепеней исхода и захода для вершин соответствующего графа,

представляющего первый набор данных «Волонтеры». На рисунке 3(a) представлено распределение полустепеней захода, на рисунке 3(b) — распределение полустепеней исхода. Для других наборов данных графики выглядят подобным образом.

Волонтеры
N = 23 651
Компонент
связности: 77



Шесть
рукопожатий
N = 23 979
Компонент
связности 399



Письма добра
N =10464
Компонент
связности: 121

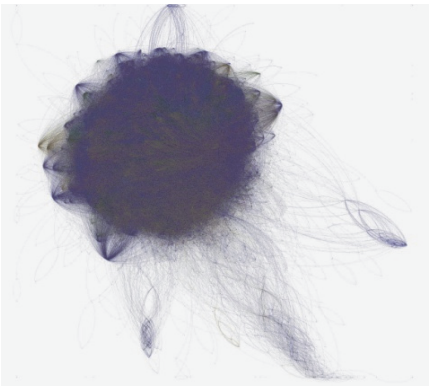
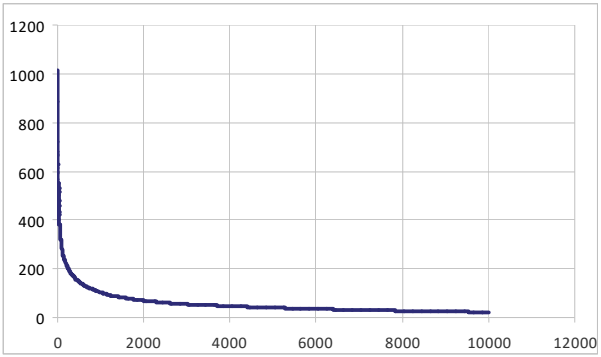
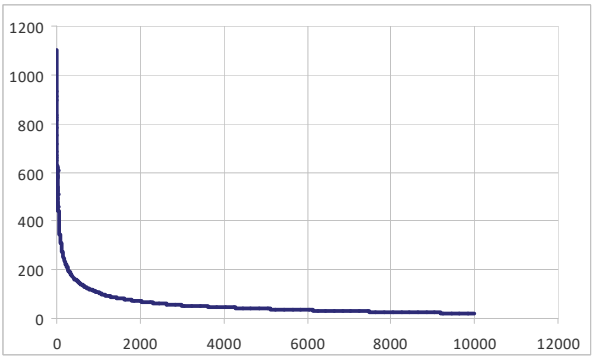


Рис. 2. Визуальное представление структуры трех групп с использованием алгоритма «ForceAtlas 2»



a)



b)

Рис. 3. Распределение полустепеней захода (a) и исхода (b) вершин графа «Волонтеры»

4.1. Анализируемые алгоритмы ранжирования. В качестве алгоритмов ранжирования рассматриваются три алгоритма, которые были использованы для построения иерархий в различных ситуациях при анализе структуры ссылочного пространства глобальной сети или структуры социальных сетей. В работе анализируется популярный и подробно описанный в литературе алгоритм ранжирования PageRank, локальный алгоритм Twitter, и глобальный алгоритм Freeman, осуществляющий рекуррентный пересчет рангов, но дающий похожие результаты. Краткое описание этих алгоритмов дано в работе [15].

В общем виде задача ранжирования, решаемая в рассматриваемых далее подходах, применяется к моделям, описывающим взаимодействие объектов глобальной сети, которые представляют акторов социальной сети либо страницы сети Интернет.

Имеющиеся данные представляют акторов социальной сети, для которых будут предложены формулы, справедливые для определения рангов вершин, исходя из весов дуг и модифицированные формулы, которые учитывают такие стандартные действия участников групп, как «Понравилось» («Лайки»).

Первый алгоритм построения иерархии ранжирования вершин ориентированного взвешенного графа использует только характеристики их локального окружения и с точки зрения нашей терминологии является локальным [16]. Оригинальный подход решает задачу вычисления формально определенной меры доверия между пользователями микроблога Twitter. На основании этой меры моделируется и предсказывается последовательность переходов при использовании сервиса. В этом случае степень доверия пропорциональна количеству взаимодействий между двумя акторами и определяется только параметрами связывающих их дуг, то есть является локальной характеристикой в принятых нами обозначениях. Расчет меры доверия актора *j* актору *i* осуществляется по формуле (2):

$$r_{ij} = \frac{wS_{ij} + (1-w)F_{ij}}{N_i}, \tag{2}$$

где $S_{ij}=1$, если актор j подписан на актора i, и $S_{ij}=0$ в противном случае; $F_{ij}=1$, если актор i подписан на актора j, и $F_{ij}=0$ в противном случае; N_i — общее количество подписок и подписчиков актора i; w — коэффициент степени влияния S_{ij} и F_{ij}.

При проведении вычислений в работе рекомендуется принять значение w в диапазоне от 0,5 до 1. При использовании значения $w = 0,5$ приравнивается степень влияния наличия собственных подписок и подписчиков на итоговый результат.

Для учета влияния «лайков» постов и комментариев необходимо дополнительно изменить формулу (2) расчета ранга. Для этого добавляются новые слагаемые в числитель формулы (2):

$$r_{ij} = \frac{wS_{ij} + (1-w)F_{ij} + cL_{ij} + (1-c)M_{ij}}{N_i}, \tag{3}$$

где c — коэффициент, показывающий степень влияния одного «лайка» на уровень доверия между двумя вершинами; L_{ij} — количество «лайков», которые актор i поставил под постами и комментариями j; M_{ij} — количество «лайков», которые актор j поставил под постами и комментариями i.

В рамках проведенного вычислительного эксперимента коэффициент c принят равным 0,25.

Для вычисления ранга актора i суммируются степени доверия ему всех акторов, которые с ним взаимодействуют. Таким образом, ранг пользователя будет рассчитываться по формуле:

$$r_i = \sum_{\forall j \in N_i} r_{i,j}. \tag{4}$$

Далее обозначаем алгоритм ранжирования вершин, основанный на формуле (4), как ***TW***.

В работе [17] предлагается способ итеративного построения иерархии ранжирования, основанный на вероятностных характеристиках взаимодействия в социальных сетях. В этом случае учитываются не только взаимодействие между актором сети и его локальным окружением, но и текущий усредненный уровень рангов вершин из этого окружения. То есть на значение ранга на каждой итерации алгоритма влияют и внешние по отношению к его окружению интеракции. Следующая формула (5) используется для подсчета ранга вершины i:

$$r_i = k\frac{F_i - S_i}{N_i} + Q_i, \tag{5}$$

где k — постоянный коэффициент, значение которого в цитируемой статье принято равным 2; F_i — количество подписчиков актора i; S_i — количество подписок актора i; N_i — общее количество подписок и подписчиков актора i (рассчитывается как количество уникальных акторов, которые являются подписчиками и тех, на кого подписан актор); Q_i — усредненный ранг всех акторов, с которыми взаимодействует актор i.

Итерации алгоритма продолжаются до тех пор, пока не будет достигнута нужная точность (максимальная разница между значениями ранга вершины на двух последовательных итерациях не будет превышать наперед заданного значения ε). Если алгоритм не сходится, и разница между значениями рангов в последовательных итерациях не стремится к 0, то необходимо ограничить количество итераций. Максимальное количество итераций в нашем алгоритме было установлено в значение 500, ε = 0,1. Формальное определение сходимости будет дано далее.

Для учета «лайков» комментариев и постов акторов добавляется дополнительное слагаемое в числителе формулы (5):

$$r_i = \frac{k(F_i - S_i) + \frac{k}{2}(M_i - L_i)}{N_i} + Q_i, \tag{6}$$

где L_i — количество лайков, которые поставил актор; M_i — количество лайков, которые были поставлены на записи и комментарии текущего актора; S_i , F_i — множество всех вершин, с которыми взаимосвязана вершина I.

Далее алгоритм ранжирования, основанный на формуле (6), обозначается как ***FR***.

Третий использованный в работе алгоритм ранжирования PageRank описан в работе [18]. Итеративный алгоритм подсчитывает ранг каждой вершины по следующей формуле:

$$r_i = c\sum\frac{r_j}{N_i} + cE_i, \tag{7}$$

где r_i — ранг i-ой вершины; c — коэффициент нормализации; F_i — множество вершин, ссылающихся на вершину i; S_i — множество вершин, на которые ссылается вершина i; $N_i = S_i \cup F_i$; E_i — некоторое первоначальное значение ранга вершины. При описании алгоритма утверждается, что оптимальными значением этого параметра является 0,15.

Далее алгоритм ранжирования, основанный на формуле (7), обозначается как ***PR***.

Алгоритм ***PR*** и его модификации подробно рассматриваются в литературе, проанализированы преимущества и недостатки его использования. Но особенное внимание в последнее время уделяется практике использования модификаций ***PageRank*** для моделей данных, имеющих большую размерность. Для этого случая разработаны специальные алгоритмы, использующие только локальные фрагменты всей модели и распределенную среду вычисления [19, 20]. В работе [21] рассматривается вычисление иерархии ранжирования в случае, когда граф может содержать «определенные» и «неопределенные» дуги. Авторами применяется традиционная для таких задач терминология, имеющая отношение к центральности. «Определенные» дуги являются традиционными для ориентированного графа с одной начальной и одной конечной

вершинами. «Неопределенные» дуги имеют единственную вершину в качестве начальной, но множественные варианты конечной вершины, называемые «потенциальными» конечными. В качестве допустимого варианта потенциальной конечной рассматривается и вершина, которая отсутствует в заданном графе, но имеет специальное обозначение и добавлена в формальную математическую модель графа. Предложенный алгоритм ориентирован на динамический инкрементальный пересчет рангов при изменении графа в противовес статическому, когда вся совокупность рангов вычисляется заново при любом изменении.

Подсчет рангов вершин графов с использованием алгоритма ***PR*** производился с помощью программы Gephi 0.82 beta. Два других индекса (***TW*** и ***FR***) рассчитывались с использованием программного обеспечения собственной разработки.

4.2. Сходимость алгоритма ранжирования для заданного графа. Дадим определение понятия сходимости алгоритма ранжирования для заданного набора данных. Такое определение необходимо для корректного обсуждения временной сложности обсуждаемых алгоритмов, которая зависит не только от размерности исходных графов, но и от количества итераций, выполняемых алгоритмами. Последнее определяется либо наперёд заданным числом, либо выполнением некоторого условия, позволяющего прекратить вычисления.

Предполагаем, что алгоритм выполняет последовательное вычисление всей совокупности рангов вершин. Пусть $r^i = \left(r_1^i, r_2^i, ..., r_n^i\right)$ — набор рангов для вершин заданного графа G, полученный на i-ой итерации алгоритма. В качестве меры расстояния D^i между двумя последовательными наборами r^i и r^{i+1}, рассматриваемыми как векторы, удобно использовать метрику Манхэттена, определенную формулой (8):

$$D^i = d(r^i, r^{i+1}) = \sum_{j=1}^{n} \left| r_j^i - r_j^{i+1} \right|. \tag{8}$$

Более «мягким» вариантом метрики, которая предъявляет менее жесткие условия к процедуре вычисления рангов, является метрика Чебышева, определяемая формулой:

$$D^i = d(r^i, r^{i+1}) = \max_{j=1,..,n} \left| r_j^i - r_j^{i+1} \right|. \tag{9}$$

Определение 2. Говорим, что алгоритм ранжирования сходится для заданного графа *G,* если:

$$\lim_{i\to\infty} D^i = 0. \tag{10}$$

Очевидно, что проблема сходимости (поточечной или равномерной) алгоритма ранжирования для заданного графа актуальна только для алгоритмов ***FR*** и ***PR***, так как они используют рекуррентную процедуру вычисления рангов. Алгоритм ***TW*** при вычислении рангов вершин использует только характеристики их локального окружения и не требует пересчета всей совокупности рангов при локальном изменении в графе.

4.3. Результаты вычислительного эксперимента. Диапазоны полученных рангов для трех алгоритмов, которые вычисляются по представленным выше формулам, для одних и тех же данных различаются. Например, при использовании алгоритма ***FR***, ранги могут становиться отрицательными из-за структуры формул (5), (6). По этой причине значения всех рангов были приведены к диапазону [0,1] преобразованием (11):

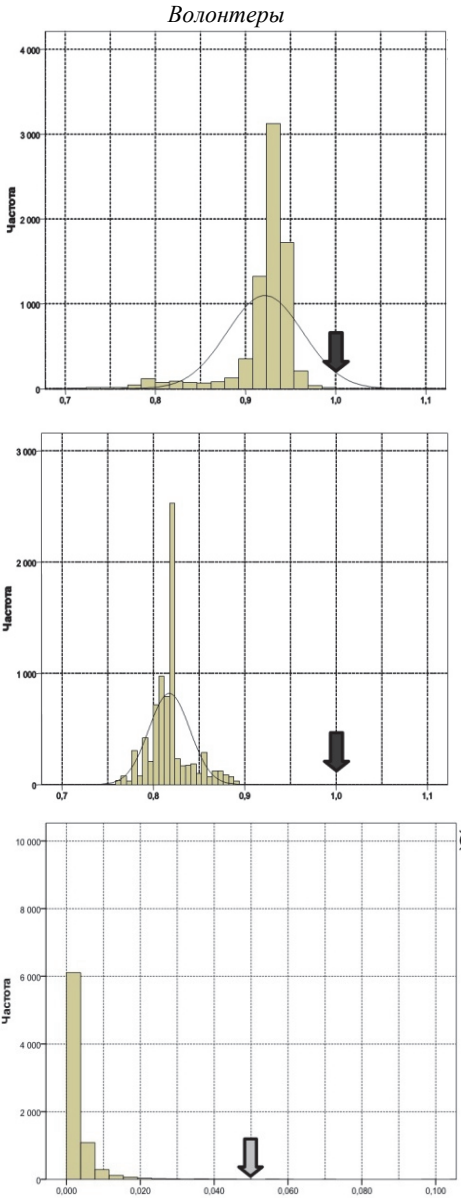
$$r_i' = \frac{r_i - \min\limits_{i=1,N}(r_i)}{\max\limits_{i=1,N}(r_i) - \min\limits_{i=1,N}(r_i)}, \tag{11}$$

где максимум и минимум берутся по всему множеству рангов.

На рисунке 4 (a, b, c) показаны гистограммы распределений рангов вершин для графа, представляющего группу «Волонтеры».

Гистограммы сформированы для совокупности рангов, полученных алгоритмами ***FR***, ***TW*** и ***PR***. На графиках по оси абсцисс отложено нормализованное значение ранга вершин, на оси ординат — частота появления рангового диапазона. Для наборов данных «Шесть рукопожатий» и «Письма добра» получены похожие результаты, демонстрирующие такие же закономерности в распределении полученных рангов. Гистограммы распределений для этих графов показаны на рисунках 5 и 6 в уменьшенном масштабе, но позволяют подтвердить выявленные и описанные далее закономерности. Отдельно подчеркнем существенные различия в структурных характеристиках полученных графов (количество компонент связности, статистики степенных последовательностей и некоторые другие), описанных более подробно в работе [14].

Подобная задача решается в работе [22], в которой рассматриваются структурные характеристики виртуальных сообществ, представленных графами.



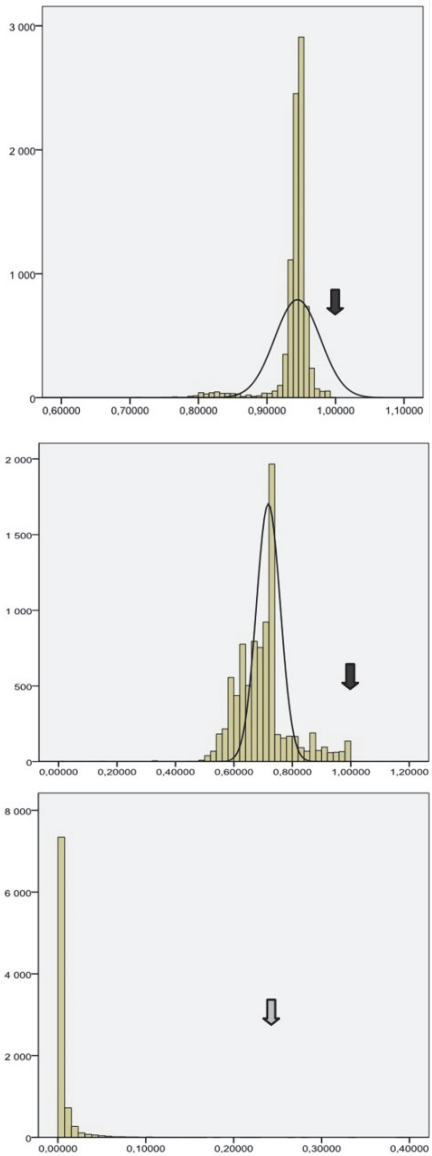
FR *(a)*

TW *(b)*

PR *(c)*

Рис. 4. Гистограммы распределений рангов вершин графа «Волонтеры» для ***FR***(a), ***TW***(b), ***PR***(c). На a, b стрелка показывает положение 1, c — позиция значения 0,05

Письма Добра

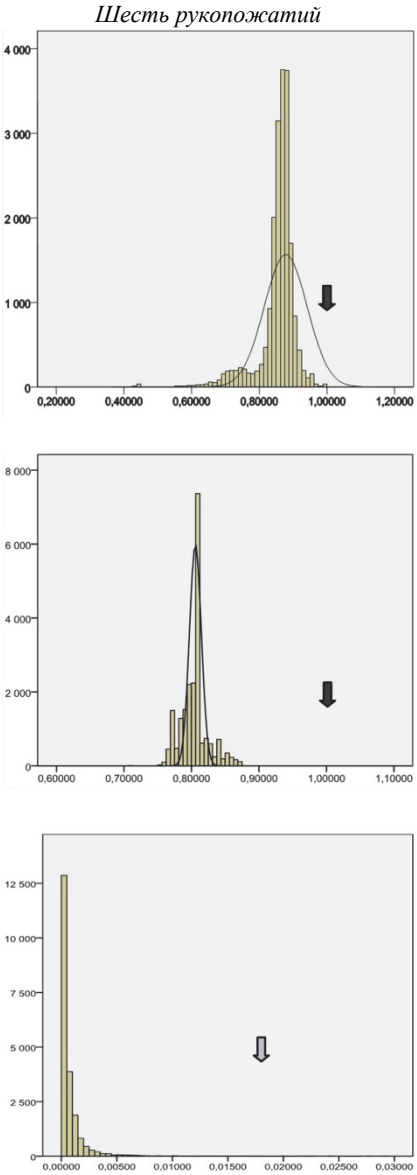


FR (a)

TW *(b)*

PR *(c)*

Рис. 5. Гистограммы распределений для наборов данных «Письма добра» при использовании алгоритмов ***FR***(a), ***TW***(b), ***PR***(c). На a, b стрелка показывает положение 1, c — позиция значения 0,25



FR (a)

TW (b)

PR (c)

Рис. 6. Гистограммы распределений для наборов данных «Шесть рукопожатий» при использовании алгоритмов ***FR***(a), ***TW***(b), ***PR***(c). На a, b стрелка показывает положение 1, c — позиция значения 0,0175

Распределение полученных рангов подчиняются сходным закономерностям, которые могут быть описаны следующими пунктами.

1. В описанных в разделе 3 публикациях предложены решения оптимизационных задач, которые используют предварительное решение задачи ранжирования вершин. Этот подход позволяет повысить эффективность алгоритмов, улучшить их временные оценки.

2. Имеется существенное отличие распределений рангов, генерируемых различными алгоритмами ранжирования. Если алгоритмы ***FR*** и ***TW*** в качестве результата строят ранги, имеющие распределение близкое к нормальному, то ранги, полученные алгоритмом ***PR***, имеют распределение близкое к распределению, которое может быть описано как экспоненциальное (см. рисунки 4-6) или степенное с определенным набором взаимосвязанных параметров. Информация о типе распределения позволяет корректно выбирать параметрические тесты для анализа совокупности рангов.

3. Алгоритм ***FR*** дает совокупность рангов, распределение которых после нормирования смещено в сторону максимального значения 1, алгоритм же ***TW*** даёт меньшие значения распределения с меньшим разбросом значений. Этот вывод подтверждается данными таблица 1, которая позволяет сравнить средние значения, стандартное отклонение и скошенность распределений полученных рангов. Это различие можно объяснить тем, что алгоритм ***FR*** увеличивает ранги за счет дополнительного слагаемого Q_i в формулах 5, 6, что приводит к сдвигу центра распределения в сторону максимального значения 1.

Таблица 1. Статистические параметры распределения рангов

		FR	***TW***	***PR***
Волонтеры	Среднее	0,921	0,818	0,004
	Медиана	0,930	0,818	0,002
	Ст. отклонение	0,039	0,022	0,014
	Скошенность	-3,783	0,905	-
Шесть рукопожатий	Среднее	0,854	0,804	0,001
	Медиана	0,866	0,808	0,000
	Ст. отклонение	0,057	0,021	0,002
	Скошенность	-2,697	0,553	-
Письма добра	Среднее	0,938	0,703	0,006
	Медиана	0,946	0,702	0,007
	Ст. отклонение	0,050	0,089	0,015
	Скошенность	-13,552	1,051	-

4. Алгоритмы ***FR*** и ***PR*** используют рекуррентную схему вычисления следующего приближения рангов, что актуализирует проблему сходимости алгоритма для заданного набора данных. Фактически, в вычислительном эксперименте условием остановки процедуры вычисления рангов стало достижение наперед заданного максимального числа итераций. Алгоритм ***TW*** при вычислении ранга вершины использует только параметры ее локального окружения и при динамическом изменении графа (добавление/удаление дуги, изменение ее атрибутов) потребует только пересчета рангов двух вершин, которые соединены этой дугой. Все остальные ранги остаются неизменными. Это обстоятельство позволяет использовать эффективные инкрементальные/декрементальные процедуры вычисления рангов, сложность которых зависит только от размера локальных окружений вершин и сложности процедуры коррекции глобальной совокупности рангов.

5. Заключение. В работе на примере графов большой размерности проанализированы результаты применения алгоритмов ранжирования вершин. Показано, что процедура ранжирования имеет большое прикладное значение для графов большой размерности, представляющих отношения в социальных сетях. Для анализа применимости результатов в решении оптимизационных задач самой разнообразной природы необходимы дополнительные исследования, учитывающие особенности предметной области, используемых дискретных моделей.

Проанализированные алгоритмы ранжирования имеют различные вычислительную сложность и распределения полученных рангов. Если при решении прикладной задачи к распределению рангов предъявляются какие-либо требования (например, нормальность распределения является обязательным при использовании некоторых параметрических тестов), необходимо предварительно провести анализ результатов работы алгоритма, определить статистические характеристики полученного распределения рангов.

При использовании предварительного ранжирования вершин графа в случае большой размерности модели исходных данных важным моментом является возможность использования инкрементальных/декрементальных алгоритмов, которые минимизируют сложность обработки данных, как правило, учитывают при построении глобального решения только локальные изменения, не требуют пересчета всех рангов. В этом смысле алгоритм ***TW***, вычисляющий ранги только по данным локальной окрестности вершин, выгодно отличается от алгоритмов ***FR*** и ***PR***, которые требуют пересчета всей совокупности рангов даже при локальном изменении

графа. Это обстоятельство не позволяет использовать последние два алгоритма в случае, когда размерность модели является большой, необходимо быстро получить решение, не пересчитывая всю совокупность рангов. Распределенная модель вычислений смягчает это требование, но не устраняет его принципиально.

В качестве дискуссионного момента можно выделить вопрос о свойствах графа, которые гарантируют сходимость алгоритма ранжирования. Алгоритмы ***FR*** и ***PR*** в проведенном вычислительном эксперименте качестве решения выдавали ранги после предельного, наперед заданного числа итераций. В статье предложено определение сходимости алгоритма ранжирования, оставляя за скобками интересный теоретический вопрос о критериях сходимости, определении нетривиальных структурных характеристик графов, наличие которых гарантирует достижение заданной точности вычисления рангов при выбранной мере расстояния, существование предела рекуррентной последовательности.

Этот же вопрос возникает и по поводу структуры формул, используемых для последовательного вычисления рангов. Ясно, что если второе слагаемое в Q_i в формулах (5) и (6) для алгоритма ***FR***, будет монотонно убывающей последовательностью в процессе итераций, то соответствующий алгоритм будет сходиться. Регулируя скорость уменьшения значений Q_i можно получить управляемую скорость сходимости всего процесса для графов любой структуры.

Литература

1. *Кочетов Ю.А., Хмелев А.В.* Гибридный алгоритм локального поиска для задачи маршрутизации разнородного ограниченного автопарка // Дискретный анализ и исследование операций. 2015. Т. 22. № 5. С. 5–29.
2. *Кривошеин Д.Ю., Марченко А.М.* Алгоритмы пересчёта кратчайших путей в графе при изменении весов ребер // Проблемы разработки перспективных микро– и наноэлектронных систем. 2012. № 1. С. 263–266.
3. *Demetrescu C., Finocchi I., Italiano G.F.* Dynamic graphs. URL: www.diku.dk/PATH05/CRC-book1.pdf (дата обращения: 01.02.17).
4. *Pearce D.J., Kelly P.H.J.* A dynamic topological sort algorithm for directed acyclic graphs // Journal of Experimental Algorithmics (JEA). 2007. vol. 11. pp. 1–7.
5. *Deepak A., Tobias F.* Average-Case Analysis of Incremental Topological Ordering // Discrete Applied Mathematics. 2010. vol. 158. no. 4. pp. 240–250.
6. *Nicoara D., Kamali S., Daudjee K., Chen L.* Hermes: Dynamic Partitioning for Distributed Social Network Graph Databases // EDBT. 2015. pp. 25–36.
7. *Ammar A.B.* Query optimization techniques in graph Databases // International Journal of Database Management Systems (IJDMS). 2016. vol. 8. no. 4. 14 p.
8. *Курейчик В.В., Жиленков М.А.* Муравьиный алгоритм для решения оптимизационных задач с явно выраженной целевой функцией // Информатика, вычислительная техника и инженерное образование. 2015. № 2. С. 1–12.

9. *Agarw S.* Ranking on Graph Data // Proceedings of the 23rd International Conference on Machine Learning. 2006. pp. 25–32.
10. *Розенберг И.Н.* Использование нечетких представлений данных при определении медиан графа // Известия Южного федерального университета. Технические науки. 2001. № 4. С. 64–72.
11. *Adaikalam A., Manikandan S., Rajamani V.* Fuzzy graph based shortest path ranking method for optical network // Optical and Quantum Electronics. 2017. vol. 49. no. 9. 296 p.
12. *Barman A., Shah S.K.* SHaPE: A Novel Graph Theoretic Algorithm for Making Consensus-Based Decisions in Person Re-identification Systems // IEEE International Conference on Computer Vision (ICCV). 2017. pp. 1124–1133.
13. *Roffo G., Melzi S., Castellani U., Vinciarelli A.* Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach // IEEE International Conference on Computer Vision (ICCV). 2017. pp. 1407–1415.
14. *Печенкин В.В., Решетников Д.С., Ярская-Смирнова В.Н.* Визуализация сетевой структуры групповых отношений // 4М. Методология, методы, математическое моделирование. 2014. № 39. С. 40–61.
15. *Королёв М.С., Решетников Д.С.* Подходы к задаче ранжирования вершин в теории графов // Проблемы управления в социально-экономических и технических системах. 2017. С. 138–141.
16. *Lumbreras A., Gavaldà R.* Applying trust metrics based on user interactions to recommendation in social networks // IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2012. pp. 1159–1164.
17. *Jameson K.A., Appleby M.C., Freeman L.C.* Finding an appropriate order for a hierarchy based on probabilistic dominance // Animal Behaviour. 1999. vol. 57. no. 5. pp. 991–998.
18. Easy visualizations of PageRank and Page Groups with Gephi. URL: https://searchengineland.com/easy-visualizations-pagerank-page-groups-gephi-265716 (дата обращения: 07.02.2017).
19. *Sarma A.D., Molla A.R., Pandurangan G., Upfal E.* Fast distributed pagerank computation // Theoretical Computer Science. 2015. vol. 561. pp. 113–121.
20. *Dai L., Freris N.M.* Fully distributed PageRank computation with exponential convergence // 2017. arXiv preprint arXiv:1705.09927. 5 p.
21. *Nathan E., Fairbanks J., Bader D.* Ranking in Dynamic Graphs using Exponential Centrality // International Workshop on Complex Networks and their Applications. 2017. pp. 378–389.
22. *Рыков Ю.Г., Кольцова О.Ю., Мейлахс П.А.* Структура и функции онлайн-сообществ: сетевая картография ВИЧ-релевантных групп в социальной сети «ВКонтакте» // Социологические исследования. 2016. № 8. С. 30–42.

Печенкин Виталий Владимирович — д-р социол. наук, профессор, профессор кафедры прикладных информационных технологий института прикладных информационных технологий и коммуникаций, Саратовский государственный технический университет имени Гагарина Ю.А.. Область научных интересов: оптимизационные задачи на графах и сетях, визуализация графов, методы статистической обработки многомерных данных, многомерное шкалирование, анализ социальных сетей, использование математических и компьютерных подходов в социологических исследованиях. Число научных публикаций — 110. pechenkinvv@mail.ru; Политехническая, 77, Саратов, 41005; р.т.: +7(8452)99-87-15.

Королёв Михаил Сергеевич — аспирант кафедры прикладных информационных технологий института прикладных информационных технологий и коммуникаций, Саратовский государственный технический университет имени Гагарина Ю.А., ассистент кафедры кафедры прикладных информационных технологий института прикладных информационных технологий и коммуникаций, Саратовский государственный технический университет имени Гагарина Ю.А.. Область научных интересов: алгоритмы и методы оптимизации, виртуальная реальность, дополненная реальность, 3D-проектирование, 3D-моделирвоание, архитектурная визуализация, компьютерное зрение, компьютерный дизайн.. Число научных публикаций — 8. koroliow.mikhail@yandex.ru; Политехническая, 77, Саратов, 41005; р.т.: +7(8452)99-87-15.

Димитров Любомир Ванков — д-р физ.-мат. наук, профессор, проректор по учебной деятельности и аккредитации, Технический университет - София, профессор машиностроительного факультета, Технический университет - София. Область научных интересов: мехатроника, автоматика, микроэлектронные модули и системы и их применение (MEMS). Число научных публикаций — 230. lubomir_dimitrov@tu-sofia.bg; бул. Св. Климент Охридски, 8, София, 1756, Болгария; р.т.: +359 2 965-25 60.

DOI 10.15622/sp.61.4

V.V. Pechenkin, M.S. Korolev, L.D. Dimitrov

APPLIED ASPECTS OF RANKING ALGORITHMS FOR ORIENTED WEIGHTED GRAPHS (ON THE EXAMPLE OF SOCIAL NETWORK GRAPHS)

Pechenkin V.V., Korolev M.S., Domotrov L.D. **Applied Aspects of Ranking Algorithms for Oriented Weighted Graphs (on the Example of Social Network Graphs).**

Abstract. The article deals with the applied aspects of the preliminary vertices ranking for oriented weighted graph. In this paper, the authors observed the widespread use of this technique in developing heuristic discrete optimization algorithms. The ranking problem is directly related to the problem of social networks centrality and large real world data sets, but as shown in the article ranking is explicitly or implicitly used in the development of algorithms as the initial stage of obtaining a solution for solving applied problems. Examples of such ranking application are given. The examples demonstrate the increase of efficiency in solving some optimization applied problems, which are widely used in mathematical methods of optimization, decision-making not only from the theoretical development point of view but also their applications. The article describes the structure of the first phase of the computational experiment, which is associated with the procedure of obtaining test data sets. The obtained data are presented by weighted graphs that correspond to several groups of the social network Vkontakte with the number of participants in the range from 9000 to 24 thousand. It is shown that the structural characteristics of the obtained graphs differ significantly in the number of connectivity components. Characteristics of centrality (degree's sequences), as shown, have exponential distribution. The main attention is given to the analysis of three approaches to graph vertices ranking. We propose analysis and comparison of the obtained set of ranks by the nature of their distribution. The definition of convergence for graph vertex ranking algorithms is introduced and the differences of their use in considering the data of large dimension and the need to build a solution in the presence of local changes are discussed.

Keywords: ranking, oriented graph, weighted graph, incremental algorithm, local algorithm.

Pechenkin Vitaly Vladimirovich — Ph.D., Dr. Sci., professor, professor of applied information technologies department of school of applied information technology and communication, Yuri Gagarin State Technical University of Saratov (SSTU). Research interests: optimization problems on graphs and networks, graph visualization, statistical processing of multidimensional data, multidimensional scaling, social networks analysis, application of mathematics and computer science in sociological studies. The number of publications — 110. pechenkinvv@mail.ru; 77, Politechnicheskaya, Saratov, 410054, Russia; office phone: +7(8452)99-87-15.

Korolev Mikhail Sergeevich — Ph.D. student of applied information technologies department of school of applied information technology and communication, Yuri Gagarin State Technical University of Saratov (SSTU), assistant of applied information technologies department of school of applied information technology and communication, Yuri Gagarin State Technical University of Saratov (SSTU). Research interests: optimization algorithms and methods, virtual reality, augmented reality, 3d-technology, computer vision, computer design.. The number of publications — 8. koroliow.mikhail@yandex.ru; 77, Politechnicheskaya, Saratov, 410054, Russia; office phone: +7(8452)99-87-15.

Dimitrov Lyubomir Vankov — Ph.D., Dr. Sci., vice-rector of learning activity and accreditation, Technical University of Sofia, professor of engineering faculty, Technical

University of Sofia. Research interests: mechatronics, automation, microelectronic modules and systems and their application(MEMS). The number of publications — 230. lubomir_dimitrov@tu-sofia.bg; 8, Kliment Orhidski Boulevard, Sofia, 1756, Bulgaria; office phone: +359 2 965-25 60.

