НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ
ИНСТИТУТ ПРОБЛЕМ РЕГИСТРАЦИИ ИНФОРМАЦИИ НАН УКРАИНЫ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И БЕЗОПАСНОСТЬ

МАТЕРИАЛЫ XIX МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ

ВЫПУСК 19

Киев – 2019

Рекомендовано к печати Ученым советом
Института проблем регистрации информации НАН Украины
(протокол № 3 от 24 декабря 2019 г.)

Информационные технологии и безопасность. Материалы XIX Международной научно-практической конференции ИТБ-2019. – К.: ООО "Инжиниринг", 2019. – 236 с. ISBN 978-966-2344-72-1

В сборник вошли материалы докладов, представленных на XVI Международной научно-практической конференции «Информационные технологии и безопасность» (ИТБ-2019, 28 ноября 2019 года, г. Киев, Украина).

В сборнике представлены статьи, посвященные вопросам безопасности живучести критических инфраструктур, моделирования и противодействия информационным операциям, информационных технологий в управлении, методов и способов информационной поддержки принятия решений, комп'ютерного моделирования систем организационного управления, информационно-аналитических исследований на основе открытых источников информации, сценарного анализа при обеспечения информационной поддержки принятия решений, актуальным проблемам обеспечения информационной и кибернетической безопасности.

Для специалистов в области информационных технологий, информационной безопасности, информационного права а также для аспирантов и студентов старших курсов высшей школы соответствующих специальностей.

ISBN 978-966-2344-72-1

SURVIVABILITY OF ORGANIZATIONAL MANAGEMENT SYSTEMS AND THE MAINTENANCE OF CRITICAL INFRASTRUCTURE SECURITY

Oleksandr G. Dodonov, Olena S. Gorbachyk,
Maryna G. Kuznietsova
Institute for Information Recording of the National Academy of Sciences of Ukraine, Kyiv, Ukraine
dodonov@ipri.kiev.ua, ges@ipri.kiev.ua, margle@ipri.kiev.ua

Abstract. The paper is devoted to the issue of improving the security of critical infrastructures functioning using the capabilities of their automated organizational management systems. It's substantiated, that security of critical infrastructures functioning depends on the level of survivability of automated organizational management systems (OMS). Increasing the survivability of automated organizational management systems is an essential element of a secure risk management system for critical infrastructures. The survivability property of automated OMS is defined as their ability to retain their functionality by performing the set of functions necessary to achieve the goal of functioning with a given quality, in the context of accumulation of component damages and loss of resources, by changing the behavior of the system. The survivability states of automated OMS are classified. A model is proposed to investigate the survivability of an automated OMS regarding a set of functions aimed at ensuring the security of critical infrastructure functioning. The methodological aspects of the development and implementation of the automated OMS are highlighted, that will function in the conditions of permanent changes of the environment and modernization of the components of the OMS. The criteria of estimation of system qualities of OMS and its components - automated workplaces are offered. An integrated survivability index has been proposed to evaluate the survivability and functional degradation of the OMS. Time constraints for the fulfillment the procedures for building information infrastructure in the OMS are formulated, to ensure the implementation of the functions supporting the critical infrastructure security. The expediency of creation a specialized modeling complex for OMS automation for the development of basic system, design and technological solutions, development of management decisions for the basic processes of organizational management is substantiated. At the specialized modeling complex it is possible to analyze and improve the existing methods of maintaining the security of critical infrastructure functioning, to develop templates for managers' actions in the event of undesirable changes during the critical infrastructure functioning, occurrence and development of emergency situations on the objects of critical infrastructure.

Keywords: survivability, security, critical infrastructure, automated organizational management system.

1 Introduction

Ensuring the security of critical infrastructures is, first and foremost, a reduction to an acceptable level of risk of harm to the environment, the individual, society, and the country. Critical infrastructure objects must be guaranteed to maintain a certain level of security, avoid emergency situations, prevent their transition to dangerous conditions.

Low survivability systems collapse quickly and this can lead to cascading accidents and significant material losses, while systems with high survivability break down gradually, retaining in part functionality, limited performance, and time-consuming adoption for switching to safe mode of operation, emergency shutdown, isolation of damage, preventing their spread, etc. An important part of the security risk management system of critical infrastructures is to increase the survivability of organizational management systems.

Today, all chains of control of critical objects and infrastructures involve complexes of hardware and software, information systems and telecommunication networks, intended to support the solution or solve the problems of operational management and control over various processes and technical objects within the organization of production or technological processes. Computer tools have become an integral part of various management systems, components of complex technical, administrative, economic and other systems of regional and global scale. Computer systems and technologies provide advanced communication tools, support a complex structure of resources, production and management processes automation, various technologies for information processing. Security of management objects depends on the technologies of automated organizational management systems, which are complex sociotechnical systems, that include technical and technological subsystems, relevant systems of activity (systems of roles and functions of service and management personnel), an environment that actively interacts with others compound.

Effective use of the sociotechnical systems components capabilities, taking into account the synergistic effect allows ensure the functional safety of critical infrastructures, to reduce the risk of accidents.

2 Critical infrastructure security dependence on the survivability of organizational management systems

Critical infrastructure security will mean the independence from unacceptable risk [1], that is, the ability of infrastructure, as a system, to minimize the risks of disaster.

Automated organizational management systems (OMS) for critical infrastructures should ensure not only the proper functioning of the infrastructure and the desired end result, but also the adequate response to security incidents occurring at critical infrastructure objects [2]. Tools of the automated OMS should ensure timely recognition of the threat and the moment of occurrence of a critical (emergency) situation, determine an adequate level of their processing, initiate processes of counteraction, compensation or adaptation to continue the functioning of the critical infrastructure in full or in part, and, if necessary, procedures for slow gradual degradation and safe shutdown must be activated.

An analysis of the current trends in the development of automated OMS shows that there is an increasing number of functions that are critical to the security of critical infrastructures that rely on information and communication technologies and computers. Even today, thanks to automation, the implementation of organizational management processes occurs in such a way as to prevent the transition of infrastructure or its components into potentially dangerous states [2-4]. As a rule, the shutdown of a technical object in case of occurrence or realization of a threat of its transition to a dangerous (emergency) state is automatic. Intelligent software products for predicting, assessing and minimizing the security risks of critical objects and structures are available to support and develop management solutions at various levels.

Under the survivability of an organizational management system, we will understand its ability to retain its functionality by performing the set of functions necessary to achieve the goal of functioning with a given quality, in the context of accumulation of component damage and loss of resources, by changing the behavior of the system.

In the general case, the survivability of the OMS depends on the set of parameters that characterize the system, the functions performed by it, the effects of the environment and the type, extent and dynamics of interaction with it. If the OMS is in a "status of survivability," the system performs the set of functions $\phi = (\phi_1, \phi_2, ..., \phi_n)$ with the specified quality and required efficiency, that is, the purpose of the function is achieved. The “status of survivability” is characterized by the stability and predictability of the functioning of the system, that is, the critical infrastructure OMS fulfills all managerial functions.

There are three types of system survivability status $|S| = \{|S|_t\}, t = 1, 2, 3$, to which the OMS can go, namely [2, 5]:

- system survivability status type $|S|_1$, in which the OMS provides all the functions of the set $\phi = (\phi_1, \phi_2, ..., \phi_n)$ with given quality and required efficiency or with poor quality and less efficiency in any of the states $w_j \in W$, that is

$$\prod_{i \in I} x(\phi_i) = 1, \qquad x(\phi_i) = \begin{cases} 1, \textit{ if } \phi_i \textit{ is fulfilled} \\ 0, \textit{ otherwise} \end{cases}$$

- system survivability status type $|S|_2$, whereby only a certain subset of the functions are provided by the OMS $\phi^* \subset \phi$ in any of the states $w_j \in W$, that is

$$\prod_{\phi_i \in \phi^*} x(\phi_i) = 1$$

- system survivability status type $|S|_3$, whereby at least one of the functions of the set is performed in the OMS $\phi = (\phi_1, \phi_2, ..., \phi_n)$ in any of the states $w_j \in W$, that is

$$\sum_{i \in I} x(\phi_i) \geq 1.$$

Transition of the OMS into "status of survivability" types $|S|_2$ and $|S|_3$ means that there are violations in the functioning of the system (functional failures of components or "wrong actions" of managers), and there is a narrowing of the functionality of the OMS.

Among the functions of the set $\phi = (\phi_1, \phi_2, ..., \phi_n)$ identify the functions of the OMS, aimed at

ensuring the security of critical infrastructure, $\phi^S \subset \phi,\ \phi^S = (\phi_1^s, \phi_2^s, ..., \phi_r^s)$. Functions from the set ϕ^S can be both independent and information related. The ability of the OMS to maintain the secure functioning of the critical infrastructure can be characterized by the OMS 's survivability with respect to a set of functions ϕ^S.

In automated OMS, the automated workstations of managers (AWM) are functional components. Each AWM is a subsystem, the structure of which is determined by its functional purpose. AWM specialization is done by installing the appropriate software and establishing links between system components. The modular principle of software development makes it quite easy to form the required configuration of the AWM, as a subsystem of the OMS, to perform certain management functions. AWM functionality can be expanded as needed by connecting new software modules. This creates a flexible scalable environment for implementing management functions.

To study the survivability of automated OMS with respect to a set of functions ϕ^S, to ensure the critical functioning of the critical infrastructure, the following model can be applied:

$$\overline{\Im} = \left\langle G, \phi^S, \mathrm{Tm}, \mathrm{Can}, \mathrm{Tt} \right\rangle,$$

where G – a graph that describes the information and communication links in the OMS and may be changed during the operation of the OMS or in the case of changing the functionality of the individual AWM; $\phi^S = (\phi_1^s, \phi_2^s, ..., \phi_r^s)$ – a set of functions implemented in the OMS to maintain the security of critical infrastructure, Tm – some of the time-deficient functions of the last argument, Can – a matrix of functionalities of the totality of AWM, which actually represent an automated OMS; Tt – vector that characterizes the load of the AWM.

Let us denote the set of managerial tasks performed in the OMS of the implementation of the set of functions $\phi^S = (\phi_1^s, \phi_2^s, ..., \phi_r^s)$ with the required quality and the required efficiency through

$$F = \bigcup_{i \in I} F_i = \{ f_1, f_2, \ldots, f_n \},$$

while the AWM Φ_k can potentially perform a number of managerial tasks $\varphi_н : \{1, 2, \ldots, p\} \to P(F)$, where $P(F)$ – the set of all subsets F.

If $\varphi_н(k) = \left\{ f_{i_1}, f_{i_2}, \ldots, f_{i_j} \right\}$, $1 \le i_r \le n$, then the functional component of the AWM Φ_k can perform managerial tasks $f_{i_1}, f_{i_2}, \ldots, f_{i_j}$.

Suppose that all the necessary information and communication links between the AWM of the OMS to perform the security support functions of the critical infrastructure can be provided.

At each specific moment of time specific AWM Φ_k implements a subset of managerial tasks $\varphi_{men} : \{1, 2, \ldots, p\} \to P(F)$.

At the AWM of the OMS the decision of managerial tasks $f_1, f_2, \ldots, f_n$ is supported, if $\varphi_{men}(k) = \left\{ f_{i_1}, f_{i_2}, ..., f_{i_j} \right\}$. If $\varphi_{men}(k) = \varnothing$, than AWM of the OMS Φ_k does not perform managerial tasks, the solution of which requires the implementation of functions ϕ^S, which support the security operation of critical infrastructure.

Assuming every managerial task $f_i \in F$ is characterized by some performance efficiency c_i, you can define the performance function for the entire automated OMS on the performance of functions ϕ^S which support the security operation of critical infrastructure:

$\psi_{еф} : F \times \{1, 2, ..., p\} \times P(F) \to C$, where C is a certain number set.

If the AWM of the OMS Φ_k is focused on managerial tasks $\varphi_{men}(k) = \left\{ f_{i_1}, f_{i_2}, ..., f_{i_j} \right\}$ and performance when fulfilling $f_i \in \left\{ f_{i_1}, f_{i_2}, ..., f_{i_j} \right\}$ is equal to c_{i_k}, than: $\psi_{еф}\left(f_i, k, \varphi_{men}(k) \right) = c_{i_k}$.

To implement by automated OMS the functions ϕ^S, which support the security operation of critical infrastructure, with the efficiency not lower than the specified, the managerial tasks $f_1, f_2, \ldots, f_n$ must be performed with appropriate efficiency, that is, the following conditions must be met [5]:

$$\bigcup_{k=1}^{p} \varphi_{н}(k) \supseteq F\,, \tag{1}$$

$$\varphi_{men}(k) \subseteq \varphi_{н}(k), \quad \forall k = \overline{1,p} \tag{2}$$

$$\sum_{k=1}^{p} \psi_{еф}\left(f_i, k, \psi_{men}(k)\right) \geq c_i\,,\ \forall i = \overline{1,n}\,, \tag{3}$$

Let us define as a functional failure the impossibility of fulfilling at least one of the managerial tasks in the OMS $f_i \in F$. In the case of functional failure, the status of some AWM Φ_k changes and the corresponding function φ_{men} also changes. Although condition (2) cannot be violated by functional failure AWM Φ_k, but its violation may be caused by errors in management. If the narrowing of the functionality leads to a violation of conditions (1) - (3), then the means of ensuring the survivability of the OMS must be activated and the system must be adjusted so that conditions (1) - (3) are again fulfilled.

When reconfiguring the OMS, it is advisable to minimize the number of AWM of the OMS involved in failure compensation procedures, i.e. to minimize the number of changes φ_{men}. It is because of the conditions (1) - (3) that the minimum quantity of φ_{men} is changed, the optimality of OMS behavior can be characterized, and the number of compensated functional failures may serve as a criterion for system survivability.

3 Development and implementation automated high-survivability OMS for critical infrastructure

The problem of ensuring the safety of the functioning of critical objects and infrastructures is complex, but the quality and properties of their automated OMS are of utmost importance, since the security and safety of critical infrastructures depends on management decisions, especially in the event of an emergency situation, i.e. in conditions when there is no possibility of a clear prediction of the results of management impacts. Functional stability of the OMS itself becomes a factor and condition for security and safety of critical infrastructure objects.

Already at the initial stage of development and implementation of automated OMS, it is necessary to define criteria for assessing systemic qualities, in particular, [6, 7]:

- criteria for compliance of the OMS and the individual AWM with the specified indicators of quality of functioning and/or assessment of the degree of its functional degradation;
- criteria for evaluating the performance of dynamic reconfiguration and reallocation of resources;
- criteria for assessing the extent of system recovery after glitches and failures due to mechanisms of reorganization or reconstruction;
- criteria that characterize changes in performance, reactivity, system sensitivity in the conditions of system resources degradation;
- criteria for assessing the adaptability of the system to external and internal changes;
- cost-effectiveness criteria for the use of available resources.

Analyzing the survivability of an automated OMS that operates in a constantly changing external environment and often undergoes modernization, one can obtain the most objective and adequate indicator of the quality of its functioning, because it is in the study of survivability that the system's ability to perform its functions over a long period is revealed, and not just the possibility of continuation of functioning upon recovery after individual glitches or failures. Quantitative assessment of survivability is generally performed on the basis of specific metrics that characterize the loss of functionality (functional degradation) over a certain period of time. Various methodological approaches to calculating such metrics are possible, in particular through quantitative assessments of the system's ability to perform mission-critical functions, through the degree of system degradation, and the like. For example, for the analysis of survivability and the assessment of functional degradation, it is possible to use the integral survivability index, determined by the weighted average of the estimates of performance indicators in the following form [7]:

$$\xi = \frac{1}{N}\sum_{j=1}^{N} z_j(r)\,,$$

where the values of the normalized indicators $z_j(r),\ j = \overline{1,N}$ are calculated as

$$z_j(r)=\begin{cases} a_j\dfrac{q_j^*(r)-q_j^{TB}}{q_j^{TB}},\ j=\overline{1,l},\ \textit{for technical requirements (TR) of the form } q_j \ge q_j^{TB} \\ a_j\dfrac{q_j^{TB}(r)-q_j^*}{q_j^{TB}},\ j=\overline{l+1,N}\ \textit{for TR of the form } q_j \le q_j^{TB} \end{cases}$$

where a_j – weighting factor characterizing the degree of significance of the j-th index of survivability for the integrated assessment of the quality of functioning of the automated OMS; r - the number of accumulated functional failures in the OMS over a given period of time (taking into account the recovery); $q_j \subset Q=\{q_1,q_2,...q_S\}$ – element of a set of metrics that should be within the appropriate range, which are defined by the technical requirements, which are formulated, as a rule, in the terms of reference for the development of automated OMS; $q_j^*(r)$ – the "worst" in understanding the fulfillment of the terms of reference of the j-th indicator value of the quality of functioning when r components failure accumulated in the system.

If there is a restriction for all given survival rates for a given period of time $q_j \ge q_j^{TB}$ or $q_j \le q_j^{TB}$, $j=\overline{1,N}$, then $\min\limits_j z_j \ge 0,\quad j=\overline{1,N}$, and, accordingly, the value of the integral index ξ will be no lower than some critical lower limit ξ_{kp}, the specific value of which may be specified when determining the functionality of the system for a certain period of operation, or as the initial value of the integral index ξ. In this case, a quantitative assessment of the degree of degradation of the system's capabilities may be, for example, the value:

$$S_d=\frac{\xi_{n.}-\xi_{nom.}}{\xi_{n.}}\times 100\%=\frac{\xi_{вmp.}}{\xi_{n.}}\times 100\%\ ,$$

where $\xi_{n.}$ – quantitative assessment of the initial (design) functionality of an automated OMS, taking into account the weighting coefficients of the significance of the survivability indicators; a $\xi_{nom.}$, $\xi_{вmp.}$ – quantification for current (existing) and lost OMS functionality, respectively.

The automation of the OMS involves the introduction of information and communication technologies for the collection, processing, accumulation, systematization, storage, retrieval and dissemination of information [8]. The functioning of an automated OMS can be modeled with help of network model, the nodes of which are the functional components (AWM), and the arcs are the different communication channels (wired, wireless, combined).

The implementation of information and communication technology is ensured by the parallel and sequential operation of a set of functional components that interact with each other and with the external environment through communication channels. Technology should provide the development of management decision, for example, over time T_z, which does not exceed $T_{\partial on}$ (the maximum time allowed for the collection and processing of information, which depends on the requirements of the subject area). Therefore,

$$T_z=(T_f+T_{o\delta})\le T_{\partial on},$$

where T_f – time spent for processing information by functional components, $T_{o\delta}$ – time spent on information interaction.

In the case of undesirable influences on the system or communication channels, the time for implementation of information technology to develop a management decision may increase by $T_{\partial o\partial}$. The survivability criterion of the OMS may be the feasibility of building the necessary information infrastructure as a set of functional components and communication channels under restrictions

$$(T_f+T_{o\delta}+T_{\partial o\partial})\le T_{\partial on}.$$

For comparison of different variants of automated OMS implementation for the purpose of choosing the most functionally sustainable one can use the survivability index - the number of information infrastructures

that allow to implement information technology of management decision making, reducing the risks of occurrence and development of emergency situations in critical infrastructures.

The practical experience of designing and implementing automated OMS shows that all basic system, design, software and technological solutions for the created OMS should be pre-tested, and the managers have to go through the re-education and training stage. It is advisable to make adjustments and approbations of the AWM on the specialized modeling complex. The architecture of the complex depends on the features of the critical infrastructure, systems models and processes involved in the operation of the objects and infrastructure as a whole. Each management task that arises in the operation of critical infrastructure can be reproduced in the modeling complex as a separate functional task, the execution of which in the OMS generates a separate management process. The input for this process can be either the initial management impact, or the output or intermediate data of some other management process. The execution of the management process involves the preparation and development of decisions on the actions ordering, necessary to perform a functional task, into a certain sequence of operations implemented within the relevant technology, determining what people (employees), at what time, what technological processes (operations) perform to ensure the secure functioning of the critical infrastructure. The implementation of the technological process requires not only the specialists with the appropriate level of qualification, but also the technical means, techniques and instructions for their application, software, information and other services, necessary and sufficient for the fulfillment of the functional tasks of management. [5].

On the specialized modeling complex could not only be worked out the basic processes of organizational management, but also carried out the selection and testing of practically suitable formalized methods of maintaining the security of critical infrastructure functioning with high promptness of justified management decisions development, clarity of results of management and taking into account the existing system-based subordination and interaction in OMS.

Traditionally in the process of working out and making decisions, managers use "subjective" knowledge of certain events, informal experience of experts, who are involved in the assessment of the current situation in the critical infrastructure.

When working on the specialized modeling complex of basic processes of organizational management, the transition from intuitive estimates to quantitative is possible, that significantly objectifies management decisions and helps to improve their quality. Preliminary analysis of emergency situations on the modeling complex, crashes in the operation of critical infrastructure and erroneous management decisions allows create specific templates of managers' actions, which is important in the conditions of time resource criticality in accidents at the management object. [8].

In the future, the specialized modeling complex can become an analytical resource for the OMS, its toolkit can be used to develop strategic management decisions and substantiate current management decisions.

4 Conclusions and recommendations

The use of automated high-survivability OMS in critical infrastructures will reduce the risks of infrastructure transition to disaster states, as such OMS are guaranteed to perform their functions over a long period in the face of permanent environmental change and many frequent upgrades.

The analytical resource developed during the AWM OMS creation, formalized methods of maintaining the security of the functioning of critical infrastructures will allow to increase the efficiency and validity of management decisions, the clarity of the results of management even in the context of unwanted changes in the system of subordination and interaction in OMS, caused by changes in the functioning of critical infrastructure.

References

1. Dodonov, O., Gorbachyk, O., Kuznietsova, M. Increasing the survivability of automated systems of organizational management as a way to security of critical infrastructures. In: XVIII International Scientific and Practical Conference «Information Technologies and Security» (ITS 2018), CEUR Workshop Proceeding (ISSN 1613-0073). 2018. Vol.2318. P.261-270. [Online]. Available: http:// ceur-ws.org/Vol-2318//.
2. Kharchenko, V.S., Yakovlev, S.V., Gorbachyk, O.S. , et al. Provision of Functional Safety of Critical Information-control Systems. Kharkov: Konstanta, 2019. 272 p. Ukr.
3. Kuznietsova, N. V. Information Technologies for Clients' Database Analysis and Behaviour Forecasting. CEUR Workshop Proceeding (ISSN 1613-0073). 2017. Vol. 2067. P.56-62 [Online]. Available: http://ceur-ws.org/Vol-2067/.
4. Churyumov, G., Tkachov, V., Tokariev, V., Diachenko, V. Method for Ensuring Survivability of Flying Ad-hoc Network Based on Structural and Functional Reconfiguration. In: XVIII

International Scientific and Practical Conference «Information Technologies and Security» (ITS 2018), CEUR Workshop Proceeding (ISSN 1613-0073). 2018. Vol.2318. P.64-76. [Online]. Available: http:// ceur-ws.org/Vol-2318//.
5. Dodonov, O., Kuznietsova, M. Gorbachyk, O. Complex Systems Survivability: Analysis and Modeling. Kyiv: NTUU "KPI", 2009. 264p. Ukr.
6. Dodonov, O., Gorbachyk, O., Kuznietsova, M. System Research of Survivability and Safety for Complex Technical Systems // Data Rec., Storage & Processing. 2010 Vol.12, N 2. P 202-208. Ukr.
7. Dodonov, O., Gorbachyk, O., Kuznietsova, M. Organizational Management Systems: Information Technology and Security In: XIII International Scientific and Practical Conference «Information Technologies and Security» (ITS 2013). V.13. P.5-11.
8. Dodonov, O. Computer Modelling of the Process of Organizing Management // Visnyk NANU.2016, N1. P. 69 – 77.

ANALYSIS OF TESTING APPROACHES TO ANDROID MOBILE APPLICATION VULNERABILITIES

Mykhailo Antonishyn[0000-0002-2665-0066], Oleksii Misnik[0000-0002-4654-9125]

Pukhov Institute for Modelling in Energy Engineering National Academy of Sciences of Ukraine, Kyiv, 03164, Ukraine
antonishin.mihail@gmail.com, alexmisnik91@gmail.com

Abstract. In the first part of the article we discuss international, industrial and national standards and methodologies that describe the process of testing for application vulnerabilities, including mobile applications for Android OS. The following standards and methodologies were taken for the study: ISO/IEC 27034. Information technology. Security techniques. Application Security, NIST 800-163 Vetting the Security of Mobile application, National In-formation Assurance Partnership and Mobile application security verification standard. Also, we have compared methods itselves and methods of testing for vulnerabilities of mobile software applications for operating system Android. The analysis of the stages by which testing for vulnerabilities is carried out. The second part of the article presents statistics on vulnerabilities that were published by vendors – Google security statistics and Quick Heal, as well as statistics, which were formed by the authors of the publication. For statistics, the test results were taken from an online store, two crypto exchanges and two crypto wallets. The conclusions to the article summarize the results of a study of standards and statistics for conducting subsequent research on the subject of scientific work.

Keywords: mobile application, security assessment, security testing, Open Web Application Security Project , National Institute of Standards and Technology, ISO/IEC 27034, National Information Assurance Partnership.

Introduction

According to BetaNews, among the 30 best applications with more than 500,000 downloads, 94% contain at least 3 average risk vulnerabilities, while 77% contain at minimum two high-level vulnerabilities. Among the 30 applications 17% were vulnerable to Man-In-The-Middle (MITM) attacks exposing all data to interception by malicious users. Furthermore, 44% of applications contain confidential data with strict encryption requirements, including passwords or Application Programming Interface (API) keys, while 66% utilize functional abilities which can compromise users' confidentiality. This is exactly why mobile devices are subject to numerous security discussions [1].

1 Application security standards

1.1 ISO/IEC 27034. Information technology. Security techniques. Application Security

ISO/IEC 27034 offers guidance on information security to those specifying, designing and programming or procuring, implementing and using application systems, in other words business and Information

Technology (IT) managers, developers and auditors, and ultimately the end-users of Information and Communication Technology (ICT). The aim is to ensure that computer applications deliver the desired or necessary level of security in support of the organization's Information Security Management System, adequately addressing many ICT security risks.

This multi-part standard provides guidance on specifying, designing/selecting and implementing information security controls through a set of processes integrated throughout an organization's Systems Development Life Cycle (SDLC). It is process oriented [2-5].

It covers software applications developed internally, by external acquisition, outsourcing/offshoring or through hybrid approaches. It addresses all aspects from determining information security requirements, to protecting information accessed by an application as well as preventing unauthorized use and/or actions of an application. The standard is SDLC-method-agnostic: it does not mandate one or more specific development methods, approaches or stages but is written in a general manner to be applicable to them all. In this way, it complements other systems development standards and methods without conflicting with them. One of the key driving principles is that it is worth investing more heavily in specifying, designing, developing and testing software security controls or functions if they are reusable across multiple applications, systems and situations, albeit at the risk of propagating vulnerabilities more widely than might otherwise be the case. In a nutshell, "Do it properly, do it once, and reuse it". The approach may seem a little idealistic, but some far-sighted organizations are already successfully using it: it is more than just an academic interest [3-6].

Section 8.5 "Security Audit" of this standard consists of verification and formal confirmation of evidence that the application that is being developed is at the required level of security, which is written in the technical documentation. An application security audit can be performed at any time during the development and operation life cycle. The sixth part of the standard ISO / IEC 27034-6:2016. Information technology. Security techniques. Application security. Part 6. Case studies does not describe how and by what means it is necessary to conduct testing. It shows just what needs to be tested [7, 8].

1.2 NIST 800-163 Vetting the Security of Mobile application

This document defines an app vetting process and provides guidance on planning and implementing an app vetting process, developing security requirements for mobile apps, identifying appropriate tools for testing mobile apps and determining if a mobile app is acceptable for deployment on an organization's mobile devices. An overview of techniques commonly used by software assurance professionals is provided, including methods of testing for discrete software vulnerabilities and misconfigurations related to mobile app software [9, 10].

Standards includes security checks according to which mobile applications are tested [10]:

1.2.1 Incorrect Permissions. Permissions allow accessing controlled functionality such as the camera or Global Positioning System (GPS) and are requested in the program. Permissions can be implicitly granted to an app without the user's consent.

1.2.2 Exposed Communications. Internal communications protocols are the means by which an app passes messages internally within the device, either to itself or to other apps. External communications allow information to leave the device.

1.2.3 Exposed Data Storage. Files created by apps on Android can be stored in Internal Storage, External Storage, or the Keystore. Files stored in External Storage may be read and modified by all other apps with the External Storage permission.

1.2.4 Potentially Dangerous Functionality. Controlled functionality that accesses system-critical resources or the user's personal information. This functionality can be invoked through API calls or hard coded into an app.

1.2.5 App Collusion. Two or more apps passing information to each other in order to increase the capabilities of one or both apps beyond their declared scope.

1.2.6. Obfuscation. Functionality or control flows that are hidden or obscured from the user. For the purposes of this appendix, obfuscation was defined as three criteria: external library calls, reflection, and native code usage.

1.2.7. Excessive Power Consumption. Excessive functions or unintended apps running on a device which intentionally or unintentionally drain the battery.

1.2.8. Traditional Software Vulnerabilities. All vulnerabilities associated with traditional Java code including: Authentication and Access Control, Buffer Handling, Control Flow Management, Encryption and Randomness, Error Handling, File Handling, Information Leaks, Initialization and Shutdown, Injection, Malicious Logic, Number Handling, and Pointer and Reference Handling [2-7, 10].

1.3 National Information Assurance Partnership (NIAP)

This document presents functional and assurance requirements found in the Protection Profile for Application Software which are appropriate for vetting mobile application software ("apps") outside formal Common Criteria (ISO/IEC 15408) evaluations. Common Criteria evaluation, facilitated in the U.S. by the National Information Assurance Partnership (NIAP), is required for IA and IA-enabled products in National Security Systems according to Committee on National Security Systems (CNSS) Policy #11. Such evaluations, including those for mobile apps, must use the complete Protection Profile. However, even apps without IA functionality may impose some security risks, and concern about these risks has motivated the vetting of such apps in government and industry [2].

Security Functional Requirements [3]:

1.3.1. Random Bit Generation Services. If implement Deterministic Random Bit Generator (DRBG) functionality is chosen, then additional security requirements elements shall be included in the ST. In this requirement, cryptographic operations include all cryptographic key generation/derivation/agreement, Initial Vector`s (IVs) (for certain modes), as well as protocol-specific random values.

1.3.2. Storage of Credentials. This requirement ensures that persistent credentials (secret keys, Public Key Infrastructure (PKI) private keys, or passwords) are stored securely. The assurance activity implicitly restricts which selections can be made, on per-platform basis. For example, if a platform provides hardware-backed protection for credential storage, then the third selection cannot be indicated. If implement functionality to securely store credentials is selected, then the following components must be included in the Security Target (ST). If other cryptographic operations are used to implement the secure storage of credentials, the corresponding requirements must be included in the Security Target.

1.3.3. Access to Platform Resources. The intent is for the evaluator to ensure that the selection captures all hardware resources which the application accesses, and that these are restricted to those which are justified. On some platforms, the application must explicitly solicit permission in order to access hardware resources. Seeking such permissions, even if the application does not later make use of the hardware resource, should still be considered access. Selections should be expressed in a manner consistent with how the application expresses its access needs to the underlying platform. For example, the platform may provide location services which implies the potential use of a variety of hardware resources (e.g. satellite receivers, WiFi, cellular radio) yet location services are the proper selection. This is because use of these resources can be inferred, but also because the actual usage may vary based on the particular platform. Resources that do not need to be explicitly identified are those which are ordinarily used by any application such as central processing units, main memory, displays, input devices (e.g. keyboards, mice), and persistent storage devices provided by the platform. Sensitive information repositories are defined as those collections of sensitive data that could be expected to be shared among some applications, users, or user roles, but to which not all of these would ordinarily require access.

1.3.4. Network Communications. This requirement is intended to restrict both inbound and outbound network communications to only those required, or to network communications that are user initiated. It does not apply to network communications in which the application may generically access the filesystem which may result in the platform accessing remotely mounted drives/shares.

1.3.5. Encryption Of Sensitive Application Data. Any file that may potentially contain sensitive data (to include temporary files) shall be protected. The only exception is if the user intentionally exports the sensitive data to non-protected files.

1.3.6. Supported Configuration Mechanism. Configuration options that are stored remotely are not subject to this requirement.

1.3.7. Secure by Default Configuration. Default credentials are credentials (e.g., passwords, keys) that are automatically (without user interaction) loaded onto the platform during application installation.

Credentials that are generated during installation using requirements laid out in ST are not by definition default credentials. The precise expectations for file permissions vary per platform but the general intention is that a trust boundary protects the application and its data.

1.3.8. Specification of Management Functions. This requirement stipulates that an application needs to provide the ability to enable/disable only those functions that it actually implements. The application is not responsible for controlling the behavior of the platform or other applications.

1.3.9. User Consent for Transmission of Personally Identifiable Information (PII). This requirement applies only to PII that is specifically requested by the application; it does not apply if the user volunteers PII without prompting from the application into a general (or inappropriate) data field. A dialog box that declares intent to send PII presented to the user at the time the application is started is sufficient to meet this requirement.

1.3.10. Use of Supported Services and APIs. The definition of documented may vary depending upon whether the application is provided by a third party (who relies upon documented platform APIs) or by a platform vendor who may be able to guarantee support for platform APIs.

1.3.11. Anti-Exploitation Capabilities. Requesting a memory mapping at an explicit address subverts address space layout randomization (ASLR). Requesting a memory mapping with both write and execute permissions subverts the platform protection provided by Data Execution Prevention (DEP). If the application performs no just-in-time compiling, then the first selection must be chosen.

1.3.12. Integrity for Installation and Update. This requirement is about the ability to "check" for updates. The actual installation of any updates should be done by the platform. This requirement is intended to ensure that the application can check for updates provided by the vendor, as updates provided by another source may contain malicious code.

1.3.13. Use of Third-Party Libraries. The intention of this requirement is for the evaluator to discover and document whether the application is including unnecessary or unexpected third-party libraries. This includes adware libraries which could present a privacy threat, as well as ensuring documentation of such libraries in case vulnerabilities are later discovered.

1.3.14. Protection of Data in Transit. Application should transmit sensitive data only via encryption channel.

1.4 Mobile application security verification standard (MASVS).

The MASVS is a community effort to establish a framework of security requirements needed to design, develop and test secure mobile apps on iOS and Android [4].

MASVS contains three parts (see Fig. 1):

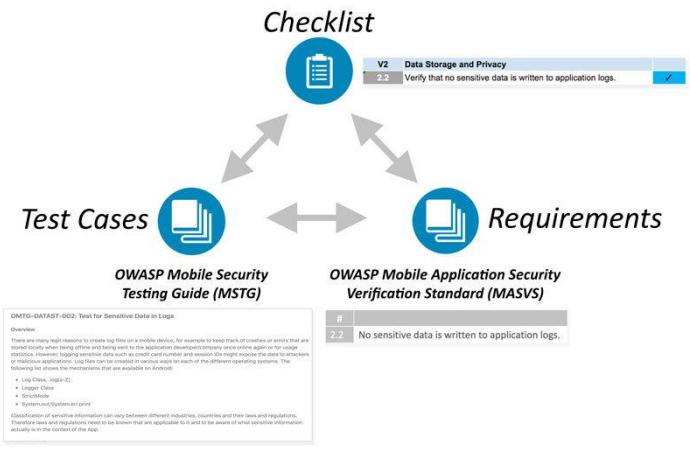


Fig. 1. MASVS Systems.

The Mobile Application Security Verification Standard (MASVS): This standard document defines a mobile app security model and lists generic security requirements for mobile apps. It can be used by architects, developers, testers, security professionals, and consumers to define what a secure mobile application is [4].

Check controls:

V1: Architecture, Design and Threat Modeling Requirements.
V2: Data Storage and Privacy Requirements.
V3: Cryptography Requirements.
V4: Authentication and Session Management Requirements.
V5: Network Communication Requirements.
V6: Platform Interaction Requirements.
V7: Code Quality and Build Setting Requirements.
V8: Resilience Requirements.

The Mobile Security Testing Guide (MSTG): The MSTG is a manual for testing the security of mobile apps. It provides verification instructions for the requirements in the MASVS along with operating-system-specific best practices (currently for Android and iOS). The MSTG helps ensure completeness and consistency of mobile app security test. It is also useful as a standalone learning resource and reference guide for mobile application security testers [4, 5].

Mobile App Security Checklist: A checklist for tracking compliance against the MASVS during practical assessments. The list conveniently links to the MSTG test case for each requirement, making mobile penetration app testing a breeze [4].

The MASVS defines two security verification levels (MASVS-L1 and MASVS-L2), as well as a set of reverse engineering resiliency requirements (MASVS-R). MASVS-L1 contains generic security requirements that are recommended for all mobile apps, while MASVS-L2 should be applied to apps handling highly sensitive data. MASVS-R covers additional protective controls that can be applied if preventing client-side threats is a design goal (see Fig. 2).

Fulfilling the requirements in MASVS-L1 results in a secure app that follows security best practices and doesn't suffer from common vulnerabilities. MASVS-L2 adds additional defense-in-depth controls such as SSL pinning, resulting in an app that is resilient against more sophisticated attacks - assuming the security controls of the mobile operating system are intact, and the end user is not viewed as a potential adversary. Fulfilling all, or subsets of, the software protection requirements in MASVS-R helps impede specific client-side threats where the end user is malicious and/or the mobile OS is compromised [4].

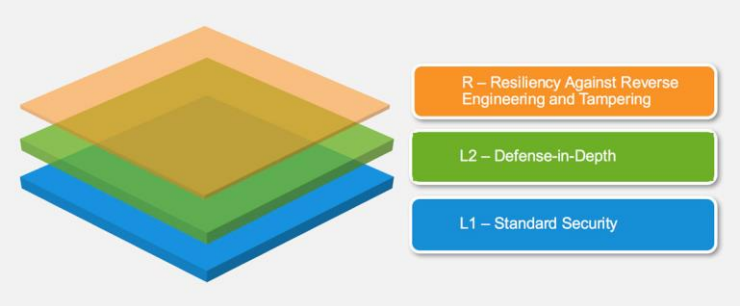


Fig. 2. MASVS security level [3, 4]

1.4.1. MASVS-L1: Standard Security. A mobile app that achieves MASVS-L1 adheres to mobile application security best practices. It fulfills basic requirements in terms of code quality, handling of sensitive data, and interaction with the mobile environment. A testing process must be in place to verify the security controls. This level is appropriate for all mobile applications [4].

1.4.2. MASVS-L2: Defense-in-Depth. MASVS-L2 introduces advanced security controls that go beyond the standard requirements. To fulfill MASVS-L2, a threat model must exist, and security must be an integral part of the app's architecture and design. Based on the threat model, the right MASVS-L2 controls should have been selected and implemented successfully. This level is appropriate for apps that handle highly sensitive data, such as mobile banking apps [4].

1.4.3. MASVS-R: Resiliency Against Reverse Engineering and Tampering. The app has state-of-the-art security, and is also resilient against specific, clearly defined client-side attacks, such as tampering, modding, or reverse engineering to extract sensitive code or data. Such an app either leverages hardware security features or sufficiently strong and verifiable software protection techniques. MASVS-R is applicable to apps that handle highly sensitive data and may serve as a means of protecting intellectual property or tamper-proofing an app [4].

Notes:

I: Although OWASP recommend implementing MASVS-L1 controls in every app, implementing a control or not should ultimately be a risk-based decision, which is taken/communicated with the business owners.

II: Note that the software protection controls listed in MASVS-R and described in the OWASP Mobile Security Testing Guide can ultimately be bypassed and must never be used as a replacement for security controls. Instead, they are intended to add additional threat-specific, protective controls to apps that also fulfill the MASVS requirements in MASVS-L1 or MASVS-L2 [4-6].

2 Statistics

2.1 Us perform testing statistics

Us personal statistics of vulnerability assessment Android mobile application (see Tables 1 – 5). We perform 5 tests on real mobile application and use MASVS for describe vulnerabilities [4-6].

Table 1. Cryptocurrency exchanger 1

High level vulnerabilities	Medium level vulnerabilities	Low level vulnerabilities	Information level vulnerabilities
Sensitive data in logs	SMS spam	Session fixation	Application uses old library
Brakeforce password	Mobile phone number enumeration	SSL certificate potential vulnerable	Cross-origin resource sharing
	OTP return in response	No Certificate and Public Key Pinning	Vulnerability in old version of WebView
	Absence of source code obfuscation	Application data can be backup	

Table 2. Cryptocurrency exchanger 2

High level vulnerabilities	Medium level vulnerabilities	Low level vulnerabilities	Information level vulnerabilities
		No Certificate and Public Key Pinning	Application can run on rooting application
		SSL certificate potential vulnerable	
		Application can be backup	

Table 3. Cryptocurrency wallet 1

High level vulnerabilities	Medium level vulnerabilities	Low level vulnerabilities	Information level vulnerabilities
Sensitive data in logs	Absence of source code obfuscation	Backup private key explicity visible	Application uses old library
Sensitive data saves in local files		Insecure communication – application uses HTTP	
		No Certificate and Public Key Pinning	

2.1 Google security statistics

In 2018, 0.04% of all downloads from Google Play were Potentially Harmful Applications (PHAs). In 2017, the number was 0.02%. This increase is due to the change in methodology of upgrading the severity level of click fraud applications from policy violations to PHAs. If we omit the addition of click fraud for a comparison, 2018 is at 0.017% which is still a reduction from 2017 (see Fig. 3). Now we look for click fraud inside and outside of Google Play and warn users about these apps. All other PHA categories have declined each year or increased at low levels [11].

Table 4. Cryptocurrency wallet 2

High level vulnerabilities	Medium level vulnerabilities	Low level vulnerabilities	Information level vulnerabilities
Root and developer mode bypass	Absence of source code obfuscation	Application data can be backup	Vulnerability in old version of WebView
Critical bug in money transfer	Check modify source code	No Certificate and Public Key Pinning	
Personal data in logs	User Enumeration		

Table 5. Mobile marketplace.

High level vulnerabilities	Medium level vulnerabilities	Low level vulnerabilities	Information level vulnerabilities
Two vulnerabilities – OTP value return in response	Absence of source code obfuscation	Unrestricted user creation	
Four vulnerabilities – Insecure direct object	Cleartext password submission	No Certificate and Public Key Pinning	
	User`s info enumeration	Password changed attack	

2.2 Quick Heal

Quick Heal Annual Threat Report 2019 brings forth insights and intelligence gathered by Quick Heal Security Labs about all that unfolded in the realm of cybersecurity in 2018 – divided into two sections viz (see Fig. 4). Windows and Android. The threat report begins with significant cyber-attack predictions made by Quick Heal Security Labs in 2018 that proved to be true, flagging off the possibility for future cyber-attacks. The report also sheds light on detection highlights of 2018 for both Windows and Android, with a breakup of detections made per day, per hour, per minute, and the entire year, along with a list of top 10 Windows and Android malware [12].