References

1. Kochetov Y.A., Khmelev V.A. [A Hybrid algorithm of local search for routing problem with heterogeneous fleet limited]. *Diskretnyj analiz i issledovanie operatsij – Journal of Applied and Industrial Mathematics.* 2015. Issue 22. vol. 5. pp. 5–29. (In Russ.).
2. Krivoshein D.Yu., Marchenko A.m. [Algorithms for recalculating shortest paths in a graph when changing edge weights]. *Problemy razrabotki perspektivnykh mikro- i nanoehlektronnykh sistem* [Problems of perspective micro- and nanoelectronic systems development]. 2012. vol. 1. pp. 263–266. (In Russ.).
3. Demetrescu C., Finocchi I., Italiano G.F. Dynamic graphs. Available at: www.diku.dk/PATH05/CRC-book1.pdf (accessed: 01.02.17).
4. Pearce D.J., Kelly P.H.J. A Dynamic Topological Sort Algorithm for Directed Acyclic Graphs. *Journal of Experimental Algorithmics (JEA)*. 2007. vol. 11. pp. 1–7.
5. Deepak A., Tobias F. Average-Case Analysis of Incremental Topological Ordering. *Discrete Applied Mathematics.* 2010. vol. 158. no. 4. pp. 240–250.
6. Nicoara D., Kamali S., Daudjee K., Chen L. Hermes: Dynamic Partitioning for Distributed Social Network Graph Databases. EDBT. 2015. pp. 25–36.
7. Ammar A.B. Query optimization techniques in graph Databases. *International Journal of Database Management Systems (IJDMS).* 2016. vol. 8. no. 4. 14 p.
8. Kureychik V.V., Zhilenkov M.A. [Ant algorithm for solving optimization problems with explicit objective function information, computing and engineering education]. *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie – Computer science, computer engineering and engineering education*. 2015. vol. 2. pp. 1–12. (In Russ.).
9. Agarw S. Ranking on Graph Data. Proceedings of the 23rd International Conference on Machine Learning. 2006. pp. 25–32.
10. Rosenberg I.N. [The use of fuzzy representations of data when determining the medians of the graph]. *Izvestiya YUFU. Tekhnicheskie nauki – Izvestiya SFedU. Engineering sciences.* 2001. vol. 4. pp. 64–72. (In Russ.).
11. Adaikalam A., Manikandan S., Rajamani V. Fuzzy graph based shortest path ranking method for optical network. *Optical and Quantum Electronics*. 2017. vol. 49. no. 9. 296 p.
12. Barman A., Shah S.K. SHaPE: A Novel Graph Theoretic Algorithm for Making Consensus-Based Decisions in Person Re-identification Systems. IEEE International Conference on Computer Vision (ICCV). 2017. pp. 1124–1133.
13. Roffo G., Melzi S., Castellani U., Vinciarelli A. Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach. IEEE International Conference on Computer Vision (ICCV). 2017. pp. 1407–1415.
14. Pechenkin V.V., Reshetnikov D.S., Yarskaya-Smirnova V.N. [Visualization of the network structure of group relations] *4M. Metodologiya, metody, matematicheskoe modelirovanie – 4M. Methodology, methods, mathematical modeling.* 2014. vol. 39. pp. 40–61. (In Russ.).
15. Korolev M.S., Reshetnikov D.S. [Approaches to the problem of ranking vertices in graph theory]. *Problemy upravleniya v sotsial'no-ehkonomicheskikh i tekhnicheskikh sistemakh* [Problems of control in socio-economic and technical systems]. 2017. pp. 138–141. (In Russ.).
16. Lumbreras A., Gavaldà R. Applying trust metrics based on user interactions to recommendation in social networks. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2012. pp. 1159–1164.

17. Jameson K.A., Appleby M.C., Freeman L.C. Finding an appropriate order for a hierarchy based on probabilistic dominance. *Animal Behaviour.* 1999. vol. 57. no. 5. pp. 991–998.
18. Easy visualizations of PageRank and Page Groups with Gephi. Available at: https://searchengineland.com/easy-visualizations-pagerank-page-groups-gephi-265716 (accessed: 07.02.2017).
19. Sarma A.D., Molla A.R., Pandurangan G., Upfal E. Fast distributed pagerank computation. *Theoretical Computer Science.* 2015. vol. 561. pp. 113–121.
20. Dai L., Freris N.M. Fully distributed PageRank computation with exponential convergence. 2017. arXiv preprint arXiv:1705.09927. 5 p.
21. Nathan E., Fairbanks J., Bader D. Ranking in Dynamic Graphs using Exponential Centrality. International Workshop on Complex Networks and their Applications. 2017. pp. 378–389.
22. Rykov YU.G., Kol'tsova O.YU., Mejlakhs P.A. [Structure and Functions of Online Communities Network Mapping of HIV-relevant groups in VK]. *Sociological Studies – Sotsiologicheskie issledovaniya*. 2016. vol. 8. pp. 30–42.

УДК 512.624.5 **DOI** 10.15622/sp.61.5

А.А. Молдовян, Н.А. Молдовян

СПОСОБЫ И АЛГОРИТМЫ ПСЕВДОВЕРОЯТНОСТНОГО ШИФРОВАНИЯ С РАЗДЕЛЯЕМЫМ КЛЮЧОМ

Молдовян А.А., Молдовян Н.А. **Способы и алгоритмы псевдовероятностного шифрования с разделяемым ключом.**

Аннотация. В качестве способа обеспечения секретности сообщений, переданных в зашифрованном виде по открытым каналам связи, при потенциальных атаках с принуждением к раскрытию ключей шифрования предложены алгоритмы и протоколы отрицаемого шифрования, которые разделяются на следующие типы: 1) с открытым ключом; 2) с разделяемым секретным ключом; 3) бесключевые. В статье описываются псевдовероятностные симметричные шифры, представляющие собой специальный вариант реализации алгоритмов отрицаемого шифрования. Обсуждается применение псевдовероятностного шифрования для построения специальных механизмов защиты информации, в том числе стеганографических каналов, носителями которых являются шифртексты. Рассмотрены способы построения поточных и блочных алгоритмов псевдовероятностного шифрования, реализующих совместное шифрование фиктивного и секретного сообщений таким образом, что формируемый шифртекст является вычислительно неразличимым от шифртекста, получаемого в результате вероятностного шифрования фиктивного сообщения. В качестве одного из критериев построения использовано требование неотличимости по шифртексту псевдовероятностного шифрования от вероятностного. Для реализации этого требования в схеме построения псевдовероянностных шифров используется шаг взаимно-однозначного отображения пар блоков промежуточных шифртекстов фиктивного и секретного сообщений в единый расширенный блок выходного шифртекста. Описаны реализации псевдовероятностных блочных шифров, в которых алгоритмы расшифровывания фиктивного и секретного сообщений полностью совпадают. Предложены общие подходы к построению псевдовероятностных протоколов бесключевого шифрования и рандомизированных псевдовероятностных блочных шифров, а также приведены конкретные реализации криптосхем данных типов.

Ключевые слова: криптография, отрицаемое шифрование, псевдовероятностное шифрование, блочный шифр, поточный шифр, фиктивное сообщение, рандомизация шифров, бесключевое шифрование.

1. Введение. Криптографические методы и средства защиты информации широко применяются для защиты информации [1, 2], аутентификации сообщений и пользователей [3, 4] в информационно-телекоммуникационных системах. Они также лежат в основе технологии электронной цифровой подписи [5, 6] и решения задачи обеспечения анонимности пользователей [7] в технологиях тайного электронного голосования и электронных денег. При рассмотрении стойкости криптографических алгоритмов и протоколов обычно принимаются модели потенциального нарушителя, в которых секретный ключ неизвестен. Сравнительно новым направлением в области криптографии является разработка протоколов передачи сообщений, обеспечивающих стойкость к атакам со стороны нарушителя, который принуждает абонентов

сеанса секретной связи раскрыть секретный ключ после того как шифртекст был передан по каналу связи [8-11]. Криптосхемы, обеспечивающие защищенность секретного сообщения при указанных принуждающих атаках, называются протоколами и алгоритмами отрицаемого шифрования (ОШ). Интерес к криптосхемам ОШ связан с решением задач обеспечения информационной безопасности распределенных вычислений [12], защиты технологий тайного электронного голосования от скупки голосов [13, 14] и расширением класса алгоритмических средств защиты информации, используемых в составе комплексных средств обеспечения информационной безопасности [15-17].

Прикладной интерес представляют предложенные недавно реализации схем ОШ на основе механизмов разделения секрета [18, 19] и одноразовых ключей [20], а также протоколы отрицаемой аутентификации [21], ориентированные на применение в системах электронного голосования.

Протоколы ОШ могут быть разделены на следующие три класса: криптосхемы с разделяемым секретным ключом [8, 22], криптосхемы с открытым ключом [23, 24] и бесключевые криптосхемы [25].

В статье [25] впервые рассмотрены протоколы ОШ, основанные на коммутативных функциях шифрования и обсуждается реализация протоколов бесключевого ОШ. Однако предложенная в [25] конкретная реализация протоколов последнего типа не в полной мере соответствует термину «бесключевой», поскольку в ней дополнительно к локальным ключам (которые вырабатываются каждой стороной протокола самостоятельно и не передаются другой стороне) используется вспомогательный ключ, разделяемый получателем и отправителем секретного сообщения.

В рамках первого класса недавно был предложен подход к построению алгоритмов ОШ, в которых выполняется одновременное зашифровывание фиктивного и секретного сообщений в единый шифртекст, вычислительно неразличимый от шифртекста, формируемого как результат вероятностного шифрования фиктивного сообщения [15, 22]. Алгоритмы и протоколы, удовлетворяющие этому критерию, называются псевдовероятностными (ПВ) криптосхемами. В настоящее время известны отдельные публикации, связанные с разработкой алгоритмов ПВ шифрования, однако тематика ПВ шифрования как самостоятельная область прикладной криптографии не была рассмотрена.

В настоящей статье обобщаются известные результаты в области ПВ шифрования, выделяются общие приемы построения криптосхем данного типа, предлагаются новые блочные и поточные ПВ шифры, рассматривается механизм рандомизации ПВ шифров и представ-

лен новый протокол бесключевого ПВ шифрования, в котором устранен недостаток, состоящий в использовании заранее согласованного вспомогательного ключа.

2. Типы псевдовероятностных шифров с разделяемым ключом и модель нарушителя. Известные ПВ криптосхемы с разделяемым ключом делятся на алгоритмы ПВ шифрования следующих типов: поточные [15], блочные [22], алгебраические [25]. Общими моментами, используемыми при их построении, являются:

1) совместное зашифровывание фиктивного и секретного сообщений в единый выходной шифртекст;

2) выполнение критерия вычислительной неотличимости по шифртексту от вероятностного шифрования;

3) предъявление алгоритма вероятностного шифрования, для которого множество шифртекстов, соответствующих фиктивному сообщению, включает шифртекст, полученный в результате ПВ шифрования.

Указанный алгоритм вероятностного шифрования называется ассоциируемым (с алгоритмом ПВ шифрования), поскольку его наличие служит доказательством выполнимости критерия вычислительной неотличимости формируемого шифртекста от криптограммы, вырабатываемой при вероятностном шифровании фиктивного сообщения по фиктивному ключу.

Вероятностное шифрование находит применение как способ повышения стойкости криптографического преобразования, поэтому в случае атаки с принуждением к раскрытию ключа шифрования пользователи, предъявляя атакующему ассоциированный алгоритм, могут правдоподобно утверждать, что при шифровании они использовали криптографическое преобразования с включением случайных значений.

На практике возможны различные варианты потенциальных атак, в рамках которых атакующий получает секретный ключ, использованный для выполнения шифрования. Например, это может произойти в результате подкупа, хищения ключевых носителей, выполнения криптоанализа, предварительной несанкционированной установки программ-закладок и так далее. При рассмотрении протоколов ОШ для обобщения таких атак рассматривается модель принуждающей атаки (или атаки с принуждением), в рамках которой предполагается, что атакующий имеет некоторый ресурс принуждения отправителя сообщения, получателя или одновременно их обоих к раскрытию секретного ключа. При этом принимается предположение, что атакующий перехватил все сообщения, переданные в ходе реализации коммуникационного протокола.

В случае ПВ шифрования как частного случая ОШ стойкость к принуждающим атакам обеспечивается тем, что одновременно зашифровываются два или более сообщения и по крайней мере одно из них

является фиктивным. Атакующему раскрывается ключ, по которому выполнение процедуры восстановления исходного текста по шифртексту приводит к получению фиктивного сообщения. При этом алгоритм расшифровывания не должен содержать признаков, по которым атакующий мог бы сделать обоснованный вывод о возможности восстановления из шифртекста и некоторого другого сообщения. При формулировке требований к алгоритмам ПВ шифрования в качестве указанных признаков рассматриваются следующие признаки:

– неполнота использования шифртекста в ходе выполнения процедуры расшифровывания;

– зависимость процесса расшифровывания от ключа (например, наличие ветвлений в алгоритме расшифровывания, зависящих от ключа);

– неравномерность влияния битов криптограммы на биты восстановленного сообщения.

При разработке алгоритмов ПВ шифрования следует обеспечить отсутствие перечисленных признаков. Потенциально возможны атаки, в которых атакующий может измерить время выполнения процедуры расшифровывания секретного сообщения, а также получить доступ к машинному коду, соответствующему программе алгоритма расшифровывания, или непосредственно к самой программе. Для обеспечения стойкости к атакам последнего типа при разработке алгоритмов ПВ шифрования принимается дополнительный критерий построения, формулируемый следующим образом: *алгоритмы восстановления фиктивного и секретного сообщений должны полностью совпадать.* Этот критерий означает, что один и тот же алгоритм должен восстанавливать фиктивное и секретное сообщения в зависимости от задаваемого ключа расшифровывания.

3. Особенности псевдовероятностного шифрования как механизма защиты информации. Алгоритмы ПВ шифрования позволяют реализовать криптографические обманные ловушки, с помощью которых атакующий направляется на ложный путь. Например, потенциальному нарушителя создаются условия для перехвата (или хищения) фиктивного ключа, с помощью которого из шифртекста восстанавливается фиктивное сообщение. Или размер фиктивного ключа выбирается таким, что атакующий имеет возможность его раскрыть методом полного перебора.

Псевдовероятностное шифрование можно трактовать как способ построения стеганографического канала криптографическими средствами. Действительно, благодаря вычислительной неразличимость по шифртексту алгоритма ПВ шифрования от ассоциированного алгоритма вероятностного шифрования при получении (каким-либо способом) ключа для расшифровывания фиктивного сообщения атакующий не

имеет возможности определить однозначно существование в шифртексте еще одного сообщения.

Само по себе существование способов ПВ шифрования ставит потенциального криптоаналитика перед следующей дилеммой. Предположим, ему удалось восстановить ключ, с помощью которого шифртекст расшифровывается в осмысленное сообщение, однако текущие попытки найти еще один ключ, с помощью которого из шифртекста могло бы быть восстановлено еще одно сообщение, оказываются безуспешными. Следует ли криптоаналитику продолжить вычислительно затратный процесс криптоанализа или принять решение, что перехваченный шифртекст получен в процессе вероятностного шифрования и следует прекратить попытки решения неразрешимой задачи?

Применение алгоритмов вероятностного шифрования, которые могут быть ассоциированы с некоторыми алгоритмами псевдовероятностного шифрования, для защиты передаваемых сообщений (файлов, хранимых в ЭВМ) дает возможность встраивания в отдельные шифртексты дополнительных сообщений (файлов). Прежде чем приступить к раскрытию таких криптографических стегоканалов криптоаналитику требуется решить задачу распознавания шифртекста, допускающего возможность неоднозначного расшифровывания.

В целом ПВ шифры предоставляют возможность разработки и использования новых механизмов защиты информации.

4. Псевдовероятностные блочные шифры. Общим подходом к построению псевдовероятностных блочных шифров, описанных в работах [15, 22], является выполнение следующих трех обобщенных шагов преобразования:

1) разбиение фиктивного и секретного сообщения на блоки данных;

2) независимое зашифровывание пары соответствующих друг другу блоков фиктивного и секретного сообщений на различных ключах;

3) совместное зашифровывание пары блоков промежуточных шифртекстов, полученных на шаге 2, в единый блок выходного шифртекста с помощью обратимой процедуры преобразования.

При этом используемая на шаге 3 процедура зашифровывания задается таким образом, что обратное ей преобразование выполняется как независимое восстановление блоков промежуточных шифртекстов, соответствующих фиктивному и секретному сообщениям, осуществляемое по одним и тем же математическим формулам. Рассмотрим конкретные варианты реализации описанного общего подхода.

Алгоритм ПВ шифрования с использованием процесса решения системы линейных сравнений как процедуры биективного отображения пары блоков промежуточных шифртекстов в единый блок выходного шифртекста описывается следующим образом.

Зададим выполнение совместного шифрования двух различных сообщений $M = (M_1, M_2, \ldots, M_z)$ и $T = (T_1, T_2, \ldots, T_z)$, представленных в виде последовательности n-битовых блоков данных M_i и T_i ($i = 1, 2, \ldots, z$), по ключам (K_1, p_1) и (K_2, p_2) соответственно, причем K_1 и K_2 — ключи некоторого блочного шифра E с n-битовым входом; p_1 и p_2 — взаимно простые числа, удовлетворяющие условиям $2^{n+1} > p_1 > 2^n$ и $2^{n+1} > p_2 > 2^n$:

1. Используя алгоритм блочного шифрования E и ключ K_1, зашифровать i-й блок сообщения T_i: $C_{T_i} = E_{K_1}(T_i)$.

2. Используя блочный шифр E, зашифровать i-й блок сообщения M_i по ключу K_2: $C_{M_i} = E_{K_2}(M_i)$.

3. Используя промежуточные шифртексты C_{T_i} и C_{M_i} и подключи p_1 и p_2, вычислить блок выходного шифртекста C_i как решение следующей системы сравнений:

$$\begin{cases} C_i \equiv C_{T_i} \bmod p_1 \\ C_i \equiv C_{M_i} \bmod p_2 \end{cases}, \tag{1}$$

где выходные блоки C_{T_i} и C_{M_i} функции шифрования E рассматриваются как n-битовые двоичные числа. Криптограмма C, содержащая в себе в скрытом виде сообщения T и M, формируется в виде следующей последовательности блоков шифртекста C_i размером $2n + 2$ бит: $C = (C_1, C_2, \ldots, C_z)$.

В соответствии с китайской теоремой об остатках решение системы линейных сравнений (1) описывается формулой:

$$C_i = \left[C_{T_i} p_2 \left(p_2^{-1} \bmod p_1\right) + C_{M_i} p_1 \left(p_1^{-1} \bmod p_2\right)\right] \bmod p_1 p_2.$$

При выполнении вычислений по этой формуле наибольший вклад в вычислительную трудоемкость расчета блока криптограммы C_i вносят две операции инверсии по модулям p_1 и p_2 и операция деления на число $p_1 p_2$. Вычисление значений $p_2\left(p_2^{-1} \bmod p_1\right)$ и $p_1\left(p_1^{-1} \bmod p_2\right)$ может быть осуществлено на этапе генерации секретных ключей. В этом случае основной вклад в трудоемкость вычисле-

ния значения C_i вносит операция деления значения в квадратных скобках на модуль p_1p_2, которую надо выполнять при формировании каждого нового блока криптограммы, объединяющей два текущих блока промежуточных шифртекстов C_{T_i} и C_{M_i}.

С описанным алгоритмом ПВ шифрования ассоциируется следующий алгоритм вероятностного шифрования фиктивного сообщения M:

1. Разбить сообщение M на n-битовые блоки данных M_i: $M = (M_1, M_2, \ldots, M_z)$.

2. Каждый i-й ($i = 1, 2, \ldots, z$) блок зашифровать, выполнив следующие три шага:

2.1. Зашифровать блок данных M_i по ключу K_2 с использованием n-битового блочного алгоритма шифрования E по формуле $C_{M_i} = E_{K_2}(M_i)$.

2.2. Сгенерировать случайное число $R < 2^n$ и простое случайное значение $r \neq p_2$, удовлетворяющее условию $2^n < r < 2^{n+1}$.

2.3. Вычислить i-й блок криптограммы C_i как решение следующей системы сравнений:

$$\begin{cases} C_i \equiv C_{M_i} \bmod p_2 \\ C_i \equiv R \bmod r \end{cases}. \qquad (2)$$

Легко можно увидеть, что в шифртексте C каждый i-й блок C_i потенциально может быть получен как результат преобразования блока фиктивного сообщения M_i в соответствии с ассоциированным алгоритмом вероятностного шифрования. Причем это реализуется при выборе многих различных пар значений $R < 2^n$ и $r < 2^{n+1}$. Для произвольного простого числа r, удовлетворяющего условию $rp_2 < C_i$, по формуле $R \equiv C_i \bmod r$ находим число R, при котором для пары значений r и R решение системы (2) совпадает со значением C_i. Это показывает, что шифртекст C потенциально может быть получен в результате вероятностного шифрования фиктивного сообщения по фиктивному ключу (K_2, p_2).

Чтобы доказать, что шифртекст C содержит не только фиктивное, но и секретное сообщение T, потенциальному криптоаналитику потребуется вычислить ключ (K_1, p_1) и восстановить по нему из шифртекста C сообщение T. Однако последнее даже при известном значении p_2 не проще взлома алгоритма блочного шифрования E. Действительно, по известному p_2 можно вычислить шифртекст, формируемый

на выходе функции блочного шифрования E при шифровании сообщения T по ключу K_1, то есть имеем стандартные условия, при которых блочные шифры должны быть стойкими.

Расшифровывание криптограммы $C = (C_1, C_2, \ldots, C_i, \ldots, C_z)$ по фиктивному ключу (K_2, p_2) выполняется следующим образом:

1. Каждый i-й ($i = 1, 2, \ldots, z$) блок C_i расшифровать, выполнив следующие два шага:

1.1. Вычислить блок промежуточного шифртекста $C_{M_i} = C_i \bmod p_2$.

1.2. Расшифровать блок C_{M_i} по ключу K_2, используя функцию блочного расшифровывания $D = E^{-1}$: $M_i = D_{K_2}(C_{M_i})$.

2. Объединить восстановленные блоки данных M_i в единое сообщение $M = (M_1, M_2, \ldots, M_i, \ldots, M_z)$.

Для восстановления секретного сообщения T из криптограммы $C = (C_1, C_2, \ldots, C_i, \ldots, C_z)$ используется ключ (K_1, p_1) и тот же алгоритм расшифровывания:

1. Преобразовать каждый i-й блок шифртекста C_i:

1.1. Вычислить значение $C_{T_i} = C_i \bmod p_1$.

1.2. Расшифровать блок промежуточного шифртекста C_{T_i} по формуле: $T_i = D_{K_1}(C_{T_i})$.

2. Объединить восстановленные блоки данных T_i в единое сообщение $T = (T_1, T_2, \ldots, T_i, \ldots, T_z)$.

По аналогии с рассмотренным алгоритмом блочного ПВ шифрования может быть построен ПВ блочный шифр с использованием вычислений над двоичными многочленами (подключи и преобразуемые блоки данных рассматриваются как двоичные многочлены, представленные упорядоченным набором коэффициентов последнего). При таком подходе получаем следующий алгоритм совместного шифрования сообщений T и M по ключам (K_1, η_1) и (K_2, η_2), где подключи η_1 и η_2 — взаимно неприводимые двоичные многочлены степени n:

1. Вычислить n-битовый блок промежуточного шифртекста C_{T_i} по формуле $C_{T_i} = E_{K_1}(T_i)$.

2. Вычислить n-битовый блок промежуточного шифртекста C_{M_i} по формуле $C_{M_i} = E_{K_2}(M_i)$.

3. Сформировать $2n$-битовый блок выходного шифртекста C_i как решение следующей системы сравнений:

$$\begin{cases} C_i \equiv C_{T_i} \bmod \eta_1 \\ C_i \equiv C_{M_i} \bmod \eta_2 \end{cases}, \tag{3}$$

в которой блоки C_{T_i} и C_{M_i} промежуточных шифртекстов трактуются как двоичные многочлены числа, а размер блока шифртекста C_i равен $2n$.

Решение системы линейных сравнений (3) задается формулой:

$$C_i = \left[C_{T_i}\eta_2\left(\eta_2^{-1} \bmod \eta_1\right) \oplus C_{M_i}\eta_1\left(\eta_1^{-1} \bmod \eta_2\right)\right] \bmod \eta_1\eta_2,$$

где $\oplus$ — операция сложения двоичных многочленов (поразрядное суммирование битовых строк по модулю два). Так же как и в случае шифра-налога, вычисление значений $\eta_2\left(\eta_2^{-1} \bmod \eta_1\right)$ и $\eta_1\left(\eta_1^{-1} \bmod \eta_2\right)$ может быть осуществлено на этапе генерации секретных ключей, что позволяет значительно повысить производительность процедуры шифрования.

Заслуживает внимания вариант реализации блочного ПВ шифра с различным размеров входных блоков данных M_i и T_i. Например, для большей степени скрытности криптографического стегоканала секретное сообщение предварительно сжимается с устранением его избыточности и существенным сокращением размера текста $T = (T_1, T_2, \ldots, T_i, \ldots, T_z)$. Если размеры блоков данных задаются равными значениям n_1 и n_2, то, соответственно, следует задать степени многочленов η_1 и η_2 равными n_1 и n_2. Размер блока выходного шифртекста в точности равен сумме $n_1 + n_2$.

Рассмотрим построение блочного ПВ шифра, в котором в качестве процедуры преобразования пар блоков промежуточных шифртекстов в единый блок выходного шифртекста используется решение системы уравнений в конечном поле. В данном случае в отличии от предыдущего алгоритма размер входных блоков M_i и T_i должен быть одинаковым: $n_1 = n_2 = n$. Пусть, например, $n = 128$ и блоки промежуточного шифртекста формируются путем зашифровывания блоков фиктивного и секретного сообщений с помощью 128-битовой функции блочного шифрования E и 256-битовых ключей $K = (K_1, K_2)$ и $Q = (Q_1 , Q_2)$, каждый из которых разбит на два 128-битовых подключа. Генерацию ключей K и Q выполним как генерацию пар равноверо-

ятных случайных 128-битовых строк, рассматриваемых как двоичные многочлены и удовлетворяющих условию $K_1 Q_2 \oplus K_2 Q_1 \neq 0 \bmod \eta$, где η — неприводимый двоичный многочлен степени 128.

Процедуру совместного шифрования сообщений T и M зададим в виде следующих шагов:

1. Разбить сообщения T и M на 128-битовые блоки T_i и M_i.

2. Каждый i-й блок T_i ($i = 1, 2, \ldots, z$) и каждый i-й блок M_i зашифровать, выполнив следующие два шага:

2.1. Зашифровать блок данных T_i по ключу Q: $C_{T_i} = E_Q(T_i)$.

2.2. Зашифровать блок данных M_i по ключу K: $C_{M_i} = E_K(M_i)$.

2. Для каждого значения $i = 1, 2, \ldots, z$ сформировать 256-битовый блок криптограммы $C_i = (C'_i, C''_i)$ в виде конкатенации двух 128-битовых двоичных многочленов C'_i и C''_i, являющихся решением следующей системы линейных уравнений с неизвестными C'_i и C''_i:

$$\begin{cases} K_1 C'_i \oplus K_2 C''_i \equiv C_{M_i} \bmod \eta \\ Q_1 C'_i \oplus Q_2 C''_i \equiv C_{T_i} \bmod \eta \end{cases}, \tag{4}$$

Ассоциируемый алгоритм вероятностного шифрования имеет вид:

1. Разбить сообщение M на 128-битовые блоки M_i.

2. Каждый i-й блок M_i ($i = 1, 2, \ldots, z$) зашифровать, выполнив следующие два шага:

2.1. Зашифровать блок данных M_i по ключу K: $C_{M_i} = E_K(M_i)$.

2.2. Сгенерировать случайные двоичные многочлены λ и ρ степени 127.

2.3. Вычислить i-й 256-битовый блок шифртекста $C_i = (C'_i, C''_i)$ как решение следующей системы сравнений:

$$\begin{cases} K_1 C'_i \oplus K_2 C''_i \equiv C_{M_i} \bmod \eta \\ C'_i \oplus \lambda C''_i \equiv \rho \bmod \eta \end{cases}. \tag{5}$$

При фиксированном ключе K и фиксированном блоке промежуточного шифртекста C_{M_i} один и тот же блок C_i криптограммы в общем случае может быть получен с помощью ассоциированного алгоритма вероятностного шифрования при различных парах значений многочленов λ и ρ. Действительно, выбор произвольного многочлена λ од-

нозначно определяет значение ρ, при котором система уравнений (5) в качестве своего решения будет иметь пару многочленов C'_i, C''_i, таких, что $C_i = (C'_i, C''_i)$.

Расшифровывание криптограммы $C = (C_1, C_2, \ldots, C_i, \ldots, C_z)$ по фиктивному ключу $K = (K_1, K_2)$ выполняется следующим образом:

1. Каждый i-й ($i = 1, 2, \ldots, z$) 256-битовый блок $C_i = (C'_i, C''_i)$ расшифровать, выполнив следующие два шага:

1.1. Вычислить 128-битовый блок промежуточного шифртекста по формуле $C_{M_i} \equiv K_1 C'_i \oplus K_2 C''_i \bmod \eta(x)$;

1.2. Расшифровать 128-битовый блок C_{M_i} промежуточного шифртекста по ключу K, используя функцию блочного расшифровывания $D = E^{-1}$: $M_i = D_K(C_{M_i})$.

2. Объединить все восстановленные блоки данных M_i в единое сообщение $M = (M_1, M_2, \ldots, M_i, \ldots, M_z)$.

Секретное сообщение восстанавливается из шифртекста $C = (C_1, C_2, \ldots, C_z)$ по ключу $Q = (Q_1, Q_2)$ с использованием идентичного алгоритма:

1. Каждый 256-битовый блок $C_i = (C'_i, C''_i)$ расшифровать, выполнив следующие два шага:

1.1. Вычислить 128-битовый блок промежуточного шифртекста по формуле $C_{T_i} \equiv Q_1 C'_i \oplus Q_2 C''_i \bmod \eta$.

1.2. Расшифровать 128-битовый блок C_{T_i} промежуточного шифртекста по ключу Q, используя функцию блочного расшифровывания $D = E^{-1}$: $M_i = D_Q(C_{T_i})$.

2. Объединить все восстановленные блоки данных T_i в единое сообщение $T = (T_1, T_2, \ldots, T_z)$.

5. Псевдовероятностные поточные шифры. Алгоритмы поточного шифрования представляют интерес для обеспечения защиты информации, передаваемой по открытым каналам связи [26, 27], поэтому значительный интерес представляет рассмотрение подходов к построению поточных ПВ шифров. В разделе 3 представлены алгоритмы блочного ПВ шифрования, в которых выполняется совместное преобразования двух сообщений. Если сообщения разбить на блоки данных малого размера (например, 4, 8 или 16 бит), рассматриваемых как знаки текста, то эти алгоритмы фактически будут задавать процесс поточного ПВ шифрования. Однако для получения высокого уровня

стойкости требуется решить задачу смены ключей шифрования при переходе от одного шифруемого знака к другому.

За счет смены ключей по псевдослучайному закону стойкое шифрование может быть обеспечено использованием достаточно простых операций при формировании знаков промежуточных шифртекстов. Данная идея детерминистического изменения ключей, используемых для шифрования пар знаков шифруемых сообщений, потенциально обеспечивает существенное повышение скорости шифрования по сравнению с алгоритмами поточного ПВ шифрования [15], в которых используется переборный механизм нахождения текущего знака шифртекста. Так же как и в случае поточных алгоритмов [15], для безопасного шифрования многих пар входных сообщений без изменения базовых секретных ключей K и Q требуется задать зависимость значений сменяемых ключей от несекретного вектора инициализации V, который направляется получателю вместе с шифртекстом.

Последовательно сменяемые ключи будем рассматривать как элементы ключевой гаммы. Пусть дана стойкая функция блочного шифрования E и требуется выполнить совместное шифрование сообщений $M = (m_1, m_2, \ldots, m_i, \ldots, m_z)$ и $T = (t_1, t_2, \ldots, t_i, \ldots, t_z)$, имеющих вид последовательности u-битовых знаков t_i и m_i. Шифрование сообщений M и T зададим, соответственно, по фиксированным секретным ключам K и Q, с помощью которых генерируются следующие две ключевые гаммы:

$$\Gamma = \{(\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_i, \beta_i), \ldots, (\alpha_z, \beta_z)\} \text{ и}$$

$$\Gamma' = \{(\alpha'_1, \beta'_1), (\alpha'_2, \beta'_2), \ldots, (\alpha'_i, \beta'_i), \ldots, (\alpha'_z, \beta'_z)\},$$

элементами которых являются пары u-битовых ключевых знаков (α_i, β_i) и (α'_i, β'_i). Ключевые знаки α_i, β_i, α'_i и β'_i используются для совместного преобразования знаков m_i и t_i и вычисляются в зависимости от ключей K и Q, номера i и вектора инициализации V.

Процедуру генерации i-ых пар ключевых элементов (α_i, β_i) и (α'_i, β'_i) зададим следующем виде:

1. Вычислить пару u-битовых ключевых знаков $(\alpha_i, \beta_i) = E_K(V\|i) \bmod 2^{2u}$, где $\|$ — операция конкатенации (присоединения битовых строк); E — блочный шифр с входным блоком данных размером 128 бит и значения V и i задаются в виде 64-битовых строк.

2. Вычислить пару u-битовых ключевых знаков $(\alpha'_i, \beta'_i) = E_Q(V\|i) \bmod 2^{2u}$.

3. Присоединяя слева единичный бит к u-битовому знаку β_i, получить значение $\lambda = (1||\beta_i)$.

4. Добавляя слева единичный бит к u-битовому знаку β'_i, сформировать битовую строку $\eta = (1||\beta'_i)$. Если наибольший общий делитель НОД$(\eta, \lambda) \neq 1$, где битовые строки η и λ интерпретируются как двоичные многочлены, то модифицировать β'_i по формуле $\beta'_i \leftarrow (\beta'_i + 1) \bmod 2^u$, где битовая строка β'_i рассматривается как двоичное число, и вернуться к началу шага 4.

5. Взять пары ключевых знаков (α_i, β_i) и (α'_i, β'_i) в качестве i-ых элементов ключевых гамм Γ и Γ' соответственно.

Поточный алгоритм ПВ шифрования фиктивного сообщения M и секретного сообщения T описывается следующим образом:

1. Каждую i-ю ($i = 1, 2, \ldots, z$) пару знаков m_i и t_i входных сообщений преобразовать в $2u$-битовый знак c_i шифртекста, осуществляя следующие три шага:

1.1. Сгенерировать i-е элементы (α_i, β_i) и (α'_i, β'_i) ключевых гамм Γ и Γ' соответственно.

1.2. Сформировать двоичные многочлены λ и η степени u по формулам $\lambda = (1||\beta_i)$ и $\eta = (1||\beta'_i)$, где ключевые знаки β_i и β'_i трактуются как двоичные многочлены.

1.3. Вычислить $2u$-битовый знак c_i как решение следующей системы линейных сравнений:

$$\begin{cases} c_i \equiv \alpha'_i \oplus t_i \bmod \eta \\ c_i \equiv \alpha_i \oplus m_i \bmod \lambda \end{cases}, \tag{6}$$

где ключевые знаки α_i и α'_i и знаки исходных текстов m_i и t_i трактуются как двоичные многочлены, заданные в виде двоичного вектора. Шаг 4 процедуры генерации элементов ключевых гамм задает выполнимость условия взаимной неприводимости двоичных многочленов η и λ, поэтому система линейных сравнений (6) имеет единственное решение по модулю многочлена, равного произведению $\eta\lambda$, которое представляет собой многочлен, степень которого не превышает значения $2u - 1$, то есть битовую строку длины $2u$. Решение системы (6) находится по следующей формуле:

$$c_i = \left[(\alpha'_i \oplus t_i)\lambda\left(\lambda^{-1} \bmod \eta\right) \oplus (\alpha_i \oplus m_i)\eta\left(\eta^{-1} \bmod \lambda\right)\right] \bmod \eta(x)\lambda(x).$$

2. Объединяя все знаки c_i, сформировать выходной шифртекст $C = (c_1, c_2, \ldots, c_i, \ldots, c_z)$.

Ассоциируемый алгоритм вероятностного шифрования фиктивного сообщения M по фиктивному ключу K и вектору инициализации V выполняется следующим образом:

1. Каждый i-й ($i = 1, 2, \ldots, z$) знак m_i исходного текста M преобразовать в $2u$-битовый знак c_i криптограммы, выполнив следующие три шага:

1.1. Используя 128-битовый блочный шифр E, сгенерировать i-й элемент (α_i, β_i) ключевой гаммы Г по формуле $(\alpha_i, \beta_i) = E_K(V||i) \bmod 2^{2u}$.

1.2. Сгенерировать случайный многочлен ρ, степень которого не превышает значения $(u - 1)$, и случайный многочлен η степени u, такой, что НОД$(\eta, 1||\beta_i) = 1$, где битовая строка $1||\beta_i$ рассматривается как двоичный многочлен.

1.3. Вычислить $2u$-битовый знак c_i как решение следующей системы линейных сравнений:

$$\begin{cases} c_i \equiv \alpha_i \oplus m_i \bmod \lambda \\ c_i \equiv \rho \bmod \eta \end{cases},$$

где $\lambda = 1||\beta_i$.

2. Объединяя все знаки c_i, сформировать выходной шифртекст $C = (c_1, c_2, \ldots, c_i, \ldots, c_z)$.

Алгоритм расшифровывания фиктивного сообщения:

1. Каждый i-й ($i = 1, 2, \ldots, z$) $2u$-битовый знак c_i шифртекста преобразовать в i-й u-битовый знак m_i исходного сообщения M, выполнив следующие два шага:

1.1. Используя 128-битовый блочный шифр E, вычислить i-й элемент ключевой гаммы Г: $(\alpha_i, \beta_i) = E_K(V||i) \bmod 2^{2u}$.

1.2. Вычислить u-битовый знак m_i по формуле $m_i = c_i \oplus \alpha_i \bmod \lambda$, где $\lambda = 1||\beta_i$.

2. Объединяя все знаки m_i, сформировать восстановленное сообщение $M = (m_1, m_2, \ldots, m_i, \ldots, m_z)$.

Расшифровывание секретного сообщения T выполняется по такому же алгоритму с использованием секретного ключа Q:

1. Каждый i-й ($i = 1, 2, \ldots, z$) $2u$-битовый знак c_i шифртекста преобразовать в u-битовый знак t_i исходного сообщения T, выполнив следующие четыре шага:

1.1. Вычислить i-й элемент гаммы Γ': $(\alpha'_i, \beta'_i) = E_Q(V||i) \bmod 2^{2u}$.

1.2. Вычислить i-й элемент гаммы Γ: $(\alpha_i, \beta_i) = E_K(V||i) \bmod 2^{2u}$.

1.3. Если НОД$(1||\beta_i, 1||\beta'_i) \neq 1$, то модифицировать битовую строку β'_i по формуле $\beta'_i \leftarrow (\beta'_i + 1) \bmod 2^u$, где битовая строка β'_i рассматривается как двоичное число, и перейти в начало шага 1.3.

1.4. Вычислить u-битовый знак t_i по формуле $t_i = c_i \oplus \alpha'_i \bmod \eta$, где $\eta = 1||\beta'_i$.

2. Объединяя все знаки t_i, сформировать восстановленное сообщение $T = (t_1, t_2, \ldots, t_i, \ldots, t_z)$.

Сравнение двух последних алгоритмов показывает, что в рассмотренном поточном ПВ шифре не выполняется критерий идентичности алгоритмов расшифровывания криптограммы по фиктивному и секретному ключам, которому удовлетворяют поточные ПВ шифры [15]. Действительно, при восстановлении фиктивного сообщения генерируются только элементы ключевой гаммы Γ, а в случае восстановления секретного сообщения требуется вычислять элементы двух гамм Γ и Γ'. Это связано с тем, что на некоторых шагах процедуры зашифровывания входных знаков t_i и m_i осуществляется модифицирование ключевого знака β'_i (см. шаг 4 процедуры генерации) для реализации условия взаимной неприводимость модулей в системе линейных сравнений (6). При построении скоростных поточных ПВ шифров выполнимость указанного критерия остается открытой задачей.

6. Рандомизация блочных псевдовероятностных шифров. Представленные в разделе 3 блочные алгоритмы ПВ шифрования задают детерминистическое преобразование, поэтому наблюдение со стороны потенциального атакующего повторяющихся шифртекстов (случай выполнения повторного шифрования некоторых входных сообщений) даст ему возможность уличить в обмане отправителя и/или получателя сообщения, утверждающих в момент принуждающей атаки, что ими использовался алгоритм вероятностного шифрования. Для задания изменения шифртекста при повторном шифровании можно использовать рандомизацию процесса ПВ шифрования, то есть задать зависимость шифртекста от случайных значений.

Рандомизация блочных ПВ шифров может быть реализована путем добавления в систему линейных сравнений (линейных уравнений)

дополнительного сравнения (уравнения) со случайными коэффициентами. Это модифицирование схемы совместного шифрования фиктивного и секретного сообщений в целом сохраняет ее исходное построение. Изменения связаны с тем, что в алгоритме зашифровывания добавляется один дополнительный шаг — шаг генерации двух или трех случайных значений, а на шаге вычисления шифртекста решается система из трех сравнений (уравнений) вместо решения систем из двух линейных соотношений в случае детерминистических блочных ПВ шифров.

В случае ПВ шифра с использованием системы сравнений (3) соответствующая ему рандомизированная версия включает формирование блока шифртекста в виде решения следующей системы:

$$\begin{cases} C_i \equiv C_{T_i} \bmod \eta_1 \\ C_i \equiv C_{M_i} \bmod \eta_2 \\ C_i \equiv \lambda \bmod \rho \end{cases},$$

где λ и ρ — случайные двоичные многочлены, такие что ρ является взаимно неприводимым с многочленами η_1 и η_2, а степень λ меньше степени ρ. Размер блока шифртекста увеличивается на число битов, равное степени многочлена ρ.

В случае ПВ шифра с использованием системы уравнений (4) соответствующая ему рандомизированная версия включает формирование блока выходного шифртекста в виде решения следующей системы уравнений:

$$\begin{cases} K_1 C_i' \oplus K_2 C_i'' \oplus K_3 C_i''' \equiv C_{T_i} \bmod \eta \\ Q_1 C_i' \oplus Q_2 C'' \oplus Q_3 C_i''' \equiv C_{M_i} \bmod \eta \\ \lambda_1 C_i' \oplus \lambda_2 C_i'' \oplus C''' \equiv \rho \bmod \eta \end{cases},$$

где λ_1, λ_2 и ρ — случайные двоичные многочлены, степень которых меньше степени η; ключи K и Q имеют вид $K = (K_1, K_2, K_3)$ и $Q = (Q_1, Q_2, Q_3)$. Блок шифртекста имеет вид $C = (C_1, C_2, C_3)$ и его длина увеличивается на число битов, равное степени многочлена η.

7. Псевдовероятностный протокол бесключевого шифрования. Функции коммутативного шифрования (коммутативные шифры) лежат в основе протоколов бесключевого шифрования, которые решают задачу передачи секретного сообщения по открытому каналу связи без выполнения процедуры обмена ключами между получателем и отправителем сообщения. Участники сеанса связи выполняют проце-

дуры коммутативного зашифровывания и расшифровывания по локальным ключам, которые они выбирают произвольным образом без согласования с другой стороной.

При использовании коммутативного шифра, стойкого к атаке на основе известного исходного текста, протоколы бесключевого шифрования при соответствующем выборе параметров алгоритма коммутативного шифрования обеспечивают произвольную наперед заданную стойкость к атакам со стороны пассивного нарушителя. Протоколы данного типа обеспечивают секретность, но не аутентификацию отправителя и получателя сообщения, поэтому они не могут применяться в условиях потенциальной возможности активных атак, когда атакующий может навязать пользователям ложный сеанс секретной связи.

При рассмотрении протоколов ОШ, как правило, оценивается стойкость к принуждающим атакам со стороны пассивного нарушителя. В рамках модели пассивных принуждающих атак, представляет интерес задача построения бесключевого протокола ОШ, то есть протокола безопасной передачи секретного сообщения T по открытому каналу, где не используются предварительно распределенные по защищенным каналам секретные ключи или открытые ключи, подлинность которых подтверждена до осуществления сеанса передачи секретного сообщения.

Построение протокола бесключевого ОШ может быть выполнено в виде схемы ПВ бесключевого шифрования, в которой формируемые шифртексты вычислительно неотличимы от шифртекстов, формируемых в процессе вероятностного бесключевого шифрования. При таком подходе требуется разработать протокол вероятностного шифрования по локальным ключам, который будет ассоциироваться с протоколом ПВ бесключевого шифрования.

Задача построения вероятностного протокола бесключевого шифрования может быть решена путем включения в протокол этапа согласования разового общего секретного ключа, реализуемого в соответствии с широко известной схемой Диффи — Хеллмана. На данном этапе пользователи обмениваются разовыми открытыми ключами, по которым каждый из них может вычислить одно и то же секретное значение Z. Используя данный параметр участники протокола выполняют вероятностное шифрование данных, подлежащих передаче другой стороне на текущем шаге протокола. Поскольку обе стороны знают значение Z, то каждая из них может правильно восстановить направляемые ей сообщения.

Представленная обобщенная схема вероятностного бесключевого шифрования преобразуется в протокол ПВ бесключевого шифрова-

ния путем выполнения дополнительной процедуры зашифровывания и расшифровывания некоторого второго (фиктивного) сообщения M с использованием дополнительных локальных ключей. При этом локальные ключи для шифрования сообщений M и T являются независимыми друг от друга, а шифр тексты, полученные в результате шифрования секретного сообщения T, используются как случайные значения в протоколе вероятностного бесключевого шифрования.

В случае принуждения отправителя и получателя к раскрытию переданных в ходе сеанса связи сообщения и локальных ключей каждый из них раскрывает свой дополнительный локальный ключ и фиктивное сообщение. При этом они заявляют, что в ходе сеанса передачи сообщения M ими использовался протокол вероятностного бесключевого шифрования. Имея в наличии раскрытые параметры, атакующему вычислительно невозможно доказательно опровергнуть последнее утверждение.

В качестве коммутативной функции шифрования целесообразно использовать экспоненциальный шифр, в котором процедуры зашифровывания и расшифровывания представляют собой операцию возведения в степени e и d по простому модулю p достаточно большой разрядности. Зашифровывание сообщения $M < p$ состоит в вычислении шифртекста $C = M^e \bmod p$, и для правильного расшифровывания значения C используется значение ключа расшифровывания d, удовлетворяющее условию $ed = 1 \bmod p - 1$, благодаря чему выполняется равенство $M = C^d \bmod p$, которое справедливо для любого исходного значения $M < p$.

Пусть удаленный пользователь А желает послать секретное сообщение $T < p$ удаленному пользователю В, используя протокол бесключевого шифрования таким образом, что в случае принуждающей атаки, осуществляемой пассивным атакующим после перехвата всех значений, переданных по каналу связи, пользователи могут раскрыть локальные ключи K_A и K_B, сохраняя секретность сообщения T. Для решения этой задачи может быть использован способ шифрования, описываемый в обобщенном виде следующим образом.

1. В соответствии с криптосхемой Диффи — Хеллмана пользователи генерируют сеансовые (разовые) открытые ключи, обмениваются ими и вычисляют разовый (действующий в рамках текущего сеанса связи) общий секретный ключ Z.

2. Пользователь А генерирует фиктивное сообщение $M < p$.

3. Пользователи А и В выполняют процедуру ПВ бесключевого шифрования сообщений T и M одновременно, причем каждый из пользователей использует различные локальные ключи для шифрования сообщений T и M.

При этом формируемые в ходе процедуры бесключевого шифрования шифртексты, которые передаются по открытому каналу, вычислительно неотличимы от шифртекстов, получаемых в ходе вероятностного бесключевого шифрования фиктивного сообщения *M*. Наличие такого вероятностного протокола позволяет пользователям в случае принуждающей атаки раскрыть только локальные ключи, использованные для преобразования фиктивного сообщения. Атакующему вычислительно невозможно уличить пользователей в обмане, поскольку перехваченные им шифртексты могли быть на самом деле получены путем шифрования сообщения *M* по предоставленным ему локальным ключам при определенных значениях случайных параметров вероятностного коммутативного шифрования.

Рассмотрим конкретную реализацию протокола вероятностного бесключевого шифрования сообщения $T < p$:

1. Пользователь A генерирует случайное значение $k_A < p - 1$, играющее роль его разового личного секретного ключа, вычисляет свой разовый открытый ключ $R_A = \alpha^{k_A} \bmod p$ и направляет значение R_A пользователю B.

2. Пользователь B генерирует случайное значение $k_B < p - 1$, играющее роль его разового личного секретного ключа, вычисляет свой разовый открытый ключ $R_B = \alpha^{k_B} \bmod p$ и направляет значение R_B пользователю A.

3. Пользователь A генерирует локальный ключ $K_A = (e_A, d_A)$, где $d_A = e_A^{-1} \bmod p - 1$, вычисляет разовый общий секрет $Z = R_B^{k_A} \bmod p$, генерирует случайное значение ρ_1 и вычисляет шифртекст $C_1 = (C'_1, C''_1)$ как решение следующей системы линейных уравнений относительно неизвестных C'_1 и C''_1:

$$\begin{cases} C'_1 + C''_1 = \rho_1 \bmod p, \\ C'_1 + ZC''_1 = T^{e_A} \bmod p. \end{cases}$$

Затем A направляет шифртекст C_1 пользователю B.

4. Пользователь B генерирует локальный ключ $K_B = (e_B, d_B)$, где $d_B = e_B^{-1} \bmod p - 1$, вычисляет разовый общий секрет $Z = R_A^{k_B} \bmod p$ и значение $S_1 = M^{e_A} \bmod p = \left(C'_1 + ZC''_1\right) \bmod p$, генерирует случайное

значение ρ_2 и вычисляет шифртекст $C_2 = (C'_2, C''_2)$ как решение следующей системы уравнений относительно неизвестных C'_2 и C''_2:

$$\begin{cases} C'_2 + C''_2 = \rho_2 \bmod p, \\ C'_2 + ZC''_2 = S_1^{e_B} \bmod p. \end{cases}$$

Затем B направляет шифртекст C_2 пользователю A.

5. Пользователь A генерирует случайное значение ρ_3, вычисляет значение $S_2 \equiv S_1^{e_B} \equiv \left(C'_2 + ZC''_2\right) \bmod p$ и шифртекст $C_3 = (C'_3, C''_3)$ как решение следующей системы уравнений относительно неизвестных C'_3 и C''_3:

$$\begin{cases} C'_3 + C''_3 = \rho_3 \bmod p \\ C'_3 + ZC''_3 = S_2^{e_A} \bmod p \end{cases}.$$

Затем A направляет шифртекст C_3 пользователю B. Получив значение C_3, пользователь B вычисляет сообщение T:

$$T = \left(C'_3 + ZC''_3\right)^{d_B} \bmod p.$$

С учетом описанного конкретного протокола вероятностного бесключевого шифрования легко составить следующий конкретный протокол ПВ бесключевого шифрования, обеспечивающий секретность сообщения $T < p$ в случае пассивной принуждающей атаки:

1. Отправитель сообщения T генерирует случайный разовый секретный ключ k_A, вычисляет свой разовый открытый ключ $R_A = \alpha^{k_A} \bmod p$ и направляет R_A получателю.

2. Получатель генерирует случайный секретный ключ k_B, вычисляет свой разовый открытый ключ $R_B = \alpha^{k_B} \bmod p$ и направляет значение R_B пользователю A.

3. Отправитель генерирует локальные ключи $K_A = (e_A, d_A)$, где $d_A = e_A^{-1} \bmod p - 1$ и $Q_A = (\varepsilon_A, \delta_A)$, где $\delta_A = \varepsilon_A^{-1} \bmod p - 1$, вычисляет разовый общий секрет $Z = R_B^{k_A} \bmod p$, формирует фиктивное сообщение $M < p$ и вычисляет шифртекст $C_1 = (C'_1, C''_1)$ как решение следующей системы уравнений относительно неизвестных C'_1 и C''_1:

$$\begin{cases} C_1' + Z^2 C_1'' = T^{\varepsilon_A} \bmod p, \\ C_1' + Z C_1'' \quad = M^{e_A} \bmod p. \end{cases}$$

Затем отправитель направляет шифртекст C_1 получателю.

4. Получатель генерирует локальные ключи $K_B = (e_B, d_B)$, где $d_B = e_B^{-1} \bmod p-1$, и $Q_B = (\varepsilon_B, \delta_B)$, где $\delta_B = \varepsilon_B^{-1} \bmod p-1$, вычисляет разовый секрет $Z = R_A^{k_B} \bmod p$, значения $S_1 \equiv M^{e_A} \equiv \left(C_1' + ZC_1''\right) \bmod p$ и $U_1 \equiv T^{\varepsilon_{\mathbf{A}}} \equiv \left(C_1' + Z^2 C_1''\right) \bmod p$. После этого он вычисляет шифртекст $C_2 = (C'_2, C''_2)$ как решение следующей системы уравнений относительно неизвестных C'_2 и C''_2:

$$\begin{cases} C_2' + Z^2 C_2'' = U_1^{\varepsilon_B} \bmod p, \\ C_2' + ZC_2'' \quad = S_1^{e_B} \bmod p. \end{cases}$$

Затем получатель направляет шифртекст C_2 отправителю.

5. По полученному шифртексту C_2 отправитель вычисляет значения $S_2 \equiv S_1^{e_B} \equiv \left(C_2' + ZC_2''\right) \bmod p$ и $U_2 \equiv U_1^{\varepsilon_B} \equiv \left(C_2' + Z^2 C_2''\right) \bmod p$ и шифртекст $C_3 = (C'_3, C''_3)$ как решение следующей системы уравнений относительно неизвестных C'_3 и C''_3:

$$\begin{cases} C_3' + Z^2 C_3'' = U_2^{\delta_A} \bmod p, \\ C_3' + ZC_3'' \quad = S_2^{e_A} \bmod p. \end{cases}$$

Значение C_3 направляется получателю.

По шифртексту C_3 получатель вычисляет сообщения T и M:

$$T = \left(C_3' + Z^2 C_3''\right)^{\delta_B} \bmod p; \quad M = \left(C_3' + ZC_3''\right)^{d_B} \bmod p.$$

Доказательство корректности протокола ПВ бесключевго шифрования состоит в том, что устанавливается справедливость следующих двух соотношений.

1. Восстановление секретного сообщения:

$$\left(C_3' + Z^2 C_3''\right)^{\delta_B} \equiv \left(U_2^{\delta_A}\right)^{\delta_B} \equiv \left(U_1^{\varepsilon_B}\right)^{\delta_A \delta_B} \equiv \left(T^{\varepsilon_A}\right)^{\varepsilon_B \delta_A \delta_B} \equiv T \bmod p.$$

2. Восстановление фиктивного сообщения:

$$\left(C_3' + ZC_3''\right)^{d_B} \equiv \left(S_2^{d_A}\right)^{d_B} \equiv \left(S_1^{e_B}\right)^{d_A d_B} \equiv \left(M^{e_A}\right)^{e_B d_A d_B} \equiv M \bmod p.$$

Подвергаясь принудительной атаке, отправитель и получатель сообщения раскрывают фиктивное сообщение M и ключи k_A, R_A, k_B, R_B, Z, (e_A, d_A) и (e_B, d_B). При этом они заявляют, что для передачи сообщения M они использовали протокол вероятностного бесключевого шифрования. Благодаря наличию такого протокола, ассоциируемого с протоколом бесключевого ОШ, атакующий не имеет практической возможности уличить в обмане хотя бы одну из сторон сеанса защищенной передачи секретного сообщения. Действительно, в рамках ассоциированного протокола раскрытые параметры корректно связаны со всеми значениями, переданными по открытому каналу связи.

Для того чтобы показать отличие значений $\rho_i = \left(C_i' + C_i''\right) \bmod p$ $(i = 1, 2, 3)$ от случайных, требуется вычислить один из локальных ключей Q_A и Q_B, что позволит атакующему восстановить сообщение T. Однако для этого нужно решить задачу дискретного логарифмирования по простому модулю p, который выбирается таким, что решение этой вычислительной задачи является практически неосуществимым.

Таким образом, в описанном протоколе выполнено требование вычислительной неотличимости по шифртексту процедуры ОШ от процедуры вероятностного шифрования, то есть он действительно может быть отнесен к протоколам ПВ шифрования. Построенный протокол вероятностного бесключевого шифрования, ассоциированный с разработанным протоколом, имеет также самостоятельное значение в случаях, когда требуется обеспечить достаточную стойкость к атакам на основе специально подобранных текстов.

Бесключевые протоколы различного типа обеспечивают безопасную передачу сообщений по открытым каналам относительно атак пассивного нарушителя. В случаях, когда требуется обеспечить стойкость к принуждающим атакам со стороны активного нарушителя, выдающего себя за отправителя или получателя секретного сообщения, описанный протокол должен быть дополнен механизмами проверки аутентичности передаваемых в ходе протокола разовых открытых ключей и шифртекстов.

8. Заключение. На основе обобщения результатов в области разработки алгоритмов ПВ шифрования выделены общие приемы их построения и рассмотрены особенности реализации механизмов защи-

ты информации с использованием ПВ шифров. Предложены новые алгоритмы симметричного ПВ шифрования, обладающие существенно более высокой производительностью по сравнению с известными в литературе аналогами, и общий подход к рандомизации ПВ шифров, позволяющей расширить класс ПВ шифров и обеспечивающей повышение стойкости к принуждающим атакам. Показано, что критерий неотличимости по шифртексту от вероятностного шифрования может быть использован также и для построения протоколов бесключевого ПВ шифрования, не требующий наличия у участников протокола заранее согласованных ключей.

Дальнейшее развитие данного направления прикладной криптографии, относящейся к разработке и анализу ПВ шифров, связано с разработкой новых способов алгоритмического задания функций взаимно-однозначного отображения пар блоков промежуточных шифртекстов в блоки выходного шифртекста, включая случай разбиения фиктивного и секретного сообщений на блоки различного размера. Для приложений в области защиты информации также представляет интерес разработка коммутативных ПВ шифров и протоколов бесключевого ПВ шифрования, основанных на вычислительной трудности задачи дискретного логарифмирования на эллиптической кривой.

Литература

1. *Жуков К.Д.* Обзор атак на AES-128: к пятнадцатилетию стандарта AES // Прикладная дискретная математика. 2017. № 35. С. 48–62.
2. *Sirwan A., Majeed N.* New Algorithm for Wireless Network Communication Security // International Journal on Cryptography and Information Security. 2016. vol. 6. no. 3/4. pp. 1–8.
3. *Agievich S.V.* EHE: nonce misuse-resistant message authentication // Прикладная дискретная математика. 2018. № 39. С. 33–41.
4. *Nikolaev M.V.* On the complexity of two-dimensional discrete logarithm problem in a finite cyclic group with efficient automorphism // Математические вопросы криптографии. 2015. Т. 6. № 2. С. 45–57.
5. *Алексеев Е.К., Ошкин И.Б., Попов В.О., Смышляев С.В.* О криптографических свойствах алгоритмов, сопутствующих применению стандартов ГОСТ Р 34.11-2012 и ГОСТ Р 34.10-2012 // Математические вопросы криптографии. 2016. Т. 7. № 1. С. 5–38.
6. *Николаев В.Д.* Атаки на схемы электронной подписи, не учитываемые традиционными определениями стойкости, и меры противодействия им // Математические вопросы криптографии. 2016. Т. 7. № 1. С. 93–118.
7. *Verma G.K.* A Proxy Blind Signature Scheme over Braid Groups // International Journal of Network Security. 2009. vol. 9. no 3. pp. 214–217.
8. *Canetti R.*, *Dwork C.*, *Naor M.*, *Ostrovsky R.* Deniable Encryption // Annual International Cryptology Conference. 1997. vol. 1294. pp. 90–104.
9. *Ibrahim M.H.* A Method for Obtaining Deniable Public-Key Encryption // International J. of Network security. 2009. vol. 8. no 1. pp. 1–9.
10. *Dachman-Soled D.* On minimal assumptions for sender-deniable public key encryption // International Workshop on Public Key Cryptography. 2014. vol. 8383. pp. 574–591.

11. *Asif A.M.A.M., Hannan S.* A Review on Classical and Modern Encryption Techniques // International Journal of Engineering Trends and Technology. 2014. vol. 12. no. 4. pp. 199–203.
12. *Ishai Yu. et al.* Efficient non-interactive secure computation // Annual International Conference on the Theory and Applications of Cryptographic Techniques. 2011. vol. 6632. pp. 406–425.
13. *Meng B.* A secure Internet voting protocol based on non-interactive deniable authentication protocol and proof protocol that two ciphertexts are encryption of the same plaintext // Journal of Networks. 2009. vol. 4. pp. 370–377.
14. *Barakat T.M.* A New Sender-Side Public-Key Deniable Encryption Scheme with Fast Decryption // KSII Transactions on Internet and Information Systems. 2014. vol. 8. no. 9. pp. 3231–3249.
15. *Moldovyan N.A. et al.* Deniability of Symmetric Encryption Based on Computational Indistinguishability from Probabilistic Ciphering // Information Systems Design and Intelligent Applications. 2018. vol. 672. pp. 209–218.
16. *Hong X., Wang B.* A Non-interactive Deniable Authentication Scheme in the Standard Model // Journal of Electrical and Electronic Engineering. 2017. vol. 5. no. 2. pp. 80–85.
17. *Yoon E.J.* Security Analysis of Kar's ID-based Deniable Authentication Protocol // Contemporary Engineering Sciences. 2015. vol. 8. no. 17 pp. 765–771.
18. *Hata M.M., Ali F.H.M., Aljunid S.A.* Secret Sharing Deniable Encryption Technique // International Conference on Information Science and Applications. 2017. vol. 424. pp. 347–357.
19. *Amrutiya V., Baskaran A., Iyengar N.* Deniable Encryption using One Time Pads // Proceedings of the International Conference on Advances in Information Communication Technology & Computing. 2016. 49 p.
20. *Talouki M.A., Dastjerdi A.B.* Anonymous electronic voting protocol with deniable authentication for mobile ad hoc networks // International journal of Multimedia and Unbiquitous Engineering. 2014. vol. 9. no. 1. pp. 361–366.
21. *Moldovyan N.A. et al.* Pseudo-probabilistic block ciphers and their randomization // Journal of Ambient Intelligence and Humanized Computing. 2018. pp. 1–8.
22. *Wang C., Wang J.* A shared-key and receiver-deniable encryption scheme over lattice // Journal of Computational Information Systems. 2012. vol. 8. no. 2. pp. 747–753.
23. *O'Neil A., Peikert C., Waters B.* Bi-deniable public-key encryption // Annual Cryptology Conference. 2011. vol. 6841. pp. 525–542.
24. *Moldovyan N.A., Shcherbacov A.V., Eremeev M.A.* Deniable-encryption protocols based on commutative ciphers // Quasigroups and related systems. 2017. vol. 25. no. 1. pp. 95–108.
25. *Zou M.H. et al.* Scan-based attack on stream ciphers: A case study on eSTREAM finalists // Computer science and technology. 2014. vol. 29. pp. 646–655.
26. *Hwang T., Gope P.* Robust stream-cipher mode of authenticated encryption for secure communication in wireless sensor network // Security and communication networks. 2016. pp. 667-679.

Молдовян Александр Андреевич — д-р техн. наук, профессор, главный научный сотрудник лаборатории безопасности информационных систем, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: компьютерная безопасность, криптография, безопасность компьютерных сетей, управление политиками безопасности, разграничение доступа, аутентификация, анализ защищенности, обнаружение компьютерных атак, межсетевые экраны, защита от вирусов и сетевых червей, анализ и верификация протоколов безопасности и систем защиты информации, защита программного обеспечения от взлома. Число научных публикаций —

200. maa1305@yandex.ru; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7(812)328–5185.

Молдовян Николай Андреевич — д-р техн. наук, профессор, главный научный сотрудник лаборатории безопасности информационных систем, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: компьютерная безопасность, криптография, симметричные и асимметричные криптосистемы, электронная цифровая подпись, аутентификация, блочные шифры, псевдовероятностные шифры. Число научных публикаций — 250. nmold@mail.ru; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7(812)328–5185.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проект № 18-57-54002-Вьет_а).

DOI 10.15622/sp.61.5

A.A. MOLDOVYAN, N.A. MOLDOVYAN

METHODS AND ALGORITHMS FOR PSEUDO-PROBABILISTIC ENCRYPTION WITH SHARED KEY

Moldovyan A.A., Moldovyan N.A. **Methods and Algorithms for Pseudo-Probabilistic Encryption with Shared Key.**

Abstract. As a method for providing security of the messages sent via a public channel in the case of potential coercive attacks there had been proposed algorithms and protocols of deniable encryption. The lasts are divided on the following types: 1) schemes with public key, 2) schemes with shares secret key, and 3) no-key schemes. There are introduced pseudo-probabilistic symmetric ciphers that represent a particular variant of implementing deniable encryption algorithms. It is discussed application of the pseudo-probabilistic encryption for constructing special mechanisms of the information protection including steganographic channels hidden in ciphertexts. There are considered methods for designing stream and block pseudo-probabilistic encryption algorithms that implement simultaneous ciphering fake and secret messages so that the generated ciphertext is computationally indistinguishable from the ciphertext obtained as output of the probabilistic encryption of the fake message. The requirement of the ciphertext indistinguishability from the probabilistic encryption has been used as one of the design criteria. To implement this criterion in the construction scheme of the pseudo-probabilistic ciphers it is included step of bijective mapping pairs of intermediate ciphertext blocks of the fake and secret messages into a single expanded block of the output ciphertext. Implementations of the pseudo-probabilistic block ciphers in which algorithms for recovering the fake and secret messages coincide completely are also considered. There are proposed general approaches to constructing no-key encryption protocols and randomized pseudo-probabilistic block ciphers. Concrete implementations of the cryptoschemes of such types are presented.

Keywords: cryptography, deniable encryption, pseudo-probabilistic encryption, block cipher, stream cipher, fake message, randomization of ciphers, no-key encryption.

Moldovyan Alexandr Andreevich — Ph.D., Dr. Sci., professor, chief researcher of laboratory of information systems security, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: computer security, cryptography, network security, including security policy management, access control, authentication, network security analysis, intrusion detection, firewalls, deception systems, malware protection, verification of security systems. The number of publications — 200. maa1305@yandex.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328–5185.

Moldovyan Nikolay Andreevich — Ph.D., Dr. Sci., professor, chief researcher of laboratory of information systems security, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: computer security, cryptography, symmetric and asymmetric cryptosystems, digital signature, authentication, block ciphers, pseudo-probabilistic ciphers. The number of publications — 250. nmold@mail.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-812-328-5185.

Acknowledgements. This research is supported by the Russian Foundation for Basic Research (project No. 18-57-54002-Viet_a).

References

1. Zhukov K.D. [Review of attacks on AES-128: by the fifteenth anniversary of the AES standard]. *Prikladnaya diskretnaya matematika – Applied discrete mathematics*. 2017. vol. 35. pp. 48–62. (In Russ.).
2. Sirwan A., Majeed N. New Algorithm for Wireless Network Communication Security. *International Journal on Cryptography and Information Security*. 2016. vol. 6. no. 3/4. pp. 1–8.
3. Agievich S.V. [EHE: nonce misuse-resistant message authentication]. *Prikladnaya diskretnaya matematika – Applied discrete mathematics*. 2018. vol. 39. pp. 33–41.
4. Nikolaev M.V. [On the complexity of two-dimensional discrete logarithm problem in a finite cyclic group with efficient automorphism]. *Matematicheskie voprosy kriptografii – Mathematical items of cryptography*. 2015. Issue 6. vol. 2. pp. 45–57.
5. Alexeev E.K., Oshkin I.B., Popov V.O., Smyshlyaev S.V. [On the cryptographic properties of algorithms that accompany the application of standards GOST R 34.11–2012 and GOST R 34.10–2012]. *Matematicheskie voprosy kriptografii – Mathematical items of cryptography*. 2016. Issue 7. vol. 1. pp. 5–38. (In Russ.).
6. Nikolaev V.D. [Attacks on digital signature schemes, which are not taken into account by traditional security definitions, and countermeasures against them]. *Matematicheskie voprosy kriptografii – Mathematical items of cryptography*. 2016. Issue 7. vol. 1. pp. 93–118. (In Russ.).
7. Verma G.K. A Proxy Blind Signature Scheme over Braid Groups. *International Journal of Network Security*. 2009. vol. 9. no. 3. pp. 214−217.
8. Canetti R., Dwork C., Naor M., Ostrovsky R. Deniable Encryption. Annual International Cryptology Conference. 1997. vol. 1294. pp. 90–104.
9. Ibrahim M.H. A Method for Obtaining Deniable Public-Key Encryption. *International J. of Network security*. 2009. vol. 8. no. 1. pp. 1–9.
10. Dachman-Soled D. On minimal assumptions for sender-deniable public key encryption. International Workshop on Public Key Cryptography. 2014. vol. 8383. pp. 574–591.
11. Asif A.M.A.M., Hannan S. A Review on Classical and Modern Encryption Techniques. *International Journal of Engineering Trends and Technology*. 2014. vol. 12. no. 4. pp. 199–203.
12. Ishai Yu. et al. Efficient non-interactive secure computation. Annual International Conference on the Theory and Applications of Cryptographic Techniques. 2011. vol. 6632. pp. 406–425.
13. Meng B. A secure Internet voting protocol based on non-interactive deniable authentication protocol and proof protocol that two ciphertexts are encryption of the same plaintext. *Journal of Networks*. 2009. vol. 4. pp. 370–377.
14. Barakat. T.M. A New Sender-Side Public-Key Deniable Encryption Scheme with Fast Decryption. *KSII Transactions on Internet and Information Systems*. 2014. vol. 8. no. 9. pp. 3231–3249.
15. Moldovyan N.A. et al. Deniability of Symmetric Encryption Based on Computational Indistinguishability from Probabilistic Ciphering. Information Systems Design and Intelligent Applications. 2018. vol. 672. pp. 209–218.
16. Hong X., Wang B. A Non-interactive Deniable Authentication Scheme in the Standard Model. *Journal of Electrical and Electronic Engineering*. 2017. vol. 5. no. 2. pp. 80–85.
17. Yoon E.J. Security Analysis of Kar's ID-based Deniable Authentication Protocol. *Contemporary Engineering Sciences*. 2015. vol. 8. no. 17. pp. 765–771.
18. Hata M.M., Ali F.H.M., Aljunid S.A. Secret Sharing Deniable Encryption Technique. International Conference on Information Science and Applications. Springer. 2017. vol. 424. pp. 347–357.

19. Amrutiya V., Baskaran A., Iyengar N. Deniable Encryption using One Time Pads. Proceedings of the International Conference on Advances in Information Communication Technology & Computing. 2016. 49 p.
20. Talouki M.A., Dastjerdi A.B. Anonymous electronic voting protocol with deniable authentication for mobile ad hoc networks. *International journal of Multimedia and Unbiquitous Engineering*. 2014. vol. 9. no. 1. pp. 361–366.
21. Moldovyan N.A. et al. Pseudo-probabilistic block ciphers and their randomization. *Journal of Ambient Intelligence and Humanized Computing*. 2018. pp. 1–8.
22. Wang C., Wang J. A shared-key and receiver-deniable encryption scheme over lattice. *Journal of Computational Information Systems*. 2012. vol. 8. no. 2. pp. 747–753.
23. O'Neil A., Peikert C., Waters B. Bi-deniable public-key encryption. Annual Cryptology Conference. 2011. vol. 6841. pp. 525–542.
24. Moldovyan N.A., Shcherbacov A.V., Eremeev M.A. Deniable-encryption protocols based on commutative ciphers. *Quasigroups and related systems*. 2017. vol. 25. no. 1. pp. 95–108.
25. Zou M.H. et al. Scan-based attack on stream ciphers: A case study on eSTREAM finalists. *Computer science and technology*. 2014. vol. 29. pp. 646–655.
26. Hwang T., Gope P. Robust stream-cipher mode of authenticated encryption for secure communication in wireless sensor network. *Security and communication networks*. 2016. pp. 667–679.