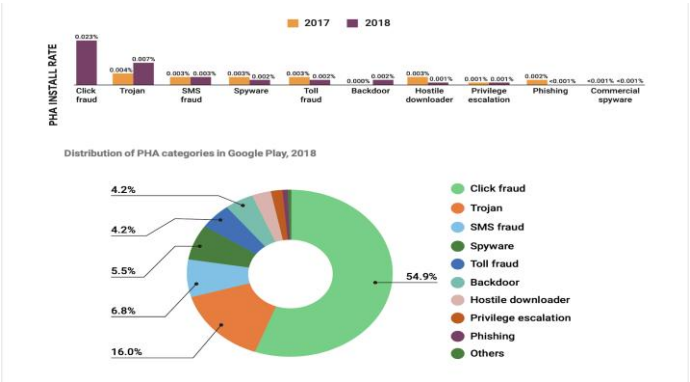


Fig. 3. Google security statistics.

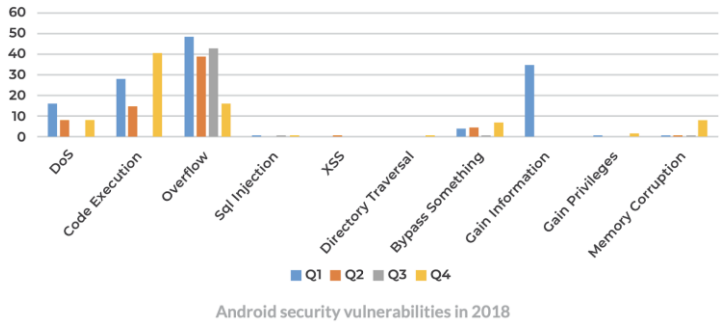


Fig. 4. Quick Heal statistics.

Conclusions

Based on the research results, it can be concluded that the ISO/IEC 27034 standard regulates that vulnerability testing should be carried out, but it is not specified how and what should be tested for vulnerabilities. NIST and NIAP both refer to OWASP MASVS and contain controls by which the mobile application is tested, mainly focusing on vulnerabilities that relate to vulnerabilities in data storage and authorization. This is confirmed by statistics provided by Digital Security. The most recognized is MASVS. One of the parts of MASVS describes what and how to test.

It should be noted that all standards rather weakly assess vulnerabilities that relate to interaction with the API. As it can be seen from the tests described in Section 2.2, the most critical vulnerabilities are vulnerabilities that are associated with interaction with the application server.

References

1. Google publishes its Android Security & Privacy 2018 Year in Review, https://betanews.com/2019/03/30/google-android-security-and-privacy-2018-year-in-review/, last accessed 2019/03/30.
2. National Information Assurance Partnership, Protection Profile for Mobile Device Fundamentals, Version 3.1, June 16, 2017, https://www.niap- ccevs.org/MMO/PP/pp_ md_v3.1.pdf, last accessed 2019/10/21.
3. National Information Assurance Partnership, Requirements for Vetting Mobile Apps from the Protection Profile for Application Software, Version 1.2, April 22, 2016, https://www.niap-ccevs.org/MMO/PP/394.R/pp_app_v1.2_table- reqs.htm, last accessed 2019/10/21.
4. OWASP Foundation, Mobile AppSec Verification, Version 1.1.3, January 2019, https://github.com/OWASP/owasp-masvs/releases/download/1.1.3/OWASP_Mobile_ AppSec_Verification_Standard_1. 1.3_Document.pdf, last accessed 2019/10/01.
5. OWASP Foundation, Mobile Security Testing Guide (MSTG), 1.1.0 Release, November 30, 2018, https://www.owasp.org/index.php/owasp_mobile_security_testing_guide, last accessed 2019/10/21.
6. OWASP, Man-in-the-middle attack, https://www.owasp.org/index.php/Man-in-the-middle_attack, last accessed 2019/10/21.
7. S. Dent, “Report finds Android malware pre-installed on hundreds of phones,” Engaged, May 24, 2018, https://www.engadget.com/2018/05/24/report- finds-android-malware-pre-installed-on-hundreds-of-phones, last accessed 2019/10/01.
8. ISO/IEC 27034-6. Information technology. Security technique’s Application Security, first edition 2016, https://www.iso.org/standard/60804.html, last accessed 2019/10/01.
9. Digital Security, Checking for information security vulnerability, august 2018. Available at https://dsec.ru/wp-content/uploads/2018/08/checklist.pdf, last accessed 2019/10/01.
10. NIST 800-163. Vetting the Security of Mobile application, https://doi.org/10.6028/NIST.SP.800-163r1, last accessed 2019/10/01.
11. Android Security & Privacy 2018 Year In Review, https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=2ahUKEwiLgsa9zMLmAhVXAhAIHcxKBLEQFjACegQIARAC&url=https%3A%2F%2Fsource.android.com%2Fsecurity%2Freports%2FGoogle_Android_Security_2018_Report_Final.pdf&usg=AOvVaw0q4HKyiHNyxRsWeebYfxbP, last accessed 2019/10/01.

12. Quick heal annual threat report 2019. Available athttps://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=2ahUKEwi5wfeSzcLmAhUai8MKHZybBWAQFjACegQIAhAC&url=https%3A%2F%2Fwww.quickheal.co.in%2Fdocuments%2Fthreat-report%2FQH-Annual-Threat-Report-2019.pdf&usg=AOvVaw0OtWUuMBxp0Txjy0ExKPWN, last accessed 2019/10/01.

DEFINING POTENTIAL ACADEMIC EXPERT GROUPS BASED ON JOINT AUTHORSHIP NETWORKS USING DECISION SUPPORT TOOLS

Iryna Balagura, Serhii Kadenko, Oleh Andriichuk, Ivan Gorbov

Institute for Information Recording of the NAS of Ukraine

Detection of academic communities is a relevant task while choosing experts for evaluation of scientific research works, solving topical problems in certain areas, and searching for partners to cooperate with. Besides that, in scientometrics it is important to understand the processes that take place during academic collaboration. Academic community structure, intensity of interaction in it, its leaders: these and other aspects led to emergence of a whole new research area: the Science of Team Science (SciTS) [1]. In order to study the key trends of academic cooperation and detect “reach people’s clubs” as well as the most highly communicative academics, co-authorship networks are used [2]. Usage of social networks featuring specialists’ profiles, such as ResearchGate (https://www.researchgate.net/) and LinkedIn (https://www.linkedin. com/), simplifies the task of looking up specific researchers [3]. Scientific profiles can be found in Google scholar, Scopus, Web of science, and other databases. Besides that, there are resources for unification of information on academics from different databases, such as ORCID (https://orcid.org/), “Bibliometrics of Ukrainian science” (http://www.nbuviap.gov.ua/bpnu/), “Scientists of Ukraine” (http://irbis-nbuv.gov.ua), AMiner (https://aminer.org/), and others. Academic publication databases represent the most thorough resource to look for academic research groups.

In order to single out research groups in co-authorship networks such approaches as cluster analysis and modularity are used. Methods of network centrality and citation indices are used to evaluate separate researchers in a network and to study a specific research group [4].

We propose to use decision support methods to define potential academic expert groups and academic research schools in co-authorship networks, and demonstrate the application of “ordinal factorial analysis-based” approach to calculation of relative weights of different centrality indicators of complex networks. For this purpose we use data from reference databases “Ukrainika naukova” and Scopus. We consider an example of unification of rankings of centrality and citation measures for researchers in the area of informatics. Additionally, the approach allows us to verify the degree of dependence between different centrality measures among themselves and in comparison with citation indicators.

References

1. Valerio Leone Sciabolazza, Raffaele Vacca, Therese Kennelly Okraku, Christopher McCarty Detecting and analyzing research communities in longitudinal scientific networks // PLOS one, August 10, 2017, https://doi.org/10.1371/journal.pone.0182516

2. Балагура І.В. Визначення експертних груп на основі аналізу реферативної бази даних «Україніка наукова» / І.В. Балагура // Реєстрація, зберігання і обробка даних, 2015. – 17, № 3 – С.75–82.

3. C. Wei, W. Lin, H. Chen, W. An, and W. Yeh. Finding experts in online forums for enhancing knowledge sharing and accessibility, Computers in Human Behavior, 2015, vol.51, pp.325-335.

4. Горбов І.В., Каденко С.В., Балагура І.В., Манько Д.Ю., Андрійчук О.В. Визначення потенційних експертних груп науковців в мережі співавторства з використанням методів підтримки прийняття рішень / // Реєстрація, зберігання і обробка даних. – 2013, - Т.15, №4 –С.87-97.

DEVELOPMENT OF THE MULTISPECTRAL VOLUME RECORDING METHODS

Beliak Ie.V., Kryuchyn A.A.
Institute for Information Recording of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

Urgency of the research. Optical information recording by forming of microrelief information elements on the substrate is considered as the most efficient method of long-term data storage.

Target setting. However, density of the optical information recording does not meet the requirements of the modern digital recording systems because diffraction limit significantly reduce the resolution of optical systems.

Actual scientific researches and issues analysis. Scientific research in the area of optical recording shows the high potential of multi-layer photoluminescent media. Volumetric recording is much preferable than development of subdiffraction optical systems due to technical simplicity.

Uninvestigated parts of general matters defining. Typical problems of multilayer photoluminescent recording are low signal-noise ratio and low readout signal level, which makes this method to be inappropriate one for long-term data storage.

The research objective. In this paper it was proposed method of volumetric optical information recording in multilayer, optically homogeneous media with photoluminescent information elements which spectra of photoluminescence differs from layer to layer .

The statement of basic materials. To determine the optimal parameters of multilayer photoluminescent storage with multispectral recording medium the mathematical model of the photoluminescent multilayer recording process was developed.

Conclusions. The task of finding the optimal parameters of the multilayer photoluminescent storage with multispectral recording medium was solved by finding the maximums of the objective functions.

Keywords: optical recording of information; long-term data storage systems; multiispectral medium; photoluminescent multilayer media; readout signal; objective function; function maximum.

Urgency of the research. Optical information recording by forming of microrelief information elements on the reflecting substrate and information layers is considered as the most efficient method of non-volatile long-term data storage. It could be indicated by the relevance of work in the field of optical recording of information and long term data storage, as opposed to work in the field of magnetic and solid-state storage which are not considered as archival media.

Target setting. However, density of the optical information recording does not meet the requirements of the modern digital recording systems because diffraction limit significantly reduce the resolution of optical systems. It should be noticed that developers of the modern "Blu-ray" media (BD) has sacrificed the reliability of the optical system. It can be shown that airy disk which determines resolution of the optical system partially overlaps the information elements of adjacent tracks, which leads to the appearance of a parasitic signal.

Actual scientific researches and issues analysis. Basically, there are three ways to solve the diffraction resolution, which can be divided into three groups:

1. optical recording within the diffraction limit [1];
2. development of subdiffraction optical systems [2];
3. development of volumetric optical recording methods [1-6].

In order to increase the optical information recording density within the diffraction limit of optical system it is necessary to achieve a decrease in the laser radiation wavelength λ and an increase in the numerical aperture of the objective NA. It should be noticed that developers of BD have shown the limit on the laser radiation wavelength for the visible range $\lambda = 405$ nm and the $NA = 0.85$ is also close to the limit of the aperture angle. Further decrease of the Airy disk NA could be achieved by the ultraviolet lasers, vacuum systems and technologically sophisticated, superaccurate immersion recording methods which are inappropriable for the mass production and application of the long term storage [1].

Development of the subdiffraction optical systems are also proved to technologically sophisticated, superaccurate immersion recording methods. It was shown that they are significantly reduces the speed of data reading and more suitable for optical microscopy than for information recording [2].

Therefore, high potential of volumetric recording is much preferable than development of subdiffraction optical systems due to technical simplicity [1-6]. The most promising method of volumetric optical recording is the development of optically transparent, homogeneous and anisotropic recording media with multilayer microrelief structures of photoluminescent information elements [1-6].

Uninvestigated parts of general matters defining. It's necessary to notice that main disadvantages of multilayer photoluminescent information recording method leads to the decrease of the reliability of optical media due to low readout signal level associated with the photoluminescent response loss and low signal-noise ratio due to parasitic signal from the multilayer structure [1, 2]. There are different approaches to solve mentioned problems but no uniform methodology for determining the optimal architecture of the optical reading system and storage parameters.

The research objective. In this paper it was proposed method of volumetric optical information recording in multilayer, optically homogeneous media with photoluminescent information elements. Additionally it was propose to develop multispectral recording media for further increase of signal-noise ratio by separating of the photoluminescent signal from different layers. The task of the multilayer photoluminescent storage optimal parameters finding with multispectral recording medium has to be solved by finding the maximums of the objective functions.

The statement of basic materials. To determine the optimal parameters of multilayer photoluminescent storage with multispectral recording medium the mathematical model of the photoluminescent multilayer recording process has to be developed.

The statement of basic materials. Multilayer photoluminescent storage (MPS) includes substrate and sandwich-structure of data layers (DL) and intermediate layers (IL) which should be transparent and optically homogeneus (fig. 1). Thereby further parameters of this media type could be defined:

- geometrical form and linear sizes of layers and substrate;
- data layer thickness d_{DL};
- intermediate layers thickness d_{IL};
- quantity of the information layers N.

Most important stage of MPS development is determination of the microrelief information structure. Usually developers take as a basis information structure of optical discs (CD, DVD, BD and UMD) that includes information elements (pits) situated in a spiral. Information is coded by pit length and length of the spaces between the pits (lands). Thereby microrelief information structure could be defined by further parameters:

- linear sizes of the pits: set of the pits' lengths $\{l_p^{\min}; l_p^{\max}\}$, pits' width w_p and pits' depth d_p;
- linear sizes of the lands: set of the lands' lengths $\{l_l^{\min}; l_l^{\max}\}$, lands' width w_l and lands' depth d_l;
- track width w_t;
- refraction coefficient of the layers and substrate n.

MPS readout process implies photoluminescent response, therefore it's important to define recording medium characteristic:

- photoluminescent spectrum (or spectra for multispectral recording);
- absorption spectrum (or spectra for multispectral recording).

And finally should be defined parameters of MPS optical system:

- laser beam wavelength λ;
- numerical aperture of the objective lens NA;
- objective lens movement distance h_O;
- type of objective lens and diaphragm.

Obviously d_{DL} depends on the on the d_p, while d_{IL} depends on NA and desired value of signal-noise ratio. Thereby d_{IL} must be chosen big enough to divide photoluminescent signal from different layers. While all the structure of the disc is transparent and homogeneous the parasitic signal will be caused mostly by NA and absorption of pits areas lighted by unfocused laser beam. For big enough value of the N parasitic signal will surpass useful signal. It was proposed to distinguish readout signal as a variable one which is possible if pits at the lighted area does not change during readout process.

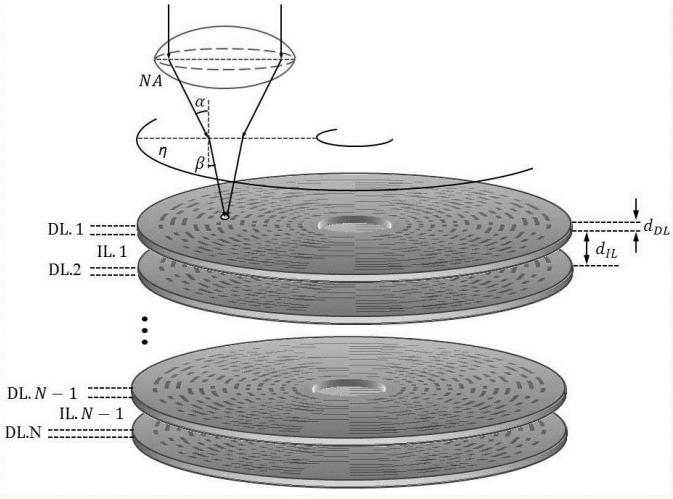


Fig. 1. Structure of the multilayer photoluminescent storage layers.

To avoid undesirable absorption and thereby decrease maximal noise intensity with its variable component it was proposed to record information only by the lands while pits should be recorded by single pulse of laser and can be simulated in mathematical model as cylinders. Furthermore data layer should peripheral area (inner and outside peripheral areas for optical disk structure) which uphold a stable level of parasitic signal during readout from the edges of the disc as its shown at Fig. 2.

While photoluminescent readout signal is spatially isotropic it can be read only part of the probing beam energy within receipt angle $d\Omega$. Finally readout signal power P as a function of probing laser pulse P_0 depends on quantum yield η, absorption factor k_A, receiver system's loss coefficient k_R and out of pit exposure area loss coefficient k_l:

$$P = P_0 \cdot \eta \cdot \left(\frac{d\Omega}{4\pi} \right) \cdot k_A \cdot k_R \cdot k_l . \qquad (1)$$

Thus the only solving of the low signal-noise ratio is synthesis of the dye with a high photoluminescence quantum yield. It's also important to get luminophor with minimal relaxation time to increase data readout rate.

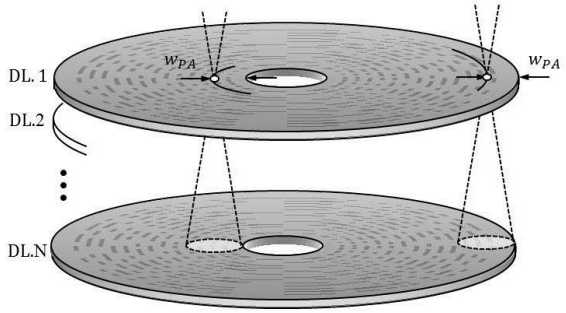


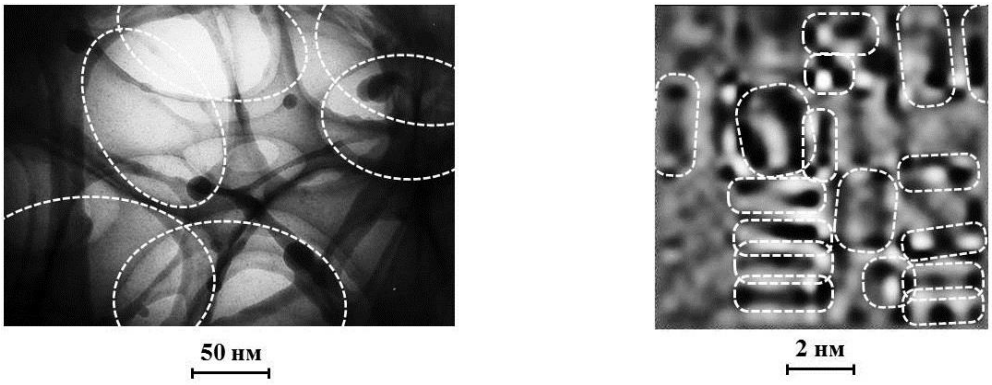
Fig. 2. Determination of inner and outside peripheral areas' width.

There was synthesized class of the pyrazoline dyes which are based on base pyrazoline ultraviolet (UV) dye and pyrazoline "orange-red" with inclusion of polymethylmethacrylate and polystyrene (Table 1). Luminophors were proved to be effective recording media with a quantum yield of photoluminescence value of 60-70%, relaxation time less than 100 ns and wide spectrum of the photoluminescence [1, 2].

Table 1.

Class of the synthesized pyrazoline dyes with inclusion of PMMM and polystyrene [1, 2].

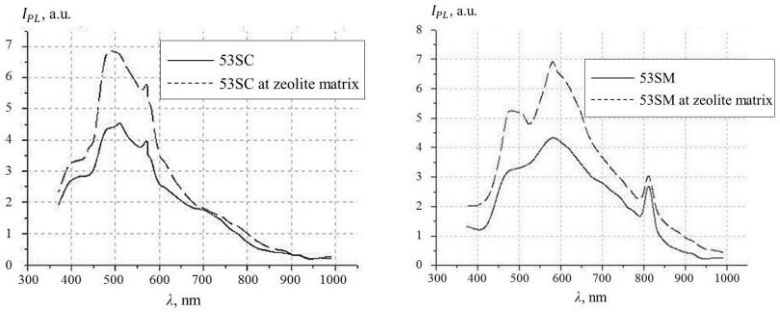
	5% of polymethylmethacrylate	5% of polystyrene
Base pyrazoline ultraviolet dye	53SM	53 SC
Pyrazoline "orange-red" dye	59HM	59HC



(а) (б)

Fig. 3. AFM photographs of pyrazoline luminophor introducing in the white zeolite matrix

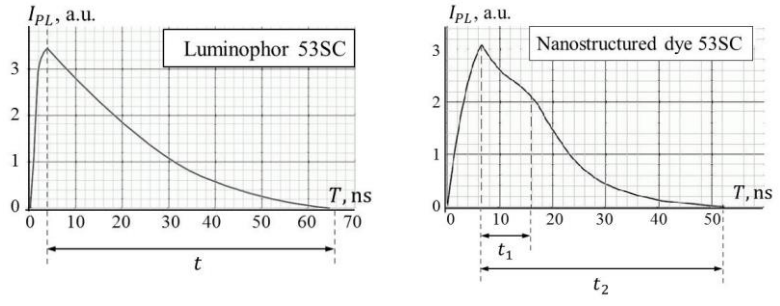
Further research demonstrated potential of the improvement of luminophor parameters by introducing it in the white zeolite matrix [1,2]. The series of experiments confirmed theoretical consideration that zeolite submicron- and nanopores will divide bulk of the dye to the submicron particles (Fig. 3-a: 100-350 nm) and nanoparticles (Fig. 3-b: 1.5-2.5 nm) with a variety of improved optical characteristics.



(a) (b)

Fig. 4. Photoluminescent spectra of basic nanostructured pyrazoline luminophors 53SC (a) and 53SM (b) spectra.

Fig. 4 shows for 53SC and 53SM nanostructured pyrazoline dyes that photoluminescence intensity main peak growth could be within 20-30%. Additional laser annealing increases this parameter causing complete inclusion luminophor at nanoscale pores of white zeolite (Table 2). The growth of photoluminescence quantum yield is caused by the quantum size effects which change molecular energy structure of the luminophor. Pyrazolyne luminophor absorption value growth is also concerned with an appearance of the new allowed transitions. In white zeolite porous structure exited molecules relax to the lower levels and thus absorb larger quantities of the laser beam energy. Such effects also cause the rise of additional peaks of the photoluminescence, which was demonstrated for some types of the pyrazoline dyes 53SC and 53SM luminophors. For the structure of complex synthesized pyrazoline dyes is important to consider $\pi\pi^*$ cross-linking with an active energy hydrogen bond. It causes the effect of transmission spectrum infrared shift which was also confirmed by experiment.



(a) (b)

Fig. 4. Photoluminescent relaxation time of basic (a) nanostructured pyrazoline luminophor 53SC (b) spectra

To achieve readout rate of MPS close to BD readout system it's necessary to get recording medium with photoluminescence relaxation time lower than 100 ns. Experiments showed that for pyrazoline dyes relaxation time value is within measures t =60…100 ns (Fig. 4-a). Inclusion of the luminophors at the white zeolite matrix allowed to decreas PL photoluminescence relaxation time t =40…60 ns. Application of the complementary laser annealing decreased this value up to t =20…40 ns. Moreover complex structure of the photoluminescense kinetics graph (Fig. 4-b) shows that composite pyrazoline dye has not completely filled nanopores and further annealing will help to get value t =10…20 ns.

Table 2.

Improvement of pyrazoline luminophor parameters by introducing it in the white zeolite matrix

Pyrazoline luminophor	Photoluminescence main peak growth		PL relaxation time decrease	
	before annealing	after annealing	before annealing	after annealing
59HM	19 %	43 %	21 %	47 %
59HC	28 %	47 %	33 %	45 %
53SM	38 %	55 %	36 %	55 %
53SC	45 %	63 %	22 %	29 %

While there are different photoluminescence main peak wavelength but close values of intensity for different luminophors of synthesized dyes class allows to develop multispectral MPS where spectra of photoluminescence differs from layer to layer spectra of photoluminescence differs from layer to layer. It was proposed to analyse possibility of this type of media and optimal parameters of the multispectral MPS by finding the maximums of the objective functions of information capacity and reliability.

Mathematical model was based on simulation of optical media readout system focused laser beam by Gauss function [7, 8]:

$$\begin{cases} I(r,z) = I_0 \left(\dfrac{\omega_0}{\omega(z)} \right)^2 \cdot \exp\left(-\dfrac{2r^2}{\omega^2(z)} \right) \\ \begin{cases} \omega(z) = \omega_0 \cdot \sqrt{1 + \left(1 / z_R\right)^2} \\ z_R = \pi\omega_0^2 / \lambda \end{cases} \end{cases}, \qquad (2)$$

where I is time average function of the electromagnetic field intensity distribution, I_0 is intensity of laser beam at the focus point, ω_0 is the radius of the Airy disk, z is the vertical distance from the focal plane, r is the radial distance from the perpendicular to the focal plane.

At the output data of the mathematical model that simulates the readout process includes:

- I_{SN} as a photoluminescence amplitude of the useful signal during the focusing of the laser beam on the pit (this value summarize useful and parasitic signal);
- I_N as a photoluminescence amplitude of the signal during the focusing of the laser beam on the land (pure parasitic signal);
- $\Delta I_{SN}^{\max}$ as a maximal deviation in the amplitude of the signal when focusing the laser beam on the land (variable part of the parasitic signal).

At the next stage output data was used as indicators and objective functions to define optimal parameters of multispectral MPS:

- k_S is indicator of the useful signal, as the ratio of the useful signal to the maximum possible signal, which occurs when focusing on the pit of the first information layer;
- k_C is contrast indicator as ratio of the useful signal of the averaged photoluminescence signal value received by the readout system;
- k_{SNR} signal-noise ratio as a ratio of the useful signal of a variable component of parasitic signal that cannot be distinguished during readout process.

$$\begin{cases} k_S = \dfrac{I_{SN} - I_N}{I_S^{\max}} \\ k_S = \dfrac{I_{SN} - I_N}{I_{SN}} \\ k_{SNR} = \dfrac{I_{SN}}{\Delta I_{SN}^{\max}} \end{cases}. \qquad (3)$$

Defined coefficients allow specifying for multispectral MPS the concept of reliability at the mathematical level (Fig. 5).

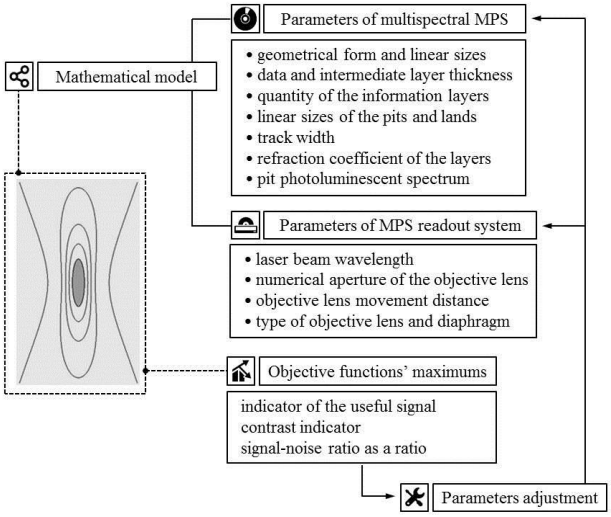


Fig. 5. Algorithm of optimal parameters of the multispectral multilayer photoluminescent storage estimation by finding the maximums of the objective functions

At the Fig. 6 shown results of the developed mathematical model of multispectral MPS readout process work for target functions:

- indicator of the useful signal as a function of depth value of readout layer (Fig 1-a for MPS and Fig 1-b for multispectral MPS);
- contrast indicator as a function of depth value of readout layer (Fig 2-a for MPS and Fig 2-b for multispectral MPS);
- signal-noise ratio as a function of depth value of readout layer (Fig 3-a for MPS and Fig 3-b for multispectral MPS).

The results indicate an improvement for $k_C(H)$ and $k_{SNR}(H)$ target functions, while predictably $k_S(H)$ dependence remained unchanged.

Conclusions. To make a proper analysis of multispectral multilayer photoluminescent optimal characteristics it was developed mathematical model and computer simulation of the readout laser beam propagation process. The model included parameters of the storage construction and readout system as input data. Thereby the task of the optimal parameters of the multilayer photoluminescent storage with multispectral recording medium finding was solved by finding the maximums of the objective functions. It was shown that development of multispectral media allows to increase reliability of multilayer photoluminescent media.

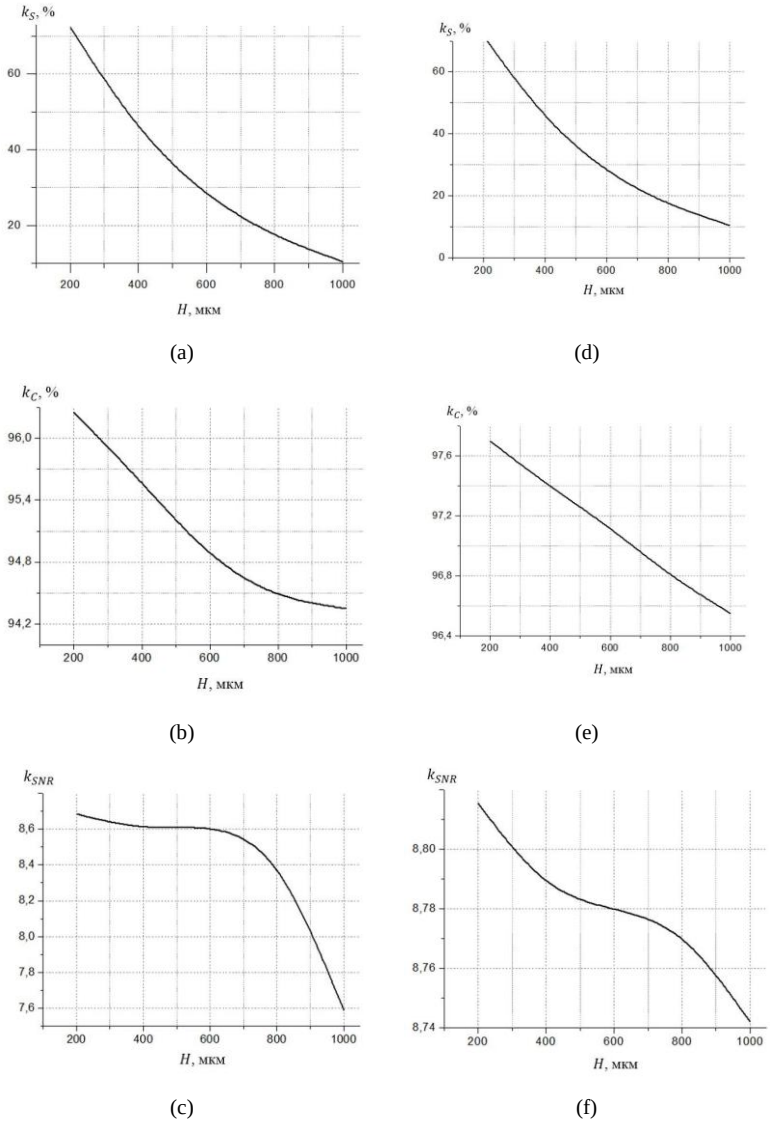


Fig. 6. Indicator function of readout layer depth value for MPS (a, b, c) and multispectral MPS (d, e, f)

1. V.V. Petrov, Zichun Le., A.A. Kryuchyn, S.M. Shanoylo, M.Fu, Ie.V. Beliak, D.Yu.Manko, A.S. Lapchuk, E.M. Morozov. Long-term storage of digital information July 2018 DOI: 10.15407 / Akademperiodyka. 360.148ISBN: 9789663603605.
2. Petrov, V.V., Kryuchyn, A.A., Beliak, Ie. V., Lapchuk, A.S. (2016). Multi photon microscopy and optical recording. Nat. acad. of sciences of Ukraine, Inst. for information recording, Akademperiodyka: Kyiv.
3. Riesen, N., Pan, X., Badek, K., Ruan, Y., Monro, T. M., Zhao, J., Riesen, H. (2018). Towards rewritable multilevel optical data storage in single nanocrystals. Optics Express, 26(9), 12266. doi:10.1364/oe.26.012266
4. Kallepalli, D. L., Alshehri, A. M., Marquez, D. T., Andrzejewski, L., Scaiano, J. C., & Bhardwaj, R. (2016). Ultra-high density optical data storage in common transparent plastics. Scientific Reports, 6(1). doi:10.1038/srep26163
5. Kazansky, P. G., Cerkauskaite, A., Drevinskas, R., & Zhang, J. (2016). Eternal 5D optical data storage in glass. Optical Data Storage 2016. doi:10.1117/12.2240594
6. Tanaka, Y., Ogata, T., & Imagawa, S. (2014). Decoupling direct tracking control system for super-multilayer optical disk. Optical Data Storage 2014. doi:10.1117/12.2064244
7. Roden J.A. Convolution PML (CPML): An efficient FDTD implementation of the CFS-PML for arbitrary media / S.D. Gedney // Microwave and Optical Technology Letters. – 2000. – № 27 – p. 334–339.
8. Rylander T. Stable FDTD-FEM hybrid method for Maxwell's equations / Bondeson A. // Computer Physics Communications. – 2000. – 125 – p. 75–82.

METHOD OF FORMATION SHIFT INDEXES VECTOR BY MINIMIZATION OF POLYNOMIALS

Liubov Berkman[1], Serhiy Otrokh[2], Valeriy Kuzminykh[2] and Olena Hryshchenko[1]
[1] *State University of Telecommunications, Kyiv, Ukraine*
[2] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine*

Abstract. In this article, the 2 methods of formation shift indexes vector of the ring code are proposed. This article presents the matrix of the ring code in the form of binary elements (0,1) and in the form of polynomials, which consist of the sum of the arguments x of a certain category corresponding to the binary value of 1 code sequences of the ring code. With these two methods, shift indexes vectors are created using an example of a ring code of 7 × 7, each row containing 4 units and 3 zeros. The first method makes it possible to form VPS by implementing logical transformations (XOR, OR, or AND) over the binary elements of the generating matrix of the ring code. The second method makes it possible to form VPS by implementing logical transformations (XOR, OR, or AND) over the arguments x of a certain category corresponding to the binary value of 1 code sequences of the ring code. According to the proposed second method, as a result of the logical transformations of the XOR, OR, or AND arguments of polynomials of the matrix G, similar shift indexes vectors were formed. The first method allows developing an algorithm and implementing a program implementation of the formation shift indexes vector. The second method is more visual, by means of which you can mathematically describe the sequence of the formation shift indexes vector of the ring code.

Keywords: matrix of ring code, polynomial, logical transformations of the XOR, OR, or AND arguments, shift indexes vector.

1 Introduction

The ring code is constructed on the principle of block-cyclic codes [8], the rows of the forming matrices of which are linked by the condition of cyclicity.

Cyclic codes have been widely used due to their efficiency in detecting and correcting errors [1, 2].

Cyclic codes are a subset of linear block codes. Linear *(n, k)* - Code C is called cyclic if the cyclic shift of any word c is also a word of that code. The cyclic shift of the code word $c=(c_0, c_{1,\dots} c_{n-1})$ corresponds to the shift of all elements of the word one position to the right, resulting in a code word $c^{(1)}= (c_{n-1}, c_0, c_{1,\dots} c_{n-2})$. As a result of the *i*- multiple shift, we get the code word $c^{(i)}= (c_{n-i,} \dots, c_{n-1}, c_0, c_{1,\dots} c_{n-i-1})$.

The schematics of the encoding and decoding devices for these codes are extremely simple and are based on conventional shift registers. An example of a feedback register is shown in fig. 1.

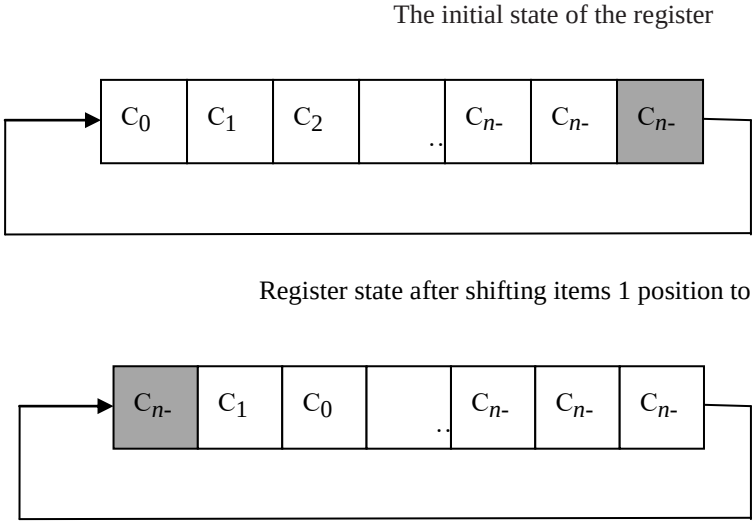


Figure 1. The feedback register

It is convenient to consider cyclic codes by presenting a combination of binary code not as sequences of zeros and ones, but as a polynomial from a dummy variable *x*.

The code word $c=(c_0, c_{1,\dots}c_{n-1})$ can be represented as the following polynomial:

$$c(x) = c_0x^0 + c_1x^1 +\dots+ c_{n-1}x^{n-1} = c_0 + c_1x +\dots+ c_{n-1}x^{n-1}. \qquad (1)$$

Where c_i – are the numbers of the given system of calculus (in binary system 0 and 1).

The above comparison of code words and polynomials is unambiguous: each word corresponds to a polynomial and each polynomial corresponds to a word that is determined by the coefficients of that polynomial.

If $c=(c_0, c_1, \ldots c_{n-1})$ is a code word belonging to code C, then the corresponding polynomial c (x) is also called a code word C [3, 4].

According to the definition of the cyclic code for constructing the forming matrix $G_{n,k}$ it is sufficient to select only one initial n-degree combination $C_i(x)$. A cyclic shift can produce $(n - 1)$ different combinations, of which any k combinations can be taken as starting points By summing the rows of the forming matrix in all possible combinations, other code combinations can be obtained.

Thus, the cyclic (n, k) - code is a set of polynomials forming a k-dimensional linear subspace of the space of all polynomials of degree n-1 closed with respect to the cyclic shift operation [5].

The cyclic shift of the combination with the unit in the higher n-th digit is identical to the multiplication of the corresponding polynomial by x with the simultaneous subtraction from the result of the polynomial $(x^n - 1)$ or $(x^n + 1)$, since the operations are performed by module two. According to [6], the code combination of the cyclic (n, k)- code can be obtained in two ways:

1) by multiplying a simple code combination of degree $(k - 1)$ by the monomial x^{n-k} and adding to this multiplication the residue obtained from dividing the resulting multiplication by the polynomial $P(x)$ of degree $(n - k)$,

2) by multiplying a simple code combination of degree $(k - 1)$ by the forming polynomial $P(x)$ of degree $(n – k)$.

According to the first encoding method, the first k symbols of the resulting code combination match the corresponding symbols of the original simple code combination.

According to the second method, in the code combination obtained, the information symbols do not always coincide with the symbols of the original simple combination. This method is easy to implement, but because the resulting code combinations do not contain information symbols in explicit form, it complicates the decoding process.

In practice, the first method of obtaining a cyclic code is usually used.

The cyclic code can be represented as a formative matrix $P_{n,k}$, which consists of two submatrices: information U_k and additional H_p :

$$P_{n,k} = |U_k, H_p|. \qquad (2)$$

Information submatrix U_k, is a square unit matrix with the number of rows and columns equal to k. The additional sub-matrix H_p contains $p = n - k$ columns and k rows and is formed by residues $R(x)$[3].

The forming matrix allows k code combinations to be obtained. Other combinations are formed by summing the modulus of two rows of the forming matrix in all possible combinations.

The forming matrix $P_{n,k}$ can also be formed by multiplying the forming polynomial $P(x)$ of degree $\rho = n - k$ by the monomial x^{k-1} of the following
k-1 shifts of the resulting combination. The number of rows and columns in the formative matrix may be arbitrary.

The forming matrix of ring code unlike the cyclic code is always a square matrix of size $N \times N$, each row of which contains m units and, accordingly, $N – m$ zeros. Each row of the forming matrix of cyclic code has the same number of elements and the same structure of unitary and null character combinations. The first row of the forming matrix of cyclic code is called the initial vector, or the initial sequence, and the last row is the final vector of the cyclic shift of the code sequence elements. In this case, the rows of the matrix seem to form a ring of a complete cycle of shift of elements of code sequences [7].

2 An example of a matrix representation of a ring code

For example is a ring code, the shift of elements of the code sequences, which is carried out, from right to left, with the leftmost symbol being transferred to the right at the end of the code sequence each time. Below are the forming of ring code matrices that contain elements of code sequences in a binary number system. The forming matrix G has a size of 7×7 and contains 4 units and 3 zeros, and the forming matrix P has a size of 9 x 9 P contains 4 units and 5 zeros. The first rows of the forming matrices are called the initial sequence.

$$G = \begin{bmatrix} 0101011 \\ 1010110 \\ 0101101 \\ 1011010 \\ 0110101 \\ 1101010 \\ 1010101 \end{bmatrix} \quad (3)$$

$$P = \begin{bmatrix} 010010011 \\ 100100110 \\ 001001101 \\ 010011010 \\ 100110100 \\ 001101001 \\ 011010010 \\ 110100100 \\ 101001001 \end{bmatrix} \quad (4)$$

The above forming matrices can also be represented in the form of polynomials that correspond to the code sequences represented in the binary number system and where x – is an argument of a particular digit corresponding to the binary value of 1 ring code. The structure of the code sequences of the forming matrix G_1 corresponds to the structure of the code sequences in the binary number system the forming matrix G, the structure of the code sequences of the forming matrix P_1 corresponds to the structure of the code sequences in the binary number system the forming matrix:

$$G_1 = \begin{bmatrix} x^5 + x^3 + x^1 + x^0 \\ x^8 + x^5 + x^2 + x^1 \\ x^5 + x^3 + x^2 + x^0 \\ x^6 + x^4 + x^3 + x^1 \\ x^5 + x^4 + x^2 + x^0 \\ x^6 + x^5 + x^3 + x^1 \\ x^6 + x^4 + x^2 + x^0 \end{bmatrix} \quad (5)$$

$$P_1 = \begin{bmatrix} x^7 + x^4 + x^1 + x^0 \\ x^6 + x^4 + x^2 + x^1 \\ x^6 + x^3 + x^2 + x^0 \\ x^7 + x^4 + x^3 + x^1 \\ x^8 + x^5 + x^4 + x^2 \\ x^6 + x^5 + x^3 + x^0 \\ x^7 + x^6 + x^4 + x^1 \\ x^8 + x^7 + x^5 + x^2 \\ x^8 + x^6 + x^3 + x^0 \end{bmatrix} \quad (6)$$

Analysis of the structure of the polynomials of the G_1 and P_1 forming matrices indicates that the number of polynomials in the forming matrices depends on the number of elements in the code sequence, and, accordingly, on the number of code sequences. At that time, the number of members in each polynomial depends only on the number of units in the code sequence and does not depend on the number of elements in the code sequence.

Using the above matrices, we form the shift indexes vectors of the ring code by performing the binary transformations of *XOR*, *AND* and *OR* over the binary elements of the code sequences of the matrices G and P, as well as over the arguments of the polynomials of the matrices G_1 and P_1.

3. The method of formation of shift indexes vector by performing logical transformations over the binary elements of the matrices G and P

The shift indexes vector is the sequence of decimal numbers formed by summing the number of units resulting from the implementation of one of the binary transformations *XOR*, *AND*, *OR* (with or without negation) of the elements of the initial sequence (first line) of the ring code and the rest of its lines.

According to [8] the shift indexes vector (hereinafter referred to as SIV) is formed in three stages:

1) Alternatively, the binary logical transformation of *XOR*, *OR*, AND elements of the first line and the next rows of the matrix of the ring code, placed at the same positions of the code sequences, is performed.