DOI 10.15622/sp.61.6

A.YU. ISKHAKOV, A.O. ISKHAKOVA, R.V. MESHCHERYAKOV,
R. BENDRAOU, O. MELEKHOVA

APPLICATION OF USER BEHAVIOR THERMAL MAPS FOR IDENTIFICATION OF INFORMATION SECURITY INCIDENT

Iskhakov A.Yu., Iskhakova A.O., Meshcheryakov R.V., Bendraou R., Melekhova O. **Application of User Behavior Thermal Maps for Identification of Information Security Incident.**

Abstract. One of the main functions of an information security system is the identification of any access subject to be able to investigate information security incidents. During executing procedures of scanning and vulnerability exploitation, qualified adversaries regularly change identifying features. Such operations can not only obfuscate logging the data in subsystems, thus, complicating the restoring of events chronology for an information security expert but also call into question the irrefutability of the evidence of participation of particular adversary to particular illegal operations.

In the paper analyses of application of modern approaches of adversary identification in web resources, which does not require authentification of main part of users, is given (fingerprinting, analysis of behavioral features).

Along with widely used in web analytics "thermal maps", user adapted profile and computer model of dynamics of "user-mouse" system, authors offer to identify the subjects of information security incident in readily available informational resources of the Internet. The main idea of the prospective approach consists of the following: when a thermal map is built, not only the density of data layout should be considered but also statistical parameters should be defined by an expert (the distance of intensity gradient, distance overlap, etc.). The authors also offer to consider the dynamics of user operations (e.g. calculation of the average duration of data entry into interactive elements). A description of each step of an appropriate technique and also information on its practical implementation are given. Robustness of the given approach is confirmed by a practical experiment. The offered technique is not a universal instrument of adversary identification. Only manual targeted attacks are considered, the cURL tools etc. used by adversaries are not taken into account. Therefore, it is recommended to use this technique exclusively in addition to working protective systems (WAF, IPS, IDS).

Keywords: identification, thermal maps, fingerprinting, biometrics, anonymizing.

1. Introduction.

Today the global Internet is one of the main tools of mass communication. Therefore, its information resources increasingly often become a target of cyber-attacks of malefactors pursuing various aims. In spite of the fact that untargeted hacker attacks dominate the Internet [1], the most serious threat is represented by target hacks.

The popular Internet resources which do not require extra authentification for the main part of users are in the risk zone. Newsfeeds, entertaining portals, Internet shops etc. can be numbered among such resources. In spite of the fact that today the market of information security software vendors provides a large number of various effective solutions for the detection of incidents and neutralization of malicious inquiries, the problem of the imperfection of user identification technologies still remains [2, 3].

Web application security becomes more difficult because of the growing interactivity, increasingly complicated scenarios and support of new protocols. It is not unknown that security of web servers is not anymore limited to the use of classical package OSI filters or stateful firewalls that trace active TCP-sessions. Today the means of traffic filtration of the application level especially focused on web applications (in particular Web Application Firewall) are considered as a traditional and effective approach to security of web resources [4]. The set of WAF functions usually includes machine learning functions and the following security mechanisms:

– protection against SQL-injections and XSS (including proprietary protection);
– signature analysis;
– protocol validation;
– integration with reputation and fraud services;
– possibility of creation of personal security rules;
– integration with other components of complex information security systems.

However, existence of the only WAF — decision isn't sufficient for complex protection of information systems according to various practical researches [3]. Especially it concerns the web-resources which aren't demanding carrying out authentication for the main user audience. The analysis of application of modern approaches to malefactors identification and also the methodical and algorithmic providing, developed by authors, are given further in the paper.

2. Relevance of the study. Web application security is, first of all, a complex of measures which is a part of a system of providing information security of a company as a whole. One of the basic procedures of a security system is the identification of any subject of access for the purpose of investigation of information security incidents. Manufacturers of modern WAF declare the function of calculation of correlations and chains of the attacks. This function allows to group similar operations and to reveal the chain of attack progress — from surveying before the theft of the important data or deployment of bookmarks. As a result, instead of a list of a thousand suspicious events, information security experts receive some tens of really important messages. However, considering a high level of competence and preparation of malefactors during a targeted attack, it is necessary to note that during scanning and vulnerability exploitation the malefactor regularly changes identifying features in order to make the subsequent investigations more complicating. Similar operations not only obfuscate the logging of the data in subsystems, making the restoring of event chronology more complex for an information security expert but also call into question the irrefutabil-

ity of the evidence base to prove the participation of the certain malefactor in certain illegal operations.

Until now widely applied logging methods of users' IP-addresses and methods of storing housekeeping information on the device of a client (Cookie technology) are not effective. When using Cookies, the subject of access possesses the complete control over contents (including legitimate possibilities of destruction and change of data). The given approaches are not capable to resist the use of basic mechanisms of e-mail address broadcasting, Proxy services, anonymizers and dynamic addressing. Besides, it is necessary to consider that even among legitimate and law-abiding visitors of websites, a considerable number of users prefer basic anonymizing resources.

The given condition not only complicates procedures of investigation of incidents but also promotes discrediting of protective mechanisms which automatically add in lists of locking arrays of IP-addresses of services VPN and Proxy, used not only by malefactors but also by quite legitimate users. Thereof, there is an actual necessity of application of such technologies which would allow on the basis of the meta-data gathering to solve the task of classification of user's sessions on real visitors.

3. The analysis of the state-of-the-art investigations in the given area. Many researchers intend to define rational attribute space which provides reliable identification of users by means of indirect characteristics (work environment parameters) or by means of processing of the statistical data on behavioural parameters (methods of dynamic biometric authentication) [6].

Fingerprinting methods have gained wide popularity. Various algorithms, methods, and techniques are actively being developed in order to obtain new results in the given area. For example in [7-9], the questions of user identification of an Internet resource by the basic set of features of the browser are considered. The paper [10] is devoted to the improvement of the reliability of the subject by means of the analysis of auxiliary meta-information on the property array of the user's software. The parameters comprising the most significant attribute space [11] include:

- a list of the installed fonts;
- a set of plug-ins of the browser provided by means of JavaScript;
- the information on OS localisation;
- SuperCookie;
- Canvas fingerprinting etc.

Some examples of the characteristics used in implementations of similar technologies are shown below.

1) Local Shared Objects (LSO) — the type of metadata that is stored as files on each user's computer; today all versions of Flash Player use LSO.

2) HTML5-repositories (localStorage, File API and IndexedDB) are intended for maintenance of constant storage of the arbitrary portions of the binary data corresponding to a specific resource.

3) Isolated Storage — isolated Silverlight storage; as with LSO, from a technical point of view, there are no barriers to storing the session identifiers.

4) The Last-Modified header (date of the cached document version).

5) Cookies — a small data fragment stored on the user's computer.

6) Browser cache objects. This mechanism was not intended to be used as random access storage. But if the service returns a JavaScript to the user document with a unique identifier inside its body and sets the value of headers "Expires / max-age" as distant future, then the identification script will be stored in the browser's cache. After such a manipulation it is possible to access this script from any page in a network, simply requesting the script download from the known URL.

7) Application cache (HTML5) — a set of functions that provides advanced caching of web application resources.

8) SDHC dictionaries. This method is a compression algorithm developed by Google, which is based on the use of the dictionaries provided by the special server. The client receives a dictionary file containing the lines that may appear in subsequent replies. After that the server can simply refer to these elements inside the dictionary, and the client will independently generate a page on their basis.

9) Abstract identifier ETag (tag of the cached document version);

10) Use of the internal DNS browser cache.

11) Other storage mechanisms (window.name or session.storage) which allow to store and request an unique identifier in such a way that it remains even after deleting all browsing history and site data.

12) Use of the protocol features. Origin Bound Certificates (persistent self-signed certificates that identify the client for an HTTPS server) - as a unique identifier, it’s possible to take a cryptographic certificate hash, provided by the client as part of a legitimate SSL handshake. TLS also has "session identifiers" and "session tickets" mechanisms that allow clients to resume interrupted HTTPS connections without performing a full handshake.

Assimilation of the information set forth above most effectively allows us to generate a unique print of the computer in the identification system database [12]. In 2010 the Electronic Frontier Foundation measured more than 18.1 bits of informational entropy which can be used for fingerprinting. However, this research was before the invention of a digital Canvas fingerprinting which added 5.7 more bits. In 2017 the method of cross-browser fingerprinting [13, 14] was introduced, allowing one to track a user from different browsers on one device.

In spite of the fact that in similar papers it is offered to use a suite of the most significant features for identification according to their authors, there is no unified correlation analysis of all features which would allow to reveal relations between them and to optimize attribute space. Considering that many found informative metrics are defined by the methods which are heavy from the computational point of view, similar approaches do not find a wide circulation in view of the necessity of maintaining a fast response of a web resource.

It is necessary to note that the mentioned techniques are effective only for usual web-browsers which do not declare the providing of user's anonymity. In specialized browsers, such as Tor Browser, the majority of developed estimation methods of the hardware and browser environment features are blocked [15]. And considering that the given research is directed on development of a technique of identification of an exclusively prepared malefactor a priori applying tools for a regular change of browser prints, the fingerprinting technology can be applied only with additional mechanisms of identification.

The identification systems based on the syntactic and morphological analysis of text data indexing messages of users with certain keywords cannot be applied in the investigated objects because of their prominent features (the majority of users do not interact with data transfer forms). The algorithms of subject authentication by keyboard handwriting for the same reason cannot be applied in the selected subject domain.

There are approaches based on the registration of features of the operation with a mouse pointing device for identification of its owner, for example [16-19]. In paper [20], the author identifies the user of a computer game on the basis of a neural network. For data processing, the state machine is used. Both the trajectory and accuracy of clicks are evaluated. The software error is 6-20%. In paper [21] the authors offered to use biometric data obtained from the analysis of the mouse use for constant (periodic) authentication of a user. The biometrics data of mouse movements is represented as a reflexion of psychological and behavioural characteristics of a user. Such data as mood and weariness are selected. In paper [22] the comparative analysis of methods conducted in similar researches (with the accuracy of identification from 84% to 99.7%) is carried out. In 2017 researchers developed a prototype of a system considering the speed of mouse movements and features of scroll wheel movements [23]. Despite a large scientific backlog in the given area, it is necessary to note that all enumerated investigations were carried out on the authentication systems applied to the solution of the "Friend or Foe" problem.

Thus, on the basis of the analysis of state-of-the-art research in the selected area, it is possible to conclude that for today there is no complex solution using modern technologies and means in aggregate, and also satisfying current development of the given problem which lies in the identification of the qualified malefactor of generally available information resource.

4. The use of thermal maps. Thermal maps are often connected to cartograms i.e. a way of the cartographical representation visually showing the intensity of any parameter within the territory on a map [24]. Data can be plotted by hatching of various densities, colouring with a certain degree of saturation, or points. Biological thermal maps are used in molecular biology and medicine for representing of data on expression of a set of genes in the various samples obtained, for example, from different patients or in different conditions from one patient. The main principle which lies in the basis in all spheres of application and construction ways of thermal maps is a representation of various values by means of colour, which provides a high level of visualization and accelerates analysis process. Classical thermal maps were used in those areas of science where input data allowed to define colour easily enough for a particular cell (temperature picture in meteorology, levels of gene expression, exchange indexes).

Today there are many different implementations of analytical maps. Below are some examples:

1) "Map of clicks". It is a tool for measuring and displaying click statistics on your web-site. The map displays clicks on all elements of the page (including those that are not links). In this case, you can see not only the interaction of visitors with one page, but also aggregated statistics on the group of pages of the site. For example, you can get statistics for a particular section (Figure 1).

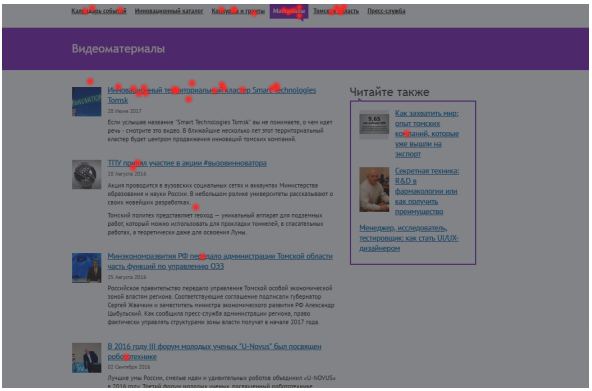


Fig. 1. Map of clicks

Several modes of map display:

– "Heat Map" — warm colours correspond to the frequent clicks, cold colors — rare.

– "Monochrome map" — the density of color corresponds to the frequency of clicks at a given point.

– "Clicks by links and buttons" — the map does not display clicks on items that are not links or buttons.

– "Transparency map" — a click map displays like a "foggy mask": the most clickable elements appear more clearly through the "fog".

– "Map of the elements" — the map displays all the elements of the web-site page.

2) "Map of links" (Figure 2). It is tool for measuring the statistics of the referrals on your site. The links on the map are highlighted in different colours depending on their popularity. When user clicks on a link, the following data is displayed:

– the number of hits by this reference;

– the percentage of following this link on the number of following the other links on web-site.

Fig. 2. Map of links

3) "Scrolling map" (Figure 3) is a tool for analysis how the attention of visitors is distributed to the certain areas of the web-site pages. The map will help to choose the optimal length of the pages and place important information correctly. The map shows the average time and number of views of a specific section of the page if the administrator hover over it. The "scrolling map" can also provide statistics for a group of pages. For example, for a separate directory.

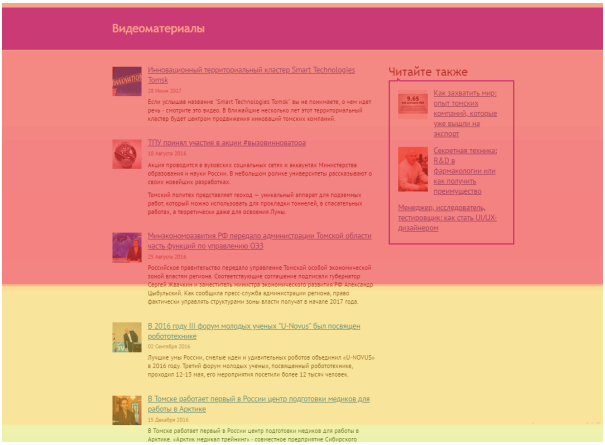
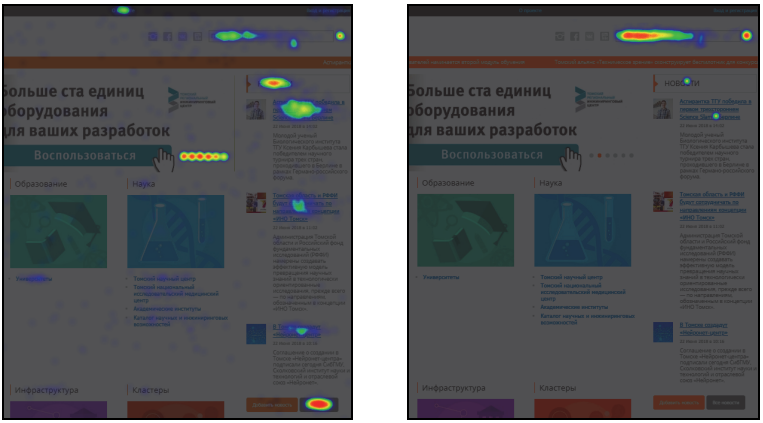


Fig. 3. Scrolling map

In case of application of thermal maps for the task of analytics of user's operations (clicks, motion step of a cursor), it is possible to define the user's activity as a dotted activity. The operations are actually linked to a particular point (pixel) on the screen which seems to be a too small area in comparison with all interface of the system for in-depth study. For the given task, it is necessary to work not with particular points but interactive elements (various data entry forms, which malefactors use for manual exploitation of vulnerabilities). In Figure 4 visualisation of a session of a typical user in comparison with a session of a malefactor is presented.



a) *b)*

Fig. 4. Thermal map *(a)* of a session of a typical user (top).Thermal map *(b)* of a session of a malefactor (bottom)

The number of thermal points and their gradient characterises the purposes and vectors of attacks by the subject of access. The map b on the right shows the interest of the malefactor in interactive elements; this interest stipulated the purpose of vulnerability exploitation in the subsystem of resource security. In this case, it can be clearly seen in the active interaction of the malefactor with the search form on an information portal.

Figures 5-7 show examples of thermal maps of the various user sessions presenting the area of the interactive form allowing users to upload news or to send a press release.

Based on the data about 117 revealed and investigated beforehand incidents of information security (4 large regional news portals) and the subsequent comparison of the facts to visualisation of the events on the data thermal maps, the authors concluded that there are certain prominent features of the use of a mouse pointing device by malefactors during attacks on information resources.

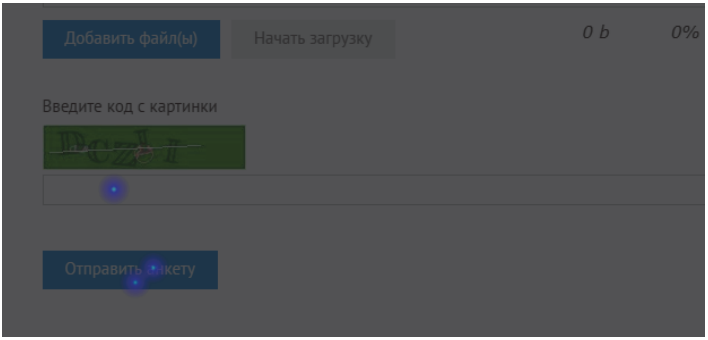


Fig. 5. Thermal map of a session of a legitimate user

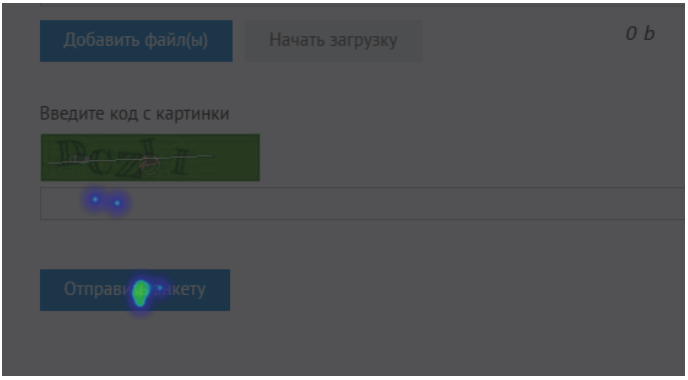


Fig. 6. Thermal map of a session of a legitimate user (some errors are found during the Turing test)

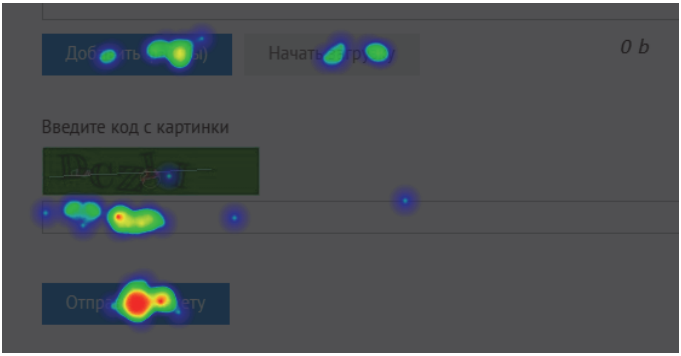


Fig. 7. Thermal map of a session of a user during which the WAF detected attempts of injections and XSS

First, the number of "warm" points in maps of the malefactor is comparable (not more than twice as much) with the number of interactive elements on the page. Under interactive elements, we will understand the data entry forms which are used by the malefactor for sending payload. The maps of sessions of legitimate users are characterised by a wide scatter of clicks on hyperlinks around all perimeter of the screen (tens of times more than the number of interactive elements). Secondly, the data from analytical tool "WebVisor" show that the average time of data entry by the malefactor is several times less than for legitimate filling in the fields with correct data. It is connected to the fact that during the attack malefactors use certain combinations of characters prepared in advance. Besides, one more informative metrics of distinction of the malefactor from the legitimate user can be the number of clicks on interactive elements (which were in advance defined by experts as possible vectors of attacks) in relation to the total number of clicks.

5. Metrics of a user's profile. On the basis of the stated suppositions, the data on functions of "thermal" maps and presumable behaviour of a malefactor, and also with application of the approach and selected in [25] features of user classification according to the pattern "user-mouse", a list of features which will allow to identify the subject of access in the considered task has been adapted.

The formed attribute space for the task of identification of the user according to his or her behavioural activity on a web resource:

1) H — the number of "warm" (activated by the user) points (areas);

2) d — the number of clicks on interactive elements in relation to the total number of clicks:

$$d = \frac{D}{N},$$

where D is the number of clicks on interactive elements, N is the total number of clicks;

3) t_{av} — average time of data entry into the interactive input fields:

$$t_{av} = \frac{1}{k}\sum_{j=1}^{k} t_j,$$

where k is the number of the detected facts of data entry into an interactive element of the page, t_j is time of j^{th} data entry in an interactive element of the page;

4) S — the length of the virtual trajectory, i.e. the distance which user's mouse has passed:

$$S = \sum_{i-1}^{n} s_i,$$

where s_i is the distance which has been passed for provisional time step of measurements according to which the trajectory is measured; the authors considered a minimum value for the personal computer Δt = 15 millisecond as a time step, n is the number of time steps (measurements of sections of a trajectory);

$$s_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2},$$

where x_i, y_i are the manipulator's coordinates on the screen in i^{th} time step, x_0, y_0 are the manipulator coordinates at the initial instant;

5) V_{max} is the maximum speed of passing of a virtual trajectory of the mouse manipulator for all time steps:

$$V_{\max} = \max_i \frac{s_i}{\Delta t},$$

6) t_{stay} is the time from the moment of aiming of the mouse on the point (t_{point}) till the moment of the click on this point (t_{press}):

$$t_{stay} = t_{press} - t_{point},$$

7) t_{hold} is the time of hold of the mouse manipulator during clicking on a point;

8) α — is the angle between the direction of initial motion (from the index point — (x_0, y_0) to the vertex 3 — (x_3, y_3)) and the line connecting the index point — (x_0, y_0) and the finite point (x_n, y_n):

$$\alpha = \operatorname{arctg}\frac{(y_3 - y_0)}{(x_3 - x_0)} - \operatorname{arctg}\frac{(y_n - y_0)}{(x_n - x_0)}.$$

6. Technique of thermal map building. The approach offered by the authors implies visualisation of the calculated characteristics in the form of a thermal map of user's clicks. During the experiment, the authors faced a problem of lack of the literature and engineering specifications on practical implementation of proprietary algorithms for thermal map building. As a result, paper [22] and initial codes of "Heatmap.js" library from the company Yandex was chosen as a basis for the experiment.

The basic idea of the prospective approach consists in the following: when building a thermal map not only density of layout of data but also static parameters defined by the expert (a distance of a gradient of intensity, an overlap distance, etc.) should be considered. Authors offer to consider the dynamics of user's actions (e.g. calculation of average duration of data entry into interactive elements).

According to [22], every element of dotted activity is represented in the form of a circle with a linear-decreasing gradient of colour intensity from centre to edges. The circle radius (intensity area) is defined by the value of the gradient intensity distance and is specified by the expert. Particular colour of each point on a thermal map is defined according to the value of its cumulative intensity, the sum of intensity values of all areas covering this point, and the selected colour scheme (graphic palette). Values of cumulative intensity are normalised within the limits from 0 to 1. The initial value of intensity gradient:

$$I_s = 1 / MQO ,$$

where MQO is the maximum quantity of circles overlapping each other around whole data area. Two or more circles are considered overlapping each other if the distance between them pairwise is less than the distance value of the overlap specified by the expert.

Finite value of the intensity gradient I_k is always equal to 0. Thanks to calculation of I_s on the basis of MQO, the reliability of lost-free visual data is ensured. At high density of dotted data, the greatest intensity is present in the separate areas actually outlined by these points, instead of where maximum cumulative intensity would be generated. For example, in Figure 8 there are shown 2 areas of intensity. The circles on the co-ordinate plane are presented, where the horizontal axis characterises the co-ordinate (position) of a point, and vertical represents its intensity. While the distance between two points is more than a half of the intensity gradient distance, cumulative intensity is less than 1 (one) as the intensity is calculated according to the linear gradient. When the distance becomes less or equal to a half of value of intensity gradient distance, cumulative intensity becomes more or equal to 1 (one) accordingly.

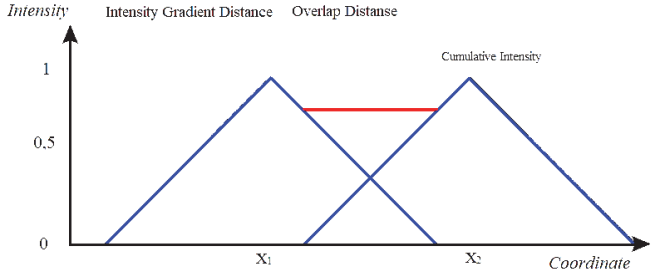


Fig. 8. Intersection of point's intensity regions

Earlier it was noted that in the given task the most interesting is the cumulative intensity of the particular active area of the interface (which, as a rule, interests the malefactor). Therefore, depending on the value of the overlap distance specified by the expert, simplification of visual analytics is up to the point admitted. For example, the clicks on an active element but not on a particular point of this element are important to the expert. If points are on the distance smaller or equal to the value of the overlap distance, there is a recalculation of initial intensity for all points.

Also at the thermal map building, the metrics t_{av} is considered, allowing one to distinguish the border of each interactive element with a characteristic colour from the selected colour graphic palette. It gives the expert an evident picture of objects into which the malefactor inserts various injections.

The method of building the thermal map of user's activity with the above-stated parameters is shown below.

Input data:

– array of dotted clicks of the user's activity (a set of points with the indication of their co-ordinates on a plane);
– distance of intensity gradient;
– overlap distance;
– array of parameters of the colour graphic palette.

Output data:

– The visualised thermal map of user's dotted activity.

Step-by-step description of the method:

Step 1. To receive input data.

The array of dotted clicks of user's activity is generated by means of a specialised script which traces clicks of the mouse on the interface and records the information in a database. The value of the gradient intensity distance, the value of the overlap distance and colour graphic palette are specified by the expert and depend on the nature of the object being analysed.

Step 2. To calculate the *MQO* value.

Step 3. To calculate I_s value according to the formula $I_s = 1 / MQO$.

Step 4. To construct on the map an area of intensity in the form of a circle with centre in the indicated point for each data item of user's activity. In so doing, to define the radius as equal to the value of gradient intensity distance and to construct a linear gradient of intensity value from circle centre to edges. To define the initial value of the gradient as equal to I_s, the finite value as equal to 0 (zero).

Step 4.1. If areas of the intensity of two or more circles are intersected, to calculate cumulative intensity as the sum of values of intensity of all areas covering this point for each point in the field of interception.

Step 4.2. If the value of cumulative intensity is more than 1, to set the value of cumulative intensity as equal to 1 (one).

Step 5. To visualise the thermal map on the basis of values of intensity for each point and the indicated graphic palette.

Step 6. To visualise the thermal map.

7. Experiment description. Yandex and Google web analytics services and the program implementation of the model [25] adapted by authors were installed on one of the popular regional news portals (average attendance more than 50,000 unique users a year). The resource functions on the basis of the "1C-Bitrix: Website management platform". Standard "Proactive Filter" with the configuration of automatic blocking of the source of attack for 30 minutes in the case of detection of suspicious inquiries was used. The imitation of the qualified malefactors' activity was conducted by the representatives of three organizations providing external penetration test of Internet resources. The procedure of testing for penetration was carried out in a stringently certain time period in the mode of greatest possible anonymization.

1) Sessions of bots and vulnerability scanners were eliminated from the access logs (if there are no clicks traced by the specialized script or User Agent null values). After the given operation, a sampling of unique IP-addresses for the reporting period was 79 records.

2) Sessions of conducted attacks (about 70 events of "Security alert" class were recorded) were selected from the WAF register of intrusions. As an automatic blocking was set up in WAF; if the identification of signatures of attacks was regular, the testers' address was added in "Stop list". Penetration testers used a Tor browser and each time when the access was denied, they activated the function of IP-address change.

3) The following task consisted in carrying out classification of the sessions of attacks on real users according to paper [26] with adapted features of users' classification. The measurement of numerical values of the selected information features was carried out. 53 sessions of legitimate users and 10 sessions of penetration testers were conducted for this purpose. The results of numerical values of the defined metrics for profiles of 10 users are brought in Tables 1-2.

Table 1. Numerical values of profiles of 10 users (Heat Map)

No	H	d	t_{av}, sec
1	21.035	0.150	4.125
2	30.154	0.161	10.195
3	25.095	0.017	0
4	16.150	0.092	10.500
5	15.986	0.014	5.116
6	19.845	0.061	6.857
7	20.213	0.181	7.500
8	20.836	0.045	0
9	17.741	0.064	0
10	19.098	0.071	5.150

Table 2. Numerical values of profiles of 10 users (Mouse manipulator)

No	S, m (for 600 sec.)	V_{max}, m/s	t_{stay}, sec	t_{hold}, sec	α
1	422.095	80.268	0.410	0.212	180.550
2	736.950	120.111	0.509	0.262	169.153
3	1022.812	196.504	0.211	0.197	144.122
4	636.568	102.870	0.360	0.290	253.689
5	698.911	60.127	0.390	0.110	190.884
6	902.100	30.006	0.543	0.411	99.117
7	1192.325	105.985	0.480	0.274	123.880
8	1011.750	41.560	0.327	0.346	140.890
9	732.890	59.998	0.399	0.281	111.400
10	442.089	56.012	0.370	0.201	150.688

The results of numerical values of the defined metrics for profiles of 10 penetration testers are shown in Tables 3-4.

Table 3. Numerical values of profiles of 10 penetration testers (Heat Map)

No	H	d	t_{av}, sec
1	4.565	0.468	1.060
2	5.965	0.655	3.458
3	4.198	0.653	3.201
4	6.065	0.719	3.787
5	6.854	0.398	2.008
6	5.782	0.687	1.989
7	4.786	0.586	2.168
8	5.865	0.350	1.469
9	5.569	0.366	1.667
10	5.068	0.387	1.318

Table 4. Numerical values of profiles of 10 penetration testers (Mouse manipulator)

No	S, m (for 600 sec.)	V_{max}, m/s	t_{stay}, sec	t_{hold}, sec	α
1	206.881	39.126	0.226	0.111	201.110
2	301.430	31.075	0.310	0.397	99.335
3	336.865	53.500	0.369	0.186	98.152
4	411.068	40.113	0.407	0.195	96.482
5	350.561	59.873	0.365	0.213	198.548
6	346.012	59.451	0.311	0.200	171.009
7	329.890	60.890	0.225	0.240	209.501
8	298.111	41.669	0.302	0.199	174.694
9	306.623	53.548	0.278	0.238	199.697
10	348.778	51.060	0.200	0.297	98.001

Figure 9 presents the numerical results of comparing the listed characteristics on the "box plot" histograms.

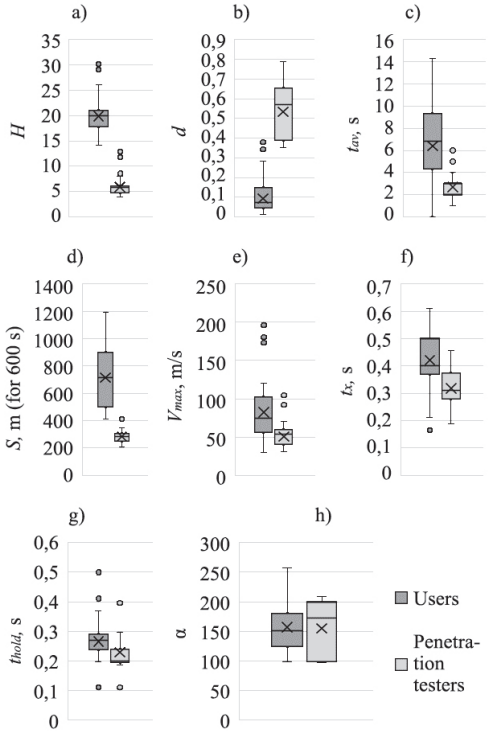


Fig. 9. Histograms (a-h) showing the characteristics numerical differences between the classes of users and penetration testers

The numerical results shown in Tables 1-4 and histograms a-h on Figure 9 allow us to notice that the average quantity of "warm" (made active by the user) areas for the legitimate user fluctuates in the range from 15 to 26, which says that to work with a resource the user uses not less than 15-17 objects in the working area. At the same time, the number of clicks on interactive elements in relation to the total number of clicks is no more than 19%; that is, various links are used, informational tools of the website are researched, and only if necessary the user transmits various interactive inquiries to the resource. Also, the average duration of data entry into interactive input fields considerably differs among certain some users. While analysing numerical parameters in Tables 1-2 as a whole, it is possible to see that deviation of the results from average values is great enough and is up to 50%. The average quantity of the distance passed in 600 seconds (the length of the mouse trajectory) depending on the user's activity fluctuates within 422 to 1192 metres; maximum speed the mouse also considerably varies: from 30 to 197 m/s. The time from the moment of pointing of the mouse on a certain point of a working area till the moment of the mouse click onto this point (object) on the average is equal to 0.399 seconds; time of the mouse button hold during a click is 0.258 seconds.

At the same time, the results of the feature values for malefactor's behaviour (simulated by a penetration tester conducting the test for penetration) have other specific mature. The average quantity of "warm" areas (made active by the malefactor) does not exceed 7; that is, during a session on the average only from 4 to 7 objects of the resource are used and not less than 35% of them are interactive. The given fact says that when searching of weak spots interactive forms for data exchange with the re-source and the analysis of the received results are used more often. The average time of data entry into interactive input fields is on average from 1 to 4 seconds, which speaks about fast data entry owing to available skills of such operations and, probably, applied automation tools for the given process. Deviation of numerical values in Tables 3, 4 from the corresponding average (on columns) is much lower than for values in Tables 1, 2 and is about 30-40%. The distance passed by penetration testers during 600 seconds vary from 206 to 351 metres that is much less than the length of the legitimate users' trajectory. The maximum speed of the mouse was from 31 to 61 m\s, which means that when simulating the actions of the malefactor — purposeful search for vulnerabilities, analysis of objects, and work with interactive fields — there are not any sharp mouse motions, a search of elements in the working area and similar actions. The period from the moment of pointing of the mouse on the object of a resource till the moment of the mouse click by penetration testers is on the average 0.299 seconds, which is lower than the corresponding parameter of legitimate users by 25%; whereas the time of the mouse button hold during a click is comparable with the results in Table 2 and is 0.228 seconds.

The obtained numerical values allow us to preliminary conclude without a special mathematical apparatus the following:

1) average values of feature H naturally differ for different user groups, for legitimate ones the value several times exceeds the value of the malefactor profile;

2) the number of click on interactive elements (input fields, fields of data sending) is significantly higher for malefactors;

3) average time of data entry also differs between user groups;

4) the vector of features for identification of a unique user according to mouse motion (S, V_{max}, t_{stay}, $t_{y\partial}$, α) shows comprehensible results according to research [26]; efficiency of the received values for identification of users should be defined additionally on the basis of the selected mathematical apparatus.

The task of identification of resource users can be presented in the form of mapping $X \rightarrow Y$ where X is a certain image of the user according to models offered in [26], i.e. a set of values of the selected features, and Y is the solution which identifies and-or characterises the user.

To test the efficiency of the offered approach, i.e. implementation of the function displaying the vector of features on an element of a set of known users and characteristics of its behaviour, an artificial neural network was applied. The application of neural networks is for today a widespread tool to make decisions of various types in the automated mode, including for problem-solving, adjacent with the solved one. Neural network learning was carried out on the basis of 75% received designed data (144 000 elementary vector sets).

Parameters of errors of the first and second type, taking into account the recognition of the identity of a user and also his or her behavioural features (detection of ill-intentioned activity) are brought in Table 5.