2) The binary elements (0, 1), resulting from the logical transformation are written to the shift indexes matrix.

3) The number of units is calculated in each row of the resulting shift indexes matrix (hereinafter referred to as the SIM). In doing so, the SIM is transformed into SIV.

Tables 1 - 3 present the sequence of transformations of ring code a 7 × 7, each row containing 4 units and 3 zeros, in the SIV using logical transformations *XOR*, *OR*, *AND*.

Table 1. The sequence of the formation of SIV by logical transformation *XOR*

Line number ring code	Structures of code sequences	Line numbers between elements of which the operation was performed *XOR*	Structures forming vectors	The sums of the single elements of the forming vectors
1	0 1 0 1 0 1 1	1 - 2	1 1 1 1 1 0 1	6
2	1 0 1 0 1 1 0	1 - 3	0 0 0 0 1 1 0	2
3	0 1 0 1 1 0 1	1 - 4	1 1 1 0 0 0 1	4
4	1 0 1 1 0 1 0	1 - 5	0 0 1 1 1 1 0	4
5	0 1 1 0 1 0 1	1 - 6	1 0 0 0 0 0 1	2
6	1 1 0 1 0 1 0	1 - 7	1 1 1 1 1 1 0	6
7	1 0 1 0 1 0 1			

Table 2. The sequence of the formation of the SIV by logical transformation *AND*

Line number ring code	Structures of code sequences	Line numbers between elements of which the operation was performed *AND*	Structures forming vectors	The sums of the single elements of the forming vectors
1	0 1 0 1 0 1 1	1 - 2	0 0 0 0 0 1 0	1
2	1 0 1 0 1 1 0	1 - 3	0 1 0 1 0 0 1	3
3	0 1 0 1 1 0 1	1 - 4	0 0 0 1 0 1 0	2
4	1 0 1 1 0 1 0	1 - 5	0 1 0 0 0 0 1	2
5	0 1 1 0 1 0 1	1 - 6	0 1 0 1 0 1 0	3
6	1 1 0 1 0 1 0	1 - 7	0 0 0 0 0 0 1	1
7	1 0 1 0 1 0 1			

Table 3. The sequence of the formation of SIV using logical OR transformation

Line number ring code	Structures of code sequences	Line numbers between elements of which the operation was performed *OR*	Structures forming vectors	The sums of the single elements of the forming vectors
1	0 1 0 1 0 1 1	1 - 2	1 1 1 1 1 1 1	7
2	1 0 1 0 1 1 0	1 - 3	0 1 0 1 1 1 1	5
3	0 1 0 1 1 0 1	1 - 4	1 1 1 1 0 1 1	6
4	1 0 1 1 0 1 0	1 - 5	0 1 1 1 1 1 1	6
5	0 1 1 0 1 0 1	1 - 6	1 1 0 1 0 1 1	5
6	1 1 0 1 0 1 0	1 - 7	1 1 1 1 1 1 1	7
7	1 0 1 0 1 0 1			

Tables 4 - 6 present the sequence of transformations of ring code a 9×9, each row containing 4 units and 5 zeros, in the SIV using logical transformations *XOR*, *OR* and *AND*.

Table 4. The sequence of the formation of the SIV by logical transformation *XOR*

Line number ring code	Structures of code sequences	Line numbers that have binary transformations between elements	Structures forming vectors	The sums of the single elements of the forming vectors
1	0 1 0 0 1 0 0 1 1	1 - 2	110110101	6
2	1 0 0 1 0 0 1 1 0	1 - 3	011011110	6
3	0 0 1 0 0 1 1 0 1	1 – 4	000001001	2
4	0 1 0 0 1 1 0 1 0	1 – 5	110100111	6
5	1 0 0 1 1 0 1 0 0	1 – 6	011111010	6

6	0 0 1 1 0 1 0 0 1	1 – 7	001000001	2
7	0 1 1 0 1 0 0 1 0	1 – 8	100110111	6
8	1 1 0 1 0 0 1 0 0	1 – 9	111011010	6
9	1 0 1 0 0 1 0 0 1			

Table 5. The sequence of the formation of SIV using logical *AND* transformation

Line number ring code	**Structures of code sequences**	**Line numbers that have binary transformations between elements**	**Structures forming vectors**	**The sums of the single elements of the forming vectors**
1	0 1 0 0 1 0 0 1 1	1 - 2	000000010	1
2	1 0 0 1 0 0 1 1 0	1 - 3	000000001	1
3	0 0 1 0 0 1 1 0 1	1 – 4	010010010	3
4	0 1 0 0 1 1 0 1 0	1 – 5	000010000	1
5	1 0 0 1 1 0 1 0 0	1 – 6	000000001	1
6	0 0 1 1 0 1 0 0 1	1 – 7	010010010	3
7	0 1 1 0 1 0 0 1 0	1 – 8	010000000	1
8	1 1 0 1 0 0 1 0 0	1 – 9	000000001	1
9	1 0 1 0 0 1 0 0 1			

Table 6. The sequence of the formation of SIV using logical *OR* transformation

Line number ring code	**Structures of code sequences**	**Line numbers that have binary transformations between elements**	**Structures forming vectors**	**The sums of the single elements of the forming vectors**
1	0 1 0 0 1 0 0 1 1	1 - 2	110110111	7
2	1 0 0 1 0 0 1 1 0	1 - 3	011011111	7
3	0 0 1 0 0 1 1 0 1	1 – 4	010011011	5
4	0 1 0 0 1 1 0 1 0	1 – 5	110110111	7
5	1 0 0 1 1 0 1 0 0	1 – 6	011111011	7
6	0 0 1 1 0 1 0 0 1	1 – 7	011010011	5
7	0 1 1 0 1 0 0 1 0	1 – 8	110110111	7
8	1 1 0 1 0 0 1 0 0	1 – 9	111011011	7
9	1 0 1 0 0 1 0 0 1			

4. The method of forming a shift indexes vector by making logical transformations over the arguments of polynomials of the matrix G_1

Formation of the SIV by making logical transformations over the arguments of polynomials is carried out as follows:

1) Alternately the binary logical transformation *XOR*, *OR*, or AND of the arguments of the first polynomial and subsequent polynomials of the ring code G_1 matrix is performed.

2) Depending on the type of logical transformation (*XOR*, *OR* or *AND*) the following steps are performed:

a) For logical *XOR* transformation:

- paired arguments of the generated polynomial with identical digits are deleted;
- counts the number of arguments left after deletion

paired arguments with equal digits;

b) For logical AND transformation:

- remove the odd arguments of the polynomial;
- paired arguments with equal digits are absorbed by one argument;
- counts the number of arguments left after deletion

and absorbing arguments;

c) For OR logical transformation:

- paired arguments with equal digits are absorbed by one argument;
- counts the number of arguments left after the arguments are absorbed.

Tables 7 - 9 show the sequence of transformations of ring code a 9 × 9, each row containing 4 units and 5 zeros by performing logical transformations over the arguments of the polynomials of the forming matrix G_1.

Table 7. The sequence of the formation of SIV by logical transformation *XOR*

Line number RC	**The ring code matrix**	**Numbers rows matrix RC**	**Shift indexes matrix**	**Result addition modulo 2**	**Total number elements SIV**
1	$x^5+x^3+x^1+x^0$	1 - 2	$x^5+x^3+x^1+x^0$ $x^6+x^4+x^2+x^1$	$x^5+x^3+x^0$ $x^6+x^4+x^2$	6
2	$x^6+x^4+x^2+x^1$	1 - 3	$x^5+x^3+x^1+x^0+$ $x^5+x^3+x^2+x^0$	x^1+x^2	2
3	$x^5+x^3+x^2+x^0$	1 - 4	$x^5+x^3+x^1+x^0+$ $x^6+x^3+x^2+x^0$	$x^5+x^1+x^6+$ x^2	4
4	$x^6+x^3+x^2+x^0$	1 - 5	$x^5+x^3+x^1+x^0+x^5+$ $x^4+x^2+x^0$	$x^3+x^1+x^4+x^2$	4
5	$x^5+x^4+x^2+x^0$	1 - 6	$x^5+x^3+x^1+x^0+$ $x^6+x^5+x^3+x^1$	x^0+x^6	2
6	$x^6+x^5+x^3+x^1$	1 - 7	$x^5+x^3+x^1+x^0$ $x^6+x^4+x^2+x^0$	$x^5+x^3+x^1+$ $x^6+x^4+x^2$	6
7	$x^6+x^4+x^2+x^0$				

Table 8. The sequence of the formation of SIV by logical transformation *AND*

Line number RC	**The ring code matrix**	**Numbers rows matrix RC**	**Shift indexes matrix**	**Result conjunctive absorption**	**Total number elements SIV**
1	$x^5+x^3+x^1+x^0$	1 - 2	$x^5 \wedge x^3 \wedge x^1 \wedge x^0 \wedge$ $x^6 \wedge x^4 \wedge x^2 \wedge x^1$	$x^1 \wedge x^1 = x^1$	1
2	$x^6+x^4+x^2+x^1$	1 - 3	$x^5 \wedge x^3 \wedge x^1 \wedge x^0 \wedge$ $x^5 \wedge x^3 \wedge x^2 \wedge x^0$	$x^5 \wedge x^{5=} x^5$ $x^3 \wedge x^{3=} x^3$ $x^0 \wedge x^{0=} x^0$	3
3	$x^5+x^3+x^2+x^0$	1 - 4	$x^5 \wedge x^3 \wedge x^1 \wedge x^0 \wedge$ $x^6 \wedge x^3 \wedge x^2 \wedge x^0$	$x^3 \wedge x^{3=} x^3$ $x^0 \wedge x^{0=} x^0$	2
4	$x^6+x^3+x^2+x^0$	1 - 5	$x^5 \wedge x^3 \wedge x^1 \wedge x^0 \wedge$ $x^5 \wedge x^4 \wedge x^2 \wedge x^0$	$x^5 \wedge x^{5=} x^5$ $x^0 \wedge x^{0=} x^0$	2
5	$x^5+x^4+x^2+x^0$	1 - 6	$x^5 \wedge x^3 \wedge x^1 \wedge x^0 \wedge$ $x^6 \wedge x^5 \wedge x^3 \wedge x^1$	$x^5 \wedge x^{5=} x^5$ $x^3 \wedge x^{3=} x^3$ $x^1 \wedge x^{1=} x^1$	3
6	$x^6+x^5+x^3+x^1$	1 - 7	$x^5 \wedge x^3 \wedge x^1 \wedge x^0$ $x^6 \wedge x^4 \wedge x^2 \wedge x^0$	$x^0 \wedge x^{0=} x^0$	1
7	$x^6+x^4+x^2+x^0$				

As noted above, polynomials of the forming matrices G_1 and P_1 consist of arguments whose degrees correspond to the degrees of the binary elements of the code sequences of the forming matrices G and P of the ring code. The analysis of the above tables indicates that the algorithm for converting the ring code represented in the form of polynomials is similar to the algorithm for converting the ring code represented as binary elements.

Table 9. The sequence of the formation of SIV by logical transformation *AND*

Line number RC	The ring code matrix	Numbers rows matrix RC	Shift indexes matrix	Result disjunctive absorption	Total number elements SIV
1	$x^5+x^3+x^1+x^0$	1 - 2	$x^5 \vee x^3 \vee x^1 \vee x^0 \vee$ $x^6 \vee x^4 \vee x^2 \vee x^1$	$x^1 \vee x^1 = x^1$ $(x^5 \vee x^3 \vee x^1 \vee x^0 \vee x^6 \vee x^4 \vee x^2)$	7
2	$x^6+x^4+x^2+x^1$	1 - 3	$x^5 \vee x^3 \vee x^1 \vee$ $x^0 \vee$ $x^5 \vee x^3 \vee x^2 \vee x^0$	$x^5 \vee x^{5=} x^5$ $x^3 \vee x^{3=} x^3$ $x^0 \vee x^{0=} x^0$ $(x^5 \vee x^3 \vee x^2 \vee$ $x^1 \vee x^0)$	5
3	$x^5+x^3+x^2+x^0$	1 - 4	$x^5 \wedge x^3 \wedge x^1 \wedge$ $x^0 \wedge$ $x^6 \wedge x^3 \wedge x^2 \wedge x^0$	$x^3 \wedge x^{3=} x^3$ $x^0 \wedge x^{0=} x^0$ $(x^5 \vee x^6 \vee x^3 \vee x^1 \vee$ $x^2 \vee x^0)$	6
4	$x^6+x^3+x^2+x^0$	1 - 5	$x^5 \wedge x^3 \wedge x^1 \wedge$ $x^0 \wedge$ $x^5 \wedge x^4 \wedge x^2 \wedge x^0$	$x^5 \wedge x^{5=} x^5$ $x^0 \wedge x^{0=} x^0$ $(x^5 \vee x^4 \vee x^3 \vee x^1 \vee$ $x^2 \vee x^0)$	6
5	$x^5+x^4+x^2+x^0$	1 - 6	$x^5 \wedge x^3 \wedge x^1 \wedge$ $x^0 \wedge$ $x^6 \wedge x^5 \wedge x^3 \wedge x^1$	$x^5 \wedge x^{5=} x^5$ $x^3 \wedge x^{3=} x^3$ $x^1 \wedge x^{1=} x^1$ $(x^5 \vee x^6 \vee x^3 \vee x^1 \vee$ $x^0)$	5
6	$x^6+x^5+x^3+x^1$	1 - 7	$x^5 \wedge x^3 \wedge x^1 \wedge x^0$ $x^6 \wedge x^4 \wedge x^2 \wedge x^0$	$x^0 \wedge x^{0=} x^0$ $(x^5 \vee x^6 \vee x^3 \vee x^4 \vee x^1$ $\vee x^2 \vee x^0)$	7
7	$x^6+x^4+x^2+x^0$				

5. Conclusions

1. The ring code can be represented in the form of a matrix, each line of which is a polynomial with arguments of a certain digit corresponding to the binary value of 1 ring code.
2. The shift indexes vector can be formed by two methods:

- by making logical transformations (*XOR*, *OR* or *AND*) over the binary elements of the ring matrix-forming matrix;
- by performing logical transformations (*XOR*, *OR* or *AND*) over the arguments of the forming matrix of polynomials

3. The first method allows you to develop an algorithm and perform software implementation of the formation of the shift indexes vector.
4. The second method is more visual and allows you to mathematically describe the sequence of formation of the shift indexes vector of the ring code.

References

1. D. Augot, E. Betti, E. Orsini, An introduction to linear and cyclic codes, Journal of Symbolic Computation (2009) 47-68.
2. L. M. J. Bazzi and S. K. Mitter, “Some randomized code constructions from group actions”. IEEE Trans. on Inf. Th., vol 52, pp. 3210–3219, 2006.
3. J. Yuan, C. Ding, Senior Member, IEEE Secret Sharing Schemes from Three Classes of Linear Codes, IEEE Trens. On Inf. Theory 52 (1) (2006) 206-212.
4. A. Vardy, “ Algorithmic complexity in coding theory and the minimum distance problem”, STOC ’97: Proceedings of the twenty-ninth annual ACM symposium on Theory of Computing, pp. 92–109, 1997.
5. A. Ashikmin, A. Barg, Minimal vectors in linear codes, IEEE Transactions on Information Theory 44 (5), (1998).
6. G. Castagnoli, J. L. Massey, P. A. Schoeller, and N. von Seeman, “On repeated root cyclic codes”, IEEE Trans. on Inf. Theory, vol. 37, pp. 337-342, 1991.
7. Gavrilko E.V. Improving the quality of the functioning of the network of the future through the use of ring codes / Gavrilko E.V., Otrokh S.I., Yarosh V.A., Grishchenko L.N. // Vesnik svyazi -no.2 –pp.60-64, (in Russian) 2018.
8. S. Otrokh, V. Kuzminykh, O. Hryshchenko. Method of forming the ring codes// CEUR Workshop Proceedings - Vol-2318, - 2018, pp. 188-198.

RATIONAL WAVELET TRANSFORM WITH REDUCIBLE RATIONAL DILATION FACTOR

Oleg Chertov[0000-0003-0087-1028] and Volodymyr Malchykov[0000-0002-1710-9171]

Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

mavr2k@gmail.com

Abstract. Wavelet analysis is very effectively used in analysis of different types of data. Mostly often dyadic wavelet transforms are used. But non-dyadic wavelet transforms allow more accurate determination of data features. Commonly used value of dilation factor for rational wavelet transforms is a irreducible fraction. In this paper we will show on an example that for reducible fraction as a dilation factor perfect reconstruction condition is satisfied.

Keywords: Wavelet Transform, Non-Dyadic Wavelet, Dilation Factor.

1 Introduction

1.1 Wavelet Transform

Wavelet transform (WT) is a very powerful tool for analyzing the data of different nature. Unlike Fourier analysis WT gives as a result time-frequency representation of source signal.

Wavelet analysis has shown its efficiency in various areas, such as image and video processing, data compression and noise reduction, solving of partial differential equations, speech recognition, processing of Electroencephalography (EEG), Electromyography (EMG), Electrocardiography (ECG) signals, etc [1-4].

1.2 Dyadic and Non-dyadic Wavelet Transform

Discrete wavelet transform is characterized by its dilation factor. It was shown [5] that as a value of dilation factor any real number greater than one can be taken.

If dilation factor equals 2 the discrete WT is called dyadic, otherwise non-dyadic. Usually dyadic WT is used due to its simple and effective implementation. But in the case when signal singularities will be located in adjoined frequency intervals the results of dyadic WT will not be eligible. Thus, it will be better to select frequency intervals in a way that they would cover such singularities. For example, non-dyadic WT are more suitable for flexible decompositions of the data [6], non-dyadic scale ratios [7],non-dyadic frequency divisions [8], or constructing non-tensor-based multi-D wavelets withcoset sum method [9].

Various approaches to non-dyadic WT were proposed by different authors.

Bratteli and Jorgensen [10] proposed a method based on the construction of the set mapping to Kunzh's algebra representation. In their approach task of non-dyadic WT filters choice reduces to unitary matrix construction. As a dilation factor only a natural number greater than one can be selected.

Pollock and Cascio [8] proposed a method for construction of packet WT with a possibility of a dilation factor choice through each level of decomposition. They generalized dyadic WT packet technique, where each frequency range at each step of wavelet decomposition is divided into two intervals. Main idea was to divide range into p intervals, where p is an arbitrary prime number.

There can be cases where integer dilation factor is not enough. Auscher [11] gave the formal definition of a rational multiresolution analysis and specified the method of corresponding orthogonal wavelet bases construction. Using his ideas, Baussard, Nicolier and Truchetet [12] developed a fast pyramidal algorithm for WT coefficients construction in a case of a rational dilation factor. Thus, they gave a generalization of the Malla algorithm for dyadic case.

Also there are some works that deal with using of irrational dilation factors. One of the first was Feauveau [13] where the value of dilation factor was .

1.3 Rational Wavelet Transform

The most general and effective approach for non-dyadic WT is rational multiresolution analysis proposed in [11] and [12].

In this approach dilation factor N is equal to a rational number p/q. At each level of wavelet decomposition we get one approximation component and p-q detail components. Rational multiresolution analysis filters construction is carried out in the frequency domain. Wavelet decomposition pyramidal algorithm with a rational dilation factor is also transferred to the frequency domain.

2 Problem Formulation

The main advantage of non-dyadic wavelet transform is that it can provide more precise separation of signal components. Very often the irreducible dilation factor is used for such decomposition. Commonly its value is taken 3/2.

The purpose of this work is to show that as value of the dilation factor for rational wavelet transform a reducible fraction 6/4 can be taken and this will allow to get the more precise detection of signal singularities. Also authors will show that perfect reconstruction condition is satisfied for this case.

3 Problem Solution

3.1 Conditions for Building Filters

In order to build filters for rational wavelet transform with dilation factor 6/4 we will use the approach described in [5] for case 3/2.

We have function

$$\varphi \in V_0 \subset V_{-1} = \overline{Span\left\{\varphi\left(\frac{6}{4}\cdot - n\right)\right\}}$$

so that

$$\varphi(x) = \sqrt{\frac{6}{4}}\sum_n h_n^0 \varphi\left(\frac{6}{4}x - n\right)$$

Orthonormality of the $\varphi(\cdot -$ implies the condition

$$\sum_n h_n^0 \overline{h_{n-6l}^0} = \delta_{ol}$$

On the other hand, $\varphi(\cdot -$ is also in V_0, and therefore can also be written as

$$\varphi(x-1) = \sqrt{\frac{6}{4}}\sum_n h_n^1 \varphi\left(\frac{6}{4}x - n\right)$$

Orthonormality of the $\varphi(x - 4l -$ and orthogonality of the $\varphi(x - 4l -$ with the respect to $\varphi(x - ‘$ give us next two conditions

$$\sum_n h_n^1 \overline{h_{n-6l}^1} = \delta_{ol}$$

$$\sum_n h_n^1 \overline{h_{n-6l}^0} = 0$$

Making the similar operations with $\varphi(x -$ and $\varphi(x -$ will give us

$$\sum_n h_n^2 \overline{h_{n-6l}^2} = \delta_{ol}$$

$$\sum_n h_n^2 \overline{h_{n-6l}^1} = 0$$

$$\sum_n h_n^2 \overline{h_{n-6l}^0} = 0$$

$$\sum_n h_n^3 \overline{h_{n-6l}^3} = \delta_{ol}$$

$$\sum_n h_n^3 \overline{h_{n-6l}^2} = 0$$

$$\sum_n h_n^3 \overline{h_{n-6l}^1} = 0$$

$$\sum_n h_n^3 \overline{h_{n-6l}^0} = 0$$

This means that we have four m_0-functions

$$m_0^0(\xi) = \sqrt{\frac{4}{6}}\sum_n h_n^0 e^{-in\xi}$$

$$m_0^1(\xi) = \sqrt{\frac{4}{6}} \sum_n h_n^1 e^{-in\xi}$$

$$m_0^2(\xi) = \sqrt{\frac{4}{6}} \sum_n h_n^2 e^{-in\xi}$$

$$m_0^3(\xi) = \sqrt{\frac{4}{6}} \sum_n h_n^3 e^{-in\xi}$$

Similar to [5] we can also get another two necessary functions

$$m_1(\xi) = \sqrt{\frac{4}{6}} \sum_n g_n^1 e^{-in\xi}$$

$$m_2(\xi) = \sqrt{\frac{4}{6}} \sum_n g_n^2 e^{-in\xi}$$

3.2 Perfect Reconstruction Conditions

Received conditions for filters coefficients are equivalent to the unitarity of matrix

$$\begin{pmatrix} m_0^0(\xi) & m_0^1(\xi) & m_0^2(\xi) & m_0^3(\xi) & m_1(\xi) & m_2(\xi) \\ m_0^0\left(\xi+\frac{\pi}{3}\right) & m_0^1\left(\xi+\frac{\pi}{3}\right) & m_0^2\left(\xi+\frac{\pi}{3}\right) & m_0^3\left(\xi+\frac{\pi}{3}\right) & m_1\left(\xi+\frac{\pi}{3}\right) & m_2\left(\xi+\frac{\pi}{3}\right) \\ m_0^0\left(\xi+\frac{2\pi}{3}\right) & m_0^1\left(\xi+\frac{2\pi}{3}\right) & m_0^2\left(\xi+\frac{2\pi}{3}\right) & m_0^3\left(\xi+\frac{2\pi}{3}\right) & m_1\left(\xi+\frac{2\pi}{3}\right) & m_2\left(\xi+\frac{2\pi}{3}\right) \\ m_0^0\left(\xi+\frac{3\pi}{3}\right) & m_0^1\left(\xi+\frac{3\pi}{3}\right) & m_0^2\left(\xi+\frac{3\pi}{3}\right) & m_0^3\left(\xi+\frac{3\pi}{3}\right) & m_1\left(\xi+\frac{3\pi}{3}\right) & m_2\left(\xi+\frac{3\pi}{3}\right) \\ m_0^0\left(\xi+\frac{4\pi}{3}\right) & m_0^1\left(\xi+\frac{4\pi}{3}\right) & m_0^2\left(\xi+\frac{4\pi}{3}\right) & m_0^3\left(\xi+\frac{4\pi}{3}\right) & m_1\left(\xi+\frac{4\pi}{3}\right) & m_2\left(\xi+\frac{4\pi}{3}\right) \\ m_0^0\left(\xi+\frac{5\pi}{3}\right) & m_0^1\left(\xi+\frac{5\pi}{3}\right) & m_0^2\left(\xi+\frac{5\pi}{3}\right) & m_0^3\left(\xi+\frac{5\pi}{3}\right) & m_1\left(\xi+\frac{5\pi}{3}\right) & m_2\left(\xi+\frac{5\pi}{3}\right) \end{pmatrix}$$

Let’s know take the proposed in [10] Littlewood-Paley rational wavelet basis and check that perfect reconstruction condition will be satisfied for this basis in the case of dilation factor 6/4.

3.3 Experimental Results

In order to illustrate that using of the reducible dilation factor can lead to more precise identifying of signal singularities we will take a look at the example which is based on model data.

This mode data is the sum of three sinusoidal signals (see Fig. 1).

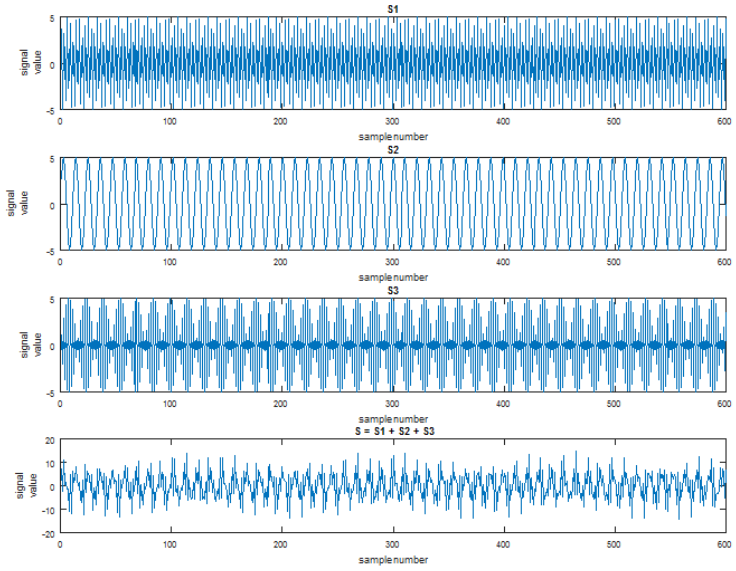


Fig.1. Model signal

Fourier spectrum of this model signal contains three peaks in frequency domain, that fall into the following frequency ranges:

- first signal – from $2\pi/3$ to $5\pi/6$
- second signal – from 0 to $2\pi/3$
- third signal – from $5\pi/6$ to π

When using the dilation factor 2/3 frequency range is divided into two parts – from 0 to $2\pi/3$ for approximation component and from $2\pi/3$ to π for detail component. Due to this rational wavelet transform will separate signal S1, but signal S2 and S3 will be mixed (see Fig. 2).

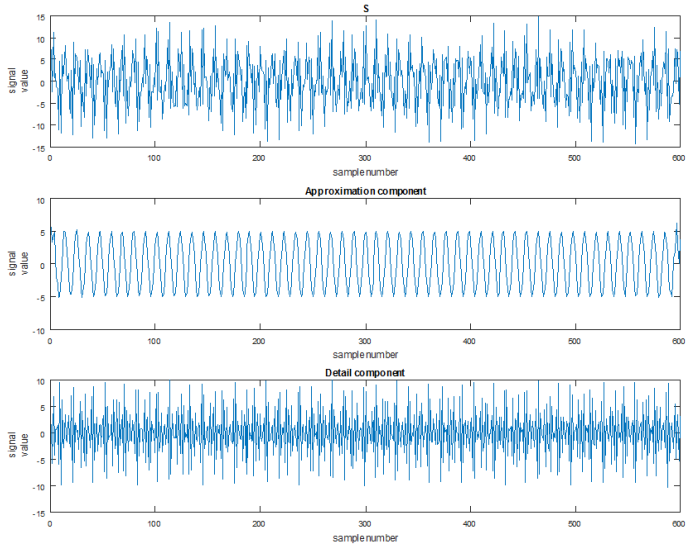


Fig.2. Decomposition of model signal with dilation factor 3/2

But if we take dilation factor 6/4 then frequency range is divided into three parts – again from 0 to $2\pi/3$ for approximation component, but now from $2\pi/3$ to $5\pi/6$ for the first detail component and from $5\pi/6$ to π for the second detail component.
Thus, in this case all three source signals will be separated (see Fig. 3).

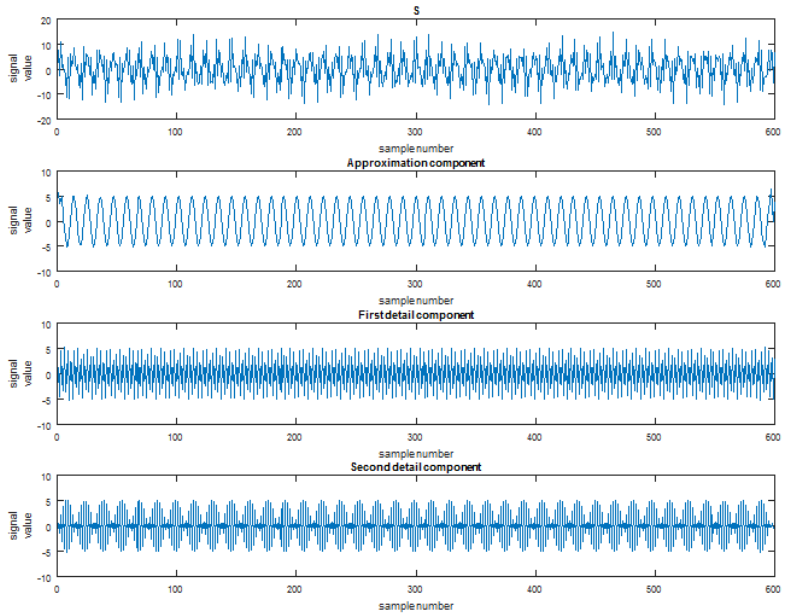


Fig.3. Decomposition of model signal with dilation factor 6/4

So, for this model signal the use of dilation factor 6/4 is preferable, since it allows to separate all three components of the original signal.

4 Conclusions

Authors proposed to use reducible rational dilation factor 6/4 except most often used 3/2. It was shown that for this value of dilation factor perfect reconstruction conditions are satisfied. An example shows that using dilation factor 6/4 allows more precise separation of signal singularities.

Further researches have to be done in order to generalize the results to the case of arbitrary reducible dilation factor.

References

1. Truchetet, F., Laligant, O.: Industrial applications of the wavelet and multi-resolution-based signal/image processing: a review. In: Eighth International Conference on Quality Control by Artificial Vision, 63560H (May 29, 2007).
2. Hussain, L., Aziz, W.: Time-Frequency Wavelet Based Coherence Analysis of EEG in EC and EO during Resting State. International Journal of Information Engineering and Electronic Business7(5), 55–61, (2015).
3. Yazdani, H.R., Nadjafikhah, M., Toomanian, M.: Solving Differential Equations by Wavelet Transform Method Based on the Mother Wavelets & Differential Invariants. Journal of Prime Research in Mathematics 14, 74-86 (2018).
4. Vishwa, A., Sharma, S.: Modified Method for Denoising the Ultrasound Images by Wavelet Thresholding. International Journal of Intelligent Systems and Applications4(6), 25–30, (2012).
5. Daubechies, I.: Ten Lectures on Wavelets. SIAM (1992).
6. Gupta, C.,Lakshminarayan, C., Wang, S., Mehta, A.: Non-dyadic haar wavelets for streaming and sensor data.In: 2010 IEEE 26th International Conference on Data Engineering, pp. 569–580. (2010).
7. Xiong, R., Xu, J., Wu,F.: A lifting-based wavelet transform supporting non-dyadic spatial scalability. In: 2006 IEEE International Conference onImage Processing, pp. 1861–1864. (2006)

8. Pollock,D.S.G., Cascio,I.L.: Non-dyadic wavelet analysis. In:Kontoghiorghes, E.J.,Gatu, C. (eds.) Optimization, Econometric and Financial Analysis: Advances In Computational Management Science, vol. 9, pp. 167–204(2007).
9. Hur, Y., Zheng, F.: Prime Coset Sum: A Systematic Method for Designing Multi-D Wavelet Filter Banks With Fast Algorithms. IEEE Transactions on Information Theory 62(11), 6580–6593 (2016).
10. Bratteli, O.,Jorgensen,P.E.T.: Isometrics, shifts, Cuntz algebras and multiresolutionwavelet analysis of scale N. Integral Equations Operator Theory 28(4), 382–443 (1997).
11. Auscher, P. Wavelet bases for $L^2(R)$ with rational dilation factor. In: Ruskai, M.B.et. al. (eds.) Wavelets and their Applications. Jones and Barlett (1992).
12. Baussard, A., Nicolier, F., Truchetet, F.: Rational multiresolution analysis and fast wavelet transform: application to wavelet shrinkage denoising. Signal Processing84(10), 1735–1747 (2004).
13. Feauveau, J.-C.: Analysemultiresolution avec un facteur de resolution $\sqrt{2}$. J. Traitement du Signal 7(2), 117–128 (1990).

МОДЕЛИРУЮЩИЕ КОМПЛЕКСЫ АВТОМАТИЗИРОВАННЫХ СИСТЕМ ОРГАНИЗАЦИОННОГО УПРАВЛЕНИЯ – СРЕДСТВО ПРЕОДОЛЕНИЯ КРИЗИСА УПРАВЛЕНИЯ

А.Г. Додонов[2][0000-0001-7569-9360], А.В. Никифоров[1][0000-0002-0207-5175],
В.Г. Путятин[2][0000-0002-5575-9159], В.А. Додонов[2][0000-0003-2031-1941]

[1] *Харьковский национальный университет Воздушных Сил им. Ивана Кожедуба, ул. Сумская 77/79, Харьков 61023, Украина*
[2] *Институт проблем регистрации информации, Национальная академия наук Украины, ул. Н. Шпака 2, Киев 03113, Украина*
aleksey.nikiforov.62@gmail.com

Аннотация. Современный кризис управления связан с недостаточно быстрой адаптацией существующих институциональных систем организационного управления к изменению управленческих концепций и парадигм, а также с несоответствием используемых математических моделей и методов принятия решений фактической сложности реальных управляемых процессов. В настоящее время во многих научных организациях ведутся исследования по созданию таких автоматизированных систем управления, которые бы позволяли осуществлять их перестройку в максимально широких пределах и в предельно сжатые сроки, обеспечивая при этом приемлемый уровень сложности (размерности) управленческих задач. В ИПРИ также ведутся такого рода исследования. При этом, в качестве рабочего инструментария, используется специальная среда проектирования, получившая название – моделирующего комплекса системы управления. С помощью данного инструментария удобно вести разработку современных автоматизированных систем организационного управления в рамках установленной концепции управления. Смена концепции заставляет изменять первоначальный проект. Трудность внесения изменений достаточно велика, что замедляет доведение системы до практического внедрения. Авторами исследованы возможности и предложены направления научных исследований по разработке теории проектирования систем организационного управления, которые были бы приспособлены к быстрым концептуальным изменениям в автоматизируемых формах деятельности. В статье представлены теоретические результаты, полученные для проектирования систем организационного управления авиационными и морскими силами. Показаны потребности по дальнейшему развитию имеющихся результатов с учётом расширения тактических норм применения сил. Сформулированы научно-исследовательские задачи по развитию имеющихся результатов на основе методологии концептуального проектирования систем и тензорного анализа сетей.

Ключевые слова: организация, организационное управление, моделирующий комплекс, система управления, управление силами, концептуальное проектирование, конструкт, тензорный анализ.

1. Кризис управления

Современный мир переживает очередной кризис своего развития. Во многом это проявляется как управленческий кризис, который проявляется в виде: нерационального использования ресурсов, задания неадекватных целей, некорректного использования производственных сил, замедленности реагирования на изменяющиеся условия, невостребованности многих управленческих функций при отсутствии функций управления, учитывающих современные социально-экономические тенденции и явления. Всё это приводит к тому, что мы всё чаще слышим о лишних миллиардах народонаселения

планеты Земля, о голодающих странах, о кризисе перепроизводства, о возрастающем национальном долге, о стагнирующих и депрессивных национальных экономиках.

Перечисленные явления носят глобальный характер, с разной степенью тяжести они проявляются во всех странах, не только на Украине. Причина этого в несоответствии современных систем управления экономикой быстроменяющемуся технологическому укладу. Создаваемые управленческие институты (организации), используемые ими процедурные регламенты, быстро устаревают, переставая соответствовать реальности. Сейчас среднестатистическому специалисту, чтобы соответствовать требованиям времени, необходимо с периодичностью один раз в пять – семь лет существенно пересматривать и дополнять свои знания о профессии или даже менять профессию. Не делающие этого, вынужденно перемещаются на периферию в своей области профессиональной деятельности, теряют социальный статус. То же самое можно сказать и об организациях. Создаваемые для решения конкретных задач [1], они должны подвергаться перманентному реформированию в смысле пересмотра (расширения) используемых форм деятельности и даже изменения самой концепции их создания. Страны, пренебрегающие такой адаптацией своих систем организационного управления, теряют субъектность (суверенность) и ввергают своё население во многие бедствия.

Формы проявления несоответствия существующих систем организационного управления характеру и сложности управляемых процессов весьма разнообразны. В [2] это описано на примере проблемы недостаточной эффективности системы управления процессами развития авиационных сил Украины. Авторами показано, что к проблемам управления здесь следует отнести:

– ресурсно-целевую несбалансированность программ развития или, иначе, разрыв между стратегическими целями развития и частными программами развития отдельных компонент сил;

– затруднённость маневрирования ресурсами при корректировке программ развития, недостаточную управляемость процессами развития авиационных сил, а также неэффективную реализацию (игнорирование) отдельных частных задач развития по причине несогласованности бизнес интересов предприятий промышленности и целей развития ВС.

Проблемы организационного управления процессами применения вооружённых сил затронуты в [3]. Отмечается, что быстроменяющаяся обстановка при ведении боевых действий приводит к острому дефициту времени на согласование и координирование действий сил при корректировке ранее принятого решения. Старые подходы не позволяют осуществлять скоординированное маневрирование силами, укладываясь в столь ограниченные временные рамки. Эффективность применения сил падает.

Академики В.С. Немчинов [4] и В.М. Глушков [5] также подчёркивали, что для систем организационного управления характерна весьма большая размерность пространства принятия решений. Это объясняется множественностью характеристик функции полезности, а также множественностью аспектов интерпретации результатов. Именно по этой причине не получилось создать автоматизированную систему управления государством, используя концепцию программно-целевого планирования на основе декомпозиции задач оптимизации. Проблема формирования критерия оптимальности и корректного агрегирования данных большой размерности так и не была решена. Создаваемые методы и модели для автоматизации функций организационного управления не позволяли отображать сложность реальных процессов в их полноте, достаточной для практики управления большими организациями.

Таким образом, кризис управления объясняется чрезвычайно большой сложностью процессов организационного управления, требующей расширения распространённой концепции программно-целевого и оптимизационного подходов, а также ускорением общественных процессов, требующих ускоренной перестройки систем управления. Естественным выходом из создавшегося положения видится дальнейшее расширение сферы применения средств автоматизации управления, с всесторонним охватом функций организационного управления, [6] и при условии совершенствования теории проектирования таких систем [7]. То есть, проблема совершенствования методов проектирования систем организационного управления, адекватных современным условиям и сложности объектов управления, приобретает особою остроту, её решение является актуальной научной задачей.

2. Моделирующие комплексы – как средство проектирования эффективных систем организационного управления

В настоящее время в научных организациях (в том числе и в Институте проблем регистрации информации (ИПРИ)), занимающихся проектированием автоматизированных систем управления, получила распространение форма исследований и проектирования, которая осуществляется в рамках разработки моделирующих комплексов автоматизированных систем управления различного назначения.

Компьютерный моделирующий комплекс с соответствующим инструментарием, как отмечается в [8], является средой разработки систем организационного управления, моделирования управленческих процессов и решения задач автоматизации этих процессов. Моделирующие комплексы дают возможность отслеживать в реальном масштабе времени динамику управления, анализировать процессы принятия решений и другие процедуры. Применение моделирующих комплексов целесообразно не только с экономической точки зрения, поскольку они, по сути, являются тренажёрами, но безальтернативно в областях, связанных с большими рисками для безопасности людей, природной и искусственной среды (военная, космическая, энергетическая сферы).

В ИПРИ до настоящего времени был проведен ряд поисковых и опытно-конструкторских работ по созданию следующих моделирующих комплексов:

- ведения мониторинга воздушного, надводного и наземного пространства [9];
- автоматизированной системы управления авиационным комплексом [8];
- командной системы управления корабельным авианосным соединением (в стадии завершения).

Подводя итоги проделанной работы в аспекте создания теории проектирования систем организационного управления, следует отметить, что в рамках перечисленных проектов получены следующие результаты.

Во-первых, была описана предметная область управления авиационными и морскими силами. Описание выполнено в виде системы понятий, связей между ними, связей между совокупностями (фреймами) понятий. Описание предметной области было зафиксировано в виде текстуального описания управляемых процессов и логических моделей баз данных моделирующих комплексов. Фрагмент логической модели данных при описании предметной области управления силами приведен на рис. 1.

Описана система исходных данных, необходимых для проведения оперативных расчётов при планировании применения сил и при оперативном управлении силами и средствами в ходе их применения. К таковым данным, например, относятся:

- тактико-технические характеристики образцов вооружения и техники;
- данные по расчёту дальности и продолжительности полёта летательных аппаратов;
- данные по расчёту эффективности применения авиационных средств поражения;
- данные по расчёту области поражения зенитных ракетных комплексов;
- данные по расчёту дальности действия радиотехнических систем и комплексов;
- данные по расчёту дальности действия средств гидролокации.

Во-вторых, сформированы информационные объекты, являющиеся результатом информационных преобразований в системе организационного управления силами. Это:

- план применения сил и средств;
- команды оперативного управления силами и средствами.

Данные объекты имеют сложную структуру и состоят из иерархически вложенных информационных элементов разных уровней детализации. Так, план применения (рис. 2) состоит из отдельных тактических задач (действий), которые имеют логико-временные связи с другими действиями и персонифицированы относительно частей (подразделений), тактических групп, отдельных кораблей, самолётов из состава группировки сил. Тактические задачи (действия) состоят из тактико-технических действий, которые образуют связи с остальными тактико-техническими действиями и персонифицированы относительно боевых расчётов, экипажей, отдельных средств, комплексов, самолётов. Тактико-технические действия представляются как совокупность технических действий, относимых к отдельным агрегатам, системам, комплексам. Каждый информационный элемент из состава плана применения характеризуется определённой системой численных метрик, с помощью которых определяется порядок, время, место и объём применения сил и средств.

В структурной таблице действий (табл. 1) содержатся:

- наименования тактических, тактико-технических и технических действий;
- сведения о действиях, выполнение которых предшествует началу (является условием начала) выполнения рассматриваемого действия;
- индекс варианта выполнения действия;
- значения параметров, характеризующих рассматриваемый вариант выполнения действия. В качестве таких параметров могут рассматриваться различные временные, пространственные, иные характеристики определяющие порядок применения сил и средств, достигаемую при этом результативность (характеристика обратной связи), а также сведения о персонификации действий (кто выполняет, с помощью чего).