Table 5. Parameters of errors of the 1st and 2nd type

Task	Errors of the 1st type	Errors of the 2nd type
Recognition of identity of a user	0.079	0.023
Detection of ill-intentioned activities	0.143	0.019

Errors of the 1st type in experiments on recognition of the identity of a user are understood as an amount of cases when the neural network made the decision that the user is not detected, though he or she was in the learning array, errors of the 2nd type are the cases when the user was falsely identified. When detecting a malefactor, the errors of the 1st type are understood as cases of mistaking of the malefactor for the legitimate user; the errors of the 2nd type are understood as mistaking of a legitimate user for a malefactor. Parameters of errors of the first and second type, taking into account the recognition of

the identity of a user and also his or her behavioural features (detection of ill-intentioned activity) are brought in Table 5.

Experimental calculations showed that the user's profile adapted with the application of thermal maps applied in the model “computer-mouse” within the limits of determination of ill-intentioned operations and identification of the subject in the web medium is an effective solution of the relevant scientific task formed above. Undoubtedly, the given method is not a universal tool of malefactor identification — only targeted manual attacks were considered, without consideration of the use cURL tools etc. by malefactors. Therefore, is recommended to use it exclusively in addition to functioning protective systems (Web Application Firewall, Intrusion Prevention System, Intrusion Detection System).

Adaptation of attribute space for the purpose of the solution of adjacent tasks can expand the sphere of "computer-mouse" application. Computing load allowed us to integrate the program implementation directly into the web application in the form of an additional unit-script. Further research has some vectors of development, including the increase of efficiency of determination of required characteristics of a web resource user, simplification of calculations, the increase the number of analysed phenomena, the extension of the spectrum of tasks being solved.

8. Conclusion. The objective opinion on the modern automated systems allows us to talk about the imperfection of existing approaches and technologies of identification of users of the freely available resources selected by malefactors as a target.

The presented method does not allow one to identify absolutely precisely and authentically a particular user but gives a chance to increase the probability of his or her detection in the combination with other indirect methods of identification which allow us with a certain probability to compare the user applying various anonymization means. In the conditions of proceeding development of the information field, the development of similar technologies can become one of priority research directions in questions of counteraction against illegal activity on the Internet.

References

1. Gerasyukova M. [The goal is captured: the most dangerous hacker attacks. Why is target hacker attacks dangerous and how to protect them]. Available at: https://www.gazeta.ru/tech/2018/02/25/11659579/targeted_attack.shtml (accessed: 20.06.2018). (In Russ.).
2. Iskhakov S.Yu., Shelupanov A.A., Mescheryakov R.V. Assessment of security systems complex networks security. Dynamics of Systems, Mechanisms and Machines (Dynamics). 2014. pp. 1–4.
3. Yankovskaya A.E., Shelupanov A.A., Hodashinsky I.A., Gorbunov I.V. Development of hybrid intelligent system of express-diagnostics for detection potential attacker. 9th International Conference on Application of Information and Communication Technologies (AICT). 2015. pp. 183–187.

4. Iskhakov A., Meshcheryakov R., Ekhlakov Yu. The Internet of Things in the security industry. Interactive Systems: Problems of Human-Computer Interaction (collection of scientific papers). 2017. pp. 161–168.
5. Romashev A. [Web Application Firewall Effectiveness Testing and Comparing]. Available at: https://www.anti-malware.ru/compare/web-application-firewall#part4 (accessed: 20.06.2018). (In Russ.).
6. Iskhakov A., Meshcheryakov R., Iskhakov S., Krainov A. Increase in security of authentication services through additional identification using optimal feature space. Proceedings of the IV International research conference “Information technologies in Science, Management, Social sphere and Medicine” (ITSMSSM). 2017. pp. 443–446.
7. Eckersley P. How Unique Is Your Web Browser? Proceedings of the 10th Privacy Enhancing Technologies Symposium (PETS). 2010. pp. 1–18.
8. Alnaami K. et al. Thuraisingham B. P2V: Effective Website Fingerprinting Using Vector Space Representations. IEEE Symposium Series on Computational Intelligence. 2015. pp. 59–66.
9. Iskhakova A., Meshcheryakov R. Automatic search of the malicious messages in the internet of things systems on the example of an intelligent detection of the unnatural agents requests. International Conference on Information Science and Communications Technologies (ICISCT). 2017. pp. 1–4.
10. Bessonova E.E., Zikratov I.A., Kolesnikov Yu.L., Roskov V.Yu. [Method of user identification in the Internet]. *Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki – Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2012. vol. 3. pp. 133–137. (In Russ.).
11. Usmonov B. et al. The cybersecurity in development of IoT embedded technologies. International Conference on Information Science and Communications Technologies (ICISCT). 2017. pp. 1–5.
12. Iskhakov A.Y., Iskhakov S.Y., Meshcheryakov R.V [Increase the security of authentication services by performing additional authentication using the optimal feature space]. *Informacionnye tehnologii v nauke, upravlenii, social'noj sfere i medicine. Sbornik nauchnyh trudov – Information technologies in science, management, social sphere and medicine.* 2017. pp. 117–122. (In Russ.).
13. Abouollo A., Almuhammadi S. Detecting malicious user accounts using Canvas Fingerprint. The 8th International Conference on Information and Communication Systems (ICICS). 2017. pp. 358–361.
14. Daud N.I., Haron G.R., Othman S.S.S. Adaptive authentication: Implementing random canvas fingerprinting as user attributes factor. IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). 2017. pp. 152–156.
15. Sistema upravlenija proektami Trac. [Project Management System Trac]. Available at: https://trac.torproject.org/projects/tor/query?status=accepted&status=assigned&status=needs_review&status=needs_revision&status=new&status=reopened&order=priority&col=id&col=summary&col=keywords&col=status&col=owner&col=type&col=priority&keywords=tbb-fingerprinting (accessed: 20.06.2018). (In Russ.).
16. Didenko S.M. [Investigation of the dynamics of the user's information handwriting model parameters]. *Vestnik Tjumenskogo gosudarstvennogo universiteta – Bulletin of the Tyumen State University*. 2006. vol. 5. pp. 170–174. (In Russ.).
17. Pilankar P.S., Padiya P. Multi-phase mouse dynamics authentication system using behavioural biometrics. International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES). 2016. pp. 1947–1950.
18. Hu S. et al. Deceive Mouse-Dynamics-Based Authentication Model via Movement Simulation. The 10th International Symposium on Computational Intelligence and Design (ISCID). 2017. vol. 1. pp. 482–485.
19. Chen X. et al. A practical real-time authentication system with Identity Tracking based on mouse dynamics. IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). 2014. pp. 121–122.

20. Kaminsky R., Enev M., Andersen E. Identifying Game Players with Mouse Biometrics. University of Washington. Technical Report. 2008. 12 p.
21. Feher C. et al. User Identity Verification via Mouse Dynamic**s**. *Information Sciences*. 2012. vol. 201. pp. 19–36.
22. Stanic M. Continuous user verification based on behavioral biometrics using mouse dynamics. The 35th International Conference on Information Technology Interfaces. 2013. pp. 251–256.
23. Identifikacija pol'zovatelej Tor Browser cherez analiz osobennostej raboty s mysh'ju. [The Tor Browser users identification through the analysis of the mouse features]. Available at: https://www.opennet.ru/opennews/art.shtml?num=44027 (accessed: 20.06.2018). (In Russ.).
24. Danilov N., Shulga T. [Constructing a heat map based on the point of the application user's activity]. *Prikladnaja informatika – Applied informatics*. 2015. vol. 2(56). pp. 49–58. (In Russ.).
25. Didenko S.M. [Development of the mathematical model of the user's information handwriting]. *Matematicheskoe i informacionnoe modelirovanie: sbornik nauchnyh trudov – Mathematical and information modelling: collection of scientific threads*. 2006. pp. 68–73. (In Russ.).
26. Shaptsev V.A., Didenko S.M. *Razrabotka i issledovanie komp'yuternoj modeli dinamiki sistemy «pol'zovatel'-mysh'»* [Development and research of the «user-mouse» system model]. Ph.D. thesis. 2007. 95 p. (In Russ.).

Iskhakov Andrey Yunusovich — Ph.D., senior researcher of cyber-physical, systems laboratory, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences. Research interests: methods of information protection, theoretical foundations of computer security, identification and authentication systems development. The number of publications — 25. iskhakovandrey@gmail.com; 65, Profsoyuznaya str., Moscow, 117997, Russia; office phone: +7 495 336-71-05.

Iskhakova Anastasia Olegovna — Ph.D., senior researcher of cyber-physical, systems laboratory, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, junior researcher of Internet of Things Security laboratory of information systems security department, Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: methods of information protection, big data processing, artificial intelligence. The number of publications — 15. shumskaya.ao@gmail.com; 65, Profsoyuznaya str., Moscow, 117997, Russia; office phone: +7 495 336-71-05.

Meshcheryakov Roman Valerievich — Ph.D., Dr. Sci., professor, head of cyber-physical, systems laboratory, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences. Research interests: system analysis, information processing, cyber-physical systems, information security. The number of publications — 240. mrv@ieee.org; 65, Profsoyuznaya str., Moscow, 117997, Russia; office phone: 7 495 336-71-05, Fax: +7 495 334-93-40.

Bendraou Reda — Ph.D., Dr. Sci., professor, Paris Nanterre University. Research interests: model driven engineering, meta-modeling, model transformations, model execution, DSL specification and code generation. The number of publications — 51. bendraou@mail.ru; 4, Place Jussieu, Paris, 75252, France; office phone: +33 1 44 27 88 6.

Melekhova Olga — Ph.D., associate professor, Paris Nanterre University. Research interests: autonomic systems. The number of publications — 33. melekhova.o@list.ru; 4, Place Jussieu, Paris, 75252, France; office phone: +33 1 44 27 88 6.

Acknowledgements. The work was partially supported by the Russian Federation President Grant for the Lead-ing Scientific Schools (grant NSh. 3070.2018.8) and the Russian Federation President Grant for the Young Russian Scientists – PhD (grant MK-6802.2018.8).

УДК 004.9 **DOI** 10.15622/sp.61.6

А.Ю. Исхаков, А.О. Исхакова, Р.В. Мещеряков, Р. Бендрау, О. Мелехова

ИСПОЛЬЗОВАНИЕ ТЕПЛОВОЙ КАРТЫ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЯ В ЗАДАЧЕ ИДЕНТИФИКАЦИИ СУБЪЕКТА ИНЦИДЕНТА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Исхаков А.Ю., Исхакова А.О., Мещеряков Р.В., Бендрау Р., Мелехова О. **Использование тепловой карты поведения пользователя в задаче идентификации субъекта инцидента информационной безопасности.**

Аннотация. Одной из основных функций системы защиты информации является идентификация любого субъекта доступа с целью возможности расследования инцидентов информационной безопасности (ИБ). В ходе выполнения процедур сканирования и эксплуатации уязвимостей квалифицированные злоумышленники регулярно производят смену идентифицирующих признаков. Подобные действия не только обфусцируют данные в подсистемах аудита, затрудняя возможность восстановления хронологии событий эксперту ИБ, но и ставят под сомнение неопровержимость доказательной базы причастности конкретного злоумышленника к конкретным противоправным действиям. В статье приводится анализ применения современных подходов идентификации злоумышленников в веб-ресурсах, не требующих проведения аутентификации для основной пользовательской аудитории (методы fingerprinting, анализ поведенческих признаков). Рассмотрены признаки пользователя, которые могут быть использованы для решения задачи его последующей идентификации.

С использованием широко применяемых в задачах веб-аналитики «тепловых карт», адаптированного профиля пользователя и компьютерной модели динамики системы «пользователь-мышь» предлагается проводить идентификацию субъектов инцидента ИБ в общедоступных информационных ресурсах сети Интернет. Основная идея предполагаемого подхода заключается в том, что при построении тепловой карты должны учитываться не только плотность расположения данных, а также определяемые экспертом статистические параметры (дистанция градиента интенсивности, дистанция перекрытия и т.д.). Предлагается учитывать и динамику действий пользователя (например, вычисление среднего времени ввода данных в интерактивные элементы). Представлено подробное описание каждого шага соответствующей методики, а также информация по ее практической реализации. Робастность данного подхода подтверждается практическим экспериментом. Предложенная методика не является универсальным средством идентификации злоумышленника — во внимание принимаются только ручные таргетированные атаки, не учитывается использование злоумышленниками cURL инструментов и так далее. Поэтому рекомендуется использовать его исключительно в дополнение к действующим системам защиты (WAF, IPS, IDS).

Ключевые слова: идентификация, тепловые карты, fingerprinting, биометрия, анонимизация.

Исхаков Андрей Юнусович — к-т техн. наук, старший научный сотрудник лаборатории киберфизических систем, Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В. А. Трапезникова Российской академии наук. Область научных интересов: методы защиты информации, теоретические основы компьютерной безопасности, развитие систем идентификации и аутентификации. Число научных публикаций — 25. iskhakovandrey@gmail.com; ул. Профсоюзная, 65, Москва, 117997; р.т.: +7 495 336-71-05.

Исхакова Анастасия Олеговна — к-т техн. наук, старший научный сотрудник лаборатории киберфизических систем, Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В. А. Трапезникова Российской академии наук, младший научный сотрудник лаборатории безопасности интернета вещей кафедры безопасности информационных систем (БИС), Томский государственный университет систем управления и радиоэлектроники (ТУСУР). Область научных интересов: методы защиты информации, обработка больших данных, искусственный интеллект. Число научных публикаций — 15. shumskaya.ao@gmail.com; ул. Профсоюзная, 65, Москва, 117997; р.т.: +7 495 336-71-05.

Мещеряков Роман Валерьевич — д-р техн. наук, профессор, профессор РАН, заведующий лабораторией киберфизических систем, Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В. А. Трапезникова Российской академии наук. Область научных интересов: системный анализ, анализ и синтез речи, информационная безопасность, обработка информации в интеллектуальных системах. Число научных публикаций — 240. mrv@ieee.org; Профсоюзная, 65, Москва, 117997; р.т.: 7 495 336-71-05, Факс: +7 495 334-93-40.

Бендрау Реда — д-р техн. наук, профессор, Университет Париж X – Нантер. Область научных интересов: моделирование, метамоделирование, преобразование моделей, исполнение моделей, спецификация DSL, генерация кода. Число научных публикаций — 51. bendraou@mail.ru; Площадь Юссие, 4, Париж, 75252; р.т.: +33 1 44 27 88 6.

Мелехова Ольга — доцент, Университет Париж X – Нантер. Область научных интересов: автономные системы. Число научных публикаций — 33. melekhova.o@list.ru; Площадь Юссие, 4, Париж, 75252; р.т.: +33 1 44 27 88 6.

Поддержка исследований. Работа выполнена при поддержке Министерства образования и науки Российской Федерации в рамках проектной части государственного задания на 2017–2019 гг. (проект № 2.3583.2017/4.6).

Литература

1. *Герасюкова М.* Цель захвачена: самые опасные хакерские атаки. Чем опасны целевые хакерские атаки и как от них защититься. URL: https://www.gazeta.ru/tech/2018/02/25/11659579/targeted_attack.shtml (дата обращения: 20.06.2018).
2. *Iskhakov S.Yu., Shelupanov A.A., Mescheryakov R.V.* Assessment of security systems complex networks security // Dynamics of Systems, Mechanisms and Machines (Dynamics). 2014. pp. 1–4.
3. *Yankovskaya A.E., Shelupanov A.A., Hodashinsky I.A., Gorbunov I.V.* Development of hybrid intelligent system of express-diagnostics for detection potential attacker // The 9th International Conference on Application of Information and Communication Technologies (AICT). 2015. pp. 183–187.
4. *Iskhakov A., Meshcheryakov R., Ekhlakov Yu.* The Internet of Things in the security industry // Interactive Systems: Problems of Human-Computer Interaction (collection of scientific papers). 2017. pp. 161–168.
5. *Ромашев А.* Тест и сравнение эффективности WAF (Web Application Firewall). URL: https://www.anti-malware.ru/compare/web-application-firewall#part4 (дата обращения: 20.06.2018).
6. *Iskhakov A., Meshcheryakov R., Iskhakov S., Krainov A.* Increase in security of authentication services through additional identification using optimal feature space

// Proceedings of the IV International research conference “Information technologies in Science, Management, Social sphere and Medicine” (ITSMSSM). 2017. pp. 443–446.

7. *Eckersley P.* How Unique Is Your Web Browser? // Proceedings of the 10th Privacy Enhancing Technologies Symposium (PETS). 2010. pp. 1–18.
8. *Alnaami K. et al.* P2V: Effective Website Fingerprinting Using Vector Space Representations // IEEE Symposium Series on Computational Intelligence. 2015. pp. 59–66.
9. *Iskhakova A., Meshcheryakov R.* Automatic search of the malicious messages in the internet of things systems on the example of an intelligent detection of the unnatural agents requests // International Conference on Information Science and Communications Technologies (ICISCT). 2017. pp. 1–4.
10. *Бессонова Е.Е., Зикратов И.А., Колесников Ю.Л., Росков В.Ю.* Способ идентификации пользователя в сети Интернет // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 3. С. 133–137.
11. *Usmonov B. et al.* The cybersecurity in development of IoT embedded technologies // International Conference on Information Science and Communications Technologies (ICISCT). 2017. pp. 1–5.
12. *Исхаков А.Ю., Исхаков С.Ю., Мещеряков Р.В.* Повышение защищенности сервисов аутентификации путем проведения дополнительной идентификации с использованием оптимального признакового пространства // Информационные технологии в науке, управлении, социальной сфере и медицине: сборник научных трудов. 2017. С. 117–122.
13. *Abouollo A., Almuhammadi S.* Detecting malicious user accounts using Canvas Fingerprint // The 8th International Conference on Information and Communication Systems (ICICS). 2017. pp. 358–361.
14. *Daud N.I., Haron G.R., Othman S.S.S.* Adaptive authentication: Implementing random canvas fingerprinting as user attributes factor // IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). 2017. pp. 152–156.
15. Система управления проектами Trac. URL: https://trac.torproject.org/projects/tor/query?status=accepted&status=assigned&status=needs_review&status=needs_revision&status=new&status=reopened&order=priority&col=id&col=summary&col=keywords&col=status&col=owner&col=type&col=priority&keywords=tbb-fingerprinting (дата обращения 20.06.2018).
16. *Диденко С.М.* Исследование модели динамики параметров информационного почерка пользователя // Вестник Тюменского государственного университета. 2006. № 5. С. 170–174.
17. *Pilankar P.S., Padiya P.* Multi-phase mouse dynamics authentication system using behavioural biometrics // International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES). 2016. pp. 1947–1950.
18. *Hu S. et al.* Deceive Mouse-Dynamics-Based Authentication Model via Movement Simulation // The 10th International Symposium on Computational Intelligence and Design (ISCID). 2017. vol. 1. pp. 482–485.
19. *Chen X. et al.* A practical real-time authentication system with Identity Tracking based on mouse dynamics // IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). 2014. pp. 121–122.
20. *Kaminsky R., Enev M., Andersen E.* Identifying Game Players with Mouse Biometrics // University of Washington. Technical Report. 2008. 12 p.
21. *Feher C. et al.* User Identity Verification via Mouse Dynamics // Information Sciences. 2012. vol. 201. pp. 19–36.

22. *Stanić M.* Continuous user verification based on behavioral biometrics using mouse dynamics // Proceedings of the ITI 2013 35th International Conference on Information Technology Interfaces. 2013. pp. 251–256.
23. Идентификация пользователей Tor Browser через анализ особенностей работы с мышью. URL: https://www.opennet.ru/opennews/art.shtml?num=44027 (дата обращения: 20.06.2018).
24. *Данилов Н.А., Шульга Т.Э.* Построение тепловой карты на основе точечных данных об активности пользователя приложения // Прикладная информатика. 2015. № 2(56). С. 49–58.
25. *Диденко С.М.* Развитие математической модели информационного почерка пользователя // Математическое и информационное моделирование: сборник научных трудов. 2006. С. 68–73.
26. *Шапцев В.А., Диденко С.М.* Разработка и исследование компьютерной модели динамики системы «пользователь-мышь» // Диссертация на соискание степени кандидата технических наук. 2007. 95 с.

DOI 10.15622/sp.61.7

S.YU. MIROSHNICHENKO, V.S. TITOV, E.N. DREMOV, S.A. MOSIN

HOUGH TRANSFORM APPLICATION TO DIGITIZE RECTANGULAR SPATIAL OBJECTS ON AEROSPACE IMAGERY

Miroshnichenko S.Yu., Titov V.S., Dremov E.N., Mosin S.A. **Hough Transform Application to Digitize Rectangular Spatial Objects on Aerospace Imagery.**

Abstract. The paper describes the method of development for the remote sensing data processing to speed up the digitizing workflow. The method is designed to digitize rectangular objects using their approximate spatial positions and provides an automatic estimation of the orientation and aspect ratio.

The paper contains a formal statement of the problem of digitizing an object with the desired geometric shape using it's apriori known spatial position on a source image. The method creates polygonal representations of rectangular spatial objects from one or a few reference points set by an operator. It is based on source image's pixels clustering using spectral bands as a feature space. The following Hough transform incorporates local direction of intensity gradient to estimate object's orientation and reduce computational complexity together with low-pass filtering within an accumulation process to improve robustness. It is shown that the developed method can be modified to digitize objects of any analytically described shape.

The method is designed to allow easy user interaction without any significant delays and to provide transparent and predictable control of an output object's polygon size.

To investigate the developed method a test dataset with more than 700 rectangular objects was used. The root-mean-square error of object's points positioning, mean rotation error in polar coordinates and a Jaccard index were used to measure a precision of the digitized objects. The experiment results demonstrate that digitizing workflow is accelerated by 25–40% using the software implementing the developed method without a significant precision loss.

Keywords: remote sensing data, automated digitizing, image processing, hough transform.

1. Introduction. Current stage of Earth remote sensing systems' (RSS) development is marked with the continuous growth of their quantity. Modern RSS sensors gain better spatial resolution and a wider range of spectral channels. Together these facts turn to a rapidly growing stream of remote sensing data (RSD) that should be processed and georeferenced to have sufficient precision for further use [1, 2]. RSD processing workflow also should be accelerated to prevent airborne and space imagery lose their actuality. Digital maps are one of the RSD secondary processing main products. Operators create and update digital maps using interactive software instruments (interactive method) that do not imply any assistance to recognize objects or detect their boundaries. An operator digitizes image by finding and recognizing objects using his skills and experience. He defines each object's shape, spatial position, orientation, type and then sequentially places vertexes of a polygon that bounds the object from others and an underlying surface. An operator's actions within the confines of digitizing workflow are guided by editorial and technical requirements that change due to spatial scale and a resulting map's field of application. These requirements regulate object's level of detail, acceptable shape variants, a minimum dis-

tance between neighboring objects, etc. for each thematic layer the map contains. A digitizing workflow is laborious; the precision of the result directly depends on operator's skills, experience, concentration and gradually reduces during work time due to his tiredness caused by monotony and diligence of performed actions.

There are some software products for automated RSD processing that provide wizards to create individual rules and step-by-step sequences for spatial objects digitizing and recognition [3-6]. These products are based on object-oriented detection [7] and are aimed to accelerate digital maps creation workflow, but, in practice, in spite of declared features their range of application comes down to greenery and hydrography objects digitizing and calculative and estimative problems solution. The limited range of applications results from the fact that automated RSD processing software has a limited ability to follow editorial and technical requirements and to transform objects' polygonal representations to a demanded geometric shape.

Figure 1 shows the result of space image automated digitizing comparing to the interactive method. Source image on Figure 1a is separated to clusters depending on contour and texture features (Figure 1b) and used further to extract the buildings and structures subset (Figure 1d) with kNN method [8]. Figure 1c displays ground truth polygons of the same objects that were created in the interactive mode according to common editorial and technical requirements. Figure 1c displays ground truth polygons of the same objects created in the interactive mode according to common editorial and technical requirements. The automated result is an object's "draft" (Figure 1e shows a difference from the interactively created ground truth polygons) rather than a digital map's item and is only suitable to estimate the number of objects of the desired type, their spatial positions, occupied area, etc. For now, Jaccard index is a de facto standard quantitative measure of RSD automated recognition. The results of automated digitizing require per-object revision and correction to satisfy the minimum precision demands and lead to even greater operator's efforts than a map's interactive creation from scratch.

Results of investigations and competitions devoted to automated RSD processing demonstrate domination of approach based on deep learning artificial neural networks. Modern neural network architectures combined with GPGPU computing technologies allow to find practical solutions of the following space imagery recognition problems: to detect objects of a priori specified type (for instance roads in [10] and vehicles in [11-12]) or several classes (up to 10 in [9, 13-16]), to find and count certain kinds of animals [17-18]. However, we should mention that all developments in the field of deep learning neural networks are in a stage of experimental investigation far from being incorporated into commercial software products and for now are available only for experts in machine learning.

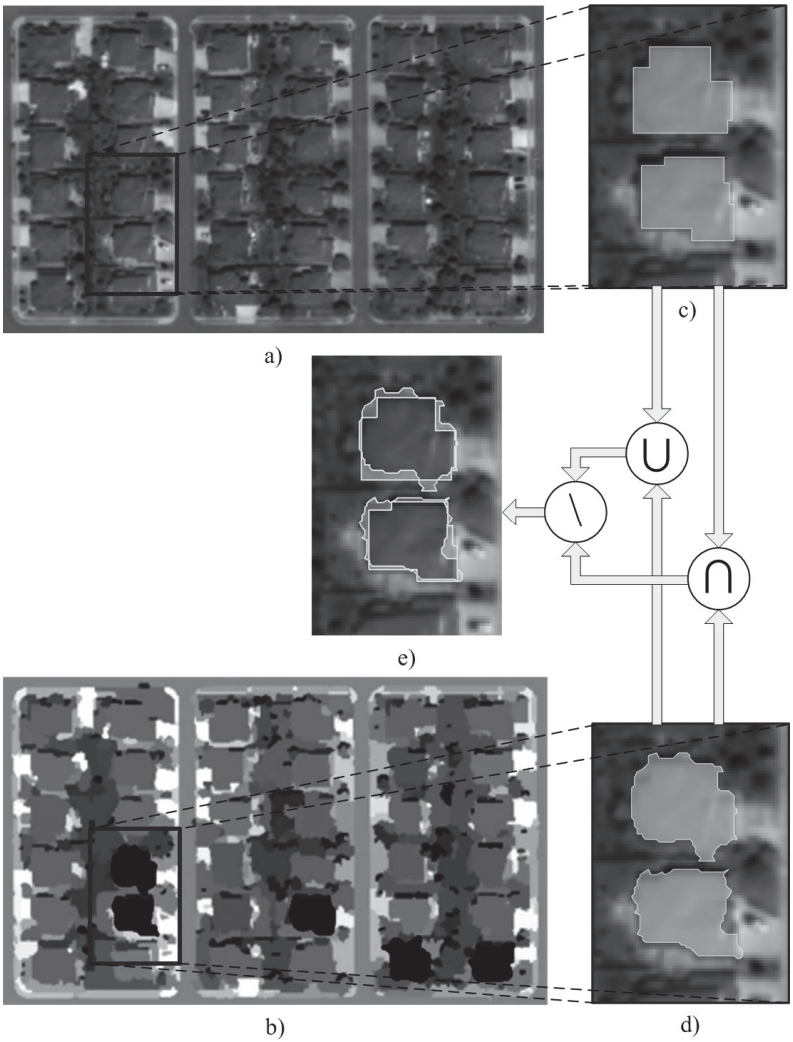


Fig. 1. The result of space image's (a) automated buildings digitizing (b, d) comparing to the interactive one (c) using Jaccard Index, visually represented on (e)

Another approach to decrease time costs for RSD processing lies in the development of automated software tools designed to be easily integrated into existing digitizing workflow and help users to perform the most laborious and tedious operations. Also, these tools can transform objects' polygons to the desired shape according to editorial and technical requirements.

ERDAS IMAGINE software contains the "Region grow" tool that in combination with a set of vector cleanup operations provide a user-controlled automated digitizing [19]. However, this tool has significant computational complexity (a created polygon appears through at least a couple of seconds after the user action). A vector cleanup process performs after the digitizing was finished for the whole layer and requires more operator's efforts to control output precision compared to the object-to-object approach. Feature Analyst contains a toolset for the assisted extraction that allows a user to digitize building by merely dragging a rectangle that encompasses it [20]. The shortcoming of this tool is that the object must be oriented strictly along the screen axes. Our paper focuses on the RSD digitizing process speedup by development and investigation of a new method designed to create rectangular polygonal representations with automatic estimation of orientation and aspect ratio based on the corresponding objects' approximate position.

2. Problem formalization. The essence of the digitizing process is to create a bounding polygon v for each recognized object on the source image $\mathbf{I}$. The bounding polygonal representation (further — polygon) is the polygon separating the object from others and an underlying surface. A priori information referring to the object is obtained by a preliminary recognition by an operator and includes the following:

– set of reference points (Figure 2a) representing an approximate spatial position of the object to be described by the polygon v;

– object's shape s that guides a selection of the corresponding digitizing method.

Hence, automated digitizing the object with the rectangular shape s_{sq} is described by:

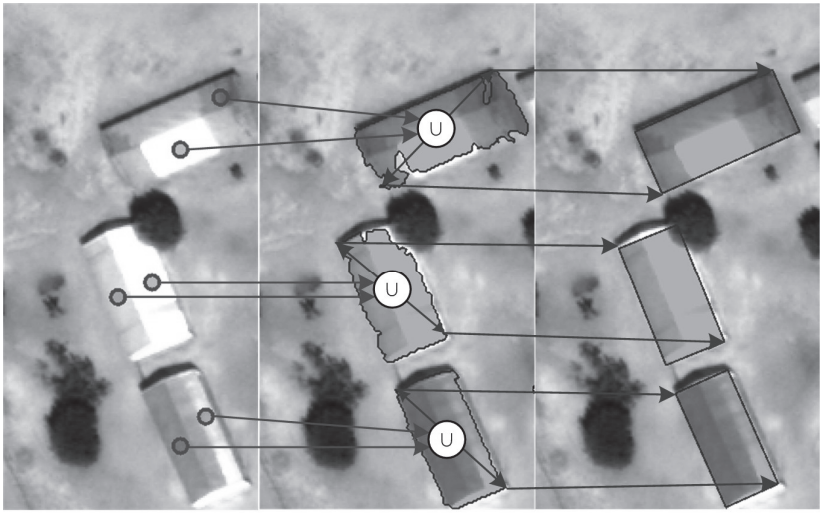
$$v = F_{sq}\left(\mathbf{I}, \mathbf{R}, t\right)\Big|_{s=s_{sq}}, \tag{1}$$

where t — is a threshold value regulating a size of a produced rectangular polygon.

Source image $\mathbf{I}$ contains K spectral bands, and each is described by a discrete brightness field with linear sizes M x N:

$$\mathbf{I} = \mathbf{f}\left(x, y\right) = \left(f_k\left(x, y\right)\right)_{k=1}^{K},\ x = \overline{1, M}, y = \overline{1, N}, \tag{2}$$

where $\mathbf{f}\left(\cdot\right)$ — is a vector of brightness values in the pixel $\left(x, y\right)$.



a) b) c)

Fig. 2. Automated digitizing process: reference points (a) are used to detect corresponding objects (b) and then to transform them to a required (rectangular in our case) shape (c)

Reference points of **R** are set in the source image local coordinate system and can be readily transformed into geodetic or projected coordinates using appropriate matrix [1]:

$$\mathbf{R} = \left\{ \left(x_j^{(\mathbf{R})}, y_j^{(\mathbf{R})} \right) \right\}_{j=1}^{N_{\mathbf{R}}}. \tag{3}$$

Object polygon v (Figure 2c) is described by a closed sequence of points in the source image local coordinate system:

$$v = \left(x_i^{(v)}, y_i^{(v)} \right)_{i=1}^{N_v}, \tag{4}$$

where N_v — is a number of points in v, which for a rectangular object equals four.

To get a ground truth digital map, we have to make a parallel projection of every terrain object to an underlying surface with a normal vector $\vec{n}$. For a perspective image (created with a sensor whose optical axis is set by a vector $\vec{o}$ having a non-zero angle with $\vec{n}$) without a digital surface model (DSM) of sufficient spatial resolution, we encounter coordinate distor-

tions $\left(\Delta x_i^{(v)}, \Delta y_i^{(v)}\right)$ shown at Figure 3a. For that reason, the paper considers only orthorectified images which have measurement properties [1] and can be used to create digital maps directly (Figure 3b).

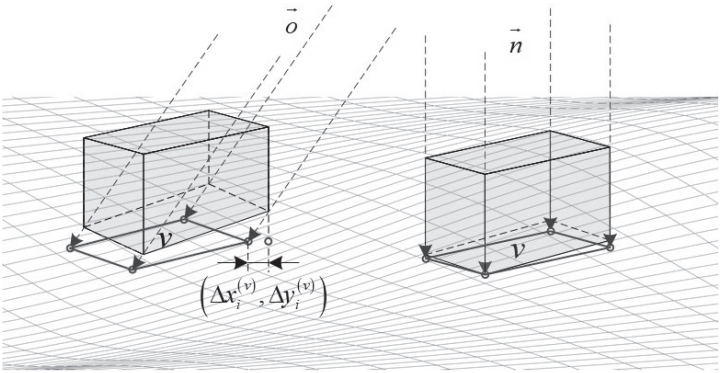


Fig. 3. Creating a polygon from a perspective image without a DSM (a) and from an orthorectified image (b)

Polygon (4) of a rectangular object contains four intersection points of four pairwise parallel and pairwise perpendicular straight lines (Figure 4). Coordinates of intersection points are calculated by a solution of the following system of linear equations:

$$\begin{aligned} &\mathbf{A}\cdot\left[x_1^{(v)}, y_1^{(v)}\right]^T = \left[r_1, r_3\right]^T,\ \mathbf{A}\cdot\left[x_2^{(v)}, y_2^{(v)}\right]^T = \left[r_1, r_4\right]^T, \\ &\mathbf{A}\cdot\left[x_3^{(v)}, y_3^{(v)}\right]^T = \left[r_2, r_4\right]^T,\ \mathbf{A}\cdot\left[x_4^{(v)}, y_4^{(v)}\right]^T = \left[r_2, r_3\right]^T, \\ &\mathbf{A} = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ \cos\left(\alpha - \frac{\pi}{2}\right) & \sin\left(\alpha - \frac{\pi}{2}\right) \end{bmatrix}. \end{aligned} \quad (5)$$

3. Developed method. The following requirements guided the method development process:

– the method has to implement functional (1) with the input parameters described by (2), (3) and output has the form of (4);

– software tool based on the developed method should allow comfortable interaction without any significant (clearly visible by an operator) delays. Hence a time of a single execution of (1) should be less than 500 ms on a workstation with Intel Core i3 processor.