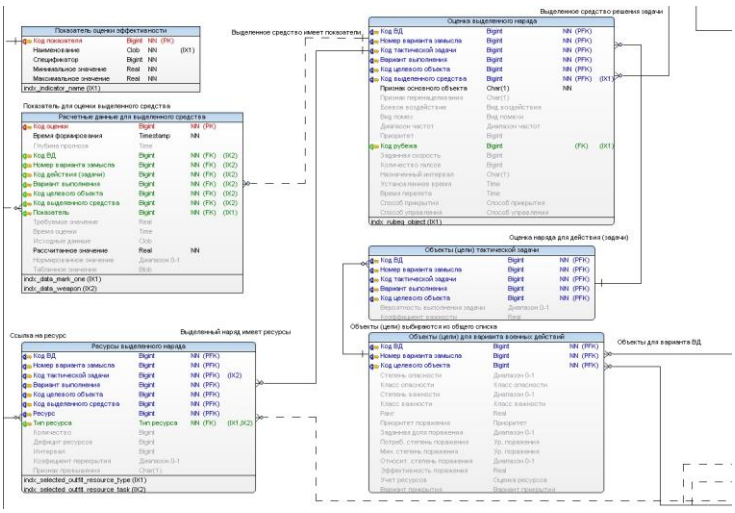


Рис.4. Фрагмент логической модели данных при описании предметной области управления авиационными и морскими силами.

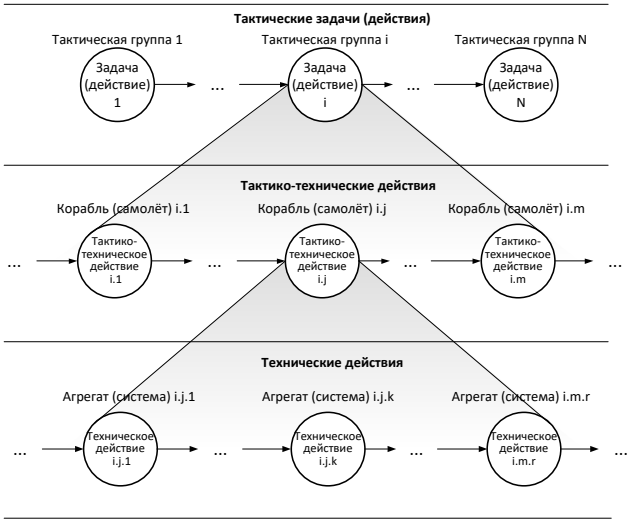


Рис.5. Структура информационных элементов, представляющих план применения сил и средств.

Таблица 1. Структурная таблица действий (тензор принятия решений).

Наименование действия	Действия-предшественники	Вариант	Параметры действий			
			x_1	x_2	x_3	x_4 …
Тактическое действие №1	-	вариант №1.1	$x_{1,1,1}$	-	-	-
- тактико-техн. действие 1.1	-	вариант №1.1.1	-	$x_{2,1,1,1}$	-	-
- тактико-техн. действие 1.2	1.1	вариант №1.2.1	-	-	$x_{3,1,2,1}$	-
		вариант №1.2.2	-	-	$x_{3,1,2,2}$	-
Тактическое действие №2	1	вариант №2.1	-	-	-	$x_{4,2,1}$
…	…	…	…	…	…	…

В-третьих, предложен многоаспектный подход принятия решений на вариативной сети большой размерности. По своей сути структурная таблица действий представляет собою сеть с варьированием связей между узлами этой сети (тактическими, тактико-техническими и техническими действиями) и варьированием отношений персонификации (распределения задач между исполнителями). При принятии решения на применение сил из множества возможных сетей, связывающих элементы плана, выбирается какой-либо один или несколько вариантов сетевых графиков последовательности и порядка действий сил. Выбор вариантов решения осуществляется на основе упорядочения вариантов

выполнения действий согласно вводимым наборам частных критериев, которые определены на множестве характеристик действий

$$K_i = K_i(x_1, x_2, ..., x_N),\ i = 1, ..., M, \quad (1)$$

где K_i – частный критерий принятия решения; M – количество учитываемых частных критериев.

Выбор того или иного набора частных критериев зависит от учитываемых аспектов управления силами:

- характера стратегии (сокрушение, истощение);
- характера действий (оборонительные, наступательные);
- типа решаемых задач (противовоздушная оборона, противолодочная оборона, поддержка наземных сил и так далее).

Следует отметить, что множество характеристик действий $\mathbf{X} = \{x_j\}$, учитывая разнообразие форм применения сил и средств, а также множественность аспектов, рассматриваемых при принятии решений, должно быть достаточно велико. Поэтому задача упорядочивания вариантов при принятии решений на применение сил является не тривиальной.

В-четвёртых, предложен подход формирования команд оперативного управления силами на множестве элементов тензора принятия решений. Команды оперативного управления представляют собою персонифицированные фрагменты выбранного при принятии решения на применение сил сетевого графика действий с внесёнными изменениями по характеристикам этих действий

$$\Psi_S = pl_S(\mathbf{X}_S + \Delta\mathbf{X}_S),\ \ pl_S \in Pl(\mathbf{X}),\ \ \mathbf{X}_S \in \mathbf{X},\ s = 1, \ldots, S, \quad (2)$$

где Ψ_S – команда для *s*-го исполнителя; pl_S – фрагмент плана, персонифицированный относительно рассматриваемого исполнителя; $\mathbf{X}_S$ – множество характеристик, относящихся к действиям, которые включены во фрагмент плана; $\Delta\mathbf{X}_S$ – изменения характеристик рассматриваемых действий; $Pl(\mathbf{X})$ – используемый план-решение при оперативном управлении силами; S – количество исполнителей.

Изменение характеристик действий осуществляется в соответствии с текущей обстановкой, с учётом фактического состояния своих сил и средств.

В-пятых, были разработаны варианты процедурных регламентов по проведению оперативных расчётов при планировании действий сил и при управлении ими в ходе применения. В результате выполнения процедур, прописанных в регламенте, на этапе планирования, формируется структурная таблица вариантов действий с установленными значениями характеристик действий и связями между ними, иначе – тензор принятия решений. На этапе оперативного управления – осуществляется выбор фрагмента структурной таблицы и корректировка значений характеристик элементов в выбранном фрагменте. По сути, предложенный процедурный регламент есть последовательность проведения оперативно-тактических расчётов в иерархии органов управления силами. В режиме планирования график последовательности решения функциональных задач имеет вид глобального алгоритма поиска рационального (оптимального) решения на множестве характеристик действий, $\mathbf{X}$. На рис. 3 показан пример такого графика для случая подготовки исходных данных для принятия решения по применению морских сил при решении задач разведки. В режиме оперативного управления график решения функциональных задач имеет вид следящего контура. На рис. 4 показан пример последовательности решения функциональных задач при оперативном управлении силами при решении аналогичных задач, плюс задач радиоэлектронной борьбы (РЭБ). На приведенных рисунках расчётные процедуры изображены в виде прямоугольников с номерами. Данные прямоугольники являются вершинами сетевого графика. Последовательность выполнения расчётов определяется с помощью рёбер сетевого графика. Суть расчётов, выполняемых с помощью отдельных процедур или их групп, представлена с помощью выносок-комментариев.

Процесс решения функциональных задач при преобразовании управленческой информации, будучи распределённым по различным органам (узлам) принятия решений, в моделирующем комплексе реализуется с помощью процедуры диспетчеризации. На рис. 3, например, задачами-диспетчерами являются процедуры под номерами 5 и 18. Процедура диспетчеризации позволяет создать единое информационное и функциональное пространство для лиц, участвующих в принятии решения. Их работа при использовании средств автоматизации, например, в режиме планирования выглядит как процесс заполнения нескольких вариантов карты командира, на которою последовательно наносятся дополнительные элементы обстановки и планируемых действий сил (рис. 5).

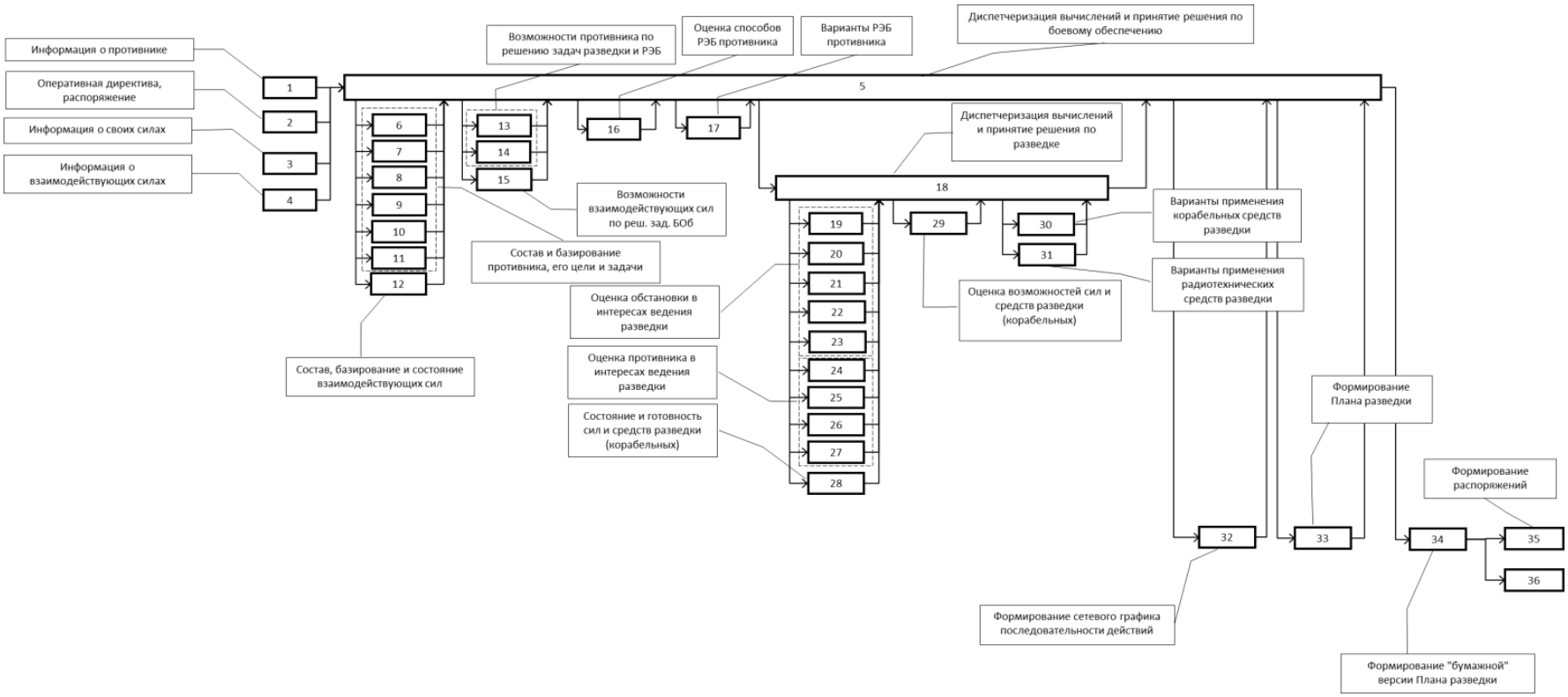


Рис.6. Последовательность решения функциональных задач специального программного обеспечения моделирующего комплекса при подготовке данных для принятия решения на этапе планирования

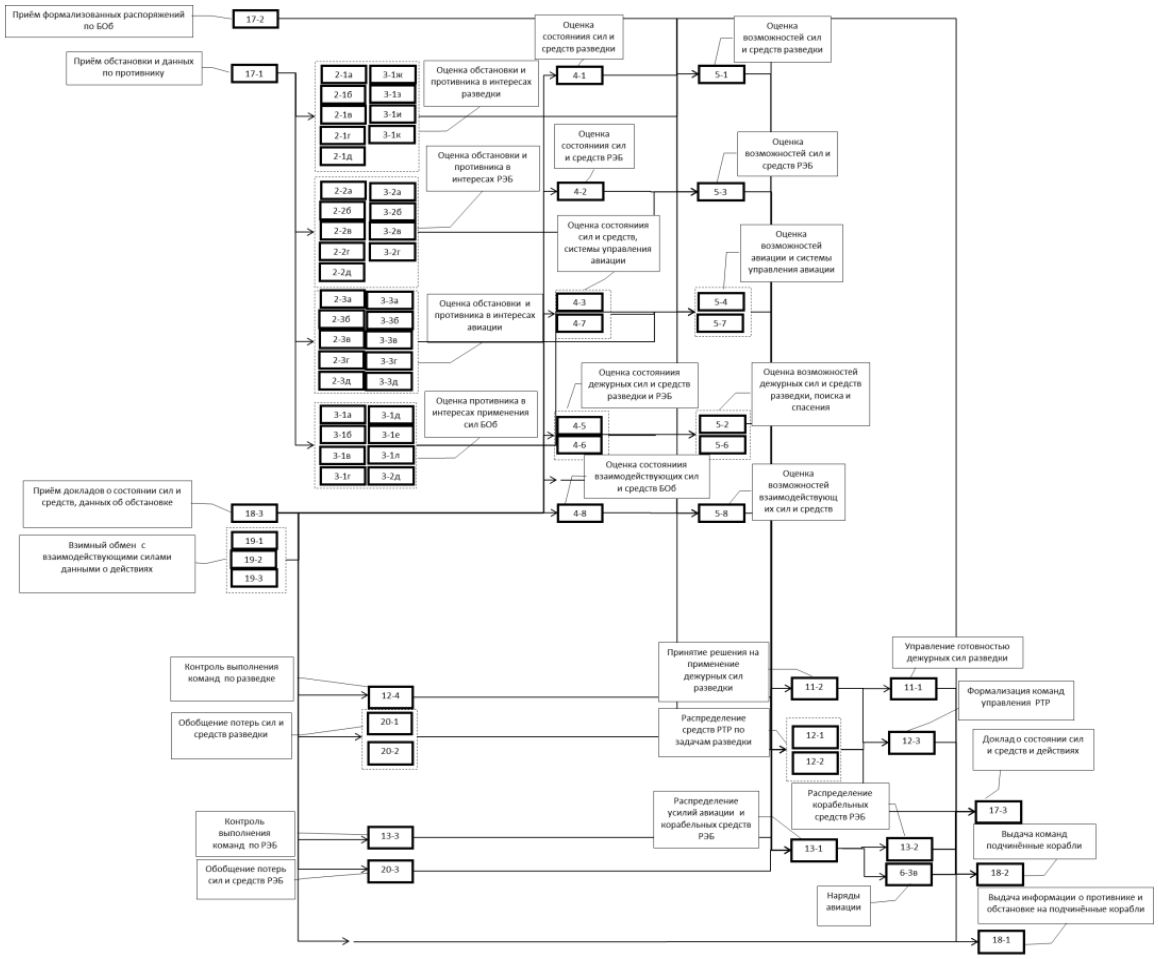


Рис.7. Последовательность решения функциональных задач специального программного обеспечения моделирующего комплекса при оперативном управлении силами.

Спроектированная система организационного управления силами функционирует следующим образом. На этапе планирования лицо, принимающее решение, (командующий) формирует замысел применения подчинённых ему сил. Замысел выглядит как некоторая последовательность выполнения укрупнённых тактических задач, определённых во времени, пространстве, и закреплённых за соответствующими исполнителями (организационно-штатными формированиями). Сформированный замысел поставлен в соответствие определённому варианту действий противника и условий применения сил. Из оглашённого замысла применения системой вычленяются эпизоды, соответствующие различным аспектам управления.

Классификация эпизодов происходит таким образом, чтобы они образовывали простую последовательность по времени (в любой момент времени, принадлежащий периоду применения сил, отрабатывается только один эпизод). Например, на рис. 6 показана последовательность эпизодов замысла применения морских сил при разгроме морского десанта противника. Это: эпизод выхода из пункта постоянной дислокации и построения походного порядка; эпизод перехода в район тактического развёртывания и эпизод развёртывания сил в боевой порядок и нанесения комплексного огневого поражения противнику.

Для каждого эпизода применения сил, при подготовке данных в решение командира, характерен свой порядок преобразования информации посредством решения функциональных задач специального программного обеспечения моделирующего комплекса. Этот порядок определяется в зависимости от аспекта управления силами. Система определяет этот аспект и ставит в соответствие рассматриваемому эпизоду график последовательности проведения расчётов. В рамках установленного порядка выполнения расчётов регулируется их вариативность и, тем самым, создаётся пространство для принятия решения.

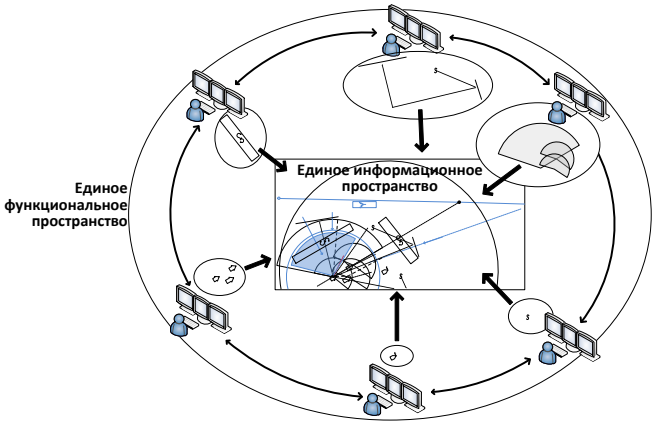


Рис.8. Автоматизированный механизм распределённого принятия решений.

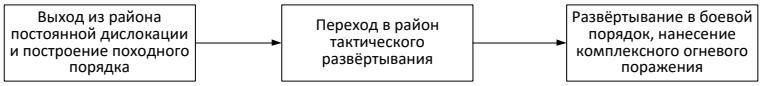


Рис.9. Эпизоды замысла применения сил.

По результатам расчётов происходит заполнение структурной таблицы действий (тензора принятия решений). Заполнение тензора осуществляется по методу динамического программирования в соответствии со схемой Беллмана. Сначала происходит решение функциональных задач для эпизода нанесения комплексного огневого поражения (третьего эпизода). Затем, с учётом характеристик, рассчитанных для третьего эпизода, тензор принятия решений дополняется данными, относящимися к эпизоду перехода в район тактического развёртывания (второму эпизоду). Далее расчёты выполняются для эпизода выхода из пункта постоянной дислокации (первого эпизода), но с учётом данных, рассчитанных для второго и третьего эпизодов применения.

При принятии решения также используется принцип динамического программирования. Сначала на множестве вариантов действий, относящихся к нанесению комплексного огневого поражения, с помощью группы частных критериев, выбирается один или несколько вариантов применения сил. Далее для каждого варианта нанесения комплексного огневого поражения выбирается по одному или по несколько вариантов перехода. Для выбора вариантов перехода используется группа частных критериев, охватывающих соответствующий аспект управления. Варианты выхода из пункта постоянной дислокации подбираются для установленных вариантов перехода при использовании заданных частных критериев.

Такое сокращённое множество вариантов действий применения сил используется для последующего анализа и окончательного принятия решения командующим.

На этапе оперативного управления ходом применения сил происходит отслеживание параметров обстановки (погодных условий, действий противника, состояния своих сил) на предмет их соответствия принятому решению. При возникновении рассогласований вносятся коррективы в параметры порядка применения сил и средств. При существенном рассогласовании плана и обстановки происходит корректировка решения с использованием ранее сформированного тензора принятия решений. Вносятся изменения в план с учётом фактической обстановки.

5 Нерешённые проблемы и актуальные направления дальнейших исследований

Как видно из представленных результатов проектирования моделирующих комплексов, в ИПРИ достигнуты определённые результаты в построении теории проектирования систем организационного управления.

Главный теоретический результат – это разработка метода формирования тензора принятия решений (структурной таблицы действий). Полученный результат включает:

– определение формы структурной таблицы;

– формализацию элементов структурной таблицы в виде тактических, тактико-технических и технических действий, характеризуемых наборами численных метрик и логико-временных связей;

– последовательность проведения расчётов по определению значений численных метрик действий, заполняющих структурную таблицу;

– синтез последовательности комплексных расчётов при формировании структурной таблицы для многоаспектного управления силами;

– методы многоаспектного принятия решений на применение сил на основе структурной таблицы действий большой размерности;

– механизм управления процессом решения функциональных задач, реализующий принцип распределённого принятия решений.

Однако, вышеперечисленное ещё далеко не полностью исчерпывает решение задачи построения эффективной теории проектирования систем организационного управления силами, приспособленной для преодоления современного кризиса управления (сложность объекта управления и адаптация системы управления к быстроменяющимся условиям). Каковы научные задачи, которые необходимо решить в данной области, чтобы приблизиться к желаемому решению обозначенной проблемы, исходя из сегодняшних представлений и знаний в этой области?

При проектировании любой системы организационного управления приходиться создавать следующие её основные элементы (рис. 7):

– систему понятий или модель системы;

– систему аксиом или начальных данных для регламента функционирования системы;

– систему процедур, которая реализуется при функционировании объекта управления и в самой системе управления.

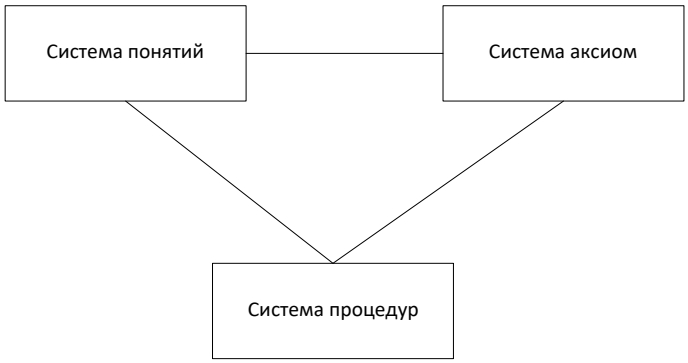


Рис.10. Составные элементы типового проекта системы организационного управления.

Указанные элементы, при завершении проекта, обретают целостность. Эта целостность имеет форму либо когнитивной, либо онтологической сети деятельности.

В представленных результатах, которые были получены в рамках проектирования моделирующих комплексов автоматизированных систем управления авиационным комплексом и корабельным соединением, система понятий (модель системы) существует в виде описания процессов управления, а также логической модели базы данных, включая и тензор принятия решений. Система аксиом представлена в виде нормативных положений по тактике применения сил, а также системы исходных данных, на основе которых проводятся оперативные расчёты при планировании и оперативном управлении силами. Тактические нормы присутствуют в виде настроек расчётных алгоритмов, а исходные данные вводятся непосредственно в базу данных моделирующего комплекса. Система процедур – это тактические, тактико-технические и технические действия сил, а также расчётные задачи по определению характеристик этих действий, увязанные в логическую сеть.

Разработанная автоматизированная система управления силами, по итогам проектирования, является корпоративной. Иными словами, она – уникальна и применима только для управления конкретными силами. При определении концепции построения системы в большой мере был использован накопленный опыт применения и управления силами соответствующего вида, а также – методы системного анализа и математического моделирования. Созданная система организационного управления является адекватной существующему на сегодняшний день уровню сложности объекта управления, то есть – эффективна. При дополнении перечней используемых понятий, введении новых тактических норм, изменении или расширении концепции управления силами, спроектированная система утратит свою эффективность, если не внести в программное обеспечение соответствующих настроек. Вносимые в какой-либо элемент программного обеспечения (в систему процедур) изменения, вероятнее всего, затронут другие элементы. Изменения этих новых элементов затронет следующие части программного обеспечения. Так по нарастающей, в геометрической прогрессии, будет расти трудность проведения модернизации первоначального проекта. Кроме того, многократные модернизации могут приводить к противоречивым изменениям. Для компенсации противоречий нужно будет вводить дополнительные контуры обработки информации. В конце

концов, под тяжестью внутренних проблем, старый проект «умрёт», потому что легче будет начать новый проект «с чистого листа», чем вносить коррекции в существующую систему.

Возможно, что с таким положением дел можно было бы и мириться. Во всяком случае, у разработчиков гарантированно будет постоянная работа. Однако не всегда такой режим перманентного сопровождения проекта возможен. Например, при работе с зарубежным заказчиком организация подобного рода работы сопряжена с большими трудностями по осуществлению межгосударственного сотрудничества и чревата имиджевыми потерями для разработчика.

В связи с этим, говоря о дальнейшем развитии теории проектирования систем организационного управления силами, хотелось бы получить методы непротиворечивого изменения системы процедур (программного обеспечения) при изменении системы понятий и системы аксиом. То есть, например, на вооружение был принят новый тип оружия, который в значительной степени отличается по возможностям от средств, учтённых при разработке расчётных задач. Использование нового вооружения расширяет тактику применения сил. Для нынешнего состояния теории проектирования, чтобы учесть фактор появления нового типа оружия, необходимо существенно дополнить и скорректировать систему расчётных задач по определению численных метрик планируемых действий. При наличии же методов непротиворечивого изменения системы процедур, достаточно внести изменения в систему понятий (модель рассматриваемой системы) и систему аксиом (тактику сил и возможности вооружений), далее система автоматически изменяет расчётные процедуры и процедуры управления. В данном случае существенно сокращается время адаптации системы управления к изменившимся условиям. Кризис управления, о котором речь шла вначале, может быть преодолён за счёт быстрой перестройки управления в изменяющихся условиях.

Конструирование процедурного регламента в автоматическом (автоматизированном) режиме можно осуществить на основе некоторого набора исходных (базовых) процедурных блоков (конструктов) [12]. К настоящему времени коллективом исследователей под руководством академика С.П. Никанорова разработаны совокупности таких конструктов, которые характерны для моделей (концепций) систем различных типов [13]. Для проектируемой системы организационного управления силами это такие типы систем, как: открытая статическая система (для режима планирования) и целенаправленная потоковая система (для режима оперативного управления). В [12] предложены модели систем и соответствующие им методы синтеза процедурных регламентов, построенные на основе наборов конструктов, характерных для этих системных моделей. В основу метода синтеза непротиворечивых систем процедур, исходя из системы аксиом и системы понятий (модели системы), положен подход тензорного анализа сетей [10].

Следовательно, дальнейшим развитием теоретических результатов, которые были получены в рамках создания упомянутых моделирующих комплексов, должно стать:

– вычленение из системы созданных вычислительных процедур фрагментов, которые могут быть интерпретированы как конкретизация конструктов из перечней, характерных для рассматриваемых систем;

– дополнение множества интерпретированных процедурных фрагментов, процедурами, интерпретирующими оставшиеся незатронутые конструкты;

– дополнение множества аксиом сведениями о тактике применения сил;

– разработка процедур синтеза систем вычислений по планированию применения сил и оперативного управления ими как конкретизации результатов конструктивного проектирования процедурного регламента;

– расширение полученных результатов на весь класс систем организационного управления силами, включая процессы управления развитием, безопасностью, подготовкой.

Для решения перечисленных научных задач целесообразно использовать следующую схему исследований (рис. 8).

На первом этапе необходимо провести идентификацию классов моделей систем, которые используются при проектировании. Для классифицированных моделей произвести их идентификацию с конкретной предметной областью. Конкретизировать понятия, их взаимосвязи, определить системы конструктов, с помощью которых могут быть описаны предметная область и процессы функционирования в организационной системе рассматриваемого класса.

Далее следует интерпретировать установленный перечень конструктов на системе расчётных процедур, вычленив из неё базовые расчётные алгоритмы, соответствующие определённым конструктам.

Также из исходного перечня расчётных процедур следует выделить аксиоматические положения по тактике применения сил, включив их в систему аксиом.

Для базового перечня расчётных алгоритмов, используя методы тензорного анализа, необходимо разработать методику синтеза процедурного регламента системы организационного управления на основании принятых аксиом управления (исходных данных по расчёту возможностей вооружений и специальной техники, базовых положений по тактике применения сил).

Конечным результатом такого рода исследований должна стать теория проектирования систем организационного управления с настраиваемой системой процедур функционирования в зависимости от изменений в концепции применения сил и управления процессами их применения.

6 Заключение

Таким образом, разворачивающийся в настоящее время кризис управления обусловлен, главным образом, недостаточно быстрой адаптацией к быстроменяющимся условиям существующих институциональных систем организационного управления, а также несоответствием используемых математических моделей и методов принятия решений фактической сложности реальных управляемых процессов.

В ИПРИ, с использованием среды проектирования моделирующих комплексов автоматизированных систем управления, ведутся работы по созданию эффективных систем организационного управления силами для зарубежных заказчиков. На сегодняшний день достигнуты значительные результаты, позволяющие говорить об автоматизации управления авиационным комплексом и морскими силами. Разработанные автоматизированные системы организационного управления адекватны существующей концепции применения сил и средств и существенно повышают эффективность работы органов управления за счёт использования методов распределённого принятия решений на основе онтологической сети большой размерности в режимах планирования применения и оперативного управления применением сил.

Тем не менее, постоянно растущие возможности современных средств ведения вооружённой борьбы, появление новых видов вооружений, в том числе на новых физических принципах, в достаточно близкой временной перспективе, приводят к пересмотру концепции применения сил и, следовательно, необходимости усовершенствования ранее спроектированных систем организационного управления для этой области.

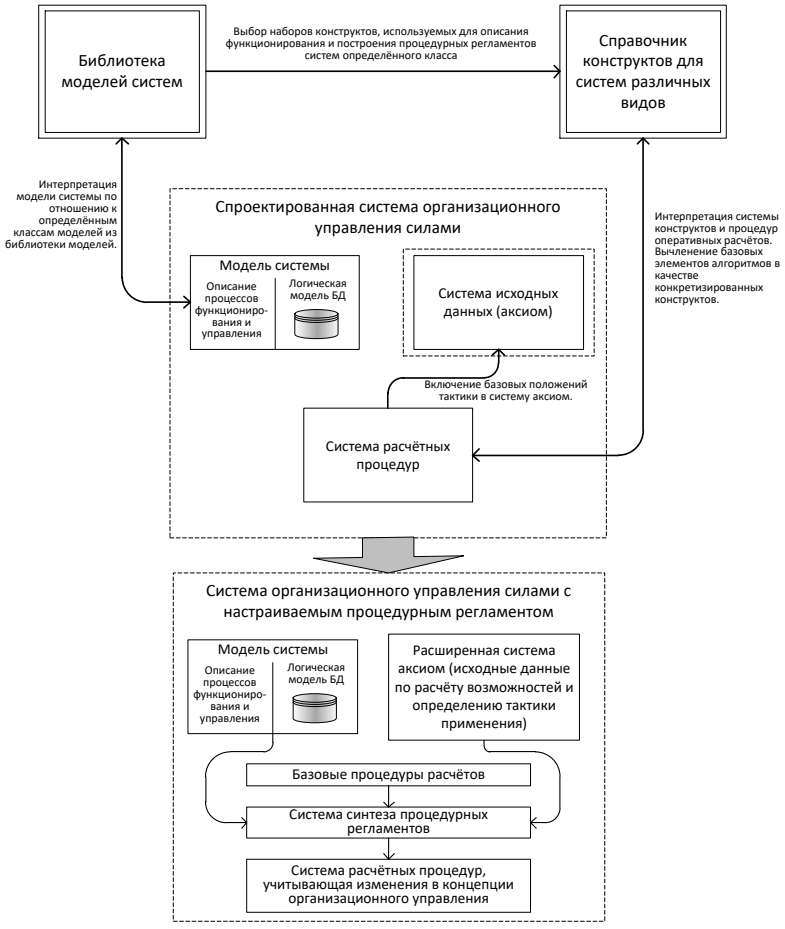


Рис.11. Направления дальнейших исследований по созданию теории проектирования систем организационного управления силами.

В данной статье авторы формулируют пути дальнейшего развития теории проектирования организационных систем на основе использования моделирующих комплексов с целью ускорения процессов адаптации программного обеспечения автоматизированных систем в условиях быстрой смены концепций и парадигм применения сил.

Методологической основой для создаваемой теории проектирования должна стать теория концептуального проектирования систем организационного управления, в том числе – методы тензорного анализа сетей.

Основным содержанием перспективных исследований должны стать задачи интерпретации и конкретизации обобщённых моделей и конструктов систем определённого класса для разработанного описания предметной области и созданной системы расчётных задач. Также, для образованного конкретизированного перечня конструктов, описания процесса и аксиом применения сил, следует разработать метод синтеза процедурного регламента системы с использованием теории тензорного анализа сетей.

Успешное решение перечисленных задач в отношении реализованных проектов автоматизированного управления силами позволит создать теорию проектирования систем организационного управления с обеспечением адаптации процедурных регламентов к изменяющимся концепциям управления.

Литература

1. Новиков Д.А. Теория управления организационными системами / Д.А. Новиков. – М.: МПСИ. – 2005. – 584 с.
2. Додонов А.Г. Проблемы построения эффективной системы управления процессами развития авиационных сил Украины / А.Г. Додонов, Б.И. Низиенко, А.В. Никифоров, В.Г. Путятин // ISSN 1560-9189 Реєстрація, зберігання і обробка даних, 2019. – Т.21. – №3. – С. 42 – 55
3. Додонов А.Г. Ситуационное управление силами на примере соединения противовоздушной обороны / А.Г. Додонов, А.В. Никифоров, В.Г. Путятин, И.В. Князь // Математические машины и системы. – №4. – 2019. – С. 17-37.
4. Немчинов В.С. Экономико-математические методы и модели / В.С. Немчинов. – М., Мысль. – 1965. – 253 с.
5. Глушков В.М. Основы безбумажной информатики. Изд. 2-е, испр. / В.М. Глушков. – М.: Наука. Гл. ред. физ.-мат. лит. – 1987. – 552 с.
6. Додонов А.Г. Тенденции и проблемы развития автоматизации управления военными силами / А.Г. Додонов, А.В. Никифоров, В.Г. Путятин // Математические машины и системы, №3. – 2019. – С. 3 – 16
7. Гвардейцев М.И. Математическое обеспечение управления. Меры развития общества / М.И. Гвардейцев, П.Г. Кузнецов, В.Я. Розенберг // Под ред. М.И. Гвардейцева. – М.: Радио и связь. – 1996. – 176 с.
8. Додонов О.Г. Комп'ютерне моделювання процесів організаційного управління. За матеріалами наукової доповіді на засіданні Президії НАН України 4 листопада 2015 року / О.Г. Додонов // ISSN 1027-3239. Вісн. НАН України, 2016. – № 1. – С. 69 - 77
9. Додонов А.Г. Организация структуры системы обработки информации и управления / А.Г. Додонов, В.Г. Путятин, А.Н. Буточнов, Н.С. Козлов, В.В. Юзефович // ISSN 1028-9763. Математичні машини і системи, 2014. – № 4. – С. 18 – 34
10. Крон Г. Тензорный анализ сетей / Г. Крон // Пер. с англ. Под ред. Л.Т. Кузина, П.Г. Кузнецова. – М.: Сов. радио. – 1978. – 720 с.
11. Евдокимова Н.А. О роли конструктивного мировоззрения Побиска Георгиевича Кузнецова в отраслевых научно-исследовательских работах (угольная отрасль) / Н.А. Евдокимова. – М.: ООО «Редакция журнала «Уголь»». – 2007. – 256 с.
12. Кононенко А.А. Технология концептуального проектирования / А.А. Кононенко, З.А. Кучкаров, С.П. Никаноров, Н.К. Никитина // ISBN 978-5-88981-079-7. – М.: Концепт. – 2008. – 580 с.
13. Иванов А.Ю. Справочник по теоретико-системным конструктам. Серия «Концептуальный анализ и проектирование». Методология и технология / А.Ю. Иванов, С.П. Никаноров, Ю.Р. Гараева // ISBN 978 5-88981-087-2. – М.: Концепт. – 2008. – 314 с.

ЗАСТОСУВАННЯ ОНТОЛОГІЧНОГО АНАЛІЗУ ДЛЯ ОБРОБКИ ВЕЛИКИХ ДАНИХ У ДОМЕНІ КІБЕРБЕЗПЕКИ

Anatoly Gladun[1], Julia Rogushina[2]

[1]*International Research and Training Center of Information Technologies and Systems of National Academy of Sciences of Ukraine and Ministry of Education and Science of Ukraine*
glanat@yahoo.com
[2]*Institute of Software systems of National Academy of Sciences of Ukraine, Kyiv, Ukraine,*
ladamandraka2010@gmail.com

Анотація. Галузь інформаційної безпеки є дуже динамічною сферою у якій швидко змінюються і удосконалюються як засоби захисту від загроз, так і самі деструктивні методи і засоби, що являється сьогодні викликом для окремих користувачів, організацій і всього суспільства в цілому. Тому інформаційні системи, пов'язані із кібербезпекою, потребують постійного надходження знань як із внутрішніх, так із зовнішніх джерел, об'єм яких постійно зростає. Неструктурована природа великих даних, їх складністю та різноманіття вимагає особливого підходу до керування життєвим циклом великих даних – застосування знань про предметну область кібербезпеки, пошуку і відбору складних інформаційних об'єктів. Тому автори статті вважають доцільним застосування методів аналітики великих даних до побудови онтологій у домені кібербезпеки.

В якості інструментів такого аналізу автори пропонують використовувати попередню обробку метаданих у домені кібербезпеки на рівні семантики з метою створення тезаурусу задачі. Цей тезаурус має забезпечити термінологічну базу для інтеграції онтологій різних рівнів, що характеризують структуру знань цієї предметної області. Домен кіберпезпеки має ієрархічну структуру, яка складається з декількох рівнів. Детальне описання знань про домен кібербезпеки потребує розроблення ієрархії онтологій починаючи від верхнього рівня до нижнього. Для побудови тезаурусу запропоновано використати стандарти кібербезпеки, описи компетенцій спеціалістів, відкриті природомовні інформаційні ресурси з кібербезпеки: словник, енциклопедії. Для покращення виокремлення семантики відношень між поняттями запропоновано використати семантично розмічені Wiki-ресурси, зовнішні тезауруси та онтології. Крім того, в онтологію цього домену ми інтегруємо знання з існуючих онтології із сфери інформаційної безпеки.

Онтологічна модель домену кібербезпеки містить його концептуалізацію, включає та інтегрує різнорідні структури даних та знань з різних підсистем кібербезпеки, а також стандарти. Розробка такої онтології формалізує взаємозв'язки між елементами даних, що в подальшому планується використати для машинного навчання та алгоритмів штучного інтелекту для адаптації до змін у середовищі, що у свою чергу підвищить ефективність аналітики великих даних для домену кібербезпеки.

Ключові слова: Аналітика великих даних, Кібербезпека, Онтологія, Тезаурус, Неструктуровані дані, Метадані, Wiki-технологія, семантична подібність.

Вступ

Сьогодні питання інформаційної безпеки (ІБ) стає наріжним каменем у діяльності кожної організації чи особистості. Під інформаційною безпекою розуміють захищеність інформації й усієї організації від навмисних або випадкових дій, що приводять до завдання збитків її власникам або користувачам. Сфера ІБ – це дуже динамічна сфера, що потребує постійного відстеження відомостей, які надходять як із внутрішніх, так із зовнішніх інформаційних ресурсів. Використовуючи дані з мереж та комп'ютерів, аналітики можуть видобути корисну інформацію з даних, використовуючи аналітичні методи та процеси. Динамізм сфери ІБ породжує великі дані, які потрібно обробляти у пакетному чи транзакційному режимі для швидкого прийняття рішень щодо ІБ. Для прийняття рішень в системах інформаційного захисту використовують аналітику великих даних[1]. Ті, хто приймає рішення, можуть приймати більш обґрунтовані рішення, скориставшись результатами аналізу, включно з тим, які дії потрібно виконати, та рекомендації для покращення політик, настанови, процедури, інструменти та інші аспекти мережних процесів.

Аналітика великих даних (Big Data) – це нова технологія аналітики, що забезпечує можливість збору, зберігання, обробки та візуалізації цих величезних обсягів даних. Така аналітика враховує основні характерні властивості великих даних [2].

1. Аналіз сучасної ситуації

Сучасна ситуація в розвитку зберіганні, обміну та обробки інформації, що характеризується інтенсивним впровадженням технологій, розповсюдженням локальних, корпоративних та глобальних мереж у всіх сферах життя цивілізованої держави, створює нові можливості та якість інформаційного обміну. У цьому контексті постає питання, щоб інформація була захищеною та безпечною одночасно [3]. Необхідність вирішення цього питання актуалізується такими факторами:

- експоненціальний ріст кількості персональних комп'ютерів, лептопів та інших автоматизованих пристроїв для обміну інформацією;
- стрімке розширення кола користувачів, що безпосередньо мають доступ до цифрових мереж та інформаційних ресурсів;
- грандіозне збільшення об'ємів інформації, що накопичується, зберігається та обробляється з допомогою засобів автоматизації;
- бурхливий розвиток апаратно-програмних засобів та технологій, що не відповідають сучасним вимогам безпеки;
- невідповідність між надшвидким розвитком засобів обробки інформації, вітчизняної теорії інформаційної безпеки та розробкою міжнародних стандартів і правових норм, що забезпечать необхідний рівень захисту інформації;
- інформаційна війна, зумовлена політичною ситуацією в країні;
- кіберзлочини, що іноді набувають державно важливого значення;
- становлення кіберполіції України.

Таким чином, перед нами постає завдання визначення відношень основними поняттями, що використовуються у пошуку та навігації серед ресурсів, що пов'язані з категоріями «інформаційна безпека», «кібербезпека» та «захист інформації».

В роботі [4] пропонується використовувати тезаурусний підхід для формалізації термінології ПрО.

Тезаурус галузі інформаційної безпеки відображає широкий спектр суттєвих властивостей, ознак та відношень, притаманних даному специфічному виду безпеки. Ми погоджуємось із дослідниками [5,6], що виділяють три групи термінів теорії інформаційної безпеки.

Терміни, що детермінують наукову основу інформаційної безпеки. До цієї групи належать терміни, що використовуються у багатьох галузях знань та є однозначними, семантично уніфікованими та стилістично нейтральними. Це: інформація, комунікація, конфлікт, вплив, загроза, небезпека, безпека, система.

Поняття в цій групі відповідають вимогам однозначності та стабільності, тобто однозначно вживаються в одній галузі знань та зберігають зміст у будь-якій іншій. Лише поняттю «інформація» притаманна специфічна властивість: у різних сферах вжитку, та навіть одній, може характеризувати предмет, явище, процес та їх властивості й відношення одночасно.

Терміни, що детермінують предметну основу інформаційної безпеки. Ця група термінів позначає поняття та їх співвідношення з іншими поняттями в рамках інформаційної безпеки як особливої галузі знань.

Терміни, що детермінують характер діяльності щодо забезпечення інформаційної безпеки. Ця група охоплює терміни, що позначають характерні для цієї сфери предмети, явища, процеси, їх властивості та відношення.

Але більш універсальний підхід, пов'язаний з засобами аналітики Big Data, потребує застосування для цього онтологічного аналізу та створення онтології ПрО кібербезпеки. На сьогодні існує вже багато розробок у цій сфері, але відкритими залишаються питання їх взаємної інтеграції та сумісності із застосовними системами.

Тезаурус кібербезпеки інтегрований із поняттями інформаційної безпеки, безпеки застосунків, мережної безпеки, безпеки Web та безпеки критичної інформаційної інфраструктури [7]. Безпека застосунків визначається відносно прикладних програмних продуктів, а також інформаційно-програмних ресурсів та процесів, що беруть участь у їхньому життєвому циклі. Безпека мереж пов'язана із проектуванням, упровадженням та використанням мереж у середині організації, між організаціями, між організаціями та користувачами. Безпека в мережі Інтернет стосується Інтернет-сервісів та відповідних систем інформаційно-комунікаційних технологій та мереж. Безпека критичної інформаційної інфраструктури характеризує захищеність від відповідних загроз, зокрема інформаційної безпеки.

Слід відмітити, що ця предметна область (ПрО) є дуже динамічною, і тому виникає проблема у відповідності термінологічної бази сучасному стану розробок та їх інтеграцією із сучасною граматикою української мови.

Аналітика великих даних в галузі кібербезпеки вивчає проблеми безпеки, пов'язані з великими даними, та надає корисні відомості, які можуть бути використані для вдосконалення сучасних практик роботи мережних операторів та адміністраторів.

Кібербезпека - це захист інформаційних систем, як апаратних, так і програмних засобів, від крадіжок, несанкціонованого доступу та розголошення, а також від навмисної чи випадкової шкоди. Вона захищає усі сегменти, що відносяться до Інтернету, від самих мереж до інформації, що передається через мережу та зберігається в базах даних, до різних застосунків та пристроїв, які керують роботою обладнання через мережні з'єднання.

Великі дані в інформаційних технологіях – це набори інформації (структурованої, слабоструктурованої або неструктурованої) настільки великих розмірів, що традиційні способи та підходи не можуть бути застосовані до них і аналізувати їх безпосередньо надто довго і важко. Для використання великих даних як зовнішнього джерела інформації, наприклад із Веб, потрібно спочатку відфільтрувати потрібну пертинентну множину наборів даних, які будуть використані для аналітики великих даних. Попередня очистка даних і підготовка їх до аналізу може бути виконана на основі аналізу метаданих (і досить часто аналіз природомовної її частини, наприклад, анотації). Метадані - дані про дані чи елементи даних, можливо, також їх описи даних, дані про право власності на дані, шляхи доступу, права доступу та зміну даних при їх обробленні.

Для того, щоб набір даних можна було вважати великими даними, він повинен володіти однією або декількома характеристиками, так званими характеристиками «5V»: об'єм; швидкість; різноманіття; достовірність; цінність [8].

Для одержання якісних результатів обробки великих даних дуже важливими характеристиками є достовірність і цінність. Достовірність стосується якості або точності даних, що може стати причиною обробки даних з метою усунути помилкові дані й шуми. Шум – це дані, котрі не можуть бути перетворені в інформацію і, отже, не мають достатньої цінності, і не можуть привести до значимої інформації.

Цінність визначається, як корисність даних для підприємства і ця характеристика інтуїтивно пов'язана з достовірністю, тому що чим вища точність даних, тим більше їх корисність для бізнесу. Цінність також залежить від часу обробки даних, оскільки аналітичні результати мають певний термін придатності. Наприклад, 20-хвилинна затримка котирування акцій практично не має значення для здійснення угоди в порівнянні з котируванням, що складає 20 мілісекунд. Як видно, цінність і час є обернено пропорційними. Чим більше часу потрібно для перетворення даних у значиму інформацію, тим менше вони корисні для бізнесу. Застарілі результати гальмують якість і швидкість прийняття обґрунтованих рішень.

Основними типами великих даних можуть бути: структуровані дані (SQL бази даних); слабоструктуровані дані (інструкції з інформаційної безпеки, дані профілю клієнтів, журнали Web-серверів, Web-сайти, тексти, електронні листи тощо) і неструктуровані (не реляційні, NoSQL бази даних) дані (аудіофайли, відеофайли, зображення, інформаційні куби тощо).

У [8] визначено головні проблеми, які існують сьогодні в технології великих даних і потребують вирішення:

1. Проблема інтеграції даних, яку можна подати як комбіновану проблему, що потребує: (1) визначення задачі, яку необхідно вирішити за допомогою великих даних; (2) виявлення (пошук) відповідних частин даних в сховищах та джерелах великих даних; (3) виконання ETL у відповідних форматах та збереження даних для подальшої обробки; (4) зняття неоднозначності даних (приміром, омонімії); (5) обробка даних для вирішення задачі.

2. Проблема подолання гетерогенності між різними наборами великих даних. Семантика може розглядатися як засіб для створення моста між гетерогенними даними.

3. Проблема зв'язування відкритих даних (Linked Open Data) для забезпечення хорошого зв'язування даних.

4. Проблема використання семантики для інтеграції даних та в розробці майбутніх СКБД. Більш того, семантика може бути використана в існуючій системі, для виявлення невідповідностей даних, генерування нових знань за допомогою машини логічного виведення або просто зв'язувати більш точно конкретні дані, що не мають відношення до методів машинного навчання.

Як показує аналіз наукових публікацій [9] сьогодні питання про актуальність метаданих, використовуваних у великих даних є найбільш гострим ніж будь-коли. Усе більше і більше організацій усвідомлюють, що для того, щоб підвищити ефективність бізнесу, використовуючи цінну інформацію від даних, необхідні робастні (працездатні) метадані для здобуття необхідного контексту та походження ключових активів даних. У той же час регулювання галузі вимагає кращої прозорості та розуміння інформації метаданих.

Хоч керування метаданими відоме уже десятки років, сьогодні розробляються нові стратегії та підходи:

- Підтримка постійного розвитку середовища упорядкування даних;

- Пошук більш ефективних шляхів керування бізнесом за допомогою метаданих.

Таким чином, нам потрібно виконати огляд стратегій та технологій роботи з метаданими, доступних для сучасної організації, а також з'ясувати питання, як побудувати успішні стратегії стосовно прийняття та використання метаданих.

Організації та компанії зацікавлені в двох типах великих даних. 1. Для оброблення дуже часто застосовують дані створені людиною, як такі, що в основному розповсюджуються через засоби Web (соціальні мережі, файли cookie, електронні листи, онлайн телебачення, онлайн мовлення тощо). 2. Існує потреба в інтеграції даних, що генеруються різними джерелами і часто є гетерогенними за своєю природою. Приміром, мережа Інтернет та системи інформаційної безпеки різних компаній генерують змішаний трафік великих даних, який використовують сумісно для передбачувального (предикативного) аналізу та отримання знань щодо розуміння, планування та упередження дій для цих систем. При цьому гостро постає питання стосовно якості даних. Насправді, оскільки великі дані характеризуються великими об’ємами, є «сирими» по своїй природі. Тому потрібне вирішення цієї проблеми.

2. Постановка задачі

Аналіз публікацій показує, що, незважаючи на високий інтерес до великих даних, їх аналітики для кібербезпеки та наявність різноманітних технологічних засобів їх збереження та обробки, на сьогодні відсутні релевантні методи на основі семантичного опису метаданих для відбору пертинентної підмножини блоків зовнішніх великих даних, які підходять для вирішення поставленої задачі.

Це пояснюється неструктурованою природою великих даних, їх складністю та різноманіттям. Традиційні метадані – це технічна інформація, яка характеризує час створення контенту, його обсяг, формати тощо, але не стосується змісту тієї інформації, яка міститься в цих даних. Дуже часто для багатьох блоків великих даних метаданими являються їх метаописи на природній мові, тобто анотації чи пояснення, які потребують подальшої семантичної обробки для прийняття відповідного рішення про підбір великих даних. Проблема підбору внутрішніх великих даних може не виникати, оскільки вони були накопичені всередині організації в процесі виконання завдань з кібербезпеки. Створення онтології домену кібербезпеки дозволить вирішити проблему аналізу природомовних анотацій і релевантності обраних блоків великих даних для вирішення поставленої задачі з кібербезпеки. Для пошуку відповідних великих даних пропонується використання тезаурусу існуючої проблеми, нові терміни якого будуть поповнювати термінологічну множину онтології домену, встановлюючи семантичні зв’язки між ними. Для побудови такого тезаурусу пропонується використання стандартів з інформаційної безпеки, які можуть бути поданими на різних мовах, використовувати описи компетенцій спеціалістів із ІБ.

3. Аналітика великих даних для кібербезпеки

Застосування аналітики великих даних у кібербезпеці сьогодні є вирішальним важливим завданням. Використовуючи дані кібербезпеки з мереж та комп'ютерів, аналітики можуть виявити корисну мережну інформацію з даних, яка дозволяє приймати більш інформативні рішення на основі результатів аналізу, включаючи те, які дії потрібно виконати, та рекомендації щодо вдосконалення політики, вказівок, процедур, інструментів та інших аспектів мережних процесів. Фахівці з аналізу даних можуть мати різний рівень компетентності у галузі кібербезпеки та різний рівень знань. Спектр тем компетенції фахівця з кібербезпеки охоплює знання: мережна криміналістика, аналіз загроз, оцінка вразливостей, візуалізація, кібертренінг тощо. Крім того, фахівці мають опанувати нові сфери ІБ, такі як IoT, хмарні обчислення, туманні обчислення, мобільні обчислення та кібер-соціальні мережі.

Мережна криміналістика (Network Forensics) - це моніторинг та аналіз трафіку комп'ютерної мережі для збору інформації, юридичних доказів або виявлення вторгнень. Системи моніторингу та аналізу трафіку здійснюють захист від широкого спектру загроз, таких як man in the middle, атаки з підбору пароля, атаки за допомогою деструктивних програм або з використанням вразливостей програм.

Останніми роками спостерігається зростання багатьох класів кібератак, починаючи від викупу програмного забезпечення до просунутих постійних загроз (APT), що становлять серйозні ризики для компаній та підприємств.

Аналітика великих даних, як нова технологія аналітики, надає можливість збирати, зберігати, обробляти та візуалізувати великі дані; отже, застосування аналітики великих даних у кібербезпеці стає вирішальною та новою тенденцією, приносить свої переваги, а її застосування у кібербезпеці має важливе значення для подолання виникаючих загроз..

У той же час кібербезпека сьогодні переживає великі виклики через зростання об'ємів великих даних, що спричинено швидкозростаючою складністю мереж (наприклад, віртуалізація, смарт пристрої, бездротові канали обміну даними, Інтернет речей тощо) та зростаюча складність загроз (наприклад, зловмисне програмне забезпечення, багатоступінчастість, просунуті стійкі загрози (APTs) тощо).

4. Використання семантичних технологій у кібербезпеці

Останнім часом у світовій практиці головні напрямки розвитку інформаційних технологій (ІТ) пов'язані зі створенням інформаційних систем, які використовують знання у відповідних предметних областях (ПрО). Ці тенденції стосуються й галузі кібербезпеки, яку дуже важливо орієнтувати на застосування сучасних інтелектуальних Web-технологій та пов'язаних з ними стандартів подання інформації

Для вирішення цих завдань по створенню інтелектуальних, динамічних, та адаптивних систем дослідниками ведуться роботи із впровадження семантичних, орієнтованих на знання алгоритмів, моделей, методів і технологій, які дозволяють значно підвищити ефективність існуючих інформаційних систем. Ці інтелектуальні інформаційні системи (ІІС) застосовують методи, запозичені зі штучного інтелекту (ШІ), які забезпечують ефективну обробку, аналіз та генерацію нових знань. Крім того, значна частина сучасних ІС орієнтована на розподілену обробку інформації та на функціювання у відкритому інформаційному просторі Web, причому частка таких систем, у загальній кількості створюваних ІІС, постійно зростає.

На базі концепції Семантичного Вебу (Semantic Web) і використання формалізованих (певним способом) знань забезпечується інтелектуалізація ІІС та застосувань у різних сферах людської діяльності, наприклад, інтелектуалізація мережних систем керування, створення прикладних інтелектуальних інформаційних систем і застосувань для медицини і е-бізнесу, для підвищення якості сервісів у мережі, оперативного вирішення задачі інформаційного захисту тощо.

Робота зі знанням – це достатньо складний багатоетапний процес, який має власний життєвий цикл. Як і будь-який процес, і тим більше складний, він потребує усвідомленого керування для забезпечення певної послідовності, погодженості, інтероперабельності, ефективності.

Через складність отримання знань важливим стає питання забезпечення їх інтероперабельності та повторного використання. Перспективним підходом до вирішення цих проблем є онтологічний аналіз: онтології базуються на ґрунтовному теоретичному базисі дескриптивних логік, для них вже створено загальноприйняті стандарти опису, мови та програмні засоби.

Керування знаннями – це сукупність процесів, що пов'язані з ефективним створенням, збереженням, поширенням і використанням знання для виконання цілей. Основними джерелами знань з кібербезпеки можуть бути відповідні міжнародні стандарти та пов'язані з ними практичні розробки.

Процес керування знаннями поширюється у чотирьох основних напрямках:
1) дослідження знань і їхня систематизація;
2) усвідомлення знань і визначення їхньої цінності;

3) планування і здійснення дій відповідно до результатів аналізу знань;

4) постійна актуалізація і переосмислення знань.

Знання не тільки являють собою самостійну цінність, але й породжують мультиплікативний ефект стосовно інших факторів виробництва, впливаючи на рівень ефективності їх застосування. Таким чином, у сучасній економіці джерелом конкурентних переваг стає не вигідна ринкова позиція, а складні для реплікації знання як активи й спосіб їх розміщення. Причому в центрі уваги тут перебуває не створення знань, а їх динамічний рух і використання в організації, яка хоче бути конкурентноспроможною.

Термін "керування знаннями " ("knowledge management") вперше було використане в 1986 р. американським вченим Карлом Віїгом на конференції у Швейцарії, яка проводилася Міжнародною організацією праці під егідою ООН, за аналогією з такими поняттями, як «керування даними» та «керування інформацією». На сьогодні існує багато досить суперечних визначень цього терміну, що відображають різні аспекти, пов'язані з ефективним використанням знань у різних областях.

Основні проблеми керування знаннями в Web пов'язані з:

- Інтеграцією знань, отриманих від різних ІР (наприклад, інтеграція онтологій кількох різних ПрО або ІР в одній ПрО);
- Пошуком протиріч між знаннями, що відображені в контенті різних ІР, оцінкою їх достовірності та надійності;
- Здобуттям нових знань з вже наявних та їх поданням у формі, зрозумілій користувачеві;
- Пошуком знань, потрібних конкретному користувачеві для вирішення певних задач;
- Автоматизацією створення метаданих, що коректно описують контент ІР (як текстових, так і мультимедійних) на рівні змісту, та ефективним пошуком у таких метаописах.

5. Класифікація вразливостей систем безпеки

Загрози інформаційної безпеки проявляються не самостійно, а через можливу взаємодію з найбільш слабкими ланками системи захисту, тобто через фактори уразливості. Загроза призводить до порушення діяльності систем на конкретному об'єкті-носії [10].

Основні уразливості виникають унаслідок дії наступних факторів:

- недосконаліть програмного забезпечення, апаратної платформи;
- різні характеристики структури автоматизованих систем в інформаційному потоці;
- частина процесів функціонування систем є неповноцінною;
- неточність протоколів обміну інформацією та інтерфейсу;
- складні умови експлуатації і розташування інформації.

Найчастіше джерела загрози запускаються з метою отримання незаконної вигоди внаслідок заподіяння шкоди інформації. Але можлива і випадкова дія загроз через недостатню міру захисту і масову дію загрозливого фактора.

Існує поділ вразливостей за класами, вони можуть бути: об'єктивними; випадковими; суб'єктивними.

Якщо усунути або як мінімум послабити вплив вразливостей, можна уникнути повноцінної загрози, спрямовану на систему зберігання інформації.

6. Онтології для класифікації та прогнозування кібернападів

Швидка реакція на мережну атаку - одна з найважливіших вимог в кіберзахисті. Коли система може ідентифікувати атаку, яка продовжується, та класифікувати атаку, то можуть бути вжиті ефективні заходи протидії. У [11] підкреслена необхідність швидкого реагування зі збільшенням швидкості комп'ютерних атак. Різні дослідники розробили онтологічні застосування для ідентифікації та класифікації мережних атак. Нижче наведено декілька прикладів таких онтологій:

• У роботі [12] авторами розроблена онтологія для розрізнення стану безпеки мережі;

• Однорангова мультиагентна розподілена система для виявлення вторгнень подана в [13];.

• Онтологія мережних атак, призначена для підтримки автоматизованої класифікації атак побудована в роботі [14].

• У роботі [15] автори розробили онтологічну систему для прогнозування потенційних мережевих атак.

7. Метадані у кібербезпеці

Для одержання якісних результатів обчислення великих даних необхідний супровід блоків даних метаданими, які фізично приєднуються до блоків великих даних і постійно їх супроводжують на всьому їх життєвому циклі. Метадані надають інформацію про характеристики та структуру набору великих даних. Відстеження метаданих має ключове значення при обробці великих даних, а також для їхнього збереження й аналізу, оскільки вони надають інформацію про походження даних, про достовірність, про джерело даних під час обробки. Складність використання великих даних для аналітики у кібербезпеці часто виникає через відсутність описів метаданих наборів даних. Для повторного використання даних потрібно якомога більше інформації про походження цих даних; повна історія методів, що використовуються для збору, супроводу та аналізу. Належні метадані збільшують шанси на те, що набори даних будуть повторно використані правильним чином – що приведе до аналітичних висновків з найменшою ймовірністю помилок.

Метадані сприяють підвищенню якості даних, яка визначається наступними характеристиками: узгодженістю (чи є представлення даних однорідним, чи існують дублікати даних, що перетинаються або конфліктують); повнотою (чи всі дані наявні); точністю (збігом збережених і фактичних значень); своєчасністю (чи є актуальним збережене значення).

Таким чином, для ефективної обробки великих даних і одержання цінних знань необхідна гнучка структура керування процесом обробки на основі метаданих, які дозволяють створити універсальне середовище для забезпечення інтероперабельності гетерогенних блоків даних, стандартизувати етапи обробки й еволюціонувати платформи обробки. Метадані можуть впливати на: стандартизацію, придбання зовнішніх даних для обробки; забезпечення конфіденційності даних і анонімність (знеособлювання) даних; на архівування джерел даних і результатів аналізу; на формування рекомендацій з очищення і фільтрації даних.

8. Онтологія домену кібербезпеки

Наука про кібербезпеку усе більше починає схилятися до використання знань про домен. Ця тенденція спричинена сучасними атаками на основі APT (розширені постійні загрози), фішингом та викупним програмним забезпеченням, які знаходяться у стадії постійного зростання, і від них усе складніше захистити організацію чи користувача. Ця ситуація змусила багатьох фахівців з кібербезпеки вкладати кошти в інтелектуалізацію платформ кібербезпеки, щоб спробувати випередити останні загрози. Однією з областей, яка починає викликати великий інтерес, є концепція створення онтології, а точніше - кіберонтології. Онтологія кібербезпеки є сукупність понять і категорій у предметній області, яка показує їх властивості та взаємозв'язки між ними.

Для створення онтології домену кібербезпеки ми поєднали уже існуючі онтології кібербезпеки та удосконалили їх. Уніфікована онтологія кібербезпеки (UCO) [16] призначена для підтримки інтеграції знань в системах кібербезпеки. Онтологія включає та інтегрує різнорідні схеми даних та знань з різних підсистем кібербезпеки та найчастіше використовуються стандарти кібербезпеки для спільного доступу та обміну. Онтологія UCO включена до ряду існуючих онтологій з кібербезпеки, а також містить концептуалізацію домену кібербезпеки. Подібно до DBpedia, яка слугує ядром загальних знань у хмарних обчисленнях, UCO може слугувати ядром знань для домену кібербезпеки, який розвиватиметься та зростатиме разом із пов’язаними наборами даних з кібербезпеки.

Онтологія UCO описує таку онтологію, яка "забезпечує загальне розуміння домену кібербезпеки та уніфікують найбільш часто використовувані стандарти кібербезпеки" Крім того, онтологія кібербезпеки, повинна також містити спеціальні словники, такі як технічний жаргон та спеціальні словники термінів використовуваних у вузькій сфері фахівців з ІБ. Як було сказано раніше, дані потрібно відфільтровувати, щоб зменшити їх обсяг. Це можна зробити за допомогою онтологій кібербезпеки, які виступають першими фільтрами даних та визначають потенційне джерело в Інтернеті.

Хоч визначення онтології кібербезпеки є відносно статичними, сьогодні побудова онтологій перетворилася на активного постачальника взаємозв'язків елементів даних, який може використовувати машинне навчання та алгоритми штучного інтелекту для адаптації до змін у середовищі[17]. Кіберонтології можуть бути адаптивним словником даних, застосунків та взаємовідносин із користувачами, який може покращити поведінковий аналіз та допомогти зупинити розповсюдження загроз до того, як розповсюдяться поширені інфекції.

При розробці онтології кібербезпеки ми використали потенційні онтології та стандарти із сфери інформаційної безпеки, які були нами використані для розширення онтології кібербезпеки в рамках

досліджуваних питань. Детальне описання знань про домен кібербезпеки потребує розроблення ієрархії онтологій починаючи від верхнього рівня до нижнього. Онтологія верхнього рівня (фундаментальна або основна) включає основні поняття домена, які уже були раніше визначені в онтологіях по цій темі. Нижче розміщені онтології, що орієнтовані на користувача, події, мережні операції та геопросторові дані, що стосуються кібербезпеки (онтології середнього рівня). І онтології нижнього рівня описують конкретні домени кібербезпеки для яких потрібне вирішення завдання для конкретної галузі.

Для створення онтології кібербезпеки потрібно розглянути усі можливі види загроз; але в плані безпеки найбільше уваги слід приділити тим загрозам, які пов'язані зі шкідливою чи іншою діяльністю людей [18]. На рис.1 подані ці загальні концепції та співвідношення. Ця схема покладена в основу онтології кібербезпеки. Терміни для кібербезпеки взяті із відкритого джерела – словника термінів кібербезпеки [19]

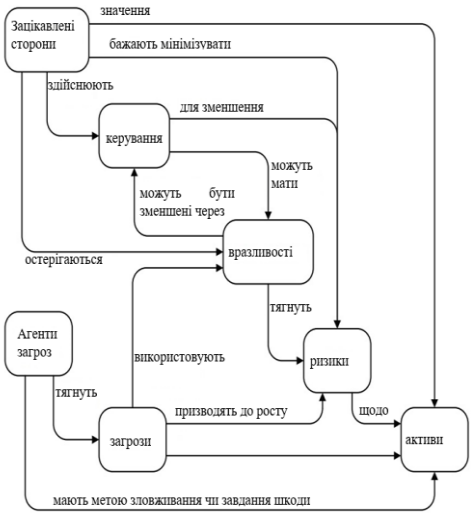


Рис.1 Терміни та відношення ПрО кібербезпеки.

Зацікавлені сторони у кіберпросторі можна розділити на такі групи: 1) споживачі: фізичні особи, як приватні, так і громадські організації; 2) постачальники, що включають, але не обмежуються такими: провайдери Інтернет-послуг; провайдери програмних сервісів.

Активом у сфері кібербезпеки може бути будь-що, що має цінність для фізичної особи чи організації. Активи охоплюють, але не обмежуються, наступним: інформація; програмні засоби, наприклад, комп'ютерні програми; фізичні активи, наприклад, комп'ютери; сервіси; люди, їхня кваліфікація, навички та досвід; нематеріальні активи, наприклад, репутація та імідж. Крім фізичних активів, віртуальні організаційні активи набувають все більшої цінності. Онлайн-бренд та інші представлення організації у кіберпросторі окреслюють її образ і є настільки ж важливими, як і її складові у реальному фізичному світі. Наприклад, URL-адреса та інформація на веб-сайті організації є її активами.

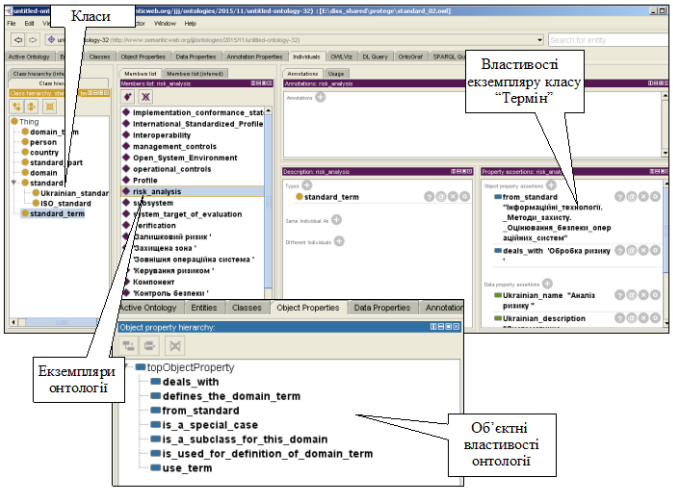


Рис.2. Онтологія "Кібербезпеки" (фрагмент)

Окремі Wiki-сторінки доцільно створити як для кожного стандарту в цілому (перелік його підрозділів), так і для окремих підрозділів та визначень. Крім того, доцільно використовувати гіперпосилання на інші Wiki-ресурси, приміром, на Вікіпедію або на електронну версію Великої української енциклопедії (e-BYE).
Таким чином, якщо вже побудована онтологія мовою OWL, тоді її досить просто використовувати в Semantic MediaWiki. Але зворотний процес не може бути цілком автоматизований. Більш того, при автоматичній генерації онтології за Semantic MediaWiki будуть загублені відомості, що містяться в OWL-онтології та стосуються характеристик класів і властивостей, що не мають аналогів у Wiki (зокрема, про еквівалентність класів і властивостей, їхній неперетин, про їхню область значення і визначення). У той же час, частина контенту Semantic MediaWiki не може бути безпосередньо трансформована в онтології [20]. Наприклад, той факт, що в сторінках використаний одні і той же шаблон, говорить про те, що на цих сторінках описані інформаційні об'єкти одного типу, але для відображення цього в онтології треба створити специфічний клас і зв'язати його з елементом сторінки. Але потім за такою онтології неможливо зрозуміти, що треба створити в Wiki – шаблон чи категорію: вибір залежить від користувача, тому що для ІО зі специфічної, але формалізованою структурою доцільно створювати шаблони, а у всіх інших випадках – сторінки, що будуть віднесені до якої-небудь категорії. Крім того, не можливо зв'язати з класом онтології не всю сторінку, а її конкретний фрагмент (крім тих випадків, коли в нього є підзаголовок).

Висновки. Використовуючи дані з мереж та комп'ютерів, аналітики можуть виявити корисну інформацію з даних, використовуючи аналітичні методи та процеси. Наука про кібербезпеку усе більше починає схилятися до використання знань про домен. Ця тенденція спричинена сучасними атаками на основі APT (розширені постійні загрози), фішингом та викупним програмним забезпеченням, які знаходяться у стадії постійного зростання, і від них усе складніше захистити організацію чи користувача. Ця ситуація змусила багатьох фахівців з кібербезпеки вкладати кошти в інтелектуалізацію платформ кібербезпеки, щоб спробувати випередити останні загрози. Одним із напрямків, який починає сьогодні швидко розвиватися, є концепція створення онтології, а точніше - кіберонтології. Онтологія кібербезпеки це сукупність понять і категорій у предметній області, яка показує їх властивості та взаємозв'язки між ними і вона призначена для підтримки інтеграції знань в системах кібербезпеки. Онтологія містить концептуалізацію домену кібербезпеки, включає та інтегрує різнорідні структури даних та знань з різних підсистем кібербезпеки, а також стандарти кібербезпеки. Кіберонтології стають адаптивним словником даних, застосунків та взаємовідносин із користувачами, який може покращити поведінковий аналіз та допомогти зупинити розповсюдження загроз до того, як розповсюдяться поширені інфекції.

Література

1. Onur Savas and Julia Deng Big Data Analytics in Cybersecurity. CRC Press, 2018.- 353p.
2. Grimes S.: Unstructured Data and the 80 Percent Rule, Clarabridge, Bridgepoints, (2008), http://breakthroughanalysis. com/2008/08/01/unstructured-data-and-the-80-percent-rule/.
3. "Are Cyber-Ontologies the Future of Cybersecurity?" by Frank Ohlhorst on July 24, 2019 - https://securityboulevard.com/ 2019/07/are-cyber-ontologies-the-future-of-cybersecurity/.
4. Сачук Ю.Є. Професійна підготовка фахівців із кібербезпеки та захисту інформації: тезаурус та онтологія// Проблеми інженерно-педагогічно ї освіти», 2018, № 59.-С.35-40.
5. Діордіца І. В. Репрезентація термінології кібербезпекової політики в текстах нормативно-правових актів України / І. В. Діордіца // Науковий вісник М іжнародного гуманітарного університету. Серія Юриспруденція. – 2017. – Вип. 29(1). – С. 64-67
6. Дубов Д. Підходи до формування тезаурусу у сфері кібербезпеки / Д. Дубов // Політичний менеджмент. –2010. – №5. – С. 19-30.
7. Каличенский А. Концепция создания Национальной информационно–коммуникационной инфраструктуры Украины [Электронный ресурс] / А. Каличенский // Региональный форум М СЭ. – Киев : НКРСИ, ОАО "Гипросвязь". – 2012. –https://www.itu.int/ITU-D/tech/ events/2012/Spectrum_CIS_Kiev_Sept12/Presentations/ Session2/A_Kalichensky _a.pdf
8. Томас Ерл, Ваджид Хаттак, Пол Булер Основы Big Data: концепции, алгоритмы и технологии.- Изд.: "Баланс Бизнес Букс", 2017.-382с.
9. Gajanan S. Kumbhar, Ajit S. Ghodke A Study of Big Data Analytics and Cybersecurity as Emerging Trends in Cyber Era//Cyber Times International Journal of Technology & Management, Vol. 11, Issue 1, 2018. – P.-19-23.
10. ДСТУ ISO/IEC 27032:2016 Інформаційні технології. Методи захисту. Настанови щодо кібербезпеки (ISO/IEC 27032:2012, IDT)

11. Balepin, I., Maltsev, S., Rowe, J. and Levitt, K. (2003). Using specification-based intrusion detection for automated response. Recent Advances in Intrusion Detection (pp. 136–154).Springer Berlin Heidelberg.
12. Bhandari, P. and Guiral, M.S. (2014). Ontology Based Approach for Perception of Network Security State. Recent Advances in Engineering and Computational Sciences (RAECS) (pp.1-6).
13. Ye, D., Bai, Q., Zhang, M., and Ye, Z. (2008). P2P distributed intrusion detections by using mobile agents. Seventh IEEE/ACIS International Conference on Computer and Information Science (ICIS 08) (pp. 259-265).
14. van Heerden, R., Leenen, L., and Irwin, B. (2013). Automated classification of computer network attacks. International Conference on Adaptive Science and Technology (ICAST 2013) (pp.157-163).
15. Salahi, A. and Ansarinia, M. (2011). Predicting Network Attacks Using Ontology-Driven Inference, arXiv prepring arXiv:1304.0913.Retrieved on Oct 15, 2014 from, http://arxiv.org/ftp/arxiv/papers/1304/1304.0913.pdf
16. Zareen Syed, Ankur Padia, M. Lisa Mathews, Tim Finin, and Anupam Joshi UCO: A Unified Cybersecurity Ontology// In Proceedings of the AAAI Workshop on Artificial Intelligence for Cyber Security, February 12, 2016 - https://ebiquity.umbc.edu/paper/html/id/722
17. Leo Obrst, Penny Chase, Richard Markeloff "Developing an Ontology of the Cyber Security Domain" – http://ceur-ws.org/Vol-966/STIDS2012_T06_ObrstEtAl_CyberOntology.pdf
18. ДСТУ ISO/IEC 15408-1:2017 Інформаційні технології. Методи захисту. Критерії оцінки. Частина 1. Вступ та загальна модель.
19. Гладун А.Я., Пучков О.О., Субач І.Ю., Хала К.О. Англо-український словник термінів з інформаційних технологій та кібербезпеки. –К.: ІСЗЗІ КПІ ім.. Ігоря Сікорського, 2018. С. 376
20. Д.В. Ланде, І.Ю. Субач, Ю.Є. Бояринова Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки, К.: ІСЗЗІ КПІ ім. Ігоря Сікорського», 2018. — 297 с.

ІНФОРМАЦІЙНА ПІДТРИМКА ЛЮДЕЙ З ПРОБЛЕМАМИ ЗОРУ НА ОСНОВІ МІКРОХВИЛЬОВОГО РАДАРУ AWR 1843

Горбатенко Анастасія, Антощук Світлана

Одеський національний політехнічний університет, проспект Шевченка, 1, Одеса, Одеська область, 65044, Україна
nastya000511@gmail.com

Анотація. У сучасному світі все більше процесів можуть бути автоматизовані за допомогою використання електроніки і комп'ютерів. Сфера технічного зору – один із головних світових трендів.

Системи технічного зору зараз вважаються невід'ємною частиною багатьох промислових процесів, тому що вони можуть пропонувати швидкі, точно відтворені можливості контролю. Наприклад, у харчовій промисловості, де система технічного зору відіграє вирішальну роль у процесах, коли швидкість і точність надзвичайно важливі і допомагають забезпечити конкурентну перевагу для виробників. Сам харчовий продукт перевіряється на предмет контролю і якості порцій, а також на якість упаковки і маркування. Крім того, на фармацевтичному ринку потрібні найвимогливіші системи бачення, які не тільки перевіряють продукти, але також перевіряють використання та налаштування систем, забезпечуючи правильне дозування і контроль за процесами виготовлення ліків.

Було запропоновано вирішити за допомогою створення системи технічного зору соціально-значущу проблему підтримки людей з проблемами зору на основі mmWave радару AWR 1843. Для людей з проблемами зору дана технологія є необхідністю та можливостю вільно пересуватись у просторі. Але існуючі рішення не є досконалими та не мають великого розповсюдженя серед людей з проблемами зору. Також у подальшому розроблений метод детектування може використовуватися у інших системах технічного зору.

Ключові слова: mmWave, AWR 1843, проблеми зору.

1. Аналіз можливостей вирішення проблеми слабозорих людей за допомогою технології технічного зору

1.1 Аналіз предметної області

Розробка системи технічного зору складний багатоетапний процес.

Комп'ю́терний зір або Комп'ютерне бачення — теорія та технологія створення машин, які можуть проводити виявлення, стеження та класифікацію об'єктів. Як наукова дисципліна, комп'ютерний зір належить до теорії та технології створення штучних систем, які отримують інформацію у вигляді зображень. Відеодані можуть бути представлені у вигляді багатьох форм, таких як відеопослідовність, зображення з різних камер або тривимірними даними з медичного сканера та інше.

1.2 Аналіз загальної ситуації у сфері технічного

Область комп'ютерного зору може бути охарактеризована як молода та різноманітна. І, хоча існують більш ранні роботи, можна сказати, що тільки з кінця 1970-х почалось інтенсивне вивчення цієї проблеми, коли комп'ютери змогли керувати обробкою великих наборів даних, таких як зображення. Однак, ці дослідження зазвичай починались з інших галузей, і, відповідно, нема стандартного формулювання проблеми комп'ютерного зору. Також, і це навіть більш важливо, нема стандартного формулювання того, як повинна вирішуватись проблема комп'ютерного зору. Замість того, існує маса методів для вирішення різноманітних строго визначених задач комп'ютерного зору, де методи часто залежать від задач і рідко коли можуть бути узагальнені для широкого кола застосування. Багато з методів та застосувань все ще знаходяться на стадії фундаментальних досліджень, але все більша кількість методів знаходить застосування в комерційних продуктах, де вони часто складають частину складнішої системи, яка може вирішувати складні задачі (наприклад, в галузі медичних зображень або вимірювання та контролю якості в процесах виробництва). У більшості практичних застосувань комп'ютерного зору, комп'ютери попередньо запрограмовано для вирішення окремих задач, але методи, що базуються на знаннях, стають все більше узагальненими.

1.3 Вирішення проблеми слабозорих людей за допомогою технології технічного зору

Актуальність теми. За даними Всесвітньої організації охорони здоров'я, в усьому світі налічується близько 39 мільйонів сліпих людей і 246 мільйонів з поганим зором.

22% становить молодь працездатнього віку, тобто практично кожен п'ятий з усіх сліпих і слабозорих.

У процентному співвідношенні, кількість інвалідів по зору по відношенню до населення Землі (за даними ООН) становить:

39000000 ÷ 7021836029 = 0,0055 повністю сліпі (або 0.55% населення) 246000000 ÷ 7021836029 = 0,035 інваліди по зору (або 3.5% населення).

Людей, які позбавлені можливості бачити від народження або через хворобу, в даний час в Україні налічується 300 тисяч.

За допомогою зору людина отримує 80-90% інформації про навколишній світ. Людей, які позбавлені такої можливості від народження або через хворобу, в даний час в Україні налічується 300 тисяч. За даними Всесвітньої організації охорони здоров'я, щороку кількість сліпих у світі зростає на 1 мільйон осіб, кожні 5 секунд втрачає зір одна доросла людина, кожну хвилину - одна дитина. Щорічно в Україні інвалідами внаслідок вад зору визнають близько 12 тисяч чоловік.

Обсяг ринку для України.

При мінімальній вартості пристрою в 100 у.о. обсяг ринку 30 млн доларів.

При вартості в 400 у.о. - 120 млн доларів.

Сьогодні в світі налічується більше 40 мільйонів сліпих: людей. Крім сліпих від народження, з'являється все більше інвалідів по зору в результаті нещасних випадків і військових дій. З кінця 90-х років спостерігається помітне збільшення числа людей, які втратили зір. Допомогти таким людям відчути себе повноцінними активними членами суспільства – найважливіше завдання реабіяітаціонних центрів і шкіл відновлення працездатності сліпих.

Справжня реабілітація не зводиться лише до створення для інвалідів деяких матеріальних благ, працевлаштування та надання можливості отримувати освіту. Проблема набагато глибше: зараз є смуга відчуження між сліпими і зрячими. Людина, позбавлена зору, особливо, тільки що втратив його, відчуває величезні психологічні труднощі, відчуває свою неповноцінність, часто дистанціюється від суспільства і втрачає віру в свої можливості. Тому перед центрами реабілітації стоїть завдання допомогти сліпим в подоланні негативних наслідків втрати зору, формуванні розумно-оптимістичного погляду на сліпоту і виробленні адекватної особистісної установки. Доля

особистості залежить не від самого дефекту, а від його соціальних наслідків. Дуже важливо не дати сформуватися комплексу неповноцінності.

Аналіз досвіду роботи реабілітаційних центрів розвинених країн показує наступне: життєздатність сучасної системи соціальної реабілітації пояснюється тим, що досягнення педагогіки вдалося з'єднати з досягненнями науково-технічного прогресу. В Україні ж, в силу серйозних економічних проблем, питання розвитку технічних засобів допомоги сліпим вирішується дуже повільно. І незважаючи на те, що більшість керівників і педагогів реабіпітаціонних центрів згодні з необхідністю впроваджувати нові розробки і винаходи в життя сліпих, поки їм доводиться будувати процес на базі вже давно відомих пристосувань.

1.4 Обзор ступення вивченості та наукової розробленості проблеми технічного зору для людей з проблемами зору

Вже з 50-х років XX століття в побуті сліпих, особливо, в розвинених західних країнах, використовуються різні детектори перешкод і сигналізатори небезпеки, використовуючі зазвичай інфрачервону або ультразвукову локацію. Найбільш складною проблемою при цьому є спосіб представлення інформації про геометрію навколишнього середовища у вигляді, який сприймається сліпими. Одним з більш простих і поширених способів побудови локатора для людей з вадами зору є так званий індикатор «вільного шляху» [1], який використовує інфрачервоне, або ультразвукове випромінювання для зондування простору попереду користувача. При появі перешкоди в зоні локації прилад формує в тому чи іншому вигляді сигнал небезпеки. Судити про розташування об'єкта і про особливості його форми такі прилади не дозволяють, але їм притаманні такі переваги як простота освоєння і дешевизна конструкції. Засновниками цього напрямку є Бенхам, Бенджамін, Рассел, Афмстронг.

Друга група – більш складних приладів, які значно виграють за обсягом одержуваної інформації – передбачає максимальне використання слуху користувача. Такі прилади вже дозволяють визначати відстань до одного або декількох перешкод одночасно, а також оцінювати швидкість їх переміщення. Локатори цієї групи, що використовують бінауральне сприйняття, дозволяють також оцінювати напрямок на перешкоду. Найбільш відомим розробником і дослідником цієї групи в західних країнах є Кей [2]. У СНД його послідовниками є Зенін В.Я. і Саприкін В.А.

Третій напрям у розвитку –уявлення фронтальних образів середовища. В таких приладах використовується телевізійна камера, а інформація про середовище постуцает через матрицю тактильних стимуляторів, розташованих на поверхні якої-небудь ділянки тіла користувача (метод Колінса), або через матрицю нейростимуляторів, імплантованих в мозок. В першому варіанті основна проблема полягає в низькій роздільній здатності, а в другому – в необхідності оперативного втручання в мозок і у високій вартості пристрою.

Аналіз існуючих рішень. Поряд з класичними методами адаптації сліпої людини (собака-поводир, тростина), сьогодні з'являється все більше електронних пристроїв подібного призначення (табл. 1).

Таблиця 1. Сучасні рішення для незрячих людей

Назва пристрою	Вхідний сигнал	Сповіщення про перегороду	Візуальне уявлення пристрою	Навігація	Ціна, у.о.
Електро-сонар	Датчик УЗ	Вібро або звуковий сигнал	Браслет на зап'ясті	–	–
Ultra Cane	Датчик УЗ	Вібросигнал	Трость	–	618
СОНАР-1	Датчик УЗ	Вібро або звуковий сигнал	Трость	–	–
СОНАР-3	Датчик УЗ	Звуковий сигнал	Підвіска на шиї	–	–
OrNavi	Координати з GPS-ГЛОНАСС модуля	Голосове сповіщення	Пульт управління	GPS-ГЛО-НАСС	360
OrCV	Відео камера x2	Голосове сповіщення	Окуляри	–	230
vOICe	Відео камера	Звукові образи	Окуляри, шолом з камерою	GPS	600

Але дані рішення зовсім не користуються популярністю серед незрячих та слабозорих людей. Причина цього аж ніяк не в консервативності людей з обмеженими можливостями. Адже вони з задоволенням користуються мобільним зв'язком, послугами спеціалізованих call-центрів та іншими досягненнями прогресу. Розгадку слід шукати в функціональності пристроїв.

2. Методика детектування перешкод на основі мікрохвильового радару AWR 1843

2.1 Основні принципи роботи мікрохвильового радару

Принцип роботи радіолокатора має на увазі наявність двох основних компонентів: приймача і передавача (рис. 1). Передавач формує високочастотний сигнал, який, відбиваючись від перешкод, фіксується приймачем [3]. За величиною затримки між переданим і прийнятим сигналом можна розрахувати відстань до об'єкта. Компанія TI для вирішення завдання виявлення об'єктів пропонує використовувати технологію mmWave (Millimeter wave) з міліметровим діапазоном довжин хвиль. Розглянемо деякі особливості її фізичної реалізації.

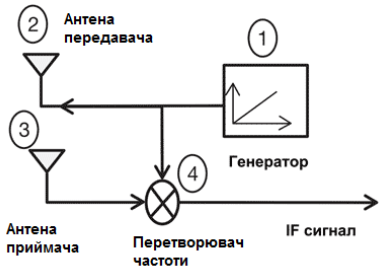


Рис. 1. Структура сенсору наближення

На відміну від традиційних радарних систем, в яких використовують імпульсну передачу сигналу, в mmWave застосовується частотно-модульований безперервний сигнал (frequency modulated continuous wave, FMCW) (рис. 2). Частота сигналу лінійно змінюється в часі від 77 ГГц до 81 ГГц і назад.