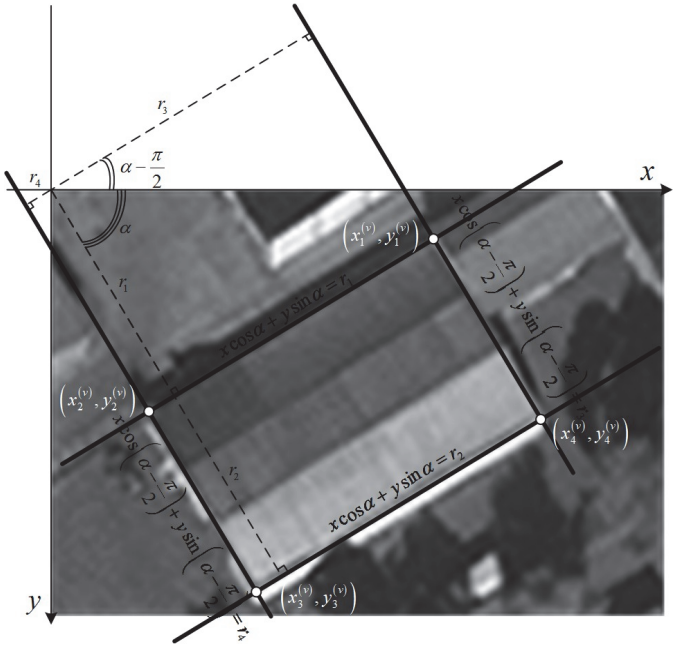


Fig. 4. Polygon of a rectangular object contains four intersection points of four pairwise parallel and pairwise perpendicular straight lines

– the method should provide transparent and predictable control of an output polygon size using a threshold: a higher t value corresponds to a greater area of a produced polygon.

Choice of Hough transform [21] to implement a rectangular objects' digitizing is explained by fusion of its ability to detect straight lines forming an output rectangles' sides with acceptable computational complexity providing an interactive operation mode for a developing software tool. However, a classic Hough transform based straight lines detector combined with edge detection [22] as an object boundaries extractor do not grant to an operator a transparent model to control sizes of output polygon (4) using a threshold value t.

To satisfy the mentioned above requirements, we have chosen a two-stage scheme consisting of the following sequentially performed functions:

1. The function $f_c(\cdot)$ creates a cluster c representing digitized object's boundaries visible on a source image $\mathbf{I}$ (Figure 2b). A function output depends on a reference points set $\mathbf{R}$ and a threshold value t:

$$c = f_c(\mathbf{I}, \mathbf{R}, t). \tag{6}$$

2. The function $f_{sq}(\cdot)$ converts a previously obtained cluster c to a rectangular polygon v using a Hough transform (figure 2c):

$$v = f_{sq}(\mathbf{I}, c). \tag{7}$$

We have knowingly divided the functional (1) into functions (6) and (7), besides the advantages described above, to gain an ability to convert a detected on an image cluster to any analytically described shape by merely replacing function (7) with the desired type of Hough transform.

3.1. Object detection. A starting point of function (6) is a reflection of a reference points set $\mathbf{R}$ into a set $\mathbf{C}^{(0)}$ of initial cluster values (Figure 5a):

$$\mathbf{R} \to \mathbf{C}^{(0)}, \mathbf{C}^{(0)} = \left\{c_i^{(0)}\right\}_{i=1}^{N_\mathbf{R}}, \; c_j^{(0)} = \left\{\left(x_0^{(c)}, y_0^{(c)}\right)\right\} = \left\{\left(x_j^{(\mathbf{R})}, y_j^{(\mathbf{R})}\right)\right\}. \tag{8}$$

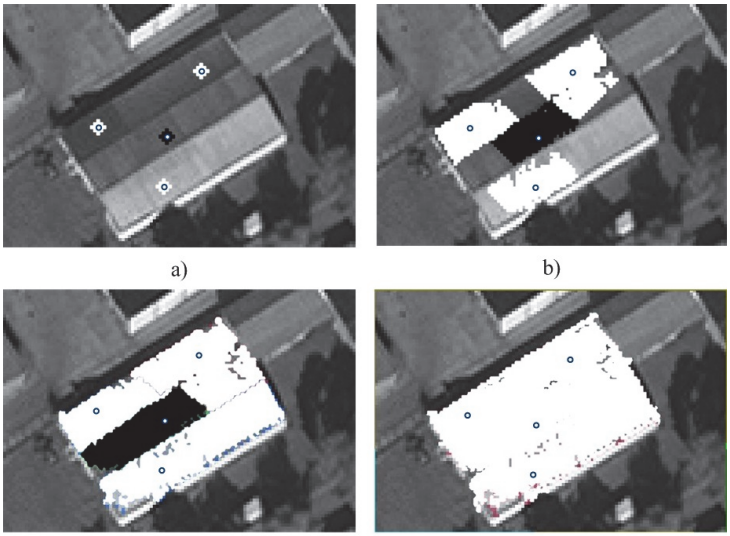


Fig. 5. An illustration of a reference points reflection into the set of initial clusters (a), their growth (b shows the result of 16th iteration, c – is the final 62nd iteration) and merging to the resulting cluster (d)

Next, an iterative cluster-growing function $f_{cg}(\cdot)$ (Figure 5b) on each m-th step to all items of $\mathbf{C}^{(m)}$ adds that pixels meet the following demands:

– do not belong to any cluster on the m-th iteration;

– situated on the current border of a cluster $c_j^{(m)}$;

– a distance value $d(\cdot)$ of a pixel in a feature space to a cluster's $c_j^{(m)}$ center does not exceed a threshold t set by an operator:

$$\mathbf{C}^{(m+1)} = \left\{c_i^{(m+1)}\right\} = \left\{f_{cg}\left(c_i^{(m)}, t\right)\right\}_{j=1}^{N_{\mathbf{C}}^{(m)}},$$

$$f_{cg}\left(c_i^{(m)}, t\right) = c_j^{(m)} \cup \left\{\begin{array}{l} \forall (x,y) \in \mathbf{I} : (x,y) \notin \mathbf{C}^{(m)} \\ \wedge \left(x \pm 1, y \pm 1\right) \in c_i^{(m)} \\ \wedge d\left(c_i^{(m)}, (x,y)\right) < t \end{array}\right\}. \quad (9)$$

We used maximum brightness difference to incorporate all K spectral bands of the image (2) as a feature space to calculate distance $d(\cdot)$:

$$d\left(c_i^{(m)}, (x,y)\right) = \max_k \left| \frac{1}{N_i^{(c)}} \sum_i f_k\left(x_i^{(c)}, y_i^{(c)}\right) - f_k(x,y) \right|_{k=1}^{K}. \quad (10)$$

Further feature space extension by edges, textures or spectrums on the one hand increases object detection precision but on the other hand — complicates distance function (10) and requires preliminary computations to obtain the mentioned features that in total do not comply with the interactivity requirement.

Iterations number of cluster-growing function (9) doesn't have predefined limit. Growing process stops in case no new image pixels can be added to any cluster within a current iteration (Figure 5c):

$$\mathbf{C} = \mathbf{C}^{(m)} : \bigcup_{i=1}^{N_{\mathbf{C}}^{(m-1)}} f_{cg}\left(c_i^{(m-1)}, t\right) \setminus c_i^{(m)} = \varnothing. \quad (11)$$

The final operation of function (6) is a merge $f_{cm}(\cdot)$ of clusters, containing in a set $\mathbf{C}$, into a resulting cluster c (Figure 5d). Clusters pair c_i, c_n merging requires them to have joint or adjacent pixels:

$$c_i = f_{cm}\left(c_i, c_n\right) = c_i \cup c_n \left| \forall (x,y) \in c_i : \left(x \pm 1, y \pm 1\right) \in c_n. \right. \quad (12)$$

If a resulting set **C** after the merge (12) contains more then one item (i.e., the threshold value is not enough to cover feature space distances (10) between all reference points to produce a single cluster), there are three options of action:

1. Accept a cluster of **C** with the maximum area as a result of the function (6):

$$c = f_c\left(\mathbf{I}, \mathbf{R}, t\right) = \arg\left[\max_i \left|c_i\right|\right]. \tag{13}$$

2. Let the result of (6) contain all produced clusters and transform each of them into a rectangular polygon using (7).

3. Continue an iterative execution of the function (9) and increase threshold value t every time the growing process stops due to condition (11) fulfill and a cluster set $\mathbf{C}^{(m)}$ contains more than one item. The developed method uses the first option because it allows an operator to concentrate on single object detection with sufficient visual control of the workflow together with the least computational complexity.

3.2. Object transformation to a rectangular polygon. To construct rectangular polygon (7), we have to find a solution of four linear equations systems (5), which, in turn, requires to obtain parameters $\alpha, r_1, r_2, r_3, r_4$ of straight lines bounding an object represented by a previously detected cluster c. As said before the mathematical basis for function (7) implementation is a Hough transform that converts source image's **I** coordinate field to an accumulator plane $h(\cdot)$ whose quantized dimensions represent parameters α, r of a straight line equation in Hesse normal form. Each pixel (x, y) appears as a curve in a plane $h(\cdot)$ by calculating the parameter r from all values α_n of the independent parameter α within a range $[\alpha_{\min}, \alpha_{\max}]$:

$$h\left(\alpha_n, r\right) = \sum_{(x,y)} \sum_{\alpha_n=\alpha_{\min}}^{\alpha_{\max}} H\left(x\cos\alpha_n + y\sin\alpha_n, r\right),$$
$$H\left(a,b\right) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases}. \tag{14}$$

For interactivity reasons, an accumulator plane $h(\cdot)$ is filled only by pixels (x, y) situated on the external boundary of the cluster c:

$$\exists(x,y) \in c \;\exists\left(x_b, y_b\right) \notin c : \left|(x,y) - \left(x_b, y_b\right)\right| \leq 1. \tag{15}$$

Transform (14) has an asymptotic described by $O\left(p_c \cdot N_\alpha\right)$, where p_c — is an external perimeter of the cluster c. To reach a precision of object's spatial position and orientation, sufficient for cartographic products, we should set a quantity N_α of parameter's α quants in proportion to p_c so, that a rotation of radius vector from coordinate system origin to any cluster boundary pixel by an angle $\left(\alpha_{\max} - \alpha_{\min}\right)/N_\alpha$ leads to shifting of its ending to a distance not exceeding one pixel.

Hence N_α is a function of p_c and furthermore an asymptotic of (14) corresponds to $O\left(p_c^2\right)$, that do not meet the interactivity requirement.

To obtain a linear asymptotic function, we reduced the number of accumulator plane $h(\cdot)$ dimensions by calculating the parameter α directly from a source image as a mode $Mo(\cdot)$ of gradient vector angle φ [22, 23] near the cluster's external boundary (15) within a range of $[0, \pi/2)$:

$$\alpha = Mo\left(\begin{cases}\varphi, \ \varphi < \pi/2 \\ \varphi - \pi/2, \ \varphi \geq \pi/2\end{cases}\right), \qquad \varphi = \angle\left(\nabla \overline{f}(x,y), \overrightarrow{Ox}\right). \tag{16}$$

Now when we have a fixed parameter α any pixel (x,y) of cluster's boundary transforms to a single value of r. Values $r_1 - r_4$ are calculated through two one-dimensional accumulators $h_\alpha(r)$ and $h_{\alpha-\pi/2}(r)$, each of what corresponds to a particular pair of straight lines:

$$h_\alpha(r) = \sum_{(x,y)} w_G\left(x\cos\alpha + y\sin\alpha - r\right), \tag{17}$$

where $w_G(\cdot)$ — is a Gauss function [24], applied as a low-pass filter to improve Hough transform robustness, and, to be more specific, to reduce an adverse effect of image's discreteness and α parameter calculation error on an output polygon precision.

Figure 6a illustrates a direct accumulation of r values similarly to (14) that leads to a r_4 parameter calculation error and the following incorrect position of one of the building's walls. Figure 6b shows an accumulation of low-pass blurred values (17) that provides more robust boundaries detection of a digitized object.

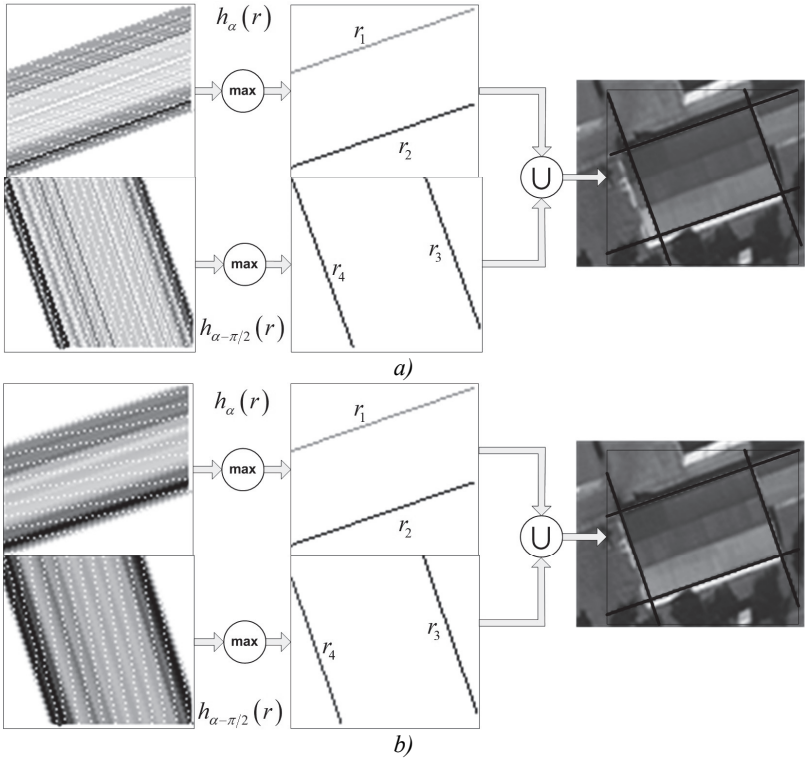


Fig. 6. An illustration of direct values accumulation (a) that leads to a parameter calculation error and to the subsequent incorrect position of building's wall compared to an accumulation of low-pass blurred values (b) that provides more robust boundaries detection

Filled up accumulators are used to select two pairs of parameter values $h_\alpha(r) \to (r_1, r_2)$, $h_{\alpha-\pi/2}(r) \to (r_3, r_4)$, and each corresponds to a maximum integral weight in view of a distance between straight lines:

$$(r_1, r_2) = \arg\left[\max_{\substack{r_i \in [r_{\max}, r_{\min}] \\ r_j \in [r_{\max}, r_{\min}]}} \left(\frac{\left(h_\alpha(r_i) + h_\alpha(r_j)\right) \cdot |r_i - r_j|}{r_{\max} - r_{\min}} \right)\right]. \quad (18)$$

Parameters $\alpha, r_1, r_2, r_3, r_4$ obtained through (16) and (18) are subsequently used to find a solution of linear equation system (5) and to create a resulting polygon in a form (4).

3.3. Method implementation. The implementation of our method has the following features and particular qualities to comply with requirements to interactivity and control predictability described in Section 3:

1. Cluster creation (6) uses only a part of the source image that is visible to an operator and has a screen zoom level. Thus sizes of a processed image part never exceed a monitor resolution that at the moment in most cases is equal to FullHD. A preliminary pyramid computation [25] provides a fast image overviewing for zoom levels that is less the original. If a working area has a zoom level more then 100%, a source image will not be scaled for processing as standard scaling algorithms would supply no additional data.

2. Only K available spectral bands of the source image **I** form a feature space to distinguish an object from its neighbors and the underlying surface. A cluster stores current mean brightness values for (10) as its attributes and updates them on adding new pixels.

3. Mode of gradient vector's angle (16) is computed by Sobel operator combining a sufficient edge detection precision with a low computational cost compared to other operators and Gaussian derivatives [26, 27]. To further maximize performance a module and an angle of gradient vector are obtained only in the neighborhood (15) of current cluster's external boundary. A histogram with a step equal to 1 deg is applied to select a mode value (we chose the step value empirically). Histogram accumulates pairs "gradient angle"→"gradient module" with the following selection of an angle value coheres with the maximum module sum.

4. To reduce an operator's need to interact with a visual interface of software tool implementing our method we transferred a threshold value control to a wheel of a mouse manipulator. Figure 7 illustrates a threshold value effect on a resulting cluster. A further increase of the value over shown on Figure 7d is redundant and leads to cluster growth beyond object's visible boundaries. After one polygon was created, a current threshold value can be applied to digitize another similar object without adjusting to further reduce the number of an operator's actions.

5. During the digitizing, the external boundary is displayed together with corresponding object's polygon as shown on Figure 8 to improve visual perception of the process by an operator and to enhance the comfort of his work.

6. The reference points' setup process must be straightforward and transparent for an operator. It is recommended to set a point for every homogeneous region of an object starting from a larger one or a central one (in case regions sizes are close). After a starting polygon has appeared operator chooses one of the following options:

– increases threshold value in case the polygon doesn't cover the whole object, and there is no clear border between its covered and uncovered regions;

– places addition reference points in case there is a clear border between the polygon and uncovered areas (mostly one point per region);

– decreases threshold value in case the polygon covers some regions that do not belong to a digitizing object.

An operator performs the described actions until the polygon's position and aspect ratio would match a digitizing object.

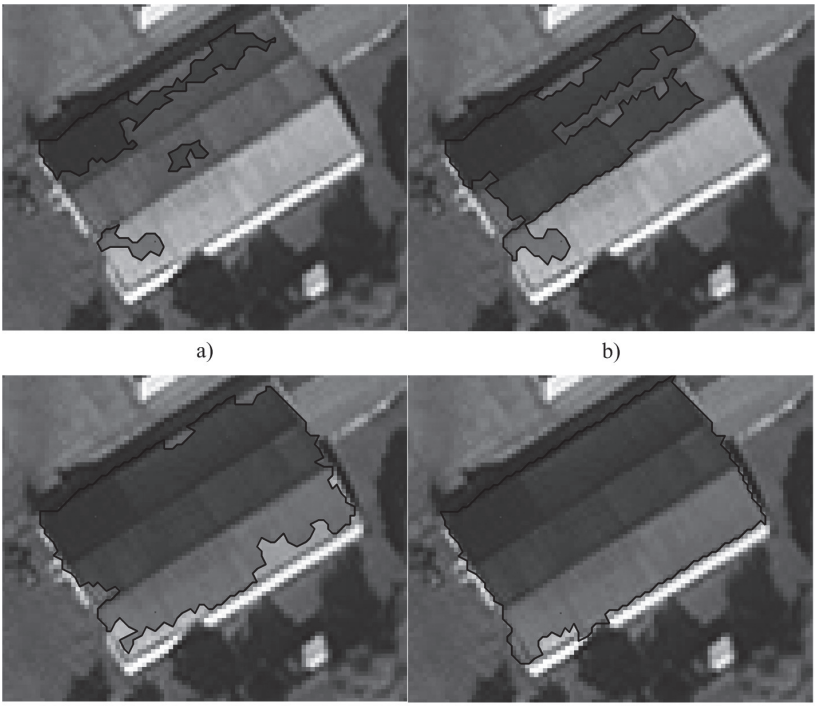


Fig. 7. An illustration of threshold value effect on a resulting cluster with $t = 4$ (a), $t = 7$ (b), $t = 10$ (c) и $t = 18$ (d)

For instance, the building with a span-roof contains two regions due to different sunlight reflection angles of its parts (the left object on Figure 8). To digitize that object an operator should set two reference points to each side of the roof.

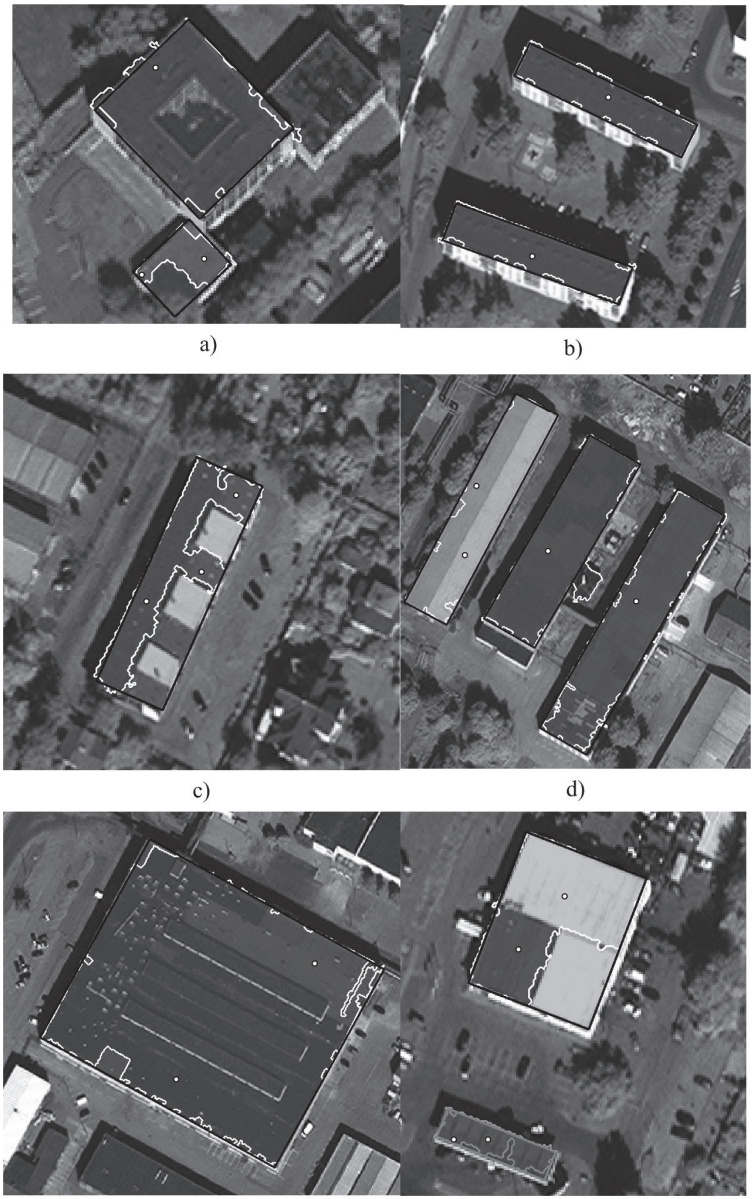


Fig. 8. A visualization of objects' digitizing with the developed method (output polygons' borders are black and the cluster borders are white)

Figure 8 displays samples of different objects' digitizing with the developed method. We set up to three reference points to detect them:

– one point per sample for buildings on Figure 8b, central and right on figure 8d and the larger one on Figure 8a;

– two points per sample were used for instances on Figure 8e, 8f, the left on Figure 8d, and the smaller one on Figure 8a;

– three points were required to digitize just the object on Figure 8c.

The developed method, on the one hand, digitizes objects that consist of several visually distinctive parts (Figure 8c, 8e, 8f) and on the other hand, correctly detects objects that are fractionally hidden or shaded by others with compensation of minor clustering errors.

4. Experiment and results. Within the experiment, we compared an output map precision and time consumption to create it using software based on the developed method and one of the most popular interactive digitizing toolsets of GIS MapInfo [28]. Dataset is represented by urban imagery with a spatial resolution of 0.5 m/pixel where mostly comprises multi-apartment residential development and industrial zones, and was used to prepare a ground truth digital map containing 777 rectangular buildings and structures (38% from all objects on the selected territory according to OpenStreetMap data [29]). Figure 10a shows a histogram of objects' sizes on this map.

We engaged two experts to carry out the experiment whose aim was to create a map similar to the ground truth using alternately the interactive toolset and the software based on the developed method. During the process experts did not have to spend time finding objects to digitize themselves – they used the specific markup layer made from the ground truth map that indicates the desired buildings by crosses as shown on Figure 9.

4.1. Precision criterions. Quantitative measure of precision for digital maps, produced by experts, is represented by the following criterions:

1. Root-mean-square error (RMSE) of object's points positioning which is expressed in pixels and equals to a Euclidian distance to the nearest point of the corresponding ground truth object.

2. Mean rotation error (MRE) expressed in degrees is measured as a minimum angle between straight lines forming the current object and the corresponding ground truth one.

3. Jaccard index (is dimensionless, expressed in percentages) represents the ratio of intersection area of a created object and a corresponding ground truth one to their union's area.

RMSE describes an absolute object positioning error comprising four types of distortions: aspect ratio mismatch, spatial shift, scaling and rotation errors. The last distortion is quantitatively expressed by MRE which was separated from RMSE to estimate a substitution precision of gradient angle mode (16) instead of the rotation angle.

Fig. 9. A specific markup layer that indicates the desired objects by crosses

Jaccard index represents a complete relative similarity measure of a created polygon to a ground truth one.

The algorithm of the created digital map precision estimation contains the following steps:

1. Find a corresponding object on the ground truth map for each one on the created map by a criterion of maximum intersection area (has to be nonzero).

2. If the ground truth object was found — calculate values of RMSE, MRE, and Jaccard index and write them to the created object's semantic fields to have the ability to analyze the results visually.

3. Calculate mean values of precision criterions for the created map in a whole.

4.2. Experiment results. In total experts created four digital maps within the experiment (two — with the interactive toolset, and two — with the developed method). Table 1 demonstrates mean RMSE, MRE, and Jaccard index values together with digitizing workflow duration. Duration of the digitizing workflow decreased by 39.9% and 26.3% for the first and the second expert respectively. Error values of the digi-

tal map created by the second expert using the developed method although are slightly greater comparing to the interactive one, but still are within the range of acceptable values. The first expert gained a more significant time economy but exceeded maximum RMSE and MRE values that for our experiment are set to 2 pixels (mean error of object's border visual detection in an original image scale without zooming) and 1 degree (determined by the a quantization of histogram used to calculate a gradient angle mode). Both experts used two reference points per object on the average.

Table 1. Mean RMSE, MRE and Jaccard index values together with digitizing workflow duration for created digital maps

Expert	Method	Objects quantity	Digitizing workflow duration, min	Time per object, sec	RMSE, pixels	MRE, deg	Jaccard index, %
Expert 1	Interactive	775	180	13.94	1.779	0.751	90.93
	Developed	773	*108*	*8.38*	2.251	1.053	88.89
	Difference		–72 (–40%)	–5.56 (–39.9%)	+0.472 (+21%)	+0.302 (+28.7%)	–2.04
Expert 2	Interactive	773	188	14.59	1.567	0.728	91.33
	Developed	770	*138*	*10.75*	1.909	0.925	90.33
	Difference		–50 (–26.8%)	–3.84 (–26.3%)	+0.342 (+17.9%)	+0.197 (+21.3%)	–1

Within the experiment, we also investigated dependencies of RMSE (Figure 10b), MRE (Figure 10c) and Jaccard index (Figure 10d) from a polygon's size to reveal features of the developed method.

RMSE has a direct dependency from a polygon's size. For objects smaller then 15 pixels RMSE value is 25% below the mean both for the interactive method and for the developed one. Similarly, RMSE value is 25% above the mean for objects with sizes higher than 39 pixels because operators subjectively pay less attention to a precision of a more large object. A higher spread of RMSE values for these objects is explained by their less quantity (not enough to get a smooth graph).

A discreetness of the source image explains an MRE reverse dependency from a polygon's size. The positioning error of 1 pixel for an object with the size of 10 pixels causes a rotation error of 6.3 degrees. In contrast, the same positioning error for an object with the size of 50 pixels causes a rotation error of 1.3 degrees. Higher MRE values of the developed method for objects with sizes above 18 pixels appear due to quantization of histogram used to calculate a gradient angle mode (16).

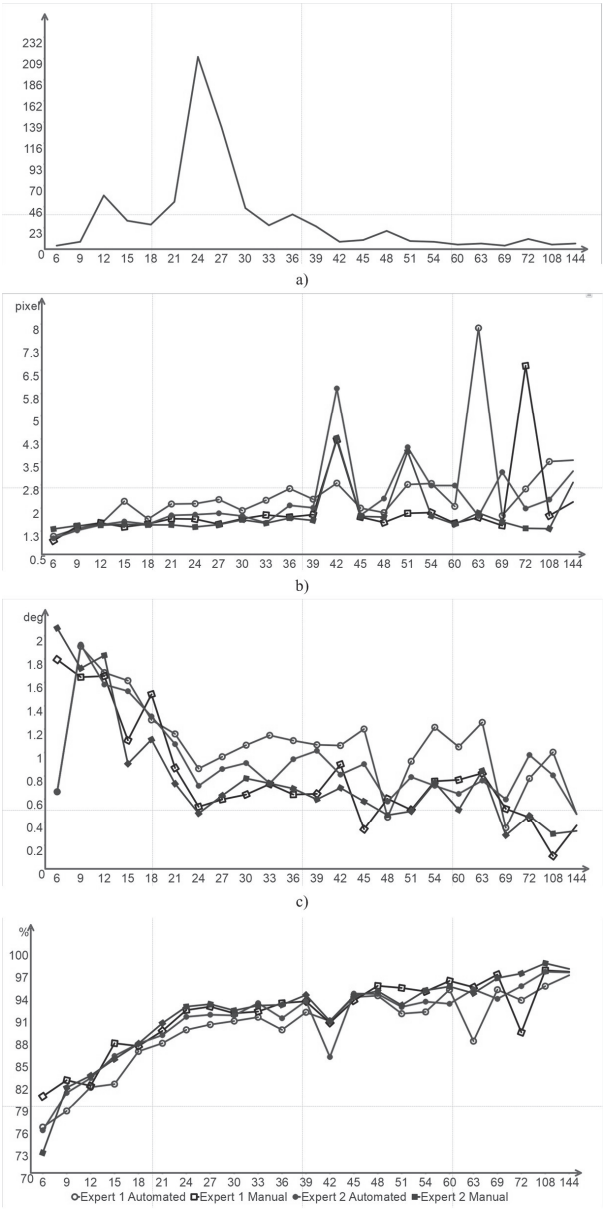


Fig. 10. Histogram of objects polygons' sizes on the experimental dataset (a), dependencies graphs of RMSE (b), MRE (c) and Jaccard index (d) from a polygon's size

Mean Jaccard index value directly depends on polygon's size similar to MRSE that is also caused by a source image's discreetness. A higher error of the interactive method is specified by a capability to reach a sub-pixel precision by digitizing objects on a source image zoomed above the original scale (an operator magnifies the image in 2-5 times and manually sets points of an output polygon between vertexes of image's discrete grid using his skills and experience).

5. Conclusion and discussion. We developed a new method for remote sensing data processing designed to digitize rectangular objects using approximate spatial positions providing automatic orientation and aspect ratio estimation. The experimental investigation of the method has shown workflow acceleration by 25–40% facing a slight precision reduction. In a shortcomings list, we should mention 20% higher MRE value comparing to the interactive method and an impossibility to reach a sub-pixel precision of a resulting polygon's points positioning.

Within a further work, we will improve our method to overcome the mentioned disadvantages, apply Hough transform for round and orthogonal polygons and further to develop a generalized method for digitizing objects of any analytically described shape, investigate a neural networks based methods and an interface design of automated digitizing tools to enhance workflow performance and comfort.

References

1. Schowengerdt R.A. Remote sensing: models and methods for image processing. Elsevier. 2007. 560 p. (Russ. ed.: Schowengerdt R.A. *Distancionnoe zondirovanie. Modeli i metody obrabotki izobrazhenij.* Moscow: Tehnosfera. 2013. 592 p.).
2. Antonushkina S.V. et al. *Sovremennye tehnologii obrabotki dannyh distancionnogo zondirovanija Zemli* [Modern technologies for processing of Earth remote sensing data]. Moscow: Physmathlit. 2015. 460 p. (In Russ.).
3. Blundell J.S., Opitz D.W. Object recognition and feature extraction from imagery: the Feature Analyst® approach. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. 2006. vol. 36. no. 4. pp. 42–47.
4. ENVI EX User's Guide. Available at: http://www.harrisgeospatial.com/portals/0/pdfs/enviex/ENVI_EX_User_Guide.pdf (accessed: 15.02.2018).
5. Manual for satellite data analysis eCognition developer. Available at: http://open_jicareport.jica.go.jp/pdf/12150314_03.pdf (accessed: 15.02.2018).
6. Kompleks avtomatizirovannogo deshifrirovanija i vektorizacii dannyx DZZ [Complex for automated recognition and digitizing of remote sensing data]. Available at: https://gisinfo.ru/products/automap.htm (accessed: 15.02.2018). (In Russ.).
7. Jebur M.N. et al. Per-pixel and object-oriented classification methods for mapping urban land cover extraction using SPOT 5 imagery. *Geocarto International.* 2014. vol. 29. no. 7. pp. 792–806.
8. Larose D.T. Discovering knowledge in data: an introduction to data mining. John Wiley & Son. 2014. 336 p.
9. Iglovikov V., Mushinskiy S., Osin V. Satellite imagery feature detection using deep convolutional neural network: A Kaggle competition. 2017. arXiv:1706.06169. Available at: https://arxiv.org/abs/1706.06169 (accessed: 15.02.2018).

10. How we participated in SpaceNet three Road Detector challenge. And how we got into top 10. Available at: https://spark-in.me/post/spacenet-three-challenge (accessed:15.02.2018).
11. Safe Passage: Detecting and Classifying Vehicles in Aerial Imagery. Available at: https://www.datasciencechallenge.org/challenges/1/safe-passage (accessed: 15.02.2018).
12. Iglovikov V. Britanskie sputnikovye snimki 2: kak vse bylo na samom dele [British Satellite Images 2: how it was really]. Available at: https://habrahabr.ru/company/ods/blog/330118/ (accessed: 15.02.2018). (In Russ.).
13. Dstl Satellite Imagery Feature Detection. Available at: https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection (accessed: 15.02.2018).
14. Kuzin A. Vtoroe pochetnoe. Zametki uchastnika konkursa Dstl Satellite Imagery Feature Detection [The second honorable place. Notes of the Satellite Imagery Feature Detection Challenge Participant]. Available at: https://habrahabr.ru/company/avito/blog/325632/ (accessed: 15.02.2018). (In Russ.).
15. Lee K. Dstl Satellite Imagery Competition, 1st Place Winner's Interview. Available at: http://blog.kaggle.com/2017/04/26/dstl-satellite-imagery-competition-1st-place-winners-interview-kyle-lee/ (accessed: 15.02.2018).
16. Otkrytyj konkurs na luchshee reshenie v oblasti sozdanija intellektual'nyh tehnologij deshifrirovanija vidovoj ajerokosmicheskoj informacii [The open competition for a best solution in a field of intellectual technologies creation for recognition of aerospace imagery in visible spectrum]. Available at: http://fpi.gov.ru/activities/konkurs/spacemap (accessed: 15.02.2018). (In Russ.).
17. NOAA Fisheries Steller Sea Lion Population Count Available at: https://www.kaggle.com/c/noaa-fisheries-steller-sea-lion-population-count (accessed: 15.02.2018).
18. Kaggle: kak nashi setochki schitali morskih l'vov na Aleutskih ostrovah [Kaggle: how our networks counted the sea lions in the Aleutian Islands]. Available at: https://habrahabr.ru/company/ods/blog/337548/ (accessed: 15.02.2018). (In Russ.).
19. Obusek F. Delineating rooftops in ERDAS IMAGINE. Available at: https://www.youtube.com/watch?v=HBcPB6Agr5c (accessed: 15.02.2018).
20. Blundell S. et al. Feature analyst v5.0. Proceedings of the ASPRS Annual Conference. 2008. pp. 28–37.
21. Ballard D.H. Generalizing the Hough transform to detect arbitrary shapes. *Pattern recognition.* 1981. vol. 13. no. 2. pp. 111–122.
22. Gonzalez R., Woods R. Digital image processing. Pearson Education. 2002. 793 p. (Russ. ed.: Gonzalez R., Woods R. *Cifrovaja obrabotka izobrazhenij.* Moscow. Tehnosfera. 2006. 1072 p.).
23. Duda R.O., Hart P.E., Stork, D.G. Pattern classification. John Wiley & Sons. 2000. 680 p.
24. van Vliet L.J., Verbeek P.W. Edge localization by MoG filters: Multiple-of-Gaussians. *Pattern Recognition Letters*. 1994. vol. 15. no. 5. pp. 485–496.
25. Burt P., Adelson E. The Laplacian pyramid as a compact image code. *IEEE Transactions on communications*. 1983. vol. 31. no. 4. pp. 532–540.
26. Emelyanov S.G. et al. *Metody i sistemy cifrovoj obrabotki ajerokosmicheskih izobrazhenij* [Methods and systems for digital processing of aerospace imagery]. Novosibirsk. Nauka. 2012. 175 p. (In Russ.).
27. Vatutin E.I., Miroshnichenko S.Yu. Titov V.S. [Software optimization of Sobel operator using SIMD-extensions of the x86 family processors]. *Telecommunikatsii — Telecommunication*. 2006. Issue 6. vol. 12. pp. 12–16. (In Russ.).
28. MapInfo Pro – Desktop GIS. Available at: https://www.pitneybowes.com/us/location-intelligence/geographic-information-systems/mapinfo-pro.html (accessed: 15.02.2018).
29. NextGIS Data. Geodata in ESRI Shape, ESRI Geodatabase, GeoJSON, Mapinfo Tab, PDF formats. Available at: https://data.nextgis.com/en/ (accessed: 15.02.2018).