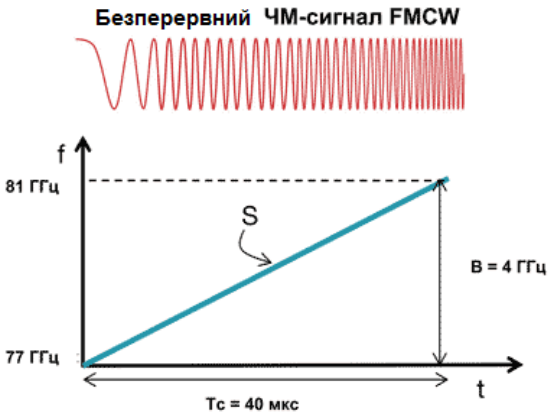


Рис. 2. Початковий сигнал в датчиках mmWave від Texas Instruments

Як і в звичайному радіолокаторі, в mmWave пряма радіохвиля досягає об'єкта, відбивається і повертається назад до датчика, де фіксується приймачем. Після цього вихідна і відбита хвилі микшируются в перетворювачі частоти. Результуючий вихідний сигнал IF (intermediate frequency) цікавий тільки в момент перекриття вихідних сигналів (рис. 3). При цьому він має постійну частоту і фазу, рівну різниці фаз вихідних сигналів. Використовуючи нескладні перетворення, за отриманими даними можна визначити відстань до об'єкта.

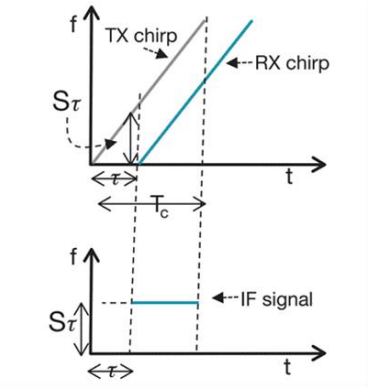


Рис. 3. Зсув прямого і відбитого сигналів і отримання вихідного сигналу постійної частоти

Технологія mmWave дозволяє виявляти кілька об'єктів одночасно. При цьому в спектрі вихідного сигналу IF будуть присутні кілька тонових частот [4]. Кожна складова спектра може бути виділена за допомогою перетворення Фур'є (рис. 4).

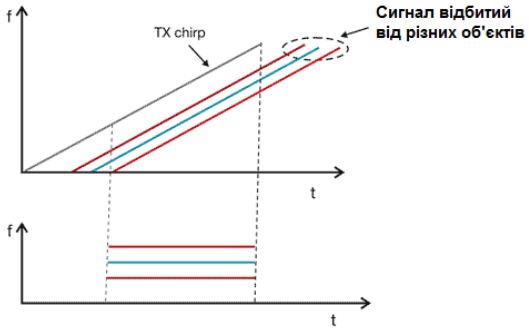


Рис. 4. Одночасне виявлення декількох об'єктів

За допомогою mmWave вдається також легко визначати частоту обертання різних об'єктів. Для цього потрібно кілька передавачів. Кожен передавач формує свій сигнал зі зміщенням один щодо одного. Детектування відбитих сигналів виробляється одним приймачем. В результаті буде спостерігатися періодичне зміщення фази вихідного сигналу IF. З цього зміщення можна судити про рух об'єкта (рис.5).

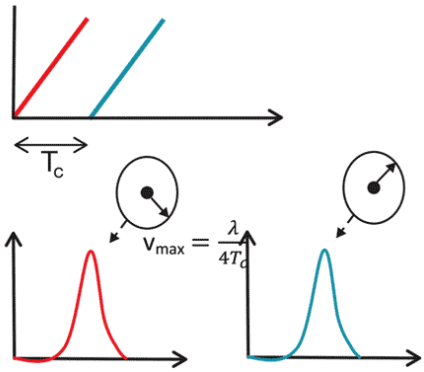


Рис. 5. Виявлення руху об'єкта

При наявності пари приймачів можна визначати не тільки відстань до об'єкта, але і його кутове положення (рис. 6). В даному випадку для розрахунків використовуються отримані значення відстаней до кожного з приймачів (d) і елементарні геометричні співвідношення.

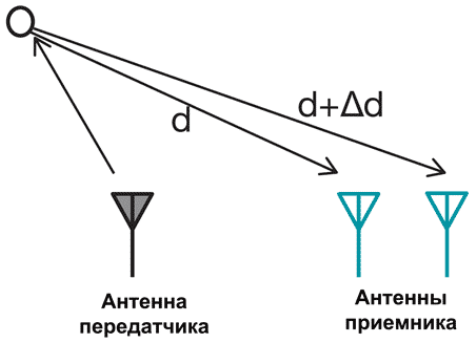


Рис. 6. Визначення кутового положення об'єкта

Запропонована технологія виявляється досить перспективною для автомобільних, промислових і комерційних додатків [5]. Також є дуже нетривіальною задачею використання даної технології для допомоги сліпим. Дана технологія є новою та пристроїв з використанням цієї технології ще не представленно на ринку.

Однак до сих пір її реалізація ускладнювалася цілою низкою цілком об'єктивних причин. Можна виділити три найбільш значущих з них:

• необхідність створення складної системи, яка об'єднує: передавачі, приймачі, синтезатор частот, ВЧ-аналогові ланцюги модуляції і демодуляції, ЦАП, АЦП, процесор;

• необхідність використання дискретних компонентів, що негативно позначалося на вартості [6];

• значні габарити системи, що не дозволяло створювати компактні мобільні пристрої.

2.2 Виявлення залежностей відбиття сигналу від типу перешкоди

Для демонстрування можливостей mmWave радару AWR 1843 та знаходження закономірностей було проведено експеримент детектування різних перешкод та зведено у таблицю 2.

Дані отримані у ході цього та інших експерементів було покладено в основу методики детектування перешкод.

Таблиця 2. Сучасні рішення для незрячих людей

Тип перешкоди	Розмір перешкоди, см	Відстань реальна (вимірена сенсором), см	Ампітуда	Примітка
Аркуш паперу	15*21	50 (52)	92	
Ключі	8*3*0,8	50(56)	80-115	Від повороту площини ключів змінюється амплітуда від практично не можливо відокремити від шуму до ярковиделенного (Амплітуда 115)
Залізо	16*12,5	50 (56)	115	
Стіна	200*400	50 (56)	112	
Пластик	12,5*14	50 (52)	120	
Тканина (бавовняна хустинка)	19*19	50 (56)	95	
Шнур пластиковий вертикаль-ний	0,5 *100	50 (-)	-	Важко виділити шнур серед шуму

На рисунку 7 наведено відбиття сигналу від металлевої пластини, що демострує якісне розпізнавання металлевих перешкод за допомогою мікрохвильового радару AWR 1843. Аналогічні експерименти булі виконанні для об’єктів з різних матеріалів та різної форми, так як на якість

розпізнавання впливає не тільки матеріал перешкоди, а також форма. Чим більше кутів має перешкода тим нижче сила відбиття сигналу та тим вище вірогідність не помітити перешкоду.

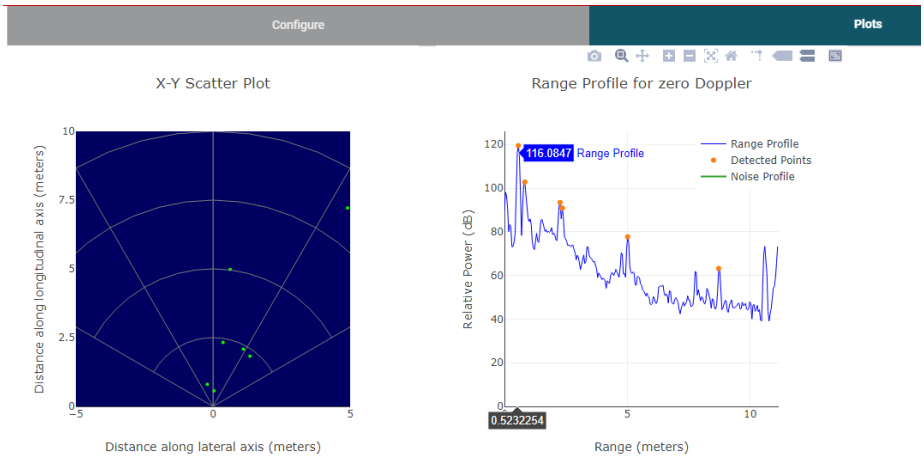


Рис. 7. Детектування металевої пластини

2.3 Архітектура для реалізації системи інформаційної підтримки людей з проблемами зору

Дуже важливо розуміти як відбувається обмін данними між компонентами нашої системи, тому створення загальної архітектури є необхідним.

Система складається з декількох головних компонентів:

– Мікрохвильовий радар AWR 1843;

– Комп'ютер (що може бути представлений як персональним комп'ютером, так і мікроконтролером з операційною системою, наприклад, Rasberry);

– Користувач.

Ми використовуємо заводську прошивку. Першочергово створюємо конфігураційний файл з параметрами та відправляємо з комп'ютеру на AWR1843, в результаті чого ми задаємо дані, які нам необхідно отримати та в якому ви-гляді. Використовується схема кодування TLV (Type-length-value) для обміну даними між AWR1843 та програмою на комп`ютері (рис. 8). Для кожного кадру надсилається пакет, що складається з фіксованого розміру заголовка кадру, а потім змінної кількості TLV.

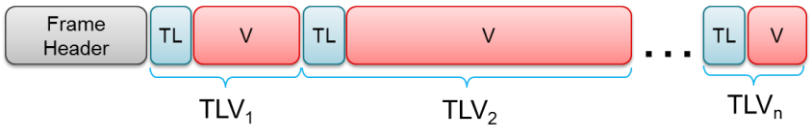


Рис. 8. Формат даних

Для отримання даних було розроблено парсер даних. Ми отримуємо масив байтів. Для кожного TLV алгоритм розбору різний. Але у всіх є одна особливість. Після кожного звернення до елементу буферу ми повинні зміститися на наступний елемент.

2.4 Парсинг списку виявлених об'єктів

Список виявлених об'єктів уявляє собою уявляє собою масив даних у форматі TLV (Тип-Довжина-Значення).

Тип: (MMWDEMO_OUTPUT_MSG_DETECTED_POINTS)

Довжина: (Кількість виявлених об'єктів) * (розмір DPIF_PointCloudCartesian_t)

Значення: Масив виявлених об'єктів.

Інформація про кожен виявлений об'єкт відповідає структурі DPIF_PointCloudCartesian_t (таблиця 3).

Таблиця 3. Представлення даних

float	**x** x - координата в метрах. Ця вісь паралельна площині датчика і складає площину азимута з віссю y.

float	**y** y - координата в метрах. Ця вісь перпендикулярна площині датчика, позитивний напрямок від датчика.
float	**z** z - координата в метрах. Ця вісь паралельна площині датчика та задає площину висоти з осю y. Позитивний напрямок z вище датчика, а негативний нижче датчика.
float	**velocity** Оцінка швидкості доплера в м / с. Позитивна швидкість означає, що ціль віддаляється від датчика, а негативна швидкість означає, що ціль рухається до датчика.

Коли кількість виявлених об'єктів дорівнює нулю, цей елемент TLV не надсилається. Максимальна кількість об'єктів, які можна виявити в підкадрі/кадрі, - DPC_OBJDET_MAX_NUM_OBJECTS.

Орієнтація осей x, y та z відносно датчика відповідає наступному рисунку 9.

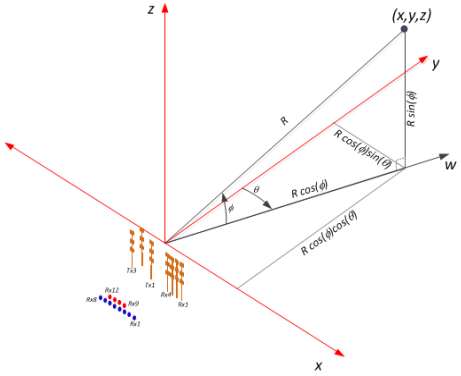


Рис. 9. Орієнтація осей x, y та z відносно датчика

Вся структура виявлених об'єктів TLV проілюстрована на рисунку 10.

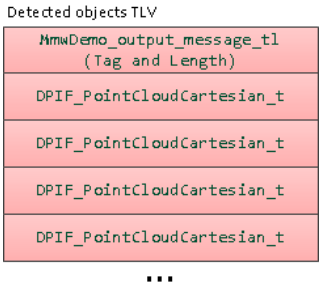


Рис. 10. Структура TLV виявлених об’єктів.

Аналогічно відбувається парсинг даних про силу сигналу на певній відстані та парсинг Heatmap.

2.5 Парсинг та аналіз Heatmap

Heatmap уявляє собою масив даних 256*8. Використовуємо полярну систему координат. 256 – це кількість рядків, це значення за замовчуванням. Один рядок відображає – 0.044 м. Тобто 256*0,044 = 11,264 м. Це значення зазначено в конфігурації. Кількість стовпців дорівнює 8, так як використовуємо 8 антен. Кожна антена працює в діапазоні 15 градусів. Тобто загальний кут огляду AWR 1843у цій конфігурації складає по 60 градусів в обидві боки (рис. 11).

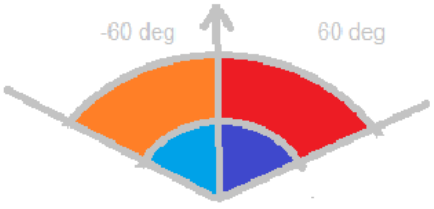


Рис. 11. Графічне представлення Heatmap

Аналізуючи Heatmap ми встановили закономірності притаманні для деяких видів перешкод. Наприклад, сходам притамані 3 сплески з силою відбиття не більше 5000 одиниць та різниця між максимумами сплесків в діапазоні між 5 та 10 юнітів. А для стіни це один сплеск зі значенням близько 7000-10000 одиниць.

Висновки

Було проведено огляд ступеня вивченості і наукової розробленості в ході якого було вибрано місце дослідження і були проаналізовані вже існуючі методики. Був проведений огляд предметної області аналіз стану сучасних систем технічного зору, детально описано актуальність та перспективність цієї технології. Огляд існуючих рішень дозволяє зрозуміти, що робляться вони без структурного підходу, дозволяючи тільки частково вирішити поставлене завдання технічного зору для сліпих. Відсутність відпрацьованих методик комп'ютерного зору перешкоджає ефективному використанню даної технології, що визначає необхідність розробки нової методики. Було обрано новітній мікрохвильовий радар для детекції перешкод у просторі людьми з проблемами зору. Було розроблено парсер даних та проведено аналіз отриманих даних, розроблено алгоритми та реалізовано програмне рішення, яке підтверджує ефективність розробленого методу. Розроблено інформаційну підтримку для людей з проблемами зору на основі AWR 1843, яка має можливість попередження про перешкоду та визначення типу деяких перешкод.

Список літератури

1. E. R. Davies. Machine Vision : Theory, Algorithms, Practicalities, pp. 14 – 16. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA (2004).
2. Batchelor B.G. and Whelan P.F. Intelligent Vision Systems for Industry. pp.84 – 85. Springer-Verlag, ISBN 3-540-19969-1 (1997).
3. Demant C., Streicher-Abel B. and Waszkewitz P. Industrial Image Processing: Visual Quality Control in Manufacturing. Springer-Verlag, ISBN 3-540-66410-6, (1999).
4. Gonzales R. C. and Wintz P. A. Digital Image Processing. Longman Higher Education, ISBN 978-0201110265, (2001)
5. Pham D.T. and Alcock R.J. Smart Inspection Systems: Techniques and Applications of Intelligent Vision. Pp. 55 – 56. Academic Press, ISBN 0-12-554157-0, (2003).
6. Berthold K.P. Horn. Robot Vision. MIT Press. ISBN 0-262-08159-8. (1986).

Survival analysis methods for churn prevention in telecommunications industry

Mariia Havrylovych[1] and Nataliia Kuznietsova[1]

[1] *Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine*

maria.babich@gmail.com, natalia-kpi@ukr.net

Abstract. This paper is devoted to the problem solving of churn prevention in real companies. This is really actual and important so modern algorithms for forecasting the probability are needed. The authors proposed approach to focus not only on the probability but also on the time period when the churn can happen. For this reason it was developed two algorithms based on the using of survival functions and giving the forecast of the time period of the churn. It was developed two different algorithms based on forecasting the time period for risk increasing based on the critical total losses or probability of survive which are defined by the company and really depends from its strategy and the situation on market. If the risk function is determined in the process of modeling through parametric, non-parametric distribution, then the calculation of time through the derived risk function is possible. Using and results of the proposed algorithm based on the set threshold of the risk probability is shown on the IBM dataset. Different types of models such as semi-parametric Cox Proportional Model and parametric Weibull and Log-normal survival models were used. The best model by the statistical criteria such as log-likelihood value was the log-normal model. Also a step-by-step outflow process in decision support system for churn detection and defining in time the most dangerous groups of clients who are thinking to churn was proposed.

Keywords: Churn, Survival Analysis, Risk Analysis, Cox Proportional Hazard Model, Telecommunication Company.

1. Introduction

Development and expansion of the company, increasing its status and importance on the market are the main tasks for each business and the part of its strategy. The success of the business really depends on the quantity of the clients, facilities and volumes of their orders, achieved profits, part of the company on the market, the benefits of the company in comparison with its main competitors. Big efforts of the client-oriented companies are going on improving the service and keeping and saving their clients from churn.

This is the man feature for such business: firstly demand of the quantity of the clients, then prevent their churn. This also characterizes the telecommunication industry where the revenue depends on a lot of customers. That's why the building forecasting models in terms of the probability of each customer to leave and classification of such customers is really important. There are a lot of different prediction tools to detect customer who is going to leave. But companies also need to know time when churn could happen. This information gives the opportunity to detect when the most probable time it could be and to use the means for its prevention. For this reason it is important what are that factors which effect on customers' churn and what is the influence - greater or lesser extent. The recent publications in such area show that the problem of customer's outflow is really important not only from budgeting, finance, marketing, logistics, but also from the using of the latest technologies and planning their loading in next moments. In [1] the modeling of customer life time value is shown. It is useful for further analysis of latent behavioral patterns and for developing the successful marketing strategies. In [2, 5, 6, 8, 10, 11] authors make survey of the survival analysis for churn prediction application and explain how these methods help to understand churn risk. Also, beside survival analysis, different machine learning techniques widely used for churn prediction (decision trees, Bayes classifier, ANN, SVM etc.). The application of the neural network techniques is discussed in the research article [4]. In the work [3] authors faced with the issue of imbalanced datasets, and in most cases different features were extracted from data, and some imbalance correction has been made as in work [3]. In [6, 7] the authors made survey on data mining approaches to understanding client's behavior, client's management and client's segmentation. In [8] using of survival analysis models for credit risks is discussed. Our research is dedicate to exploring the possibility of using survival analysis models for churn prevention for different customers segments (for this reason were build and compared the semi-parametric and parametric survival models) and retrieving useful information for churn prevention of segments or classes with higher churn risk. Also we propose some elements of decision support system for churn prevention.

2. Problem statement

To develop the specific algorithms based on survival models which give the opportunity to forecast the specific time of the churn risk growing. To give the mechanisms for detection the most marginal clients from the revenue point of view and to determine the moment of time for such losses. For giving the possibility to forecast the moment of time (many times) during data analysis develop the dynamic algorithms as the elements of informational technology for including them in decision support system in churn prevention.

3. Brief review of survival models and terms

Let's make a quick survey on survival analysis models used in this research. The basic concepts of survival analysis are such as survival function, hazard or risk function.

Survival function is determined as

$$S(t) = 1 - F(t) = P(T > t), for\ t > 0,$$

where F(t) is cumulative distribution function and T is some random non-negative variable.

Hazard function is used for describing the risk of some occasion and its statement is:

$$h(t) = \frac{f(t)}{S(t)}.$$

In our research we used different types of survival function for determining the best model which describes the churn process. For this reason we built semi-parametric model (Proportional Cox) and parametric accelerated failure time models (such as Weibull model and LogNormal model) and compare their behavior in time.

Cox Proportional Hazards Model is determined as:

$$h(t) = h_0(t)e^{(b_1X_1+\cdots+b_nX_n)}.$$

Here we can see the $h_0(t)$– baseline hazard which involves time, and exponent of linear combination of all predictors, which does not involve time and h(t) is expected hazard at time point t.

Lognormal Model:

The log-normal distribution is parametric distribution for characterizing the survival time. The log-normal distribution is denoted as $LN(\mu, \delta^2) \sim exp\{N(\mu, \delta^2)\}$.

$$F(t) = \Phi\left(\frac{log\ log\ (t) - \mu}{\sigma}\right)$$

$$f(t) = \frac{\varphi\left(\frac{log\ log\ (t) - \mu}{\sigma}\right)}{t\sigma}$$

$$h(t) = \frac{f(t)}{F(t)}.$$

Where φ is the probability density function of standard normal distribution and Φ is cumulative distribution function.

Weibull model:

Weibull distribution is denoted W(p, λ).

$$F(t) = 1 - e^{-(\lambda t)^p}$$

$$f(t) = p\lambda^p t^{p-1}\mathbf{e}^{-(\lambda t)P}$$

$$h(t) = p\lambda^p t^{p-1}, t > 0, \lambda > 0, p > 0.$$

4. Risk prediction algorithms

The main idea is that the typical groups of the clients behave really similar in time. Survival approach gives the possibility to define the probability of survive and hazard risk. In the terms of risks it is the way of the classification problem solving and forecasting the probability of the case. For real enterprises churn prediction needs not only the probability forecast the fact that client is going to make outflow but also the time period when it could happens.

Visual comparison of survival curves for different strata or groups, which is often used, is possible, but not very convenient, in predicting the time of risk. For tasks where data slices arrive at a time interval daily, hourly, minute by minute, and time prediction accuracy is critical, time setting algorithms need to be developed.

It was developed two different algorithms based on forecasting the time period for risk increasing based on the critical total losses or probability of survive which are defined by the company and really depends from its strategy and the situation on market.

There are several different approaches to this: if the risk function is determined in the process of modeling through parametric, non-parametric distribution, then the calculation of time through the derived risk function is possible.

4.1. Algorithm 1 for calculating the moment of transition to a higher degree of risk

To determine the time t, follow these steps:

1. Specify the type of initial function $\hat{\Lambda}_0(t)$. Let $\hat{\Lambda}_0(t) = exp(-a \cdot t)$ it where, but $a > 0$ it matters little so that $\hat{\Lambda}_0(t) = exp(-a \cdot t)$ it does not fall off quickly.

2. Substituting the basic hazard function for proportional risks, we obtain:
$\hat{\Lambda}(t) = exp(x^T \cdot \beta^{PHM}) \cdot exp(-a \cdot t)$.

3. Taking a derivative of the function of danger in time, we obtain: $\frac{\partial \Lambda(t \mid x)}{\partial t} = exp(x^T \cdot \beta^{PHM}) \cdot exp(-a \cdot t) \cdot (-a) = -a \cdot exp(x^T \cdot \beta^{PHM} - a \cdot t)$.

To be able to differentiate a function, it is necessary to have a derivative of this function, which cannot always be guaranteed. Therefore, it is possible to adapt the algorithm without directly calculating the derivative. Based on the definition of the derivative as the rate of change of a function, it is possible to calculate the rate of change of probability, that is, its transition to the critical probability of occurrence of risk $P_{critical}(t)$:

1. The critical probability $P_{critical}$ is set.

2. The value is calculated as the "probability margin" $P_{current}(t) - P_{critical} = \Delta P$.

3. The moment of transition of risk to critical is defined as:

$$t_{ciical} = \frac{\Delta P}{\frac{\partial P(t)}{\partial t}}.$$

If it is impossible to establish a probability of risk that is critical, then it is proposed to develop an algorithm for calculating time at a critical (or catastrophic) level of risk, that is, critical (or catastrophic) losses. This is always possible because the financial system or business operate to generate some profit and therefore can determine which risk losses are greater than the profit generated.

4.2. Algorithm 2 to determine when critical (catastrophic) level of loss risk occurs

The algorithm consist from such steps:

1. Specify the time interval $\Delta T = (0, T)$ at which the critical time will be searched.
2. Set a step to increase time $t := \Delta t$.
3. $t := 0$.
4. Calculate $Losses_{nomoч}(t)$.
5. If $Losses_{nomoч}(t) \geq Losses_{крит}$ then $t_{крит} := t$ STOP.
6. $t := t + \Delta t$.
7. If $t \geq \Delta T$, then STOP and in this interval there will be no transition of risk to critical (catastrophic).

Go to Step 4.

As defined in the previous step, such variables could be calculated: $\lambda(t \mid x)$, $P(t \mid x)$ and the expected losses EL for: acceptable, critical and catastrophic level of risk at times (t_1, t_2, t_3).

The algorithms developed allow us to determine not only the degree and level of risk, as predicted in the static evaluation, but also to predict the time when the level or degree of risk change dramatically.

5. Churn prevention as the process in decision support system

For proposing some means for churn prevention we need to understand and present the process of suspicious clients detecting: defining their types and presenting. Here we propose the main stages of detecting in decision support system (DSS) for churn prevention using of survival analysis and building the models for this. On the figure 1 we can see the pipeline of this process stages in DSS. Let's consider and

discuss deeper each step of it. First of all we need to collect the representative set of data for building the training models. Usually it is useful to use all the existing data for analysis but on the next steps it is better to build the training set based on the real assumption on the data distribution or previous modeling experience. This dataset is going to be divided in three parts: training, test and validation. While this dataset is using for solving the classification problem there must be some event column which defines if there is a churn or not for each observation, and the time column (time to event occurrence). If they don't exist these columns should be extracted from other variables. In our case the time variable is meaning the time to be staying with this company. On these data we will train some machine learning model with ability to return the probability of predicted event. For example in some cases the logistic regression returns good prediction of the probability estimation. Otherwise other models can give poor probability estimation of predicted classes, or even don't have probability predictions. If chosen machine learning model doesn't return precise probability estimation we must use calibration techniques for probability prediction. In most cases the dataset is imbalanced and predicted (important for study) event occurrence is rare. In this case the analyst should to set up the importance of making type I and type II errors, and which type of error is more crucial for the analysis. For churn prevention made in this study it is important to detect when and which clients are going to make churn. That's why the errors of type II are more important. Proceeding from this there is a need for model parameters fine tuning (for example to set the weights for classes). Then it is the stage of finding the best model for this dataset. Firstly the model is built and trained and appropriate probability threshold is set. Such threshold helps to set for all predicted labels probability which are above to one class, and rest for another. This threshold is selected by the principle of reducing the rate of errors of type 2. The next step is devoted to the training of the survival regression model. If hazard proportional assumption on data is acceptable we can use Cox proportional Hazard Model, otherwise Kaplan-Mayer or Log-normal model. In our practical task it is useful to make survival analysis for different clients' cohorts and give more precise recommendations for these classes. With the best models have been selected and with received on the previous step threshold we can give some recommendations: if the client's survival probability is close from left side or lower than this threshold, the client is going to churn and company has to use the prevention methodology (special promotion cases, advertisement, extra bonuses, etc.) to avoid this.

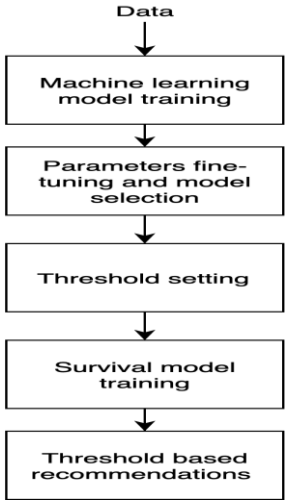


Fig.1. Churn prediction process outflow in decision support system pipeline.

In our research we use the IBM dataset for telecommunication company client's churn [9]. The data set includes information about:

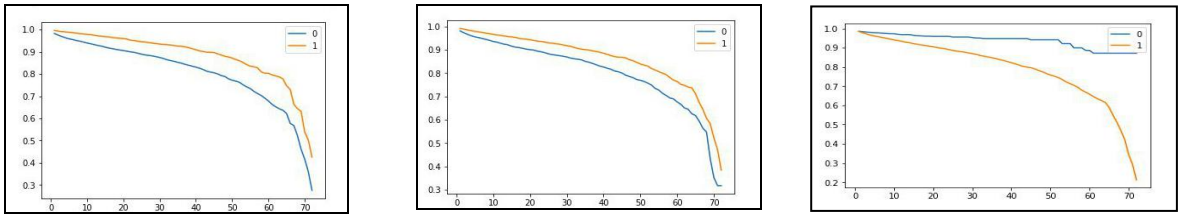
1. Customers who left the company within the last month – the column is called Churn.
2. Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies.
3. Customer account information –customer tenure, contract type, payment method, paperless billing, monthly charges, and total charges
4. Demographic info about customers – gender, age range, and if they have partners and dependents [9].

We use lifelines Python library, as duration column we use tenure (number of months client "life") and Churn as event column. Below we can see the importance of the variables in each model (results are provided in table 1). The blank cells in p-value column tell that p-value is <0.005.

Table 1. Results of Weibull, Lognormal and Cox models

Covariate name	Weibull		LogNormal		Cox Model	
	Exp (coef)	p-value	Exp(coef)	p-value	Exp (coef)	p-value
Partner	1.67		1.75		0.59	
Dependents	0.91	0.19	0.92	0.23	1.05	0.49
MultipleLines	1.74		1.87		0.59	
Internet Service	0.39		0.41		2.38	
Online Security	2.14		2.17		0.49	
Online Backup	2.01		2.08		0.49	
DeviceProtection	1.47		1.53		0.68	
TechSupport	1.72		1.65		0.62	
StreamingTV	1.18	0.01	1.26		0.86	0.01
Streaming Movies	1.29		1.28		0.78	
Contract	1.14		1.13		0.87	
PaperlessBilling	0.81		0.85	0.01	1.21	
MonthlyCharges	0.69		0.70		1.43	
automatic_payment	1.95		2.01		0.54	
Concordance	0.87		0.87		0.87	
Log-likelihood	-8921.22		-8845.61		-13889.35	

Analyzing the results, we can see that all three models show similar patterns of cohorts, what kind of customers are more likely to churn, and which are more reliable. This information is useful for this telecom company to plan in another way its promotion and marketing means and give the possibility to admit the clients who are thinking to churn and save (prevent from churn) more amount of customers. But also with survival analysis tools we are able to not only divide customers into some groups but also to define the time period when the crucial churn probability in some point of time is achieved. As described above we can set some cut-off probability with different methods: analytical, or methods based on the using of some machine learning models (for example logistic regression or ensemble model). After receiving the probability of each observation we can tune the probability threshold, where the customer with probability value below the threshold will stay and other will leave. We tune this value to decrease the quantity of errors on churn customers and at the same time we increase the errors of customers who are going to stay with the company. We are going on this step well-considered while the churn detection is more crucial for us. So after the setting the threshold parameter we can use it to detect at which point of time customer begin just thinking about leaving and give the recommendation to the telecommunication company, to whom it is reasonable to take some actions to prevent this event. We build stratified Cox Model on that features that have great impact on churn, based on previous results shown in table 1. As a reminder: the most important variables (and services for clients) were such as OnlineSecurity (so safety of the data are really important), Partner (using of family propositions and using one operator are quite often) and InternetService(providing packages of free Internet surfing, using of social networks unlimited) feature values. Results which were received during the modeling are displayed at fig 2.



а) б) с)

Fig.2. a) Online Security, b) Partner, c) Internet Service.

Table 2. Churn time (in months) and error rates of type I error and type 2 error and baseline survival time

Prob	2_type_err	1_type_err	baseline_survival
0.3	0.072	0.46	61
0.35	0.10	0.41	65
0.4	0.13	0.35	66
0.45	0.15	0.32	68
0.5	0.18	0.29	69
0.55	0.21	0.25	69
0.6	0.26	0.21	70
0.65	0.32	0.17	71
0.7	0.41	0.13	72
0.75	0.48	0.10	72
0.8	0.62	0.06	72
0.85	0.84	0.02	72
0.9	0.97	0.00	72
0.95	1	0	72

So we can use results above to find the critical time point, when client is going to leave. For finding the appropriate threshold for our task we made the modeling by logistic regression and from the point of second-type errors we set the threshold for value of probability that the client is going to leave more than 65% that means that the probability of its survive (that he is going to stay with the company) as p_{tr} = 0.35. For people without online security we have 65 months from begin. For people without partner 62 monthes. From this time points company should begin doing some prevention actions for customer retention. More detailed view is shown in table 2 below.

In tables 2-6 presented above the first column called “prob” is the probability threshold, so when we compute time of surviving we take first probability. Also there are I type error rate and II type error which depend on the probability threshold. The value that we predict was the time (number of months), when the client is doing churn. As we can see people with good quality for Internet service is less tending to churn. It is quite understandable for the telecom industry while the quantity of calls, SMS, MMS is decreasing every day. At the same time the quantity of the Internet service and the variety of online applications are increasing. That's why the most important for the client become the stable and speed Internet and reasonable price on it. Also a big impact for the clients become the paperless billing and automatic payments for charging the mobile number.

Table 3. Churn time for different InternetService cohorts

prob	internet_service	no_internet_service
0.3	72	56
0.35	72	60
0.4	72	64
0.45	72	65
0.5	72	67
0.55	72	68
0.6	72	69
0.65	72	69
0.7	72	70
0.75	72	71
0.8	72	72
0.85	72	72
0.9	72	72
0.95	72	72

Table 4. Churn time depends on MultipleLines cohorts

prob	no_multiple_lines	multiple_lines
0.3	65	59
0.35	66	64
0.4	68	68
0.45	69	69
0.5	69	70
0.55	70	71
0.6	71	72
0.65	71	72
0.7	72	72
0.75	72	72
0.8	72	72
0.85	72	72
0.9	72	72
0.95	72	72

Table 5. Churn time depends on PaperlessBilling cohorts

prob	paperless_billing	no_paperless_billing
0.3	64	59
0.35	66	64
0.4	67	66
0.45	69	67
0.5	69	68
0.55	69	69
0.6	71	70
0.65	71	71
0.7	72	71
0.75	72	72
0.8	72	72
0.85	72	72
0.9	72	72
0.95	72	72

Table 6. Churn time depends on “autopay” cohorts

prob	no_autopay	autopay
0.3	57	65
0.35	60	67
0.4	64	68
0.45	66	69
0.5	67	69
0.55	69	70
0.6	70	71
0.65	71	72
0.7	71	72
0.75	72	72
0.8	72	72
0.85	72	72
0.9	72	72
0.95	72	72

6. Conclusion

In this work we investigate the possibility of using the survival statistical models in churn prediction. We decided to go away from the classical classification task and to use survival models for prediction the time of more probable churn for the clients. We used different type of models such as: semi-parametric Cox Proportional Model and parametric Weibull and Log-normal survival models. The best model by the

statistical criteria such as log-likelihood value was the log-normal model. We propose a step-by-step outflow process in decision support system for churn detection and defining in time the most dangerous groups of clients who are thinking to churn. Thus defining the type of the clients and beforehand the period range to give the company the time slot and also the possibility to use the prevention methods and reduce the churn. By setting some threshold probability value we can find time point for different cohorts of clients, when client is going to leave. This is useful, because company can prevent churn of client in cohort with greater risk. Such approach could be used in other applications, for example, in human behavior research in critical infrastructures' organizational management system [12]. In further this churn prevention system can be more complex and sophisticated to give better and more accurate recommendations. As further research there might be used more complex algorithms to predict survival probability, for example some deep learning recurrent neural networks, which help to discover nonlinear complex relationship between data variables.

References

1. Junxiang, Lu. Modeling Customer Lifetime Value Using Survival Analysis − An Application in the Telecommunications Industry - http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.1947&rep=rep1&type=pdf, last accessed 2019/11/07.
2. Junxiang, Lu. Predicting Customer Churn in the Telecommunications Industry — An Application of Survival Analysis Modeling Using SAS - https://support.sas.com/resources/papers/proceedings/proceedings/sugi27/p114-27.pdf, last accessed 2019/11/07.
3. Clifton Phua, Hong Cao, João Bártolo Gomes, Minh Nhut Nguyen. Predicting Near-Future Churners and Win-Backs in the Telecommunications Industry https://arxiv.org/pdf/1210.6891.pdf, last accessed 2019/11/07.
4. Sharma, D. Panigrahi, P. Kumar. A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. International Journal of Computer Applications (0975 – 8887) Volume 27– No.11, August 2011 - https://arxiv.org/pdf/1309.3945.pdf, last accessed 2019/11/07.
5. Grishchenko D., Kataev A. Analysis of methods of modeling and forecasting of customers outcome. https://cyberleninka.ru/article/n/analiz-metodov-modelirovaniya-i-prognozirovaniya-ottoka-klientov, last accessed 2019/11/07.
6. Dyulicheva Yu.Yu., Ryabchenko EA About approaches to modeling and optimization of work with service consumers. http://dspace.nbuv.gov.ua/bitstream/handle/123456789/92501/56-Diulycheva.pdf?sequence=1, last accessed 2019/11/07.
7. Shin-Yuan Hung, David C. Yen b, Hsiu-Yu Wang: Applying data mining to telecom churn management. Expert Systems with Applications 31 (2006) 515–524 - http://didawiki.cli.di.unipi.it/lib/exe/fetch.php/dm/telecomchurnanalysis.pdf, last accessed 2019/11/07.
8. Kuznietsova N. V., Bidyuk P.I., Modeling of credit risks on the basis of the theory of survival. Journal of Automation and Information Sciences. 2017. N.49(11), pp. 11-24.
9. Using Customer Behavior Data to Improve Customer Retention. 2015. Homepage - https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights-in-the-telco-customer-churn-data-set/, last accessed 2019/11/07.
10. Kuznietsova N. V., Information Technologies for Clients' Database Analysis and Behaviour Forecasting. CEUR Workshop Proceeding (ISSN 1613-0073). 2017. Vol. 2067. P.56-62 [Online]. Available: http://ceur-ws.org/Vol-2067/, last accessed 2019/11/07.
11. Kuznietsova N., Bidyuk P. Forecasting of Financial Risk Users' Outflow. IEEE First International Conference on System Analysis & Intelligent Computing (SAIC), (Kyiv: 08–12 October). 2018. P.250-255. DOI: https://ieeexplore.ieee.org/abstract/document/8516782.
12. Dodonov, A., Gorbachyk, O., Kuznietsova, M. Increasing the survivability of automated systems of organizational management as a way to security of critical infrastructures. CEUR Workshop Proceeding (ISSN 1613-0073). 2018. Vol. 2318. P.261-270 [Online]. http://ceur-ws.org/Vol-2318/, last accessed 2019/11/07.

COMPARING EFFICIENCY OF EXPERT DATA AGGREGATION METHODS

Sergii Kadenko[0000-0001-7191-5636], Vitaliy Tsyganok[0000-0002-0821-4877]
and Aleksandr Karabchuk

Institute for Information Recording of the National Academy of Sciences of Ukraine

Abstract. Expert estimation of objects takes place when there are no benchmark values of object weights, but these weights still have to be defined. That is why it is problematic to define the efficiency of expert estimation methods. We propose to define efficiency of such methods based on stability of their results under perturbations

of input data. We compare two modifications of combinatorial method of expert data aggregation (spanning tree enumeration). Using the example of these two methods, we illustrate two approaches to efficiency evaluation. The first approach is based on usage of real data, obtained through estimation of a set of model objects by a group of experts. The second approach is based on simulation of the whole expert examination cycle (including expert estimates). During evaluation of efficiency of the two listed modifications of combinatorial expert data aggregation method the simulation-based approach proved more robust and credible. Our experimental study confirms that if weights of spanning trees are taken into consideration, the results of combinatorial data aggregation method become more stable. So, weighted spanning tree enumeration method has an advantage over non-weighted method (and, consequently, over logarithmic least squares and row geometric mean methods).

Keywords: Expert Data Aggregation, Decision-making Support, Estimate, Simulation, Pair-wise Comparison Matrix.

1. Introduction

Expert estimation is a powerful decision support tool for weakly structured subject domains. Weakly structured subject domain features are listed in many sources, for instance, in [1]. In our current research we propose to focus on such features of weakly structured domains as lack of benchmarks, incompleteness of information on estimated objects, and impact of human factor. These features make it difficult to define the efficiency of expert estimation methods.

While measurement of objects according to quantitative parameters (such as length, weight, duration etc), actually, means comparing them to some benchmark values or units (foot, pound, second etc), people resort to expert estimation as an alternative to measurement in cases when measurement of objects is impossible. In such cases expert estimates of objects become the only source of quantitative information about these objects. Experts can provide both direct estimates in certain scales (ordinal or cardinal, numeric or verbal, agreement scale etc) and pair-wise comparisons of objects. According to many specialists, the best way to measure a set of objects according to some "intangible" criterion is to compare them with each other. This assumption resulted in emergence of many methods based on pair-wise comparisons of objects. Particularly, we should mention the Analytic Hierarchy/Network Process (AHP/ANP) [2, 3], TOPSIS [4], "triangle" and "square" [5, 6], combinatorial method [5-7], logarithmic least squares method (LLSM) [8].

Efficiency indicators for methods which operate with determined data are mostly based on different measures of deviation of real (experimental) data from benchmark values (average mean (square) deviation, mathematical expectation of error, Euclidean distance etc). When it comes to expert data-based methods, there is always a question: "what should we compare expert data (and results of their processing) with?" (as, again, there are no benchmarks). When a decision-maker (DM) organizes an examination, (s)he heuristically assumes that there is some ground truth, i.e. "exact" values of estimates of objects and ratios between them. So, the key question is: which indicators can define the degree of accuracy and credibility of expert data and methods of their aggregation? Level of DM trust towards expert recommendations and decisions, made on their basis, will depend on these indicators.

Academic publications list relatively few approaches to determination of accuracy of expert methods (if the term "accuracy" applies at all). According to the academic school of T.Saaty, the main requirement to expert data, input into pair-wise comparison matrices (PCM), is ordinal and cardinal consistency (absence of transitivity violations) [2]. In [3] the term "legitimization" is used. Expert data-based examination result is "legitimate" if it coincides with the DM's independent choice (in [3] there is a curious example of location selection for Disneyland in China). Pankratova and Nedashkovskaya [9] demonstrate that results, obtained using ANP and original hybrid method, coincide, and this, according to the authors, confirms the credibility of hybrid method. In Elliot's research [10] experts compare several estimation scales and define which of these scales allows them to express their preferences in the most adequate way. A similar approach is used in [11], where the experts choose the result of aggregation of their estimates, which reflects their understanding of the subject domain most adequately.

The common feature of the listed approaches is absence of any benchmark estimate values (in actual expert examinations there are, indeed, no benchmarks). Conceptually different approach involves testing of expert methods on the specially generated (simulated) set of "benchmark" (model) objects, for which the exact values of their estimates according to a certain criterion are known. For example, we can mention an experiment [2], where respondents are asked to estimate the ratios of several figure squares. Exact ratios are known only to the experiment organizer. These model values are compared with values, obtained based on expert estimates using group AHP.

In Ukraine a similar approach is used in [5, 6]. The authors compare around 20 expert methods according to 3 criteria: accuracy (precision), duration of estimation process, and consistency of its results. Experts estimate 7 objects (colored figures) using different methods. Number 7 is chosen due to psycho-physiological limitations of human mind [12]. Real (benchmark, model) ratios between object weights are, again, known

only to the experiment organizer. Aggregate estimation results are compared to benchmark values. Methods that produce smaller average error are considered more accurate.