Miroshnichenko Sergey Yurievich — Ph.D., associate professor of computer sciences department, Southwest State University (SWSU). Research interests: Earth remote sensing data processing, automated recognition and digitizing of geospatial objects, machine learning. The number of publications — 90. oldguy7@rambler.ru; 94, 50 Let Oktyabrya str., Kursk, 305040, Russia; office phone: +79081277657, Fax: +7(4712)222-665.

Titov Vitalii Semenovich — Ph.D., Dr. Sci., professor, head of computer sciences department, Southwest State University (SWSU). Research interests: image analysis, pattern recognition, robotic systems, autonomous robots, machine vision systems, calibration. The number of publications — 630. titov-kstu@rambler.ru; 94, 50 Let Oktyabrya str., Kursk, 305040, Russia; office phone: +7(4712)222-665, Fax: +7(4712)222-665.

Dremov Evgenii Nikolaevich — Ph.D. student of computer sciences department, Southwest State University (SWSU). Research interests: raster topographic maps automated processing and recognition, images preprocessing. The number of publications — 16. evgenidremov@yandex.ru; 94, 50 Let Oktyabrya str., Kursk, 305040, Russia; office phone: +7(4712)222-665, Fax: +7(4712)222-665.

Mosin Sergey Aleksanrovich — Ph.D. student of computer sciences department, Southwest State University (SWSU). Research interests: automated remote sensing data recognition, artificial neural networks. The number of publications — 8. 22mosin@gmail.com; 94, 50 Let Oktyabrya str., Kursk, 305040, Russia; office phone: +7(4712)222-665, Fax: +7(4712)222-665.

Acknowledgements. This research is supported by the state assignment (project №2.9102.2017/BCh).

УДК 004.92 **DOI** 10.15622/sp.61.7

С.Ю. Мирошниченко, В.С. Титов, Е.Н. Дремов, С.А. Мосин

ВЕКТОРИЗАЦИЯ ПРЯМОУГОЛЬНЫХ ПРОСТРАНСТВЕННЫХ ОБЪЕКТОВ НА АЭРОКОСМИЧЕСКИХ ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ ПРЕОБРАЗОВАНИЯ ХАФА

Мирошниченко С.Ю., Титов В.С., Дремов Е.Н., Мосин С.А. **Векторизация прямоугольных пространственных объектов на аэрокосмических изображениях с использованием преобразования Хафа.**

Аннотация. Рассматривается метод обработки данных дистанционного зондирования, предназначенный для векторизации прямоугольных объектов по их ориентировочным положениям с автоматическим определением ориентации и соотношения сторон.

Приведена формальная постановка задачи векторизации объекта заданной геометрической формы на изображении с использованием априорной информации о его пространственном положении. Метод позволяет создавать векторные представления прямоугольных геопространственных объектов по одной или нескольким устанавливаемым оператором опорным точкам и основан на кластеризации исходного изображения с использованием имеющихся спектральных каналов в качестве пространства признаков. Последующее преобразование Хафа использует локальное направление градиента яркости для оценки пространственной ориентации объекта и снижения вычислительной сложности преобразования, а также низкочастотную фильтрацию в процессе накопления значений для повышения робастности. Показана возможность модификации метода для обеспечения возможности векторизовать объекты любой аналитически задаваемой формы.

При разработке метода учтены требования по минимизации времени векторизации для повышения комфорта работы оператора и обеспечения возможности предсказуемого контроля размера создаваемого векторного представления.

Для экспериментального исследования разработанного метода использована тестовая выборка, содержащая более 700 объектов прямоугольной формы. В качестве критериев точности векторизации использованы средняя квадратическая ошибка позиционирования точек объекта, среднее угловое отклонение объекта в полярных координатах и коэффициент схожести площадных объектов Жаккара. Результаты эксперимента показывают снижение временных затрат на векторизацию на 25–40% при использовании программной реализации разработанного метода без существенного снижения точности создаваемой картографической продукции.

Ключевые слова: данные дистанционного зондирования, автоматизированная векторизация, обработка изображений, преобразование Хафа.

Мирошниченко Сергей Юрьевич — к-т техн. наук, доцент кафедры вычислительной техники, Юго-Западный государственный университет (ЮЗГУ). Область научных интересов: обработка данных дистанционного зондирования Земли, автоматизированное дешифри-рование и векторизация геопространственных объектов, машинное обучение. Число научных публикаций — 90. oldguy7@rambler.ru; ул. 50 Лет Октября, 94, Курск, 305040; р.т.: +79081277657, Факс: +7(4712)222-665.

Титов Виталий Семенович — д-р техн. наук, профессор, заведующий кафедрой вычислительной техники, Юго-Западный государственный университет (ЮЗГУ). Область научных интересов: обработка изображений, распознавание образов, робототехнические

системы, автономные роботы, системы технического зрения, калибровка. Число научных публикаций — 630. titov-kstu@rambler.ru; ул. 50 Лет Октября, 94, Курск, 305040; р.т.: +7(4712)222-665, Факс: +7(4712)222-665.

Дремов Евгений Николаевич — аспирант кафедры вычислительной техники, Юго-Западный государственный университет (ЮЗГУ). Область научных интересов: обработка и распознавание растровых топографических карт, предварительная обработка изображений. Число научных публикаций — 16. evgeni-dremov@yandex.ru; ул. 50 Лет Октября, 94, Курск, 305040; р.т.: +7(4712)222-665, Факс: +7(4712)222-665.

Мосин Сергей Александрович — аспирант кафедры вычислительной техники, Юго-Западный государственный университет (ЮЗГУ). Область научных интересов: автоматизированное дешифрирование данных дистанционного зондирования Земли, искусственные нейронные сети. Число научных публикаций — 8. 22mosin@gmail.com; ул. 50 Лет Октября, 94, Курск, 305040; р.т.: +7(4712)222-665, Факс: +7(4712)222-665.

Поддержка исследований. Работа выполнена при финансовой поддержке государственного задания (проект №2.9102.2017/БЧ).

Литература

1. *Шовенгердт Р.А.* Дистанционное зондирование. Модели и методы обработки изображений // М.: Техносфера. 2013. 592 с.
2. *Антонушкина С.В. и др.* Современные технологии обработки данных дистанционного зондирования Земли // М.: Физматлит. 2015. 460 с.
3. *Blundell J.S., Opitz D.W.* Object recognition and feature extraction from imagery: the Feature Analyst® approach // International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. 2006. vol. 36. no. 4. pp. 42–47.
4. ENVI EX User's Guide. URL: http://www.harrisgeospatial.com/portals/0/pdfs/enviex/ENVI_EX_User_Guide.pdf (дата обращения: 15.02.2018).
5. Manual for satellite data analysis eCognition developer. URL: http://open_jicareport.jica.go.jp/pdf/12150314_03.pdf (дата обращения: 15.02.2018).
6. Комплекс автоматизированного дешифрирования и векторизации по данным ДЗЗ. URL: https://gisinfo.ru/products/automap.htm (дата обращения: 15.02.2018).
7. *Jebur M.N. et al.* Per-pixel and object-oriented classification methods for mapping urban land cover extraction using SPOT 5 imagery // Geocarto International. 2014. vol. 29. no. 7. pp. 792–806.
8. *Larose D.T.* Discovering knowledge in data: an introduction to data mining // John Wiley & Sons. 2014. 336 p.
9. *Iglovikov V., Mushinskiy S., Osin V.* Satellite imagery feature detection using deep convolutional neural network: A Kaggle competition. 2017. arXiv:1706.06169. URL: https://arxiv.org/abs/1706.06169 (дата обращения: 15.02.2018).
10. How we participated in SpaceNet three Road Detector challenge. And how we got into top 10. URL: https://spark-in.me/post/spacenet-three-challenge/ (дата обращения: 15.02.2018).
11. Safe Passage: Detecting and Classifying Vehicles in Aerial Imagery. URL: https://www.datasciencechallenge.org/challenges/1/safe-passage (дата обращения: 15.02.2018).
12. *Игловиков В.* Британские спутниковые снимки 2: как все было на самом деле. URL: https://habrahabr.ru/company/ods/blog/330118/ (дата обращения: 15.02.2018).
13. Dstl Satellite Imagery Feature Detection. URL: https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection (дата обращения: 15.02.2018).

14. *Кузин А.* Второе почетное. Заметки участника конкурса Dstl Satellite Imagery Feature Detection. URL: https://habrahabr.ru/company/avito/blog/325632/ (дата обращения: 15.02.2018).
15. *Lee K.* Dstl Satellite Imagery Competition, 1st Place Winner's Interview. URL: http://blog.kaggle.com/2017/04/26/dstl-satellite-imagery-competition-1st-place-winners-interview-kyle-lee/ (дата обращения: 15.02.2018).
16. Открытый конкурс на лучшее решение в области создания интеллектуальных технологий дешифрирования видовой аэрокосмической информации. URL: http://fpi.gov.ru/activities/konkurs/spacemap (дата обращения: 15.02.2018).
17. NOAA Fisheries Steller Sea Lion Population Count URL: https://www.kaggle.com/c/noaa-fisheries-steller-sea-lion-population-count (дата обращения: 15.02.2018).
18. Kaggle: как наши сеточки считали морских львов на Алеутских островах. URL: https://habrahabr.ru/company/ods/blog/337548/ (дата обращения: 15.02.2018).
19. *Obusek F.* Delineating rooftops in ERDAS IMAGINE. URL: https://www.youtube.com/watch?v=HBcPB6Agr5c (дата обращения: 15.02.2018).
20. *Blundell S. et al.* Feature analyst v5.0 // Proceedings of the ASPRS Annual Conference. 2008. pp. 28–37.
21. *Ballard D.H.* Generalizing the Hough transform to detect arbitrary shapes // Pattern recognition. 1981. vol. 13. no. 2. pp. 111–122.
22. *Гонсалес Р., Вудс Р.* Цифровая обработка изображений // М.: Техносфера. 2005. 1072 с.
23. *Duda R.O., Hart P.E., Stork D.G.* Pattern classification // John Wiley & Sons. 2000. 680 p.
24. *van Vliet L.J., Verbeek P.W.* Edge localization by MoG filters: Multiple-of-Gaussians // Pattern Recognition Letters. 1994. vol. 15. no. 5. pp. 485–496.
25. *Burt P., Adelson E.* The Laplacian pyramid as a compact image code // IEEE Transactions on communications. 1983. vol. 31. no. 4. pp. 532–540.
26. *Емельянов С.Г. и др.* Методы и системы цифровой обработки аэрокосмических изображений // Новосибирск. Наука. 2012. 175 с.
27. *Ватутин Э.И., Мирошниченко С.Ю., Титов В.С.* Программная оптимизация оператора Собела с использованием SIMD-расширений процессоров семейства x86 // Телекоммуникации. 2006. Т. 6. №. 12. С. 12–16.
28. MapInfo Pro – Desktop GIS. URL: https://www.pitneybowes.com/us/location-intelligence/geographic-information-systems/mapinfo-pro.html (дата обращения: 15.02.2018).
29. NextGIS Data. Геоданные в форматах ESRI Shape, ESRI Geodatabase, GeoJSON, Mapinfo Tab, PDF. URL: https://data.nextgis.com/ru/ (дата обращения: 15.02.2018).

DOI 10.15622/sp.61.8

I.S. VASILJEVIĆ, D. DRAGAN, R. OBRADOVIĆ, V.B. PETROVIĆ

ANALYSIS OF COMPRESSION TECHNIQUES FOR STEREOSCOPIC IMAGES

Vasiljević I.S., Dragan D., Obradović R., Petrović V.B. **Analysis of Compression Techniques for Stereoscopic Images.**

Abstract. Virtual Reality (VR) and Augmented Reality (AR) Head-Mounted Displays (HMDs) have been emerging in the last years and they are gaining an increased popularity in many industries. HMDs are generally used in entertainment, social interaction, education, but their use for work is also increasing in domains such as medicine, modeling and simulation. Despite the recent release of many types of HMDs, two major problems are hindering their widespread adoption in the mainstream market: the extremely high costs and the user experience issues [1]. The illusion of a 3D display in HMDs is achieved with a technique called stereoscopy. Applications of stereoscopic imagining are such that data transfer rates and–in mobile applications–storage quickly become a bottleneck. Therefore, efficient image compression techniques are required. Standard image compression techniques are not suitable for stereoscopic images due to the discrete differences that occur between the compressed and uncompressed images. The issue is that the loss in lossy image compression may blur the minute differences between the left-eye and right-eye images that are crucial in establishing the illusion of 3D perception. However, in order to achieve more efficient coding, there are various coding techniques that can be adapted to stereoscopic images. Stereo image compression techniques that can be found in the literature utilize discrete Wavelet transformation and the morphological compression algorithm applied to the transform coefficients. This paper provides an overview and comparison of available techniques for the compression of stereoscopic images, as there is still no technique that is accepted as best for all criteria. We want to test the techniques with users who would actually be potential users of HMDs and therefore would be exposed to these techniques. Also, we focused our research on low-priced, consumer grade HMDs which should be available for larger population.

Keywords: image compression, stereoscopic, wavelets, head mounted display.

1. Introduction. Stereoscopy/stereovision is a technique for making an illusion of image depth which relies on the phenomenon of stereopsis (binocular depth perception) based on the difference between the images that we see with the left and right eye, Figure 1. These are the so-called pairs of stereoscopic images [2, 3].

The images contain vast amounts of data, and the price of extra realism in stereo displays is the doubling of data (due to the simultaneous existence of two images), causing a bottleneck in the data flow 3D image is obtained only with the use of hardware or spectacles for observing stereoscopic images so that 3D information is not recorded except by implication in the difference between the two images. The technical limitations of this sort of display mean that the refresh rate of the display should be synchronized with the glasses used which may be done wirelessly as well.

It is known that because of its imperfections, the human eye cannot distinguish the entire color gamut available on modern displays. The ques-

tion arises, therefore, whether it makes sense to keep all these shades of color if the human eye is not able to see them. In addition, there are also redundancies in image content, especially in the case of frames of video. Due to limitations in data transfer, online load, and synchronization, compression is a key component in modern communication and data communication services.

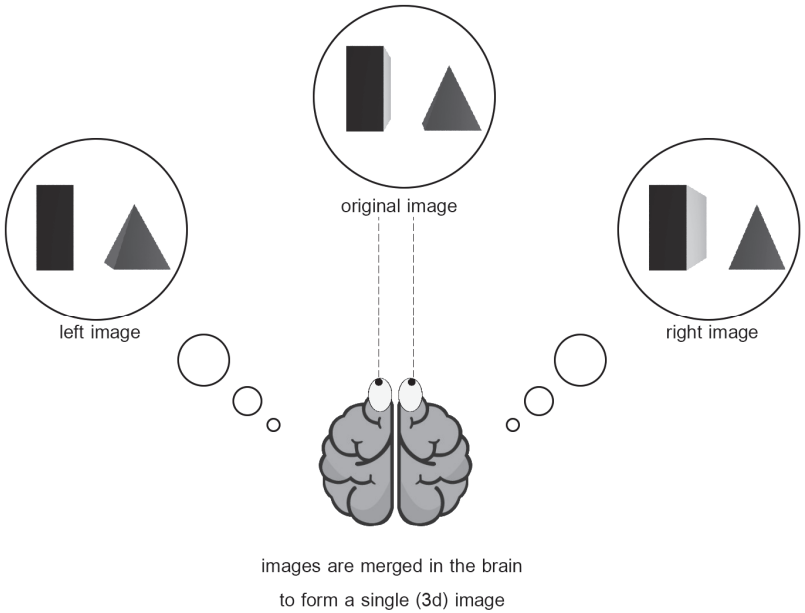


Fig. 1. Demonstration of how a pair of stereo images creates an illusion of 3D scene/objects

There are a number of standard image compression techniques such as, for example, JPEG [4] and MPEG [5, 6]. Standard image compression techniques cannot establish correlations between the left and right stereo pairs and the information they contain. In the case of standard compression techniques, it would be necessary to compress each image of a stereo pair separately, which would lead to doubling of the bandwidth in data transmission [7, 8]. Because of the discrete differences that occur between the uncompressed and compressed images, the illusion is potentially disrupted, as the vital minute differences between images may be disrupted and there is no provision in standard image compression techniques to preserve them. Due to the lack of standard compression techniques, dedicated techniques for the compression of stereoscopic images are developed. The compression

techniques of stereo images found in the literature use discrete wavelet transformation and morphological compression algorithm. The wavelet nature of the algorithm and the proposed disparity compensation provide reconstructed images without blocking artifacts and fewer annoying ringing effects. One main focus of research in stereo image coding has been disparity estimation, a technique used to reduce the coding rate by taking advantage of the redundancy in a stereo image pair. These are the reasons why the analyzed compression techniques are significant and stand out from standards compression techniques such as MPEG, JPEG and JPEG-200 [5]. Also, many of the compression techniques proposed over the time are proprietary and as such are not easily available. Therefore, we decided to limit our research on the ones whose source codes, executables, and/or results were available to us. It is our intention to expand our research and to implement more compression techniques proposed for stereoscopic imaging.

A wavelet is a mathematical function used for digital signal processing and image compression. In signal processing, wavelets are used to recover weak signals from noise. It is also useful for X-ray and magnetic resonance imaging in medical applications. In internet communication, it is used to compress images on a larger scale.

In this paper, three techniques for the compression of stereoscopic images which most frequently mentioned in the literature [9] and whose results are available to us, will be analyzed:

1. Stereoscopic image compression using discrete wavelet transformation and coding using the morphological re-presentation of coefficients, with estimation of disparities within the morphological coder, known as the Dense disparity map algorithm.

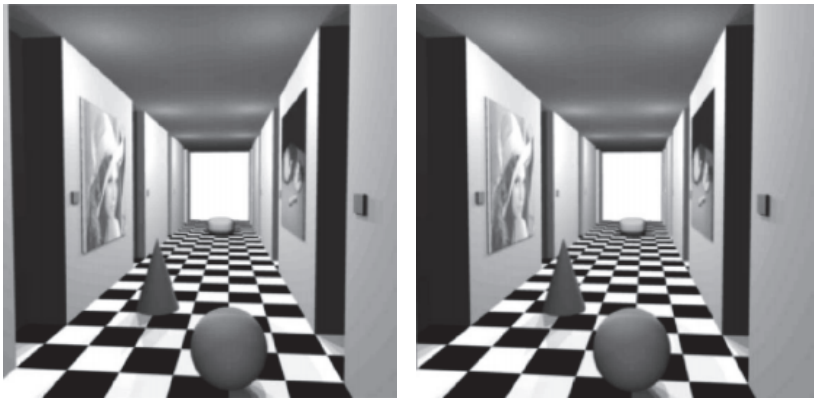
2. Stereo image compression based on quadratic analysis and morphological representation of the wavelet coefficient, known as the Disparity compensated residual algorithm.

3. Stereo image compression based on MRF (Markov Random Field) analysis for the assessment of disparities and morphological coding.

These techniques are applied to pairs of stereoscopic images. Figure 2. shows a pair of stereo images [10] used to test and compare compression techniques in this paper.

The comparison of the above-mentioned compression techniques was done by objective methods for assessing the quality found in the analyzed literature, but also by the subjective estimates of the subjects tested. There are different methods for assessing the quality of stereoscopic and digital images in general. The most common methods for estimating the quality of digital images are objective and subjective methods. Objective methods estimate the quality of images in relation to some defined parameter. Subjective methods for assessing quality, on the other hand, are based

on subjective image quality assessments. The respondents evaluate the quality according to personal feelings and observations, as was done in this paper. The interviewees observed the pictures with the use of virtual reality head-mounted display and evaluated their quality with grades of 1-5. The testing process is described in more details in the rest of the paper.



a) *b)*

Fig. 2. A pair of stereo images (256x256): a) Left image; b) Right image

The paper consists of six sections. The first section details the goal and subject of the research. In the second section, an overview of techniques for the compression of stereoscopic images will be given. The third section will give overview of devices used for display of stereoscopic images. In the fourth section, a comparison of the above techniques will be made. The results obtained by compression were also evaluated by objective and subjective metrics for image quality assessment. The fifth section will give a detailed description of the subjective analysis, as well as the analysis of the results obtained. In the last (sixth) section, the conclusions of this research will be presented.

2. Review of subjected compression techniques for stereoscopic images. This section will present an overview of techniques for the compression of stereoscopic images and a description of the notion of disparity in stereovision.

2.1. Disparity in stereovision. The problem of finding points of a stereo pair corresponding to the same point of a 3D object is called correspondence. The problem is simplified if the cameras are coplanar. The distance between the two points on the stereo pair of the image corresponding to the same point of the scene is called disparity. The estimation of this distance (disparity vector — DV) is very important because the target image

can be predicted from reference to DV. The disparity compensated difference (DCD) is estimated so that the redundant information is not encoded. Disparity compensation uses a Block Matching Algorithm-BMA and the determination of residual blocks [12, 13], equations (1):

$$DCD\left(b_{i,j}\right) = \sum_{(x,y)\in b_{i,j}} \left[{b_{i,j}}^{R}(x,y) - {\tilde{b}_{i,j}}^{L}\left(x+dv_x, y+dv_y\right)\right], \quad (1)$$

where ${b_{i,j}}^{R}$ and ${\tilde{b}_{i,j}}^{L}$ are corresponding blocks of the target and reconstructed images, and dv_x and dv_y are components of the disparity vector for the best matching [7], equation (2):

$$DV\left(b_{i,j}\right) = \arg \min_{\left(dv_x, dv_y\right)\in A} DCD \mid (b_{i,j}), \quad (2)$$

where A is the window search area, and the matching criterion is the Mean Absolute Error (MAE). The described compensation method represents a closed loop because the prediction of the target image is performed using a reconstructed reference image. Difference compensation can be performed with a reference image and is called compensation for the difference in the open loop. Open loop algorithms are simpler, but less efficient since there is no need for inverse quantization and the reverse branch transformation on the encoder side. The difference compensation process uses spatial dependence among images to remove redundant information. The blocks that do not have the appropriate blocks in the reference image are called blocked blocks. The pages of stereo pairs that cannot see both eyes, as well as the areas that arise from the overlapping of objects are the occluded areas.

2.2. Dense disparity map algorithm. The field of disparity vector is estimated by the method of pairing the pixels of the image and the pixels used for shape comparison. The algorithm is designed to produce a distribution of coefficients in each iteration in order to get the best performance. Clusters are formed in sub-groups. The target image clusters correspond to their "cousins" located on the reference image by shifting them so that they optimize the minimal absolute error. The disparity field vector is determined for the entire cluster, and the compensated disparity difference field is determined by subtracting the corresponding coefficients between the two clusters.

For the compression coefficients to be determined in a hierarchical way, a structured 3x3 element for the morphological dilation operation is used. Figure 3 shows the HL1, HL2 and HL3 subgroups of the left picture that are later divided into 3 levels. The spatial dependence of wavelet coeffi-

cients is obvious and justifies the morphological monitoring of all coefficients. The morphological dilation is done on the basis of structured 3x3 elements.

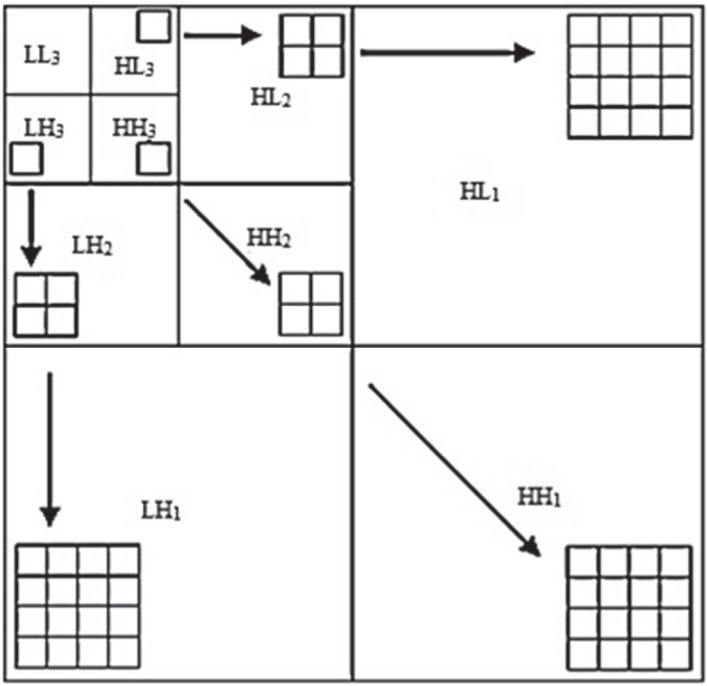


Fig. 3. Spatial dependence of wavelet coefficients between the image division subgroups

A unique step size quantizes all subgroups and, as a result, a binary image with two partitions is obtained, separating significant and insignificant parts of the image. The coefficients within a defined range are called significant. The dependence of wavelet coefficients in the formation of clusters suggests the application of morphological dilation in order to identify the "neighbors" that are significant. These coefficients are divided into groups according to the similarity of characteristics. The results of this division contain the necessary additional information for their description in the decoding phase. The algorithm starts with the first significant coefficient after quantization by a uniform quantizer. Then dilation operation is applied at each first significant coefficient and assigned to each adjacent neighbor of importance, clockwise. Darker blocks indicate significant clustering coefficients that have been collected in a predefined way. Isolated significant

coefficients that do not form clusters together with those coefficients that are not assigned as neighbors form a group of insignificant coefficients. This procedure produces a coder behavior map. The uncertainty regarding well-selected significant coefficients in the final scale is checked and solved by repeating the dilation operation on all specific structural elements.

The performance of this still image algorithm is quite good compared to other compression techniques. The division and progressive encoding of subgroups in a hierarchical sense provide an advantage for handling the estimation of disparities in the same image. This technique degrades quality, but it is computationally 'cheap.'

2.3. Disparity compensated residual algorithm. The algorithm consists of image coding based on the morphological prevalence of the coefficient of wavelet transformation and on the quadratic analysis of the disparity between the images of the stereo pair [18]. The coding unit uses a discrete wavelet transformation followed by a morphological encoder, which exploits statistically the properties of wavelet coefficients within and between subgroups in order to create entropy-reducing partitions between the significant and insignificant coefficients. The disparity compensation procedure uses a block adaptation algorithm that is based on variable-size blocks generated by the quadratic decomposition of the target image. The diagram in Figure 4 demonstrates the components of the algorithm:

– *Discrete wavelet transformation* — performs decomposition and quantization of target and residual images;

– *Morphological compression* — divides the wavelet coefficients into significant and insignificant ones in order to reduce their entropy;

– *Inverse morphological compression* — reconstructs the reference image and places it as an input to the disparity compensation units;

– *A Disparity Compensation Unit* — compares two inputs (reconstructed reference and target image), estimates the best target image prognosis, and produces a residual target image, representing target images from the best prediction (compensated disparity difference). The best prediction vectors for each block are called disparity vectors;

– *Entropy coder* — encodes the reference image, residual target image, and disparity vector.

Figure 4 shows the diagram of the proposed algorithm [9]. The right image is segmented into blocks of homogeneous intensity with quadratic decomposition with intensity difference thresholding. These blocks belong either to the same object or to the background and show homogeneous disparity characteristics. This is followed by a quadratic decomposition with a simplified speed distortion criterion, which allows division of an already existing block into 4 sub-blocks. Figure 5 shows how the reference image is divided into blocks (a) and the residual image (b) [17].

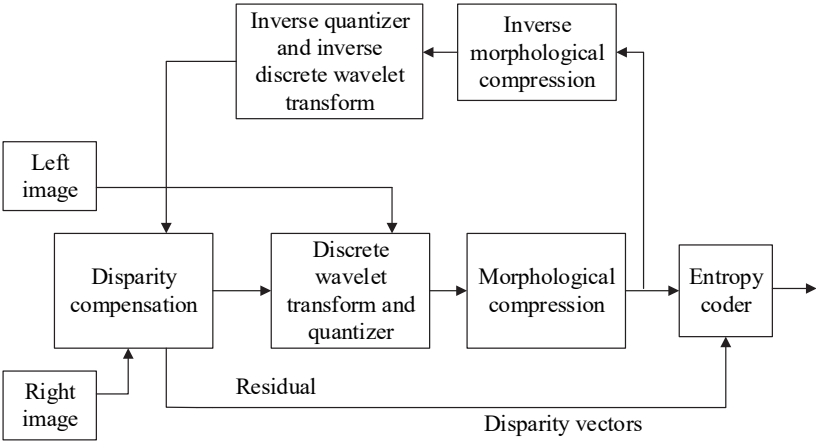
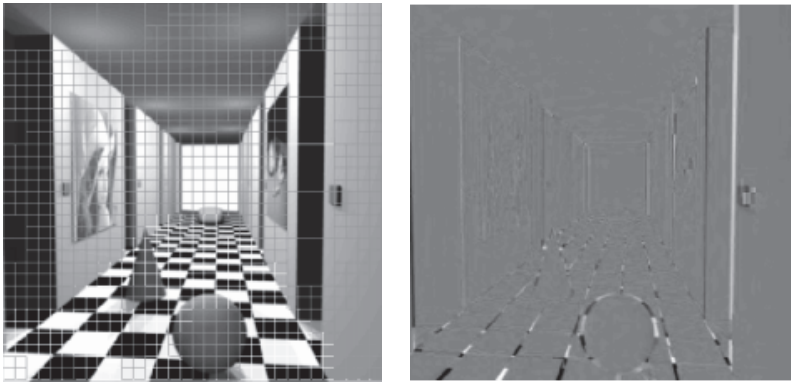


Fig. 4. Quadtree decomposition algorithm diagram



a) *b)*

Fig. 5. Disparity compensated residual algorithms: a) Target image divided; b) Residual image

2.4. Stereo image coding based on MRF analysis for the assessment of disparities and morphological coding. MRF is an algorithm that is based on the fields of disparity D and occlusion O [12, 17]. The algorithm considers the existence of irregularly distributed points or positions in the image (called nodes) which are elements of the images to be paired. Possible correspondences of each node are a discrete set of selected image properties that correspond to the characteristics of another stereo pair image, according to the allowed range of disparities. The biggest problem is only the determi-

nation of these fields. The configurations of the disparity field D and the occlusion O can be determined by the equation (3) [12]:

$$\left(\hat{d},\hat{o}\right) = \arg\min_{(d,o)\in S}\left[U\left(S^R \mid S^L, d, o\right) + U\left(d \mid O\right) + U\left(o\right)\right], \quad (3)$$

where S^L and S^R represent the reference and target image. The first equation condition is the probable energy of the target image, relative to the left or reference image, and the fields D and O. This term is called the boundary of similarity because it affects the similarity of two stereo pairs of images.

The second condition is the energy of the disparity field, in relation to the occlusion field O, and is called the obstacle. It is a smooth variation of the vector of disparity. The third condition is the energy of the occlusion field and is called the occult limit. This discards any discontinuities or closed blocks. Therefore, minimizing the total amount of these three conditions or rather optimization criteria affects the efficiency of the residual image encoding and the resulting disparity vectors. The field of the initial assessment is formed using the double threshold technique and is divided into three regions: unused, occlusive and the uncertain region depending on the estimated probability of being sorted as 'occlusive' or 'not occlusive.'

3. Overview of devices used for display of stereoscopic images. When providing a taxonomy of the current virtual reality (VR) hardware developments, the presented devices often exist only in the prototype stage; most of them are not yet commercially available and may even never be. The main category in current display technology represents the visual displays. In terms of consumer VR they are all head-mounted displays (HMDs) which are either wired or mobile [13], [14]. Other categories providing haptic and multi-sensory feedback, but are expensive and not easily available.

Mobile HMDs — it is possible to identify three subcategories in the HDMs for mobile systems. All share the property of being wireless and being usable without an additional PC. In most cases, their application areas lie in entertainment — displaying 360◦ movies or panoramas rendered from a stationary point of view or alternatively interactive walkthroughs based on gaze directed navigation.

The first subcategory in the mobile displays is called "simple casing"; these displays are basically a frame for smart phones with additional focusing lenses mounted at an appropriate distance. They fully rely on the technology of the smart phone to display and process the data. Google has developed the first devices of this kind.

The second mobile subcategory consists of ergonomically designed smart phone cases, which contain significantly better optics, possibly limited additional electronics, and are more comfortable to wear. The difference between this and the first subcategory is one of degree, not of kind and rest fundamentally on

The third subcategory are dedicated mobile HMDs. Different prototypes and examples of this sort of device exist. Gameface and Oculus Go, are stand-alone systems that do not need an additional PC or a smart phone with all necessary hardware built into the unit. This seems to be a promising approach as it allows the hardware to be tailored to the task at hand, but the question of how affordable it is going to be for wider audience remains open.

Wired HMDs — these are devices that need to be tethered to a PC or some other high-power computing device which is in charge of generating 3D graphics. Despite their name this type of device needn't be *actually* wired, and may in fact be connected wirelessly. They differ in comparison to Mobile device by being dependent on their, fixed, computing platform. The feature list of wired HMDs is diverse and they may be distinguished on the basis of not only traditional quality factors like resolution, Field of View (FOV) or weight but also specialized additional features. Some are equipped with cameras to allow for AR and can be used as video see-through displays, while others include eye tracking.

A representative of mobile HMDs is Google Cardboard. Google Cardboard follows a minimalist philosophy of design, acting as a very basic viewer, not even allowing the user to secure the device to their hear using a strap. In order to provide basic interaction the Cardboard unit is equipped with a magnet on the left side of the device. Phone sensors can detect the motion of the magnet. A vast amount of inexpensive cardboard clones exist and have disseminated the technology far and wide. They differ in such terms as lenses with a larger FOV or mounting apparatus of some kind. More advanced solutions are on the market as well, providing a simple plastic casing, back straps or hats to mount the phone [13]. Xwave glasses (used for testing) do not cost much and are easily available and work on the same principle. They do not have an integrated solution for tracking user's head in space (similarly to the professional HMDs), but they allow tracking of orientation using the accelerometer built into the phone [14]. The Xwave glasses are shown in Figure 6.

Generally, devices for displaying stereoscopic images are expensive and given our interest in the compression of stereo images and our desire for maximum reach, price is very important. Therefore, we restricted the research to a device that is relatively easy and cheap to buy and available to a large number of people [15].

Fig. 6. Xwave glasses for virtual reality

Table 1 shows the specifications of the Xwave glasses.

Table 1. Specifications of Xwave glasses

Type	3D glasses
Compatibility with phones	3.5 - 6.0 inch mobile/Andriod /Win/iOS platform
Optics	33.5mm Aspheric Optical Resin Lens
Transformations	1%-2%/1.5-2 time bigger
Simulation	1000" Screen in 3 meters
View angle	70-90°
Dimensions	198x135x110 mm
Weight	399g
Color	Black

4. Objective comparison of compression techniques for stereoscopic images. Some of the objective methods are Peak Signal to Noise Ratio, Mean Square Error MSE, Structural SIMilarity, UIQI (Universal Index Quality), and RRIQA (Reduced Reference Image Quality Assessment) [19-21].