Still another approach to defining accuracy of expert methods is based on simulation and does not require participation of experts at all. Both model object weights and expert estimates are simulated. Such simulation of PCM is used to define threshold values of consistency index (CI) and ratio (CR) (these values are constantly updated based on the number of modeled PCM [3]). Examples of PCM simulations can also be found in [13].

2. Combinatorial method: an overview

For the first time combinatorial method of pair-wise comparison aggregation was introduced in the early 2000-s; since then it went through many updates and improvements [7]. The problem (just like in AHP) is to find a vector of n priorities (object weights) based on PCM, provided by one or several experts. The key idea of the method is most thorough usage of expert data, provided in the form of PCM. In the general case, this information is redundant. That is why, all non-redundant informatively-meaningful basic pair-wise comparison sets, which can be formed from elements of a given PCM, are enumerated. Sometimes these basic sets are called spanning trees (the term is borrowed from graph theory). According to Cayley's theorem on trees [14], a complete PCM with dimensionality $n \times n$ allows us to form n^{n-2} of such spanning trees. Each basic set allows us to build a vector of relative object weights. After that we can find the aggregate priority vector as ordinary (1) or weighted (2) average.

Practical implication of combinatorial method is its usage as the primary expert estimate aggregation tool in the strategic planning technology for weakly-structured subject domains [11].

In 2010 the advantage of combinatorial method over other pair-wise comparison aggregation methods was demonstrated [15]. In 2012 the method was "re-invented" [16, 17]. In 2017 updated formulas for calculation of relative expert competence coefficients (based on the quality of expert information) were introduced [18]. During the last few years equivalence of combinatorial method, LLSM [8] (for complete and incomplete, additive and multiplicative PCM), and row geometric mean [19] methods was proved. However, equivalence holds only if ordinary (and not weighted) average formula is used for aggregation (1).

$$w_j^{aggregate} = \left(\prod_{q=1}^{T} w_j^q\right)^{1/T} \quad ; \quad j = 1...n, \tag{1}$$

where $T \le mn^{n-2}$ is the total number of basic pair-wise comparison sets (spanning trees); $\{w_j^q; j = 1..n; q = 1..T\}$ is the set of relative weights of n objects, calculated based on spanning tree number q; ($w_j^{aggregate}, j = 1..n$) are aggregate object weights; m is the number of experts.

Conceptual difference of the modified combinatory method is usage of ratings of basic pair-wise comparison sets. These ratings are based on *completeness*, *detail*, *consistency*, and *compatibility* of data, input by experts into individual PCM. Ratings of priority (relative alternative weight) vectors, obtained from ideally consistent PCM, reconstructed based on every single spanning tree (basic pair-wise comparison set), are taken into consideration. As a result, weight aggregation formula (1) assumes the following look (2):

$$w_j^{aggregate} = \prod_{k,l=1}^{m} \left(\prod_{q^k=1}^{T_k} \left(w_j^{(kq_kl)}\right)^{\frac{R_{kq_kl}}{\sum_{u,p,v} R_{upv}}}\right); \quad j = 1...n. \tag{2}$$

Additive look of basic pair-wise comparison set (spanning tree) rating is as follows:

$$R_{kql} = c_k c_l s^{kq} s^l \Big/ \ln\left(\sum_{u,v} \left| a_{uv}^{kq} - a_{uv}^{l} \right| + e\right). \tag{3}$$

Multiplicative look of basic pair-wise comparison set (spanning tree) rating is as follows:

$$R_{kql} = c_k c_l s^{kq} s^l \Big/ \ln\left(\prod_{u,v} \max\left(\frac{a_{uv}^{kq}}{a_{uv}^{l}}; \frac{a_{uv}^{l}}{a_{uv}^{kq}}\right) + e - 1\right). \tag{4}$$

In formulas (3) and (4) k,l are the numbers of experts ($k,l=1..m$), whose PCM are being compared with each other; c_k, c_l are a-priori values of relative expert competence; k and l can be equal or different; q is the number of ideally consistent PCM copy $q=1..mT_k$; s^{kq} is the relative average weight of scales in which basic pair-wise comparison set elements are input; it is calculated based on Hartley's formula [20];

$$s^{kq}=(\prod_{u=1}^{n-1}\log_2 N_u^{(kq)})^{\frac{1}{n-1}} \quad (5);\quad s^{k}=(\prod_{\substack{u,v=1\\ v>u}}^{n}\log_2 N_{uv}^{(k)})^{\frac{2}{n(n-1)}}, \quad (6)$$

s^l is the average weight of scales, in which elements of the respective individual PCM of expert number *l* are input.

In (5) and (6) N is the number of grades in the scale, in which the respective pair-wise comparison is provided. More detailed explanation of spanning tree ratings can be found in [18].

3. Numeric example

3 equally competent experts E_1, E_2, E_3 ($c_1=c_2=c_3=1$) compare 4 objects (A_1, A_2, A_3, A_4) in integer scales. We should calculate relative object weights (priorities) based on PCM, provided by the experts. Total numbers of grades in the scales, selected by experts, are given in Table 1. Table 2 provides particular grade numbers, selected by the experts. Table 3 provides the PCM, brought to the unified scale.

Table 1. Number of grades in the scales, selected by experts for pair-wise comparisons

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	9	8	7	1	3	4	5	1	9	9	8
A_2		1	6	5		1	6	7		1	3	9
A_3			1	4			1	8			1	7
A_4				1				1				1

Table 2. Numbers of specific grades of pair-wise comparison scales, selected by the experts

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	2	4	7	1	3	4	5	1	2	4	8
A_2		1	2	4		1	2	3		1	2	5
A_3			1	2			1	2			1	3
A_4				1				1				1

Table 3. Values of pair-wise comparisons, brought to the unified scale

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	2	4 1/3	8 5/6	1	7 1/2	8 1/6	8 1/2	1	2	4	9
A_2	1/2	1	2 2/7	6 1/2	1/7	1	2 2/7	3 1/2	1/2	1	3 1/2	5
A_3	2/9	3/7	1	2 5/6	1/8	3/7	1	2	1/4	2/7	1	3 1/2
A_4	1/9	1/6	1/3	1	1/8	2/7	1/2	1	1/9	1/5	2/7	1

48 ideally consistent PCM (ICPCM) ($mn^{n-2}=3\times 4^2=48$) are built based on initial 3 PCM. Each ICPCM is constructed from the respective basic pair-wise comparison set (spanning tree). For instance, the basic set of pair-wise comparisons of objects (A_1, A_2), (A_1, A_3), and (A_2, A_4) corresponds to the spanning tree, shown on Fig. 1 (clockwise). From the respective elements of the PCM provided by the expert E_2 we reconstruct the ICPCM, shown in Table 4 (basic pair-wise comparison values are highlighted in bold, while other elements are reconstructed based on transitivity rule).

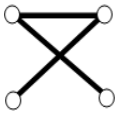


Fig 1. Spanning tree example for 4 objects

Table 4. ICPCM example

1	**7 1/2**	**8 1/6**	26 ¼
1/7	1	1	**3 ½**
1/8	1	1	3 1/5
0	2/7	1/3	1

Similarly, ICPCM are reconstructed from all basic pair-wise comparison sets, provided by each of the 3 experts. In this process, in order to verify consistency and compatibility of the initial PCM, ICPCM are compared with these initial PCM provided by the experts. For this purpose, 3 copies of each ICPCM are built. So, the total number of ICPCM to be analyzed is $T = m^2 n^{n-2} = 144$. Each ICPCM is assigned a rating, calculated according to (4). For, instance, when ICPCM, shown in table 4, is compared to the PCM of the first expert E_1, non-normalized value of the respective rating equals 1.191. When all ICPCM ratings are calculated, they are normalized by sum (see power index in (2)). From each ICPCM copy number $q = 1..144$ we reconstruct a vector of relative object weights $w_1^q, w_2^q, w_3^q, w_4^q$ (ideal consistency of the matrix allows us to use any basic set of pair-wise comparisons as priority vector; for instance, the first row). Finally, the aggregate priority (object weight) vector is calculated according to (2).

Normalized priorities w_1, w_2, w_3, w_4, calculated using the modified combinatorial method (2), based on the example data equal (0.563734299; 0.263382041; 0.120820159; 0.052063501). Values of priorities, calculated using ordinary combinatorial method (1) equal (0.590174795; 0.243658012; 0.114086692; 0.052080501).

It has been proven that ordinary combinatorial priority aggregation method (1) is equivalent to row geometric mean [19] and LLSM [8]. At the same time, as we can see from the example (and from [18]), results produced by ordinary (1) and modified method (2) are significantly different, so these two methods are not equivalent.

Both row geometric mean and LLSM have lower computational complexity than combinatorial method, and this is their advantage. The key advantage of modified combinatorial method is that it allows us to consider the quality of expert data prior to its aggregation. Consequently, results of its work more adequately reflect the level of expert competence in the issue under consideration.

The current research is an attempt to empirically confirm the advantage of the modified combinatorial method over the ordinary method (and, consequently, over row geometric mean and LLSM).

4. Available approaches to determination of efficiency of ordinary and modified combinatorial method

At the beginning of the paper we outlined several approaches used to determine the efficiency (and compare) expert methods. However, not all these approaches are applicable to our particular case.

1. Holding real expert sessions (Saaty's "legitimization" [3]) intended to empirically verify certain hypotheses is a "luxury" that an average researcher cannot afford. Finding real experts and obtaining estimates from them requires too many resources.

2. Comparing results of several methods [9] and expecting them to coincide is not our task under the circumstances. We are trying to define, which of the two methods produced *better* results.

3. Holding model examinations (for example, with students), such as ones described in [10, 11], where experts themselves would define, which results more adequately reflect their understanding of the subject domain, again would not solve the problem. We are trying to define objective characteristics of the methods that do not depend on the attitudes of the respondents, and to compare methods according to these characteristics.

4. Testing of the methods on the data of real expert estimation of specially modeled objects [5, 6] is plausible.

5. Simulation of model object weights and of expert estimates of ratios between them (priorities) [15] is plausible.

So, we propose to focus on the last two of the listed approaches: a) comparing of the two methods on real expert estimates of a set of model objects and b) simulation of both model object weights and expert estimates.

5. Comparing ordinary and modified combinatorial methods on real data of expert estimation of model objects

The model objects were figures with different numbers of colored pixels (known to expert session organizer). The number of such figures, which the experts compared according to coloring degree, amounted to 7 ($n=7$) (based on psycho-physiological constraints of human mind [12]). 18 independent pair-wise comparison sessions were conducted with real respondents (experts). Pair-wise comparisons were multiplicative ones, that is PCM transitivity (consistency) requirement looked as follows: $a_{ij}=w_i/w_j=(w_i/w_k)\times(w_k/w_j)=a_{ik}\times a_{kj}; i,j,k=1..n$, where a_{ij} is the value of pair-wise comparison of objects number i and j; w_i, w_j, w_k are the weights of objects with respective numbers. Based on expert PCM, 4 series of calculations were performed, respectively, for individual and group, ordinary (1) and modified (2) combinatorial methods. Estimates were input in a unified scale and experts were considered equally competent a priori, so the numerator in the multiplicative rating formula (4) equaled 1.

Obtained weights were compared with true values and relative estimation errors were calculated for each of the 7 objects (7):

$$\delta_k=\frac{\left|w_k-w_k^{true}\right|}{w_k^{true}};\quad k=1...n. \qquad (7)$$

Group estimation sessions were simulated as combinations of groups of 3 experts from 18 available individual estimation precedents. So, the number of such group sessions amounted to $C_{18}^3=816$. The generalized accuracy indicator was the average relative estimation error calculated across all objects:

$$\delta=\frac{1}{n}\sum_{k=1}^{n}\delta_k=\frac{1}{n}\sum_{k=1}^{n}\frac{\left|w_k-w_k^{true}\right|}{w_k^{true}}. \qquad (8)$$

In order to conduct the experiment we used the original software module (Fig. 2).

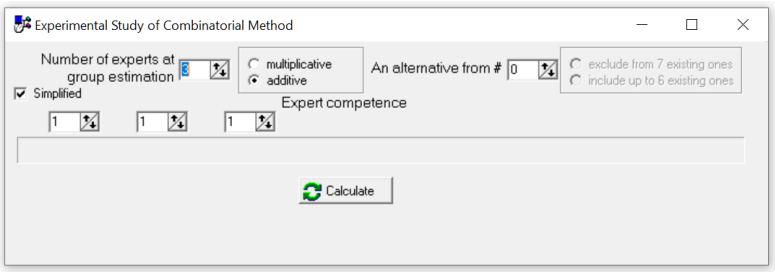


Fig. 2. Screenshot of the software module for experimental study of combinatorial method on real estimation data

Brief algorithm of the experiment (once the module is launched) is as follows.

1. Pair-wise comparison type (additive or multiplicative), as well as the number of experts in the group are selected, and their relative competence values are set. Combinatorial method modification (ordinary (1) or modified (2)) is determined. There is an opportunity to exclude some objects from the initial set of 7 model objects.

2. The module reads pair-wise comparison values from the file and performs calculations. It calculates priorities based on each estimation precedent using the selected combinatorial method modification. Calculation results (values of 7 model object weights) are written into another file in real time.

3. The module ends its work when all possible expert group variants are enumerated. If the general number of individual expert estimation precedents equals 18, and the number of experts in a group equals 3, then calculations are performed for $C_{18}^3=816$ group estimation sessions (1st session includes 1st, 2nd, and 3rd precedent; 2nd session – 1st, 2nd, and 4th precedent; … 816th session – 16th, 17th, and 18th precedent).

Calculated individual estimation error values for modified ("weighted") and ordinary (simplified) methods are shown on Fig. 3. Group estimation errors are shown on Fig. 4. On both figures X-axis denominates numbers of expert estimation sessions of 7 model objects. Y-axis denominates average mean estimation errors (8) (values range from 0 to 1). Each number of estimation session is associated with 2 average relative error values (1 for ordinary and 1 for modified method), which correspond to 2 points on coordinate plane. If for some specific estimation session, the point, corresponding to one of the two methods, lies higher (has larger Y), it means that this method produces larger error (its results are worse in comparison to true value) on the respective set of expert estimates.

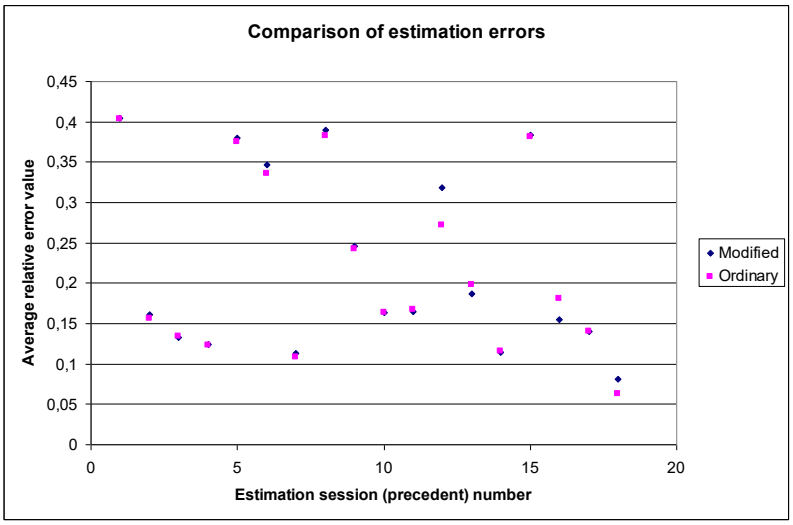


Fig. 3. Errors of estimation of 7 model objects using ordinary and modified combinatorial method (18 estimation sessions)

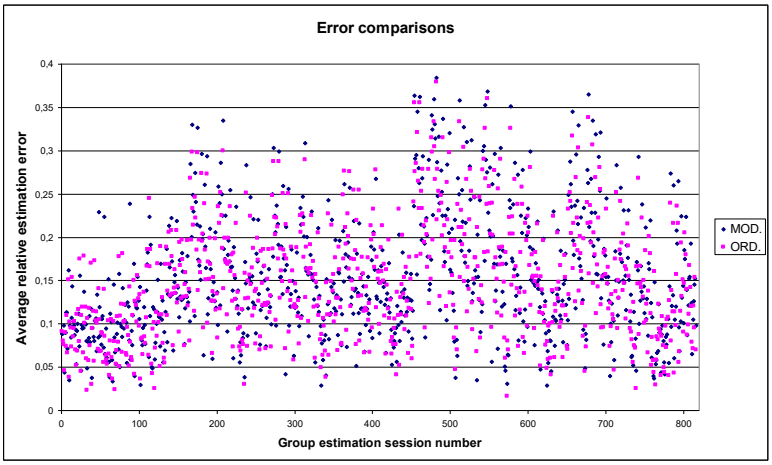


Fig. 4. Errors of estimation of 7 model objects using ordinary and modified group combinatorial method (816 estimation sessions)

Ranges and average values of errors (8) for 816 group estimation precedents are shown in table 5.

Table 5. Characteristics of modified and ordinary combinatorial methods based on real group expert estimation data

	Maximum average error	Minimum average error	Average error across all estimation precedents
Ordinary method	0.379730445	0.016795423	0.126609639
Modified method	0.384422003	0.028881538	0.130089121

Results of experimental research of ordinary and modified combinatorial method on the data of real expert estimation of model objects do not allow us to draw any definite conclusion as to advantage of one of the two methods. As we can see from figures 3 and 4, on some individual and group estimation precedents, ordinary method (that does not take the weights of basic spanning trees into account) turns out to be more accurate, while on others – the "weighted" method yields more accurate results.

The look of priority aggregation formulas (1) and (2), as well as data from 816 group estimation precedents indicate, that the modified method tends to be more efficient in the cases, when the number of accurate comparisons (closer to model true values) exceeds the number of inaccurate ones. If the majority of comparisons is inaccurate (far from true values), although consistent and compatible, they "pull" the aggregate estimate value towards themselves, as a result of weighting procedure (2). Consequently, on such precedents, the ordinary method produces better results. Moreover, both methods can produce paradoxes when averaging of inconsistent values far from true ones still produces rather accurate aggregate result.

Intermediate conclusion: it is not the chosen estimate aggregation procedure, but the estimate values themselves, that influence the results of a method. During estimation session, expert's mind "constructs" its own model priority vector. Result and relative accuracy of the method's work depend on the proximity of the expert's assumptions to the actual true vector. Hence, the difference between priorities obtained through aggregation of real expert data, and true priority values (model object weights) cannot serve as an indicator of efficiency of aggregation methods, because accurate result of a method's work confirms only the accuracy of initial expert data.

It makes sense to use relative accuracy of real expert estimates of model objects as efficiency criterion for comparing conceptually different methods (for example, methods using multiplicative and additive scales; verbal, graphic, or numeric data input; pair-wise comparison vectors, triangular, or square PCM; methods with or without feedback etc), as shown in [5, 6]. When such methods are compared, input data and aggregation procedures are substantially different. Within our current research we use one and the same expert estimate set and similar aggregation procedures. That is why usage of the approach described in [5, 6] in the context of the present research turns out to be incorrect, and produces unrepresentative results.

Human factor (subjective notion of the expert regarding ratios between objects) adds an unnecessary degree of freedom to the experiment. The only approach that would allow us to control the distance between expert estimates and true values (and, thus, neutralize the impact of human factor) is simulation of expert estimates themselves. That is why it is relevant to use the simulation-based approach [15] for comparison of methods in terms of efficiency.

6. Comparing ordinary and modified combinatorial methods through simulation of expert estimation of model objects

The key idea of data aggregation methods' efficiency evaluation is verification of their stability under fluctuations of input data (i.e. PCM). As we mentioned in the introduction, it is assumed that there is a certain true value of the object's estimate according to a given criterion (such as exact number of colored pixels in a picture). The estimate provided by an expert differs from this true value by estimation error. Let us assume that under the same expert estimation errors one aggregation method produces the result (priority vector), that is closer to the model vector (of true values) than the result produced by the other method. In this case we can state that the first method is more efficient. In order to be able to monitor expert estimation errors and deviations of priority vectors, obtained by different methods, from model values (in the context of aggregation method comparison), let us simulate expert estimation process in the following way.

1. Set true model object weights.

2. Built an ICPCM A (based on the rule $a_{ij} = w_i / w_j$ for multiplicative or $a_{ij} = w_i - w_j$ for additive comparisons, where a_{ij} is the element of A).

3. After that, add a certain "noise" to matrix A so that each element (except diagonal ones) changes as follows: $a'_{ij} = a_{ij} \pm a_{ij} \cdot \delta / 100\%$, where $\delta > 0$ is a value set in advance, which defines maximum relative deviation of pair-wise comparisons provided by the expert (i.e. elements of A) as percentage of true values. In this way expert estimation errors are simulated. In this case δ denotes potential relative error made by the expert during pair-wise comparison session.

4. "Perturbed" PCM A' is used as input data for one of aggregation methods, which produces aggregate object weights w'_i (priority vector). We propose to define the efficiency of expert data aggregation method based on maximum possible relative deviation of calculated object weight from true value of the same weight. The method producing smaller deviations should be considered more efficient.

$$\Delta = \max_i \left| \frac{w'_i - w_i}{w_i} \right| \times 100\% . \qquad (9)$$

Calculated values of Δ will depend on both δ and true relative weight values, set by experiment organizer. That is why the values of efficiency indicator for each method are presented in the form of function $\Delta(\delta)$ for each priority vector variant.

Dependence $\Delta(\delta)$ is defined for each method on an interval $\delta \in]0;100[$ (we assume that relative pair-wise comparison error made by an expert should not exceed 100%, although in the general case the function $\Delta(\delta)$ is defined in a wider range $\delta \in]0;\infty[$). We propose to find maximum possible deviation Δ for every δ using the genetic algorithm (GA) [21].

Just like in the previous section, we used an original software module to conduct the experiment. The module allows us to generate and perturb the initial PCM and launch the GA (Fig. 5).

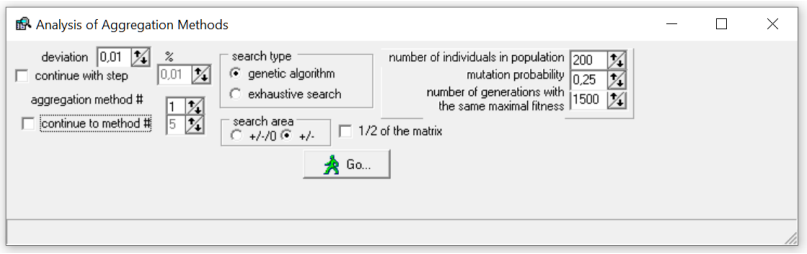


Fig. 5. Screenshot of the software module for experimental research of combinatorial method using the GA

In terms of the GA, the "individuals" are the perturbed PCM with given relative error value δ. "Fitness function" is the maximum relative deviation of resulting priorities from the true value of object weights Δ (9). The GA works as follows.

1. For given true values of object weights and δ a "population" of individuals (perturbed PCM) is generated. (Cardinality of complete enumeration of matrices of dimensionality $n \times n$, whose elements differ from true values by δ, is very large, so a population includes only a part of PCM from the complete set. However, for $n \le 5$ complete enumeration is possible).

2. Individuals with maximum fitness function value are selected from the population. That is, we are selecting PCM, which produce maximum deviation of priority vector from true value after aggregation.

3. Individuals from the selected subset are "interbred" through weighted summation (convex combination) and mutation. As a result, we get a new generation of individuals.

4. If for the new generation the value of fitness function Δ is larger than for the previous one, we should move to step 2. If during a fixed number of generations Δ does not grow, then the algorithm stops and terminates its work. As a result we get maximum Δ for a given δ.

In essence, we are looking for a maximum of a function of many variables $f(a'_{ij}); i, j = (1, n)$.Its arguments are elements of PCM A'. Values of $\Delta(\delta)$ for each aggregation method also depend on specific values of initially set true values of object weights $w_i, i = (1, n)$. Examples of functions $\Delta(\delta)$ for given true model values of object weights are shown on Fig. 6. Variants of model weight vectors are set in a way that illustrates different ratios between object weights (equal, equal in pairs, arithmetic progression, geometric progression, extreme values of scale range etc).

In [15] individual "weighted" combinatorial method was compared with several other individual pair-wise comparison aggregation methods (Fig. 6). In the context of the present paper we should note that row geometric mean (one of the methods, studied in [15]) is equivalent to ordinary combinatorial method (as proven in [19]). So, the advantage of modified (weighted) combinatorial method over row geometric mean entails its advantage over the ordinary method ([15, 18]) and LLSM (that is also equivalent to ordinary method [8]).

Consequently, we can conclude that empirical comparison data of several modifications of pair-wise comparison aggregation methods (obtained through simulation of expert estimation of model object weights) confirm the advantage of modified combinatorial method over other methods. The efficiency indicator is its stability to perturbations of the initial PCM (that is, to expert's errors).

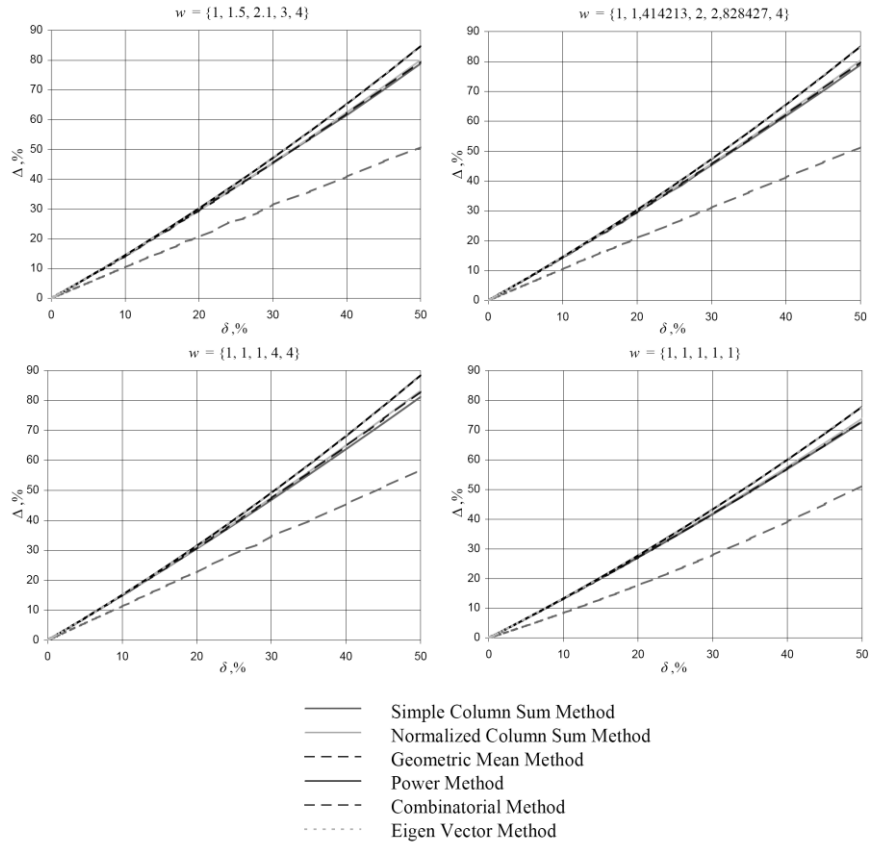


Fig. 6. Examples of dependence of Δ on δ for different model priority vectors

We should note that experimental results, shown on Fig 6, are obtained only fro individual estimation methods ($m=1$), mostly, in additive scale. That is, in the conducted experiments in formulas (1) and (2) product is replaced by summation.

$$w_j^{aggregate} = (1/T)\sum_{q=1}^{T} w_j^q,\ j = 1..n, T \in [1..n^{n-2}], \qquad (10)$$

$$w_j^{aggregate} = (1/\sum_{u,p,v} R_{upv}) \sum_{k,l=1}^{m} (\sum_{q^k=1}^{T_k} R_{kq_k l} w_j^{(kq_k l)}),\ j = 1...n. \qquad (11)$$

The next phase of the research will include simulation of group expert estimate aggregation process, where the estimates are provided in both additive and multiplicative scales. Transition from multiplicative to additive scales (particularly, during modeling) can be performed using logarithm and power operators. Both functions are monotonously increasing, so the properties of additive and multiplicative methods should be similar.

7. Conclusions

We have considered two possible ways of evaluating the efficiency of pair-wise comparison aggregation methods, namely, combinatorial methods of pair-wise comparison aggregation, where spanning tree weights are taken and not taken into account. We have shown that traditional concept of accuracy does not apply to expert data-based methods. As a result, simulation turns out to be the more correct way of evaluating the efficiency of the methods, than their testing on real expert data. We have suggested two approaches to efficiency evaluation of combinatorial method of pair-wise comparison aggregation (individual and group, additive and multiplicative methods): 1) defining relative accuracy of real expert pair-wise comparison aggregation results and 2) simulation of the whole expert examination lifecycle. We have obtained experimental results, that empirically prove the advantage of the modified combinatorial method over the ordinary method (and, consequently, over row geometric mean and LLSM).

Experimental results allow us to draw some fundamental conclusions: 1) the concepts of accuracy of expert estimates and efficiency of expert data aggregation methods should be clearly distinguished; low accuracy of some method's results is often induced by experts' errors, and not by drawbacks of the method itself; 2) it is not the accuracy, but "consistent accuracy" that counts, both within a PCM of 1 expert and in a group of experts; more consistent and compatible estimation results should be considered more credible; 3) real expert data can be used to compare conceptually different methods, using expert information of different types; if input information for several methods is the same, and only aggregation procedures are different, then in order to evaluate and compare such methods, we should simulate the whole expert session cycle, including the estimates themselves. Further research will be targeted at extended studies of modified combinatorial method through simulation of group expert estimates' aggregation for the cases when objects are estimated in both additive and multiplicative scales.

References

1. Kadenko, S.: Prospects and Potential of Expert Decision-making Support Techniques Implementation in Information Security Area. In CEUR Workshop Proceedings. Vol-1813, pp. 8-14 (2016).
2. Saaty, T.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980).
3. Saaty T.: Decision Making with Dependence and Feedback: The analytic Network Process. RWS Publicaitons, Pittsburgh (1996).
4. Hwang, C., Yoon K.: Multiple attribute decision making: methods and applications : a state-of-the-art survey. Springer-Verlag. Berlin ; New York (1981).
5. Totsenko, V., Tsyganok, V., Kachanov, P., Deev, A., Kachanova, E., Torba, L.: Experimental Research of Methods for Getting Cardinal Expert Estimates of Alternatives. Part 1. Methods without Expert Feedback. Journal of Automation and Information Sciences 35(1), 12 pages, DOI: 10.1615/JAutomatInfScien.v35.i1.40 (2003).
6. Totsenko, V., Tsyganok, V., Kachanov, P., Deev, A., Kachanova, E., Torba, L.: Experimental Research of Methods for Obtaining the Cardinal Expert Estimates of Alternatives. Part 2. Methods with Expert Feedback. Journal of Automation and Information Sciences 35(4), 11 pages, DOI: 10.1615/JAutomatInfScien.v35.i4.40 (2003).
7. Tsyganok V., Kadenko S., Andriichuk O., Roik P.: Combinatorial Method for Aggregation of Incomplete Group Judgments. In IEEE First International Conference on System Analysis & Intelligent Computing (SAIC). Kyiv 2018, pp. 25-30 (2018).
8. Bozoki, S., Tsyganok, V.: The (logarithmic) least squares optimality of the arithmetic (geometric) mean of weight vectors calculated from all spanning trees for incomplete additive (multiplicative) pairwise comparison matrices. International Journal of General Systems 48(4), 362-381 (2019).
9. Pankratova, N., Nedashkovskaya, N.: Hybrid Method of Multicriteria Evaluation of Decision Alternatives. Cybernetics and Systems Analysis 50(5), 705-711 (2014).
10. Elliott, M.: Selecting numerical scales for pair-wise comparisons. Reliability Engineering and System Safety, 95, 750–763 (2010).
11. Tsyganok, V., Kadenko, S., Andiichuk, O.: Using different pair-wise comparison scales for developing industrial strategies. International Journal of Management and Decision Making, 14(3), 224-250 (2015).
12. Miller, G.: The Magical Number Seven, Plus or Minus Two. The Psychological Review, 63, 81-97 (1956).
13. Tsyganok, V., Kadenko, S., Andriichuk, O.: Simulation of Expert Judgements for Testing the Methods of Information Processing in Decision-Making Support Systems. Journal of Automation and Information Sciences, 43(12), 21-32 (2011).
14. Cayley, A.: A Theorem on Trees. Quarterly Journal of Mathematics, 23, 376-378 (1889).
15. Tsyganok, V.: Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method. Mathematical and Computer Modeling, 52(3-4), 538–544 (2010).
16. Siraj, S., Mikhailov, L., Keane, J.: Enumerating all spanning trees for pairwise comparisons. Computers & Operations Research, 39(2), 191-199 (2012).
17. Siraj, S., Mikhailov, L., Keane, J.: Corrigendum to "Enumerating all spanning trees for pairwise comparisons [Comput. Oper. Res. 39 (2012) 191-199]". Computers & Operations Research, 39(9), 2265 (2012).
18. Kadenko, S.: Defining Relative Weights of Data Sources during Aggregation of Pair-wise Comparisons. In CEUR Workshop Proceedings. Vol-2067, pp. 47-55 (2017).
19. Lundy, M., Siraj, S., Greco, S.: The Mathematical Equivalence of the "Spanning Tree" and Row Geometric Mean Preference Vectors and its Implications for Preference Analysis. European Journal of Operational Research 257(1), 197-208 (2017).
20. Hartley, R.: Transmission of information. Bell System Technical Journal, 7, 535-563 (1928).
21. Holland, J. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor (1975).

MATHEMATICAL MODEL OF THREATS RESISTANCE IN THE CRITICAL INFORMATION RESOURCES PROTECTION SYSTEM

Korniyenko Bogdan Yaroslavovuch [1], Galata Liliya Pavlivna[2],
Ladieva Lesya Rostyslavivna[3]

[1] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, bogdanko@gmx.net*
[2] *National Aviation University, Kyiv, Ukraine, galataliliya@gmail.com*
[3] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, lrynus@yahoo.com*

Abstract. The main problems of information security of critical information resources and causes of their occurrence are considered. The main threats to security systems of critical information resources are analyzed. Mathematical model of counteraction of internal and external threats to the system of protection of critical information resources of mineral fertilizers production is developed. The process of building a mathematical model of countering the threats in the system of protection of critical information resources with the help of Markov chain. The methodology of finding current threats to data security during their processing is offered. The software module is developed in the Piton programming language. These examples calculate the probability of mathematical model of information system in one of four states (the threat did not come; the threat came but was not implemented; the threat has been implemented; the threat came, but it was reflected protection system). Examples of numerical results analysis using proposed techniques clearly show that their use helps to define threats that are relevant to the system being investigated and can be used in practice. The disadvantage of proposed methodology is the need to consider the system behaviour when it comes to each type of threat, individually and inability to define behaviour on the simultaneous action of several threats. Studying the influence of each threat separately allows to study each of its types in more detail and to determine the probability of which is the greatest occurrence. As a result of exploitation of critical information resources protection systems and substantial changes in composition and quality of modern threats, it is necessary to project and implement information security of the systems taking into account tendencies of cyber threats development.

Keywords: Mathematical Model; Threat; System of Protection; Critical Information Resources.

1. Introduction

In today's world, automated control systems are used more and more often. The need for information security in automated control systems is increasing every year by increasing the number of attacks by intruders and leakage of confidential information. But the most susceptible to security is the most automated control systems of technological processes, as when attackers attack, it can happen not only leakage of secret information, but also interference into the technological process itself, which, when malfunctions, can lead to environmental catastrophe.

Control of technological processes in enterprises of small, medium and large size is impossible without computer equipment and modern means of automation, without highly effective of automated control systems of technological processes (ACS TP).

ACS is a comprehensive system of hardware and software, which is designed for remote centralized monitoring of processes and system as a whole, as well as for automated control of engineering and technical subsystems.

Information Security (IS) is becoming more and more topical issue with the development of computing equipment and the increasing penetration of such systems into the daily lives of people. This is discussed in almost every event on the security of the IS, and the analysis of reliability and safety of the systems already commissioned is one of the mandatory conditions of State and international certification. One of the biggest problems of the security is that the majority of ACS of small and medium complexity are projected by small organizations under conditions of strict financial limitation, which eliminates or complicates the security issue.

Experts believe that the provision of IS is different from the provision of corporate information systems. Even the term "information security" is very rarely used in relation to ACS TP. The reason for this is that it is necessary to pay attention not so much the confidentiality of information, but ensuring the continuity and integrity of the technological process. Although at the same time a significant amount of attention is given to the problems of ensuring confidentiality of information, because, according to many experts, the solution of this problem leads to the automatic solution of the problem of integrity and availability of information.

Modern ACS in many cases manage difficult and dangerous technological processes, failure of which could potentially lead to accidents in production or, in the worst case, to technogenic disasters. This significantly increases the price of risk due to violations of the IS, because threats can theoretically cause damage to people, the environment, and also leads to the financial and reputational losses.

In modern times, the priority in providing the IS ACS TP is to ensure the availability and integrity of the management and configuration information on the parameters of the technological process. Particular attention should be paid to preventing unauthorized access to the system to preserve the stable functioning of ACS TP [1-5].

Despite the large number of accidents with catastrophic consequences, the problem of critical information resources protection at industrial enterprises in the chemical and energy industries has never been so acute as in recent years.

The general interest to the security of industrial systems occurred only after the incidents with the computer viruses Stuxnet, Duqu, Flame, which attacked Iranian nuclear facilities, government agencies and industrial facilities in India, China and other countries. Separately, attacks to enterprises in the Ukraine's energy sector and the banking sector should be highlighted. Before these incidents advent, it was considered that the protection system of critical information resources was been very difficult to compromise. Such representations were based on the following postulates: the software of each protection system for critical information resources is unique and closed; penetration into the system is associated with high costs of intellectual resources, and the monetary reward for the attacker is not obvious; the local network of the critical information resources protection system solves access restriction problems.

A study of software and hardware structure, that used in the protection systems of critical information resources, has shown that there have been big changes in recent years. Almost everywhere, widely known software is used, such as Windows OS, TCP/IP protocols, etc., which, together with their advantages in standardization, simplicity, and quality of use, have also brought disadvantages - vulnerabilities. Computers connected to the Internet appear in the local network, and they also present a large number of potential threats to the system [6-10].

2. Formulation of the problem

In industrial systems of critical infrastructure there are the same vulnerabilities as in most conventional IT systems. In addition, the peculiarity of industrial systems is the existence of unique vulnerabilities, which include:

Human factor. Operation Industrial and corporate systems are usually involved in various divisions and specialists. In turn, the personnel of industrial systems, as a rule, are not far from the issues of providing information security, in its composition there are no security specialists, and the recommendations of the IT staff are not distributed. The solution of technological problems arising during the operation of the system, ensuring its reliability and availability, efficiency and minimization of overhead should be one of the main tasks of specialists.

OS vulnerabilities. The vulnerabilities of OS (operating systems) are inherent for both industrial and enterprise systems, but software installations are not normally implemented in industrial systems. Uninterrupted operation of such system is the responsibility of the administrator. The establishment of pre-approved software corrections can contribute to serious problems, and there is no time or money on full testing.

Authentication. Common passwords are usually used for industrial systems. The two-factor authentication system is rarely used, and sensitive information is often transmitted in an open form.

Remote access. To control the industrial systems often used remote access by dial-up channels or by VPN channels via the Internet. If not controlled use, this may cause serious security problems.

External network connections. Lack of appropriate regulatory framework and usage is aimed more at convenience rather than security, sometimes lead to the fact that network connections are created between industrial and corporate networks. There are even recommendations for using "composite" networks that allow you to simplify administration. This can adversely affect the security of both industrial and corporate systems.

Means of protection and monitoring. Unlike enterprise systems, use IDS. Systems, etc. In the industrial system is not a common practice, the analysis of security audit logs is also not rarely bypassed.

Wireless networks. In industrial systems, various types of wireless communication are often used, including 802.11 protocols, which are known to not provide sufficient capabilities to protect the information.

Remote processors. Some remote processor classes have known vulnerabilities. Performance of these processors does not always allow to implement security functions. In addition, after installation, they try not to touch for years, during which they remain vulnerable.

Software. The software of the industrial systems usually does not have enough security functions. In addition, in most cases, it is not devoid of architectural weaknesses.

Disclosures. The owners of industrial systems are often deliberately publish information about their

architecture. Consultants and developers often share experiences and reveal useful information about employees.

Physical security. Remote processors and industrial systems equipment can be outside the controlled area. In such conditions they physically cannot be controlled by the personnel, and the only mechanism of their physical protection is the use of metal doors and locks. Such measures are not a serious obstacle to intruders.

In this way, we can conclude that there is a significant number of vulnerabilities that are both common to any information systems and specific to industrial systems. These vulnerabilities cause specific security requirements and special operating regimes for such systems [11-15].

Also, the new term “cyberwarfare” has recently appeared, which is often mentioned in the media, because there is a problem in critical information resource system protection at infrastructure facilities and hazardous industries. Thus, the widespread use of computer equipment in the management of industrial enterprises creates the need to pay more and more attention to the problems of information security of such systems.

The main problems of information security of critical information resources, according to experts, appear because there are:

- low protection against unauthorized access (passwords);
- undeclared SCADA capabilities;
- lack of control in management actions;
- using of wireless communications (crypto persistent Wi-Fi encryption);
- lack of clear boundaries between different network segments;
- untimely or incorrect software updates;
- rejection from even minimal security tools (often for convenience or performance, companies refuse to install not only, for example, anti-virus protection, but also password protection for critical assets);
- the distribution of Windows as an operating system for workstations and even servers;
- development for trusted environment of closed industrial networks;
- creating systems without taking into account the best practices in safe code development;
- human factor, poor staff discipline.

Here is an example of the main threats to critical information resource systems found after analyzing these incidents:

- attacks to SCADA;
- attacks to PLC, PLC vulnerabilities (standard password, unauthorized access to original software);
- attacks to the infrastructure and the operating system (viruses, trojans, worms, DoS and DDoS attacks, ARP spoofing);
- attacks to protocols, protocol vulnerabilities (unauthorized access, SQL injection);
- attacks such as Buffer Overflow, Information Disclose, Denial of Access, Manipulation of View.

Among all types of vulnerable components of critical information resources protection systems, the next one prevails: SCADA - 87%, systems providing human-machine interfaces - 49%, programmable controllers - 20%, protocols - 1%.

The vulnerabilities by type was divided as follows: buffer overflow - 36%, authentication / key management - 22.86%, Web application vulnerabilities - server - 10.86%, client - 9.14%, remote code execution - 13.14% [16-18].

As a result of the operation of protection systems for critical information resources and a significant change in the composition and quality of modern threats, it is necessary to design and implement information security systems taking into account trends in cyber threats development. On the other hand, it is necessary to carry out regular work to neutralize emerging or potential threats on working systems. At this level, the following information security services are implemented: access control, integrity, secure network connections, antivirus protection, security analysis, intrusion detection, information security system management (continuous state monitoring, incident detection, reaction).