MSE and PSNR methods are not complicated and are easy to understand and implement. Therefore, they are often used, perhaps most often in the assessment of image quality [22, 23]. These methods cannot give an objectively assessed quality that matches the observer's estimation for a wide range of coding and transmission parameters. This is because they compare the tested and reference data, without knowing what they actually represent. They do not take into account the characteristics of the HVS (Human Visual System), which show that HVS does not have the same sensitivity to different types of distortion and different distortion

properties. In addition, it is very important to know in which part of the image the distortion occurs, and MSE or PSNR do not take this into account. They measure the accuracy of the signal without modeling any properties of the HVS or image content and semantics.

In accordance with the literature [11, 12, 18] this research uses PSNR as the main measure of compressed image quality.

Figures 7, 8 and 9 show the stereo pairs of images compressed by the techniques described in the paper. The first proposed technique for the compression of stereoscopic images has a low complexity of the algorithm itself, but the optimization of the distortion is very complex. Therefore, it is difficult to improve the performance of the algorithm itself, taking into account the optimization criteria for the distortion. An estimation of the pixel disparity has both advantages and disadvantages in relation to the techniques based on the division of images into blocks. It affects the low complexity of the algorithm, avoiding blocking artifacts, but it has lower efficiency in terms of distortion. The proposed method also has a low degree of complexity due to the simplicity of the quantization process and the effect on the determination of the disparity within the compression scheme as a single framework. For this technique, it can be said that it is of low complexity and that it represents the technique of an alternative estimate of disparity.

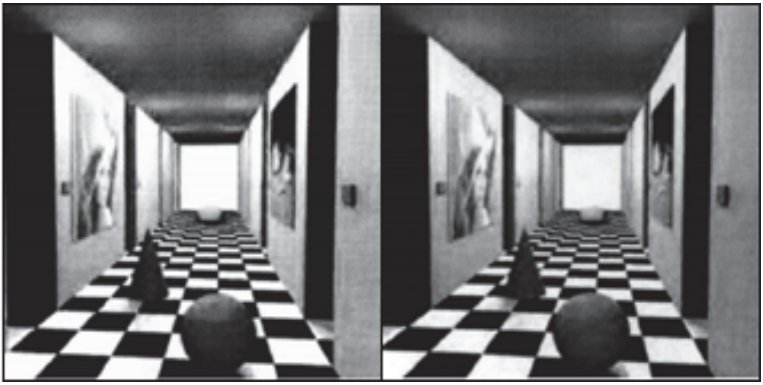


Fig. 7. Examples of stereo pairs of images (left and right) compressed with dense disparity map algorithm

Disparity compensated residual encoder is a robust encoder, which inherits all the benefits of wavelet transformation and reduces the entropy of the transferred images. The experimental evaluation of the proposed encoder [10] has shown that its performance is better than other stereoscopic image coders, since in comparison to the visual quality achieved, it can be said that the coder algorithm of low complexity.

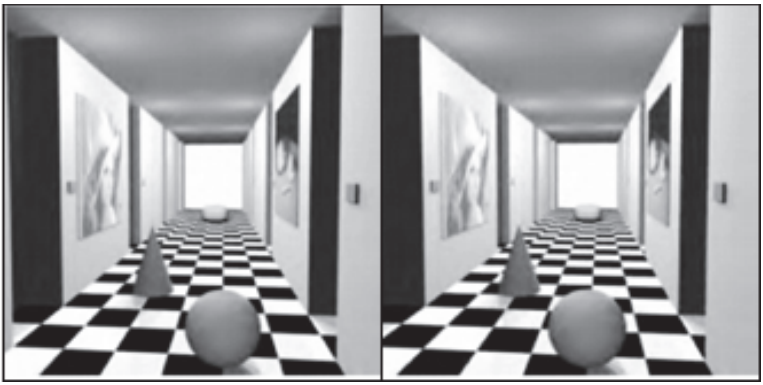


Fig. 8. Examples of stereo pairs of images (left and right) compressed with disparity compensated residual algorithm

Fig. 9. Examples of stereo pairs of images (left and right) compressed with Markov Random Field algorithm

The latest proposed technique is MRF. This algorithm obtains smooth disparate fields [24] without increasing residual energy and thus allocates fewer coding bits, and therefore takes less time to complete the algorithm in relation to the previous two suggested techniques.

Also, based on the simple visual quality of the compressed images shown in Figures 7, 8 and 9, it is can be concluded that the Dense disparity map gives the lowest quality results, while the Disparity compensated residual algorithm yields high quality results. The visual quality of the compressed images was also checked subjectively, and that is described in the next section.

5. Subjective comparison of compression techniques for stereoscopic images. In addition to objective metrics for assessing picture quality, there also exist subjective quality assessments. When the overall subjective

image quality is measured, it would be best to do tests on the entire population, but of course, this is wildly impractical. This problem can be overcome with a sampled statistical analysis. The main assumption is that this limited number of respondents is representative of the whole population. [28-30]. Accordingly, proper sampling (i.e., selection of respondents) should be ensured. Entrants participating in the testing should not have prior knowledge of the quality assessment of compressed stereoscopic images. In addition, subjects should be at the age of 18-30, because the visual system of people at this age is in optimal condition. In conclusion, the subject should be a rather young person, who will evaluate the quality of the displayed compressed images according to a subjective, personal impression.

In order to obtain statistically reliable results, the test session must be precisely carried out. Also, it is necessary to deal with practice and boredom effects. To avoid this, it is necessary to avoid showing all the test images during one test session and it is necessary to avoid showing the images in the same order as this will yield better statistical results. In the paper, the test consisted of four test images, one of which is the original image. Pairs of the pictures are presented individually to the respondents. A stereoscopic compressed image is considered as one test point. Figure 5 shows pairs of stereoscopic images seen by users. One line (a pair of images) represents one 3d image. The images have been shown in random order. Each respondent had breaks between viewing image pairs lasting five to fifteen minutes.

5.1. Description of the test procedure. In this paper, the results of image compression achieved by these techniques were shown to users on an Xwave VR display powered by a Samsung J5 2017 mobile phone. The glasses are shown in Figure 6 and the mobile phone is shown in Figure 10.

Fig. 10. Samsung J5 2017 mobile phone

The test procedure is described in Figure 11.

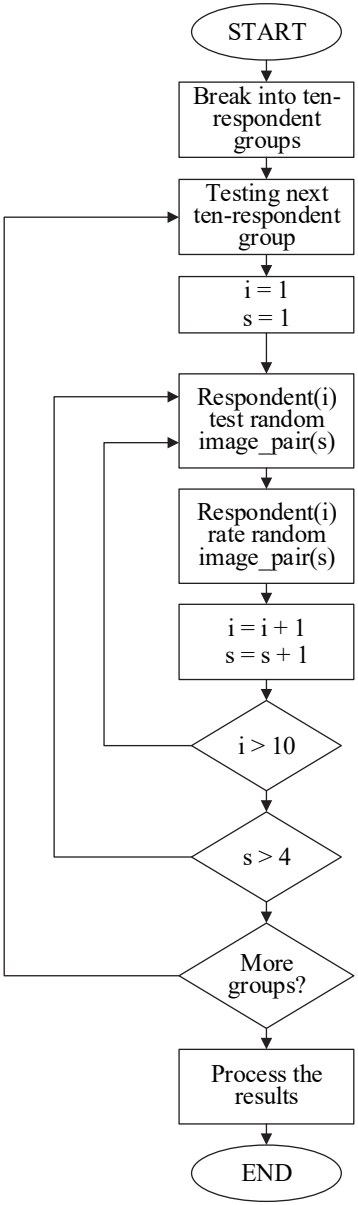


Fig. 11. Stereovision subjective test procedure

Thirty respondents participated in the test. Before the start of the testing, each of the respondents was asked to complete a survey. First, it was necessary to enter the gender, age, and field of education. Then, respondents were asked questions related to their sense of vision, which were:

– Do you have an eyeglass prescription?

– Have you been diagnosed with astigmatism?

– Have you been diagnosed with dichromatism/daltonism?

Only after they responded to all the questions in the survey, respondents were able to access the testing (Figure 12).

The respondents were divided into three groups of ten respondents. This was done in order to optimize the breaks between image pair viewing. Thus, each respondent had to wait between five to fifteen minutes. After the first respondent viewed the first image pair (which was chosen randomly), he/she had a break during which each of the rest of the nine respondents viewed their randomly chosen first image pair.

After all of the ten respondents viewed their first pair of images, the first respondent viewed the second image pair randomly chosen from the image pair he/she had not seen yet.

Fig. 12. Testing with Xwave glasses

Respondents rated the quality of the images they were shown they. The respondents rated the quality of the images they were shown. They immediately rated the quality of the image pair they viewed.

5.2. Results and discussion. As stated in the preceding section, the sample size was thirty, with a mean age of 28.6 years. The respondents were all given the same four pictures to evaluate on a five-point Likert scale. Neither the subject, nor the examiner was aware of the order in which the pictures would be shown in ensuring a double-blind experimental protocol. Given the use of a Likert scale [31], the assumption of normality is violated, especially given the relatively small n value and the heavy-tailed nature of the distribution [32] implied by treating Likert scale values on the continuous measurement level. Therefore, robust statistical methods [33] have been used to analyze the relevant data. These selected statistical methods can capture effects on a small sample. First, Figure 13 shows a bar graph of the data with 95% confidence interval error bars.

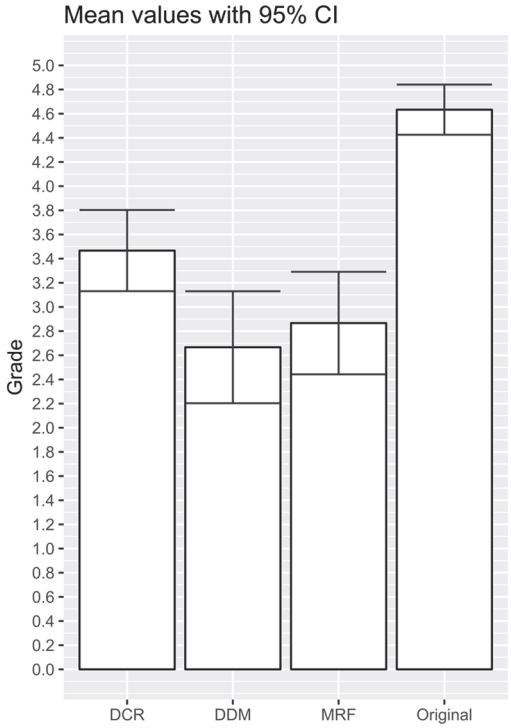


Fig. 13. Mean plot of grade values for the pictures in question. 95% error bars shown in red, n=30

While this is not a conclusive test, the error bars suggest that the means of the grades do differ between the original and the algorithms. This suggests that DCR has the superior image quality of all the algorithms but proving this may be beyond the resolution afforded by the sample.

However, this does not take into account the repeated-measures nature of the data, therefore a more careful analysis of the values is required. The first step is to perform a heteroscedastic one-way robust dependent ANOVA equivalent bootstrapped with 2000 resampling steps and 20% trimmed means. Bootstrapped statistics do not produce p-values but it should be noted that the test statistic generated is 24.1579 while the critical value is 2.9671 indicating a high degree of significance. A non-bootstrapped robust analogue produces similar values, F (2.62, 44.58) = 24.1579 with a p of 0 indicating a value too small for the system to calculate.

This permits us to reject the H_0 of the means being equal between groups which is the expected result given the nature of the graph in Figure 10. Post hoc testing is much more interesting to us, and a post-hoc analogue to the ANOVA variant used is detailed in [33] and implemented in [34] and if implemented and applied to the data produces results in Table 2.

Table 2. Results of *post-hoc* testing on mean grade values, n = 30

Comparison	$\hat{\psi}$	p	$p_{critical}$	Signifcant?
DCR vs. DDM	1.000	0.01705	0.02500	Yes
DCR vs. Original	-1.222	0.00000	0.00852	Yes
DCR vs. MRF	0.667	0.00039	0.01690	Yes
DDM vs. Original	-2.056	0.00000	0.01270	Yes
DDM vs. MRF	-0.389	0.15483	0.05000	No
Original vs. MRF	1.833	0.00000	0.01020	Yes

Given these values, it is evident that most of the individual differences are significant even after accounting for familywise error rate. The one exception is the difference between DDM and MRF algorithms whose mean grade is close enough that the difference is not significant at the level of resolution available from the present data set.

This analysis allows us to claim with a certain degree of confidence that:

– the original picture is better than the output of any of the algorithms with a p-value of 0 and an effect size of, respectively, $\hat{\xi}$=0.733, 0.818, 0.723, which according to Wilcox and Tien who defined the measure [35] is an exceptionally large effect size;

– DCR is better than any of the other algorithms with a p-value of, respectively, 0.01705, 0.01690, and an effect size of $\hat{\xi}$=0.479, 0.380, which according to [35] is a large effect.

6. Conclusion. This paper presents an overview of the compression techniques used for image compression in a stereoscopic display and analyzed the results through objective and subjective methods for assessing the quality of compressed images. As a conclusion of this analysis, it can be said that the wavelet technique of compression of pairs of stereoscopic images does not affect the minute differences between paired stereo images and thus does not affect the degree of immersion in the virtual environment created in part by the stereoscopic display, as is the case with standard image compression techniques. In the case of standard compression techniques, there is a shortcoming that sudden transitions between contrast values are not possible. It is precisely because of the discrete differences that occur between compressed and uncompressed that the 3D illusion is deteriorating. The MRF technique improves the quality of the reconstructed target images in comparison with the results that can be given by a plain technique based on the division of images into blocks.

The values from this analysis show that there are individual differences which are significant. The only exception is the difference between DDM and MRF. The difference not statistically significant on the level of the resolution available from the present data set. It shows that:

– the original picture is better than the output of any of the algorithms;

– DCR is better than any of the other algorithms.

In the future, our goal is a more detailed and more effective analysis of stereoscopic image compression techniques. A larger data set and a more involved experimental protocol are the obvious next steps and after that, the use of resulting data in the development and improvement of the new compression algorithm.

References

1. D'Angelo T. et al. Development of a Low-Cost Augmented Reality Head-Mounted Display Prototype. Examining Developments and Applications of Wearable Devices in Modern Society. 2018. pp. 1–28.
2. Ramaprabha T., Sathik M.M. Study of Performance Measurement Factors of Stereo Image Compression. *International Journal of Advanced Research in Computer Science and Software Engineering.* 2012. vol. 2. no. 7. pp. 408–410.
3. Siegel M.W., Gunatilake P., Sethuraman S., Jordan A.G. Compression of stereo image pairs and streams. Stereoscopic Displays and Virtual Reality Systems – International Society for Optics and Photonics. 1994. vol. 2177. pp. 258–269.
4. Wallace G.K. The JPEG still picture compression standard. *IEEE transactions on consumer electronics*. 1992. vol. 38. no. 1. pp. xviii–xxxiv.
5. Test Model Editing Committee. MPEG-2 Video Test Model 5. ISO/IEC JTC1/SC29/WG11 Doc N 400. 1993.

6. Test Model Editing Committee. MPEG-4 standard. MPEG 97/N1796-ISO/IEC JTC1/SC29/WG11. 1997.
7. Kumar V.V.S., Reddy M.S.I. Image compression techniques by using wavelet transform. *Journal of information engineering and applications*. 2012. vol. 2. no. 5. pp. 35–39.
8. Ramaprabha D. et al. An analytical study of image compression algorithms for stereoscopic images in non immersive virtual reality world. *International Journal of Advanced Research in Computer Science.* 2010. vol. 1. no. 4. pp. 82–86.
9. Ellinas J.N. Contribution to improvement of compression algorithms for stereoscopic images and video. Available at: http://cgi.di.uoa.gr/~phdsbook/files/OK_Ellinas.pdf (accessed: 25.08.2017).
10. Ellinas J.N., Manolis S.S. Stereo image compression using wavelet coefficients morphology. *Image and Vision Computing*. 2004. pp. 281–290.
11. Ellinas J.N., Manolis S.S. Morphological wavelet-based stereo image coders. *Journal of Visual Communication and Image Representation*. 2006. pp. 686–700.
12. Ellinas J.N., Manolis S.S. Stereo image coder based on the MRF model for disparity compensation. *EURASIP Journal on Applied Signal Processing*. 2006. vol. 2006. no. 1. pp. 35.
13. Coburn J.Q., Freeman I., Salmon J.L. A review of the capabilities of current low-cost virtual reality technology and its potential to enhance the design process. *Journal of Computing and Information Science in Engineering*. 2017. vol. 17. no. 3. pp. 031013.
14. Buń P. et al. Application of professional and low-cost head mounted devices in immersive educational application. *Procedia Computer Science*. 2015. vol. 75. pp. 173–181.
15. Anthes C. et al. State of the art of virtual reality technology. IEEE Aerospace Conference. 2016. pp. 1–19.
16. Woo W., Ortega A. Stereo image compression with disparity compensation using the MRF model. Visual Communication and Image Processing. 1996. vol. 2727. pp. 28–42.
17. Frajka T., Zeger K. Residual image coding for stereo image compression. *Optical Engineering*. 2003. vol. 42. no. 1. pp. 182–190.
18. Ellinas J.N., Manolis S.S. A novel stereo image coder based on quad-tree analysis and morphological representation of wavelet coefficients. Department of Informatics and Telecommunications. National and Kapodistrian, University of Athens. 2004. vol. 157. 84 p.
19. Brooks A.C., Zhao X., Pappas T.N. Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions. *IEEE Transactions on image processing*. 2008. vol. 17. no. 8. pp. 1261–1273.
20. Benoit A., Le Callet P., Campisi P., Cousseau R. Quality assessment of stereoscopic images. *EURASIP journal on image and video processing*. 2008. vol. 2008. no. 1. 659024 p.
21. Campisi P., Le Callet P., Marini M. Stereoscopic images quality assessment. Signal Processing Conference. 2007. pp. 2110–2114.
22. Farid M.S., Lucenteforte M., Grangetto M. Perceptual quality assessment of 3D synthesized images. IEEE International Conference on Multimedia and Expo (ICME). 2017. pp. 505–510.
23. Čanađija M. VSI metrika za objektivnu ocjenu kvalitete slike. Available at: https://repozitorij.etfos.hr/islandora/object/etfos%3A856/datastream/PDF/view (accessed: 25.08.2017).
24. Tardon L.J., Barbancho I., Alberola C. Markov random fields in the context of stereo vision. Advances in Theory and Applications of Stereo Vision. InTech. 2011. pp. 35–70.
25. Woo W., Ortega A. Stereo image compression with disparity compensation using the MRF model. Visual Communications and Image Processing'96 – International Society for Optics and Photonics. 1996. vol. 2727. pp. 28–42.
26. Yamaguchi K. et al. Continuous markov random fields for robust stereo estimation. European Conference on Computer Vision. 2012. pp. 45–58.

27. Lewandowski F., Paluszkiewicz M., Grajek T., Wegner K. Methodology for 3D Video subjective quality evaluation. *International Journal of Electronics and Telecommunications.* 2013. vol. 59. no. 1. pp. 25–32.
28. Dragan D., Petrovic V.B., Ivetic D. Methods for Assessing Still Image Compression Efficiency: PACS Example. Handbook of Research on Computational Simulation and Modeling in Engineering. 2016. pp. 389–416.
29. Dragan D., Ivetic D., Petrovic V.B. Introducing an Acceptability Metric for Image Compression in PACS-A Model. E-Health and Bioengineering Conference (EHB). 2013. pp. 1–4.
30. Dragan D. et al. An empirical study of data visualization techniques in PACS design. *Computer Science and Information Systems*. 2018. pp. 17.
31. Norman G. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*. 2010. vol. 15. no. 5. pp. 625–632.
32. Wilcox R.R. Introduction to robust estimation and hypothesis testing. Academic press. 2011. 608 p.
33. Mair P., Schoenbrodt F., Wilcox R. WRS2: Wilcox robust estimation and testing. Web site. 2016.
34. Wilcox R.R., Tian T.S. Measuring Effect Size: A Robust Heteroscedastic Approach for Two or More Groups. *Journal of Applied Statistics.* 2011. vol. 38. no. 7. pp. 1359–1368.

Vasiljevic Ivana Stojan — Ph.D. student of the Chair for Computer Graphics of Department of Fundamentals Sciences of the Faculty of Technical Sciences, University of Novi Sad. Research interests: stereoscopic virtual reality, simulations, visual effects. The number of publications — 1. ivanav@uns.ac.rs; 6, Trg Dositeja Obradovića, 21000, Novi Sad, Serbia; office phone: +381-21-485-2297.

Dragan Dinu — Ph.D., assistant professor of Department of Computing and Control Engineering of the Faculty of Technical Sciences, University of Novi Sad. Research interests: data compression, computer graphics, human computer interaction, data visualization, and computer vision. The number of publications — 48. dinud@uns.ac.rs, http://www.ftn.uns.ac.rs; 6, Trg Dositeja Obradovića, 21000, Novi Sad, Serbia; office phone: +381-21-485-2129.

Obradovic Ratko — Ph.D., professor, head of the Chair for Computer Graphics of Department of Fundamentals Sciences of the Faculty of Technical Sciences, University of Novi Sad, head of Computer Graphics - Engineering Animation Studies of the Faculty of Technical Sciences, University of Novi Sad. Research interests: computer graphics, computational geometry, computer animation, CAD, scientific visualization, simulations, virtual and augmented reality, higher education. The number of publications — 80. obrad_r@uns.ac.rs, http://www.ftn.uns.ac.rs/n1062642046/ratko-obradovic; 6, Trg Dositeja Obradovića, 21000, Novi Sad, Serbia; office phone: +381-64-200-1125.

Petrović Veljko Branko — Ph.D. student of Department of Computing and Control Engineering of the Faculty of Technical Sciences, University of Novi Sad. Research interests: visualization, human-computer interaction, large-scale high-performance computing. The number of publications — 8. pveljko@uns.ac.rs, http://www.ftn.uns.ac.rs; 6, Trg Dositeja Obradovića, 21000, Novi Sad, Serbia; office phone: +381-21-485-2415.

Acknowledgements. This research is partially supported by the Ministry of Science and Technological Development of the Republic of Serbia (project TR32044).

УДК 004.5 **DOI** 10.15622/sp.61.8

И.С. Васильевич, Д. Драган, Р. Обрадович, В.Б. Петрович

АНАЛИЗ МЕТОДОВ СЖАТИЯ СТЕРЕОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ

Васильевич И.С., Драган Д., Обрадович Р., Петрович В.Б. **Анализ методов сжатия стереоскопических изображений.**

Аннотация. В последние годы стали появляться видеошлемы виртуальной и дополненной реальностей, охватывая все больше отраслей применения. Видеошлемы обычно используются для развлечений, социального взаимодействия, образования, но вместе с тем увеличивается процент пользователей, которые применяют их для работы в таких областях, как медицина, моделирование и симуляция. Несмотря на то, что было выпущено множество видов видеошлемов, две основные проблемы препятствуют их повсеместному внедрению на основной рынок: чрезвычайно высокая стоимость и недостатки пользовательского интерфейса. Эффект трехмерного изображения в видеошлемах достигается с помощью стереоскопического изображения. Обработка и скорость передачи по сети является узким местом при работе со стереоскопическим изображением. Поэтому необходимы эффективные методы сжатия изображений. Стандартные методы сжатия не подходят для стереоскопических изображений из-за различий между сжатыми и несжатыми изображениями. Проблема в том, что потери в алгоритме сжатия изображений создает мелкие различия, которые в дальнейшем будут влиять на восприятие человеком конечного трехмерного образа. Методы сжатия стереоизображений, которые можно найти в литературе, используют дискретное вейвлет-преобразование и алгоритм морфологического сжатия, применяемый к коэффициентам преобразования.

В статье представлен обзор и сравнение доступных методов сжатия стереоскопических изображений. На основе проведенного анализа было выявлено, что до сих пор не существует метода, который по всем критериям считается наилучшим. Качество методов проверяется пользователями видеошлемов. Настоящее исследование ориентировано на недорогие доступные видеошлемы потребительского уровня.

Ключевые слова: сжатие изображений, стереоскопические, всплески, видеошлем.

Литература

1. *D'Angelo T. et al.* Development of a Low-Cost Augmented Reality Head-Mounted Display Prototype // Examining Developments and Applications of Wearable Devices in Modern Society. 2018. pp. 1–28.
2. *Ramaprabha T., Sathik M.M.* Study of Performance Measurement Factors of Stereo Image Compression // International Journal of Advanced Research in Computer Science and Software Engineering. 2012. vol. 2. no. 7. pp. 408–410.
3. *Siegel M.W., Gunatilake P., Sethuraman S., Jordan A.G.* Compression of stereo image pairs and streams // Stereoscopic Displays and Virtual Reality Systems – International Society for Optics and Photonics. 1994. vol. 2177. pp. 258–269.
4. *Wallace G.K.* The JPEG still picture compression standard // IEEE transactions on consumer electronics. 1992. vol. 38. no. 1. pp. xviii–xxxiv.
5. Test Model Editing Committee. MPEG-2 Video Test Model 5 // ISO/IEC JTC1/SC29/WG11 Doc N 400. 1993.
6. Test Model Editing Committee. MPEG-4 standard // MPEG 97/N1796-ISO/IEC JTC1/SC29/WG11. 1997.
7. *Kumar V.V.S., Reddy M.S.I.* Image compression techniques by using wavelet transform // Journal of information engineering and applications. 2012. vol. 2. no. 5. pp. 35–39.

8. *Ramaprabha D. et al.* An analytical study of image compression algorithms for stereoscopic images in non immersive virtual reality world // International Journal of Advanced Research in Computer Science. 2010. vol. 1. no. 4. pp. 82–86.
9. *Ellinas J.N.* Contribution to improvement of compression algorithms for stereoscopic images and video. URL: http://cgi.di.uoa.gr/~phdsbook/files/OK_Ellinas.pdf (дата обращения: 25.08.2017).
10. *Ellinas J.N., Manolis S.S.* Stereo image compression using wavelet coefficients morphology // Image and Vision Computing. 2004. pp. 281–290.
11. *Ellinas J.N., Manolis S.S.* Morphological wavelet-based stereo image coders // Journal of Visual Communication and Image Representation. 2006. pp. 686–700.
12. *Ellinas J.N., Manolis S.S.* Stereo image coder based on the MRF model for disparity compensation // EURASIP Journal on Applied Signal Processing. 2006. vol. 2006. no. 1. pp. 35.
13. *Coburn J.Q., Freeman I., Salmon J.L.* A review of the capabilities of current low-cost virtual reality technology and its potential to enhance the design process // Journal of Computing and Information Science in Engineering. 2017. vol. 17. no. 3. pp. 031013.
14. *Buń P. et al.* Application of professional and low-cost head mounted devices in immersive educational application // Procedia Computer Science. 2015. vol. 75. pp. 173–181.
15. *Anthes C. et al.* State of the art of virtual reality technology // IEEE Aerospace Conference. 2016. pp. 1–19.
16. *Woo W., Ortega A.* Stereo image compression with disparity compensation using the MRF model // Visual Communication and Image Processing. 1996. vol. 2727. pp. 28–42.
17. *Frajka T., Zeger K.* Residual image coding for stereo image compression // Optical Engineering. 2003. vol. 42. no. 1. pp. 182–190.
18. *Ellinas J.N., Manolis S.S.* A novel stereo image coder based on quad-tree analysis and morphological representation of wavelet coefficients // Department of Informatics and Telecommunications. National and Kapodistrian, University of Athens. 2004. vol. 157. 84 p.
19. *Brooks A.C., Zhao X., Pappas T.N.* Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions // IEEE Transactions on image processing. 2008. vol. 17. no. 8. pp. 1261–1273.
20. *Benoit A., Le Callet P., Campisi P., Cousseau R.* Quality assessment of stereoscopic images // EURASIP journal on image and video processing. 2008. vol. 2008. no. 1. 659024 p.
21. *Campisi P., Le Callet P., Marini M.* Stereoscopic images quality assessment // Signal Processing Conference. 2007. pp. 2110–2114.
22. *Farid M.S., Lucenteforte M., Grangetto M.* Perceptual quality assessment of 3D synthesized images // IEEE International Conference on Multimedia and Expo (ICME). 2017. pp. 505–510.
23. *Čanađija M.* VSI metrika za objektivnu ocjenu kvalitete slike. URL: https://repozitorij.etfos.hr/islandora/object/etfos%3A856/datastream/PDF/view (дата обращения: 25.08.2017).
24. *Tardon L.J., Barbancho I., Alberola C.* Markov random fields in the context of stereo vision // Advances in Theory and Applications of Stereo Vision. InTech. 2011. pp. 35–70.
25. *Woo W., Ortega A.* Stereo image compression with disparity compensation using the MRF model // Visual Communications and Image Processing'96 – International Society for Optics and Photonics. 1996. vol. 2727. pp. 28–42.
26. *Yamaguchi K. et al.* Continuous markov random fields for robust stereo estimation // European Conference on Computer Vision. 2012. pp. 45–58.
27. *Lewandowski F., Paluszkiewicz M., Grajek T., Wegner K.* Methodology for 3D Video subjective quality evaluation // International Journal of Electronics and Telecommunications. 2013. vol. 59. no. 1. pp. 25–32.
28. *Dragan D., Petrovic V.B., Ivetic D.* Methods for Assessing Still Image Compression Efficiency: PACS Example // Handbook of Research on Computational Simulation and Modeling in Engineering. 2016. pp. 389–416.

29. *Dragan D., Ivetic D., Petrovic V.B.* Introducing an Acceptability Metric for Image Compression in PACS-A Model // E-Health and Bioengineering Conference (EHB). 2013. pp. 1–4.
30. *Dragan D. et al.* An empirical study of data visualization techniques in PACS design // Computer Science and Information Systems. 2018. pp. 17.
31. *Norman G.* Likert scales, levels of measurement and the "laws" of statistics // *Advances in health sciences education*. 2010. vol. 15. no. 5. pp. 625–632.
32. *Wilcox R.R.* Introduction to robust estimation and hypothesis testing // Academic press. 2011. 608 p.
33. *Mair P., Schoenbrodt F., Wilcox R*. WRS2: Wilcox robust estimation and testing // Web site. 2016.
34. *Wilcox R.R., Tian T.S.* Measuring Effect Size: A Robust Heteroscedastic Approach for Two or More Groups // Journal of Applied Statistics. 2011. vol. 38. no. 7. pp. 1359–1368.

Васильевич Ивана Стоян — аспирант кафедры компьютерной графики отдела фундаментальных наук факультета технических наук, Нови-Садский университет. Область научных интересов: стереоскопическая виртуальная реальность, имитационное моделирование, визуальные эффекты. Число научных публикаций — 1. ivanav@uns.ac.rs; Трг Доситейа Обрадовича, 6, 21000, Нови-Сад, Сербия; р.т.: +381-21-485-2297.

Драган Дину — к-т техн. наук, доцент отдела вычислительной техники и управления факультета технических наук, Нови-Садский университет. Область научных интересов: сжатие данных, компьютерная графика, человеко-машинное взаимодействие, визуализация данных, компьютерное зрение. Число научных публикаций — 48. dinud@uns.ac.rs, http://www.ftn.uns.ac.rs; Трг Доситейа Обрадовича, 6, 21000, Нови-Сад, Сербия; р.т.: +381-21-485-2129.

Обрадович Ратко — к-т техн. наук, профессор, заведующий кафедрой компьютерной графики отдела фундаментальных наук факультета технических наук, Нови-Садский университет, руководитель отдела компьютерной графики - инженерной анимации факультета технических наук, Нови-Садский университет. Область научных интересов: компьютерные графика, вычислительная геометрия, компьютерная анимация, система автоматизированного проектирования, визуализация научных данных, имитационное моделирование, виртуальная и дополненная реальностью, высшее образование. Число научных публикаций — 80. obrad_r@uns.ac.rs, http://www.ftn.uns.ac.rs/n1062642046/ratko-obradovic; Трг Доситейа Обрадовича, 6, 21000, Нови-Сад, Сербия; р.т.: +381-64-200-1125.

Петрович Велько Бранко — аспирант отдела вычислительной техники и управления факультета технических наук, Нови-Садский университет. Область научных интересов: визуализация, человеко-машинное взаимодействие, крупномасштабные высокопроизводительные вычисления. Число научных публикаций — 8. pveljko@uns.ac.rs, http://www.ftn.uns.ac.rs; Трг Доситейа Обрадовича, 6, 21000, Нови-Сад, Сербия; р.т.: +381-21-485-2415.

Поддержка исследований. Работа выполнена при частичной финансовой поддержке Министерством науки и технологического развития Республики Сербия (проект TR32044).

Signed to print 26.11.2018
Printed in Publishing center GUAP, 67, B. Morskaya, St. Petersburg, 190000, Russia

The journal is registered in Russian Federal Agency for Communications
and Mass-Media Supervision, certificate ПИ № ФС77-41695 dated August 19, 2010 г.
Subscription Index П5513, Russian Post Catalog

Подписано к печати 26.11.2018. Формат 60×90 1/16. Усл. печ. л. 13,56. Заказ № 534.
Тираж 150 экз., цена свободная.
Отпечатано в Редакционно-издательском центре ГУАП, 190000, Санкт-Петербург, Б. Морская, д. 67

Журнал зарегистрирован Федеральной службой по надзору в сфере связи
и массовых коммуникаций,
свидетельство ПИ № ФС77-41695 от 19 августа 2010 г.

Подписной индекс П5513 по каталогу «Почта России»

РУКОВОДСТВО ДЛЯ АВТОРОВ

Взаимодействие автора с редакцией осуществляется через личный кабинет на сайте журнала «Труды СПИИРАН» http://www.proceedings.spiiras.nw.ru. При регистрации авторам рекомендуется заполнить все предложенные поля данных.

Подготовка статьи ведется с помощью текстовых редакторов MS Word 2007 и выше. Объем основного текста – от 20 до 30 страниц включительно. Формат страницы документа – A5 (148 мм ширина, 210 мм высота); ориентация – портретная; все поля – 20 мм. Верхний и нижний колонтитулы страницы – пустые. Основной шрифт документа – Times New Roman, основной кегль (размер) шрифта – 10 pt. Переносы разрешены. Абзацный отступ устанавливается размером в 10 мм. Межстрочный интервал – одинарный. Номера страниц не проставляются.

В основную часть допускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно рассмотрены на нашем сайте в разделе «Руководство для авторов».

AUTHOR GUIDELINES

Interaction between each potential author and the Editorial board is realized through the personal account on the website of the journal "Proceedings of SPIIRAS" http://www.proceedings.spiiras.nw.ru. At the registration the authors are requested to fill out all data fields in the proposed form.

The submissions should be prepared using MS Word 2007 text editor or higher versions, at that, only manuscripts in *.docx format will be considered. The text of the paper in the main part of it should be from 20 – 30 pages of A5 size that is 210 X 148 mm; orientation – portrait; all margins – 20 mm. The font of the main paper text is Times New Roman of 10 pt font size. The pages' headers and footers should be empty; indention – 10 mm; line spacing – single; pages are not numbered; hyphenations are allowed.

Certain figures, tables, listings and formulas are allowed in the main section, and their typography is considered by the paper template in more detail in journal web.

ISSN 2078-9181

Вып. 6(61)

ТРУДЫ СПИИРАН
proceedings.spiiras.nw.ru

С-Пб
2018