3. Analysis of recent research and publications

The protection system of critical information resources of mineral fertilizers production was chosen as the object of research [19-21]. The basis of information security models are the following theories: formal-heuristic approach; probability theory and random processes; evolutionary modeling; theory of graphs, automata, and Petri nets; game theory and conflict; catastrophe theory; theory of fuzzy sets; entropy approach. The differences between the majority of models are which parameters they use as input and which parameters are given as output after calculations. In addition, recently, modeling methods based on the informal theory of systems are widely used: structuring methods, estimation methods, and methods for finding optimal solutions [22-24].

Structuring methods is the development of a formal description, applies to organizational and technical systems. Using these methods allows us to present the architecture and process of complex system

functioning in the form that satisfies the following conditions: completeness of main parts display and their relationships; simplicity of elements organization and their relationships; flexibility - simplicity of making changes to the structure, etc. Evaluation methods allow you to determine the values of system characteristics, that cannot be measured or obtained using analytical expressions, or in the statistical analysis process, such as the probability of threats, the effectiveness of the protection system element, etc. The basis of such methods is expert assessment - an approach of attracting specialists in relevant fields to obtain some characteristics values.

Methods for finding optimal solutions is a generalization of a large number of independent, mostly mathematical, theories, in order to solve optimization problems. In the general case, this group includes methods for informally reducing a complex problem to a formal description, followed by formal approaches. Combining of methods for these three groups allows you to expand the possibilities of applying formal theories to do a complete simulation of protection systems.

The purpose of the article is the development and research of a mathematical model of threats resistance in the critical information resources protection system, obtaining transitional characteristics for system states.

4. Main research material

The mathematical model of influence resistance of internal and external threats on the protection system of critical information resources of the mineral fertilizers production is proposed. The process of constructing a mathematical model for threats resistance in the critical information resources protection system by using the Markov chain [25-27] is described in stages. The methodology for finding relevant threats to data security during their processing is proposed. The examples of calculating the probabilities of finding a mathematical model of an information system in one of four states is shown (the threat has not occurred; the threat occurred, but has not been implemented; the threat occurred, and has been implemented; the threat occurred, but has been reflected by the security system).

The system can be interpreted as a queuing system, which receives threats. First, consider the situation when threats of the same type come to the system input, assuming that the threat cannot be realized and come several times in the same time period. If these conditions are met, then the system can be in one of four states (see Fig. 1):

1. the threat has not been occurred and, accordingly, was not implemented;
2. the threat occurred, but has not been implemented;
3. the threat occurred, but has been implemented;
4. the threat occurred, but has been reflected by the security system.

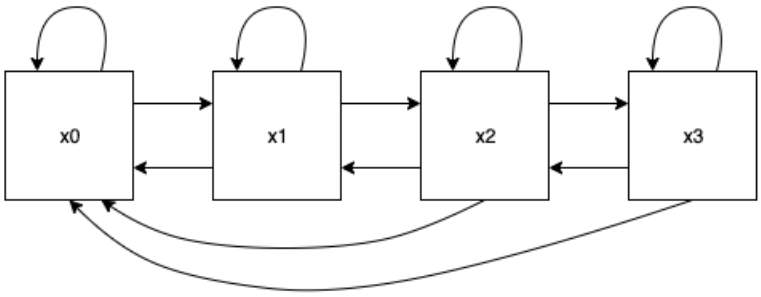


Fig.1. The graph of system states.

The system under consideration is a system with recovery, this system can go from any state to the initial one. We will consider a system with continuous time. The transition from state to state occurs according to a directed graph (see Fig. 1). To describe the transition from state to state, we will construct a transition intensity matrix:

$$p_{ij} = \begin{vmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} & 0 \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & \lambda_{24} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} & \lambda_{34} \\ \lambda_{41} & \lambda_{42} & \lambda_{43} & \lambda_{44} \end{vmatrix} \quad (1)$$

From prior approvals, as a result, the elements of this matrix have the following properties:

$$\lambda_{11} = -\lambda_{12} - \lambda_{13} \quad (2)$$

$$\lambda_{22} = -\lambda_{21} - \lambda_{23} - \lambda_{24}$$
$$\lambda_{33} = -\lambda_{31} - \lambda_{32} - \lambda_{34}$$
$$\lambda_{44} = -\lambda_{41} - \lambda_{42} - \lambda_{43}$$

To determine the probabilities of being the system in the states x_0, x_1, x_2, x_3, we construct a system of differential equations:

$$\begin{cases} \frac{dp_0(t)}{dt} = p_0(t)\lambda_{11} + p_1(t)\lambda_{21} + p_2(t)\lambda_{31} + p_3(t)\lambda_{41} \\ \frac{dp_1(t)}{dt} = p_0(t)\lambda_{12} + p_1(t)\lambda_{22} + p_2(t)\lambda_{32} + p_3(t)\lambda_{42} \\ \frac{dp_2(t)}{dt} = p_0(t)\lambda_{13} + p_1(t)\lambda_{23} + p_2(t)\lambda_{33} + p_3(t)\lambda_{43} \\ \frac{dp_3(t)}{dt} = p_1(t)\lambda_{24} + p_2(t)\lambda_{34} + p_3(t)\lambda_{44} \end{cases} \quad (3)$$

As

$$p(0) = (1,0,0,0) \quad (4)$$

is set, so vector of absolute probabilities

$$p(n) = (p_0(n), p_1(n), p_2(n), p_3(n)) \quad (5)$$

defines by the ratio:

$$p(n) = p(0)\|p_{ij}(n)\| \quad (6)$$

After research on the simulation model, two results of finding the coefficients λ_{ij} were determined, which will be given below [28-30].

Option 1. Let the transition intensity matrix λ_{ij} be as follows:

$$\begin{vmatrix} -0.040 & 0.015 & 0.010 & 0.015 \\ 0.225 & -0.250 & -0.025 & 0.050 \\ 0.625 & -0.160 & -0.855 & 0.390 \\ -0.000 & 0.075 & 0.200 & -0.275 \end{vmatrix} \quad (7)$$

We get the solution of the equations system by the fourth-order Runge-Kutta method for the time t = 50 s. To implement the solution, a software module has been developed by Python. As a result of the calculations, the probability values of being the system in each of the states were found (see Fig. 2).

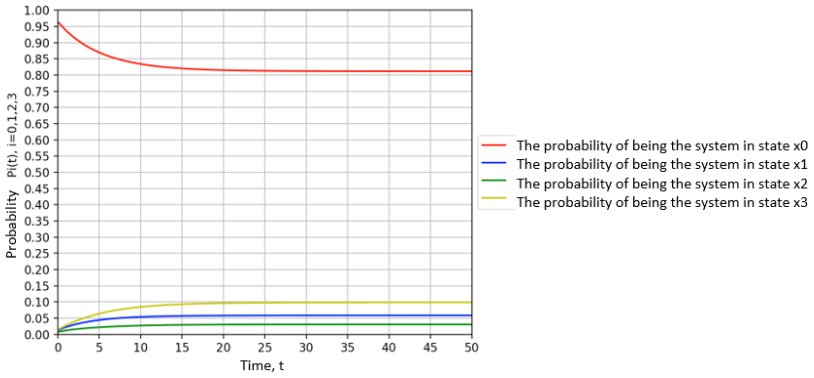


Fig.2. The probability of being the system in each state.

Option 2. We will make the calculation similar to Option 1 with the values of the transition intensity matrix λ_{ij}:

$$\begin{vmatrix} -0.040 & 0.015 & 0.010 & 0.015 \\ 0.225 & -0.250 & -0.025 & 0.050 \\ 0.575 & -0.200 & -0.875 & 0.500 \\ -0.125 & 0.145 & 0.180 & -0.200 \end{vmatrix} \quad (8)$$

As a result, we will get the following graph (see Fig. 3):

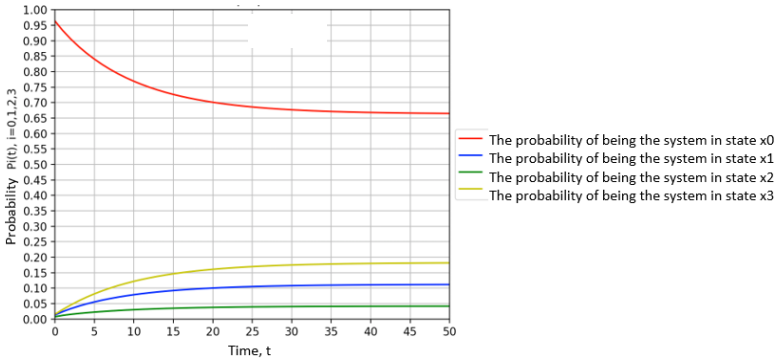


Fig.3. The probability of being the system in each state.

5. Conclusions

Based on the obtained results, it can be said that in Option 2, the system is more probable to be in a threatened state, although there is a high probability that the security system will successfully reflect the threat.

The mathematical model of threats resistance in the protection system of critical information resources of the mineral fertilizers production is proposed. A methodology for identifying actual security threats on the basis of this model has also been developed and displayed.

Examples of the numerical results analysis using the proposed methodology clearly demonstrate that their use helps to identify threats that are relevant to the research system and can be used in practice. The disadvantage of the proposed methodology is the need to consider the behavior of the system when each type of threat is exposed to it separately and the inability to determine the behavior of the system when simultaneous action of several threats is exposed to it. But on the other hand, research of the influence of each threat separately allows you more detailed study each of its types and identify those with the highest probability of occurrence.

References

[1] Kurilov F. M. Simulation of information protection systems. Graph Theory App/ Technical Sciences: Theory and Practice: Materials III Internar. Scientific. Conf. - Reading: Young Scientist Publishing House, pp. 6-9. (2016).

[2] Rosenko AP. Theoretical basis for analyzing and assessing the impact of internal threats on the security of confidential information: monograph. M.: Helios ARV, 154 p. (2008).

[3] Raphael Hertzog, Jim O'Gorman, and Mati Aharoni «Kali Linux Revealed» , Offsec Press, 347 p. (2017).

[4] Lei Chen, Hassan Takabi, Nhien-An Le-KhacJohn Wiley & Sons. Security, Privacy, and Digital Forensics in the Cloud, 360 p. (2019).

[5] Glen D. Singh, Rishi Latchmepersad. CompTIA Network+ Certification Guide, 422 p. (2018).

[6] Dijiang Huang, Ankur Chowdhary, Sandeep Pisharody. Software-Defined Networking and Security: From Theory to Practice, 328 p. (2018).

[7] Robert M. Lee. Active Cyber Defense Cycle, 651 p. (2016).

[8] Jason C. Neumann, The Book of GNS3: Build Virtual Network Labs Using Cisco, Juniper, and More 1st Edition, 274 p. (2015).

[9] Kwangjo Kim, Muhamad Erza Aminanto, Harry Chandra Tanuwidjaja. Network Intrusion Detection Using Deep Learning: A Feature Learning Approach, 79 p. (2018).

[10] Korniyenko B., Galata L., Ladieva L. Security Estimation of the Simulation Polygon for the Protection of Critical Information Resources. CEUR Workshop Proceedings, Selected Papers of the XVIII International Scientific and Practical Conference "Information Technologies and Security" (ITS 2018) Kyiv, Ukraine, November 27, 2018, Vol-2318, - pp. 176-187, urn:nbn:de:0074-2318-4 (2018).

[11] Korniyenko B., Yudin O., Galata L. Research of the Simulation Polygon for the Protection of Critical Information Resources. CEUR Workshop Proceedings, Information Technologies and Security, Selected Papers of the XVII International Scientific and Practical Conference on Information Technologies and Security (ITS 2017), Kyiv, Ukraine, November 30, 2017, Vol-2067, - pp. 23-31, urn:nbn:de:0074-2067-8 (2017).

[12] Zhulynskyi, A. A., Ladieva, L. R., Korniyenko, B. Y. Parametric identification of the process of contact membrane distillation. ARPN Journal of Engineering and Applied Sciences, Volume 14, Issue 17, pp. 3108-3112 (2019).

[13] Galata, L., Korniyenko, B. Research of the training ground for the protection of critical information resources by iRisk method. Mechanisms and Machine Science, Volume 70, pp. 227-237, (2020). doi:10.1007/978-3-030-13321-4_21
[14] Korniyenko, B. Y., Borzenkova, S. V., Ladieva, L. R. Research of three-phase mathematical model of dehydration and granulation process in the fluidized bed. ARPN Journal of Engineering and Applied Sciences, Volume 14, Issue 12, pp. 2329-2332, (2019).
[15] Kornienko, Y. M., Liubeka, A. M., Sachok, R. V., Korniyenko, B. Y. Modeling of heat exchangement in fluidized bed with mechanical liquid distribution. ARPN Journal of Engineering and Applied Sciences, Volume 14, Issue 12, pp. 2203-2210, (2019).
[16] Bieliatynskyi, A., Osipa, L., Kornienko, B. Water-saving processes control of an airport. Paper presented at the MATEC Web of Conferences, 239, (2018). doi:10.1051/matecconf/201823905003
[17] Korniyenko, B.: Informacijni tehnologii' optymal'nogo upravlinnja vyrobnyctvom mineral'nyh dobryv. K.: Vyd-vo Agrar Media Grup (2014).
[18] Korniyenko, B., Osipa, L.: Identification of the granulation process in the fluidized bed. ARPN Journal of Engineering and Applied Sciences Volume 13, Issue 14, pp. 4365-4370 (2018).
[19] Babak, V., Shchepetov, V., Nedaiborshch, S.: Wear resistance of nanocomposite coatings with dry lubricant under vacuum. Scientific Bulletin of National Mining University Issue 1, pp. 47-52 (2016).
[20] Kravets, P., Shymkovych, V. ; Hardware Implementation Neural Network Controller on FPGA for Stability Ball on the Platform 2nd International Conference on Computer Science, Engineering and Education Applications, ICCSEEA 2019; Kiev; Ukraine; 26 January 2019 - 27 January 2019 (Conference Paper), Volume 938,pp. 247-256 (2019).
[21] Arber, B. and Davey, J. The use of the CCTA risk analysis and management methodology CRAMM. Proc. MEDINFO92, North Holland, pp. 1589 –1593. (1992).
[22] Ryabko B.Y., Monarev V.A. Using information theory approach to randomness testing. Journal of Statistical Planning and Inference, Vol. 133, № 1, pp. 95-110, (2005).
[23] Chris Clymer, Ken Stasiak, Matt Neely, Stephen Marchewitz. IRisk Equatuion Available via https://securestate.en/iRisk-Equation-Whitepaper.pdf
[24] Common Vulnerability Scoring System v3.0: User Guide. Available via https://www.first.org/cvss/user-guide
[25] Loch K, Carr Houston, Warkentin M. Threats to Information Systems: Today's Reality, Yesterday's Understanding, Management Information Systems Quarterly 16.2; (1992).
[26] McCue A. Beware the insider security threat, CIO Jury; http://www.silicon.com/management/cio-insights/2008/04/17/ bewaretheinsider-security-threat-39188671/ (2008).
[27] Howard MD. LeBlanc, Writing Secure Code 2nd ed., Redmond, Washington: Microsoft Press; (2003).
[28] Rasmi M, Jantan A. Attack Intention Analysis Model for Network Forensics. Software Engineering and Computer Systems; 403-411. (2011).
[29] Ben Arfa Rabai L, Jouini M, Ben Aissa A, Mili A. A cybersecurity model in cloud computing environments. Journal of King Saud University – Computer and Information Sciences; 1: 63-75. (2012).
[30] Ben Arfa Rabai L, Jouini M, Ben Aissa A, Mili A.. An economic model of security threats for cloud computing systems. International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec); 100-105. (2012).

ЗАХИСТ ТА ПІДВИЩЕННЯ ЖИВУЧОСТІ КРИТИЧНИХ СТРУКТУР: ЗАРУБІЖНИЙ ДОСВІД ТА МОЖЛИВОСТІДЛЯ УКРАЇНИ

Костенко Н.Г[1], Броховецький І.В[1], Баришполь Д.В[1]

[1]*Харківський національний університет радіоелектроніки, м.Харків, 61000*
nikitakostenko1996@gmail.com

Анотація. Проведено дослідження щодо визначення поняття «критичні інфраструктури» та порівняльний аналіз тлумачень поняття у зарубіжних країнах. Проаналізовано досвід зарубіжних країн у сфері захисту національних критичних інфраструктур з урахуванням різних аспектів та особливостей нормативних баз та національних організаційних систем. Визначено основні відмінності у тлумаченні понять сфери захисту критичних інфраструктур та досліджено взаємозв'язок визначень з організаційними системами державного регулювання процесів захисту КІ. Також проаналізовано відмінності у визначенні офіційних списків секторів, до яких відносяться групи державних критичних інфраструктур, відмінності у системах фінансування заходів охорони КІ та систем державних органів, що відповідальні за захист КІ у різних країнах. Виконано аналіз національної системи заходів щодо охорони КІ в Україні та визначено існуючі проблеми та недоліки даної системи. Описано актуальні питання щодо вдосконалення функції охорони найбільш пріоритетних об'єктів КІ в різних сферах України. Показано, що здатність до живучості та рівень захисту КІ залежить безпосередньо від правильної організації системи заходів з

охорони КІ на державному рівні. На основі проведених аналізів створено систему рекомендацій щодо покращення живучості критичних інфраструктур в Україні з урахуванням зарубіжного досвіду та існуючої ситуації в системі охорони КІ.

Ключові слова:критична інфраструктура, захист критичної інфраструктури, аналіз критичної інфраструктури, заходи з безпеки КІ в Україні.

Вступ

Термін «критична інфраструктура» використовується у нормативних актах багатьох країн, та попри це не має установленого визначення. Узагальнюючи різноманітні підходи до тлумачення даного терміну в різних джерелах, можна визначити, що критичними інфраструктурами (КІ) називають такі об'єкти, від функціонування яких залежить порядок у суспільній сфері, стабільність економіки, а також національна безпека країни. КІ являються об'єктами, які безпосередньо задіяні у процесі забезпечення життєво важливих потреб суспільства та мають суттєвий вплив на рівень розвитку та здатність до оборони від зовнішніх ворожих дій.

Будь-яка критична інфраструктура — це багаторівнева та складна система стратегічного масштабу, що складається з сукупності багатьох елементів. Важливість та актуальність захисту та підвищення живучості такої системи визначається необхідністю застосування системного підходу, що включає нормативно-правові, регулюючі, контролюючі, організаційні та технічні засоби.

На сьогодні питання захисту та підвищення живучості критичних інфраструктур для України являється гостро актуальним. Починаючи з 2014 року на фоні загострення політико-економічної ситуації в країні, а також виникнення зовнішньої загрози для багатьох КІ у східних регіонах, постає необхідність впровадження нових методів та підходів до підвищення живучості та захисту КІ України.

Автори пропонують розглянути та проаналізувати світовий досвід у сфері захисту КІ та визначити напрямки вдосконалення українського сектору безпеки та охорони щодо покращення живучості критичних інфраструктур в Україні.

1. Аналіз систем національного захисту критичних структур у країнах світу

Визначення терміну «критична інфраструктура», як було зазначено раніше, не має єдиного затвердженого у всіх країнах визначення [1]. Тому важливо визначити відмінності у тлумаченні терміну для проведення більш ефективного аналізу зарубіжної практики у сфері захисту КІ.

Сполучені Штати Америки є світовим лідером з темпів розвитку сфери захисту КІ [2]. У законодавчих актах США критичні інфраструктури визначаються як «комплекс фізичних та віртуальних активів, мереж і систем, що відіграють життєво важливу роль для країни. Руйнування або припинення функціонування КІ та їх елементів матиме значні негативні наслідки для таких сфер, як національна безпека, економіка, безпека та здоров'я населення, або матиме будь-яку комбінацію із зазначених негативних наслідків» [2].

Ще одною країною з розвинутою системою захисту та підвищення живучості КІ є Велика Британія. Законодавчі акти даної держави містять наступне визначення КІ: « Критичними інфраструктурами являються найбільш важливі елементи державної інфраструктури, до яких відносяться об'єкти, системи, державні активи, мережі, процеси, а також ключові посадові особи, втрата або припинення функціонування яких призводить до згубного впливу на: доступність та повноту надання важливих для життя населення послуг, а також послуг, що забезпечують захист населення від виникнення нещасних випадків, негативних наслідків у економічній та соціальній сфері; функціонування національно системи безпеки та оборони країни»[3].

Нормативно-правова база Європейського союзу визначає критичну інфраструктуру як «активи, системні мережі або їх структурні частини, що розташовані в державах-членах Європейського союзу та відіграють важливу роль в функціонуванні основних, життєво важливих для населення сфер, таких як система охорони здоров'я, безпеки, економічна та політична система. КІ визначаються також об'єкти, руйнування або припинення діяльності яких здійснить значний вплив на спроможність країни виконувати свої функції» [4].

З наведених визначень видно, що відмінності у тлумаченні терміну «критична інфраструктура» в різних країнах існують, та в основному стосуються національної або організаційної специфікивикористання терміну та особливостями нормативно-правової системи в кожній країні.

Найбільш важливою, на думку авторів, є відмінність визначення терміну у країнах Європейського союзу, в яких сфера захисту критичних інфраструктур не обмежується національними об'єктами КІ, але стосується і інтересів найближчих країн-сусідів, забезпечуючи тим самим колективний, а значить і більш надійний захист об'єктів КІ.

У випадку визначення КІ нормативними актами Європейського союзу, враховуються і зарубіжні об'єкти, які безпосередньо впливають на функціонування життєво важливих державних систем.

Такими об'єктами, наприклад, можуть виступати спільні енергетичні комплекси, системи зв'язку, транспорту чи інші елементи КІ[5].

Законодавство ЕС визначає критичними інфраструктурами і ті об'єкти, «що розташовані в державах-членах Європейського союзу, якщо припинення функціонування або руйнування таких об'єктів матиме значний негативний вплив на життєво важливі сфери принаймні двох держав Європейського союзу» [4].

В більшості розвинених зарубіжних країн (США, країнах Європейського союзу), на відміну від України, поняття «захист критичної інфраструктури» є офіційно визначеним терміном.Наприклад, нормативно-правова база Республіки Польща визначає термін «захист критичної інфраструктури» як «систему заходів, що забезпечують безперебійну функціональну діяльність та цілісність структури КІ та проводяться з метою запобігання можливості виникнення зовнішніх та внутрішніх загроз, ризиків й зниження рівня вразливості КІ від факторів загрози, а також для обмеження й нейтралізації негативних наслідків їх виникнення та швидкого відновлення роботи КІ у випадку руйнування, припинення функціонування, різного виду атак чи інших дій, що порушуютьнормальне функціонування» [5].

Оскільки об'єкти критичної інфраструктури можуть значно відрізнятися за призначенням та сферою застосування, важливою умовою для забезпечення захисту КІ є групування об'єктів за секторами діяльності та визначення переліку об'єктів у кожному секторі. В зарубіжних країнах такі переліки складаються та затверджуються в офіційних нормативно-правових актах. Порівняння таких офіційних списків секторів, об'єкти яких визначаються як критичні інфраструктури у різних країнах, наведені в Таблиці 1 [2, 3].

Таблиця 1. Офіційні списки секторів КІ в зарубіжних країнах

№ з/п	Сектор	Країна				
		США	Швейцарія	Велика Британія	Японія	Канада
1.	Інформаційно-комунікаційний	+	+	+	+	+
2.	Аварійно-рятувальні служби	+	-	+	-	-
3.	Енергетика	+	+	+	-	+
4.	Фінансово-банківська сфера	+	+	+	+	+
5.	Продовольство (та сільське господарство)	+ (+)	+	+	-	+
6.	Державне управління	+	+	+	+	+
7.	Охорона здоров'я	+	+	+	+	+
8.	Транспорт	+	+	+	-	+
9.	Водопостачання (та очищення води)	+ (+)	+	+	+	+
10.	Оборонний сектор	+	-	+	-	-
11.	Ядерний сектор	+	-	+	-	-
12.	Космічна промисловість	-	-	+	-	-
13.	Хімічна промисловість	+	-	+	-	-
14.	Безпека	-	+	-	-	+
15.	Промисловість	+	-	-	-	+
16.	Повітропровідна інфраструктура	-	-	-	+	-
17.	Залізнична інфраструктура	-	-	-	+	-
18.	Електроенергетика	-	-	-	+	-
19.	Паливний сектор	-	-	-	+	-
20.	Логістика	-	-	-	+	-
21.	Утилізація відходів та очищення води	-	+	-	-	-
22.	Комерція	+	-	-	-	-
23.	Система зв'язку	+	-	-	-	-
24.	Гідротехнічна сфера	+	-	-	-	-

Як видно з Таблиці 1, офіційні списки секторів критичних інфраструктур в різних зарубіжних країнах мають суттєві відмінності. Така різниця пояснюється, насамперед, специфікою та різним рівнем розвитку секторів держав. Кількість та структурні відмінності переліку впливають безпосередньо й на розміри бюджету, який виділяється країною для фінансування захисту офіційно визначених об'єктів критичних інфраструктур. Так, наприклад, уряд США протягом 2001-2013 років витратив понад 260 мільярдів доларів США на заходи щодо захисту національних КІ [6].

Динаміка росту витрат уряду США за період 2001-2013 років на захист КІ представлена на графіку нижче (рис. 1).

Як бачимо з рис. 1, з кожним роком сума витрат на захист КІ в США зростає, що свідчить про розвиток, посилення та вдосконалення підходів щодо захисту КІ країни.

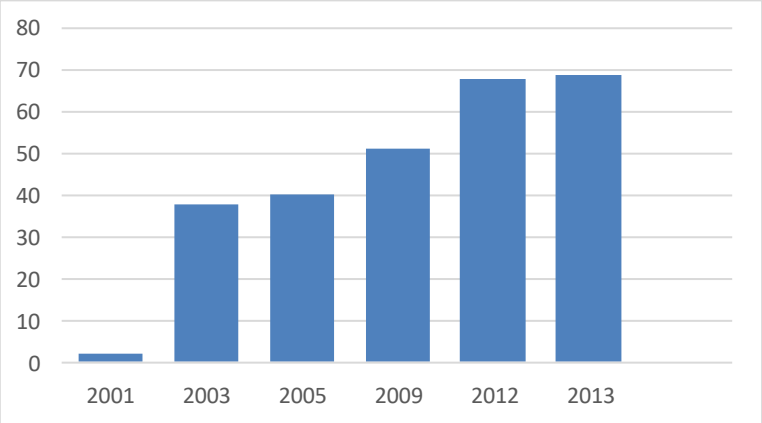


Рис. 1. Динаміка росту витрат уряду США за період 2001-2013 років на заходи щодо захисту КІ (млрд $ США) [6, 7].

Уряд Європейського союзу також щорічно виділяє значні фінансові ресурси на впровадження заходів щодо захисту та посилення живучості КІ, в тому числі на фінансування на захист науково-дослідних заходів в сфері оборони та охорони населення держав ЄС. У праці Д. Бірюквова «Концепція захисту критичної інфраструктури як елемент загальноєвропейської безпековоїполітики»[8] приводяться свідчення про здійснення понад сотні науково-дослідних проектів з метою запобігання та ліквідації надзвичайних ситуацій в країнах ЄС загальною сумою понад 45 мільйонів євро.

Стосовно наявності відповідальності та наявності спеціалізованих організаційних структур в сфері захисту КІ в зарубіжних країнах, то можна сказати, що в різних державах стосовно цього питання підходи дещо відрізняються.

Наприклад, в США відповідальним за захист та безпеку КІ несе Міністерство внутрішньої безпеки, до складу якоговходять більш ніж 22 федеральних підрозділів, відомств та окремих структур [4].

В системі урядових структур Німеччини за організацію функцій захисту КІ на національному рівні відповідає Федеральнеміністерство внутрішніх справ, в складі якого спеціалізовані підрозділи здійснюють оцінку можливих ризиків та загроз для КІ, проводять систематичний аналіз поточних умов безпеки КІ та їх елементів, а також займаються розробкою концепцій захисту КІ [5].

У Польщі відповідальність за безпеку та захист КІ несе Урядовий центр безпеки, який підпорядковується та керується безпосередньо прем'єр-міністром. Також даний центр виконує функцію створення стратегічних програм щодо захисту КІ [6].

2. Захист критичних інфраструктур в Україні

В Україні термін «критична інфраструктура» офіційно був визначений у 2016 році, хоча раніше і використовувався у нормативно-правових актах [9].

Українське законодавство визначає критичні інфраструктури як «сукупність об'єктів єдиної державної інфраструктури, які відіграють найбільш важливу роль для економічної, промислової, суспільної сфер, а також сфери безпеки та охорони населення. Припинення діяльності або руйнування об'єктів КІ має значний вплив на національну безпеку, здатність до оборони й природне середовище, та може стати причиною значних фінансових збитків та численних людських жертв»[10].

В законодавчих нормах України відсутній офіційний перелік секторів, до яких належать національні КІ. Однак, враховуючи наведене в Постанові Кабінету Міністрій «Про затвердження Порядку формування переліку інформаційно-телекомунікаційних систем об'єктів критичної інфраструктури держави» тлумачення поняття «об'єкти критичної інфраструктури»[10], можна визначити наступні сектори КІ в Україні: енергетика, хімічний сектор, транспорт, фінансово-банківська сфера, інформаційно-комунікаційний сектор, продовольча сфера, сфера охорони здоров'я та державні комунальні господарства.

Крім того, згідно з іншими правовими актами України в сфері захисту КІ, можна відокремити наступні категорії об'єктів, для яких також застосовуються заходи щодо захисту та покращення живучості КІ:

- підприємства різних форм власності, що мають стратегічне значення в сфері економіки та безпеки країни;
- державні об'єкти високого рівня важливості, такі як пункти управління органів державної влади та органів місцевого самоврядування;
- об'єкти, що можуть стати цілями для терористичних посягань;
- об'єкти охорони та оборони під час виникнення надзвичайних ситуацій, в режимі військового стану тощо;
- об'єкти, в обов'язковому порядку охороняються підрозділами Державної служби охорони;
- об'єкти, які віднесені до категорії підвищеної небезпеки;
- об'єкти, які включені до Державного реєстру потенційно небезпечних об'єктів;
- радіаційно небезпечні об'єкти;
- об'єкти категорії цивільного захисту;
- чергово-диспетчерська система екстреної допомоги населенню та екстрених служб;
- об'єкти аварійно-рятувальних служб;
- Національна система конфіденційного зв'язку;
- платіжні системи;
- об'єкти нерухомості, що входять до переліку національної культурної спадщини [1].

Як бачимо, на сьогодні в Україні відсутній єдиний, офіційний перелік об'єктів та секторів, до яких відносяться критичні інфраструктури. В нашій державі існує певна кількість нормативно-правових актів, які вказують на переліки таких об'єктів, але дані списки мають скоріш відомчий характер і не є офіційно затвердженими на законодавчому рівні. Такий підхід, на думку авторів, значно ускладнює організацію процесу захисту та підвищення живучості КІ в національних масштабах.

В українській законодавчій базі також відсутнє офіційне тлумачення поняття «суб'єкт захисту КІ». Відповідальність за функції здійснення заходів щодо захисту КІ розділяються між декількома державними органами, а саме: Єдиною системою запобігання, реагування і припинення терористичних актів та

мінімізації їх наслідків; Єдиною державною системоюцивільного захисту населення та територій; а також Єдиноюдержавною системою запобігання й реагування на надзвичайні ситуації техногенного та природного характеру [1, 6].

Така організаційна система та недосконала нормативно-правова база створюють додаткові труднощі в питаннях відповідальності у разі виникнення реальних загроз для об'єктів КІ в національному масштабі. Для існуючої системи захисту КІ в Україні характерне домінування відомчих підходів до розв'язання проблем безпеки та охорони держави та населення. Використання таких підходів формує ряд негативних наслідків та недоліків у процесі розв'язання проблем з безпеки КІ, до яких можна віднести дублювання функціональних обов'язків в одних сферах на вирішення питань безпеки та дефіцит ресурсів у інших сферах, що зумовлене дисбалансом розподілення обов'язків та ресурсів між різними органами безпеки КІ.

3. Можливості для покращення живучості критичних інфраструктур в Україні з урахуванням зарубіжного досвіду

Виходячи з аналізу досвіду зарубіжних країн у сфері захисту та підвищення живучості КІ та аналізу системи охоронних заходів КІ в Україні, авторами розроблено наступні рекомендації щодо вдосконалення організаційних, правових та інших аспектів щодо захисту критичних структур на території нашої держави:

•створення нових або вдосконалення існуючих державних нормативно-правових актів та створення офіційних визначень таких понять, як «захист КІ», «суб'єкти захисту КІ» та суміжних понять;

- удосконалення існуючої нормативно-правової бази та прийняття нових законодавчих актів для систематизації підходів, створення офіційного списку секторів та об'єктів, які належать до сфери захисту КІ;

- створення єдиної системи критеріїв та методології щодо класифікації об’єктів для визначення пріоритетних напрямків захисту КІ різних сфер призначення;
- створення на державному рівні офіційних документів, що надають єдиний перелік усіх об’єктів КІ на території України;
- створення єдиного координаційного державного органу, який буде нести відповідальність за створення, організацію та здійснення звітності щодо проведення заходів та вдосконалення системи захисту та підвищення живучості об’єктів КІ.

Крім того, важливим питанням для державних органів України залишається створення списку об’єктів найвищої пріоритетності в питаннях захисту критичних структур, що в умовах надто обмеженого фінансування дало б змогу покращити рівень безпеки КІ в національних масштабах.

Також, використовуючи досвід зарубіжних країн (зокрема, країн ЄС), варто розглянути можливості розширення партнерських відносин з найближчими державами-сусідами та створити спільні проекти щодо захисту деяких стратегічно важливих об’єктів силами декількох держав зі спільними інтересами (наприклад, об’єктів енергетичного, промислового секторів України).

Висновки

Отже, результати вивчення та аналізу зарубіжного досвіду країн світу в сфері захисту та підвищення живучості КІ може бути використаний для організації цілісної та ефективної системи охоронних заходів КІ в Україні. На основі досвіду зарубіжних країн можна не лише визначити основні напрями та пріоритетні сектори критичних інфраструктур, але й оптимізувати розміри та статті для фінансування, вдосконалити нормативно-правову базу та технології, створити чітку структуру з державних органів, які несуть відповідальність за здійснення захисту КІ на національному рівні.

Список джерел

1. Зелена книга з питань захисту критичної інфраструктури в Україні [Електронний ресурс]. – Режим доступу: http://www.niss.gov.ua/public/File/2015_table/ Green%20Paper%20on%20CIP_ua.pdf. – Назва з екрану.
2. USAPatriotActof 2001 [Електронний ресурс]. – Режим доступу: https://www.gpo.gov/fdsys/pkg/BILLS107hr3162enr/pdf/BILLS107hr3162enr. pdf.
3. The national infrastructure [Електронний ресурс]. – Режим доступу: http://www.cpni.gov. uk/about/cni/. – Назва з екрану.
4. COUNCIL DIRECTIVE 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection [Електронний ресурс]. – Режим доступу: http://eur-lex.europa. eu/LexUriServ/LexUriServ.do?uri=OJ:L: 2008:345:00 75:0082:EN:PDF. – Назва з екрану.
5. The National Critical Infrastructure Protection Programme [Електронний ресурс]. – Режим доступу: http://rcb.gov.pl/wp-content/uploads/NPOIK-2015_eng-1.pdf. – Назва з екрану.
6. Бірюков, Д.С. Захист критичної інфраструктури: проблеми та перспективи впровадження в Україні / Д.С. Бірюков, С.І. Кондратов // Аналітична доповідь. – К.: ПП „Видавництво „ФЕНІКС”, 2012. – 92 с.
7. Цыгичко В.Н. Обеспечение безопасности критических инфраструктур США (аналитический обзор) / В.Н. Цыгичко, Г.Л. Смолян, Д.С. Черешкин // Труды Института системного анализа РАН. – М.: Институт системного анализа РАН, 2006. – № 27. – С. 4-34.
8. Бірюков Д. Концепція захисту критичної інфраструктури як елемент загальноєвропейської безпекової політики / Д. Бірюков // Наукові записки. – К.: Інститут політичних і етнонаціональних досліджень імені І.Ф. Кураса НАН України, 2013. – № 6 (68). – С. 106-115.
9. Про рішення Ради національної безпеки і оборони України від 6 травня 2015 року „Про Стратегію національної безпеки України” [Текст]: Указ Президента України від 26 травня 2015 року № № 287/2015.
10. Про затвердження Порядку формування переліку інформаційно-телекомунікаційних систем об’єктів критичної інфраструктури держави [Текст]: Постанова Кабінету Міністрів України від 23 серпня 2016 р. № 563.
11. Кондратьев А. Современные тенденции в исследовании критической инфраструктуры в зарубежных странах / А.Кондратьев // Зарубежное военное обозрение. – 2012. – № 1. – С. 19-30

STANDARD ANALYTIC ACTIVITY SCENARIOS OPTIMIZATION BASED ON SUBJECT AREA ANALYSIS

O.V. Koval[0000-0003-0991-6405], V.O. Kuzminykh[0000-0002-8258-0816],
M.P. Voronko

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
vakuz0202@gmail.com, avkovalgm@gmail.com, maximvoronko@gmail.com

Abstract: The optimal scenario building task solution based on typical scenarios along with the subject area ontology analysis, using the structure of ordered by action types directed graph is proposed. This approach gives the possibility to evaluate properties of optimal scenario using the suitable metric. The evaluation of possible cost reduction for building more effective standard scenarios of analytical activity is suggested. The relevance of such problems solution in the sample domain of budget process analysis is shown. The algorithm of optimal scenarios search sequential refinement based on subject area ontology analysis along with ordered directed graph analysis is proposed.

Keywords: information gathering, ontology, graph analysis, budget analysis, typical scenario.

Preface

The information technologies development is characterized by a number of trends that require scientific comprehension and elaboration, development of new architectural solutions, including approaches to software systems design and implementation, aimed at analytical activities support. This can be explained by the current analytics systems functioning paradigm, which objectively requires increasing the intelligence of decision-making processes as well as software systems and technological components of analytical support. Considering the dynamic nature of the specific environmental requirements and the complexity of system integration tasks dictate the need of methods and tools development supporting the design of distributed information system, based on distributed analytical activities scenarios and accounts for a large number of participants [1]. Thus, the analysis reveals such phenomena and development trends.

Today there are a number of methods to select best by quality and productivity scenarios from the collection of possible or admissible scenarios, based on the factors composition and nature analysis, which affect the scenario planning process. In planning practice one can distinguish situations, differing in factors number, which are used to make decisions and define principal differences in scenarios development procedures. One of the quite difficult and urgent tasks today that require the effective scenarios development to solve is the task of constructing and further optimizing the scenarios for collecting and analyzing various processes of organizations. Today the scenarios optimizing problem in network information collecting and analysis is one of the major tasks in the domain of big data processing [2].

This is especially true for solving the budget process problems based on the financial analysis of regional budgets. During the economic crisis, the problem of increasing role of the regional budgets analysis in solving economic and social problems becomes insistent. The problems of sustainability of regional budgets gain special actuality.

The main aim of the budget financial sustainability analysis is to obtain a sustainability assessment for each subject of the analysis for a certain period of time and to conclude on the budget stability state for that entity. Based on these results one can make common conclusions about the state budget financial condition, make forecasts for the condition for future periods, and make decisions about steps, directed to improve the economic situation, gain state budget stability and independence.

Software development for state budget financial analysis and defining budget sustainability lets substantially reduce time, needed for analysis and facilitates the work of state institutions and authorities [3].

The overall structure of the network, which determines the ontological features of the problem of budget process analysis on a branched network, is quite complex in structure and has significant number of the elements.

Most usual way to solve such problems is to build scenario, based on definite ontology, which is described with the appropriate graph. Budget monitoring, as well as many other tasks in various fields, often consists in performing a series of typical actions of varying degrees of complexity. The least step, which does not need further breakdown an is usually executed by single person, or, in case of automated execution, by single piece of software, can be called elementary step. Unlike a project or workflow, in the case of a scenario, such a step, as well as compound ones, may have several outgoing results that can be assigned appropriate probabilities. Each step is characterized by a set of parameters that are used to optimize for a particular criterion.

It occurs plausible to use ontology to save steps hierarchy for definite activity, here for budget monitoring, in the ontology the list of connections between steps of the type "before-after". Unfortunately ontology is poorly suited for saving connections metrics, e.g. connection probabilities. Because of this we

need additional means to save additional data for steps pairs, connected with the relation “predecessor - successor”.

Analyst’s job begins with ontology, describing all possible steps, their hierarchy, appropriate metrics and connections of the type “before-after”, creation or development. After this in separate software the probabilities of transition between steps are set or updated. Then this or some other analyst uses steps or step blocks to build the action graph, which as outcome gives result with definite probability or set of results with probability distribution.

The criteria to be used for optimizing, depends upon task. Examples of the criteria may be threshold probability for achieving positive result, scenario execution time or cost minimizing, or even some metric combination, which is saved in ontology and additional storage of results probabilities.

In more sophisticated model some external to network factors are defined, which can alter the probability distribution among arcs or steps metrics. For example one of such factors can be the course of national currency or price of energy carriers etc. But here we will not consider this case. They can be easily implemented in future, if such need occurs.

After the criteria of optimizing is selected, one can use well defined algorithms of shortest route search on the graph, and in complex graph to define reachability of the final aim as well as other common graph algorithms.

If needed on network obtained it is possible to build optimistic, pessimistic and optimal steps series. Such approach makes it possible to supply analyst a set of alternative steps from the current node of the possible actions graph.

In sample case of budget monitoring the simplified way of building the whole process may be next:

1) Building an ontology of possible actions;
2) Automatic building or updating of edges probabilities storage, based on the ontology instances as nodes;
3) Expert manually defines or updates the probability distribution of results for each ontology object, considering the results of previous analysis;
4) Possible actions in budget monitoring graph building;
5) Shortest way search criteria selection;
6) Reachability analysis of the end node from the starting one. If the node is unreachable then building of extended graph or connections correction will be needed to achieve reachability;
7) Shortest route selection using selected criteria and selection of the set of several routes with minimal lengths;
8) Probability of the positive result definition for each of the selected routes;
9) Combined graph building as support system of the analyst activity;
10) After the current state analysis it is possible to update edges probabilities according to practical results. If some new practical steps appear – the extension of ontology is made to include new steps;

If necessary the update and recalculation of the graph is made to include additional new knowledge, obtained from practice.

There is a great number of ontology based scenario building models in subject area, which use different mathematical methods. Among other the scenario building models based on structured data storage in branching networks are usual enough; such is especially true for the problems of regional budget monitoring. Based on these models and their different modifications modern information-analytical and information-searching systems are built [3].

In this case the ontology is defined in the next way:

$O = \{T, A, R, D\}$, where (1)

T – set of terms, defining objects and concepts of the subject area,

A – set of concepts attributes,

R – set of relations (connections) between the terms,

D - set, holding definitions of concepts and relations.

In the graphical form, the ontology is represented as a network, the vertices of which are denoted by the concepts of the domain, and the edges denote the connections between them. Basic are hierarchical class-subclass and part-whole relationships that define the structure of the branching structure of the information storage network.

Thus, ontology represents the description of the subject area, giving view of the concepts set with connections between them.

Descriptive and mapping techniques based on graph theory are widely used to describe ontologies for the tasks of selecting the scenarios optimal by quality or performance criteria from the set of possible or feasible scenarios. In this case, the most widespread descriptions are in the form of hierarchical graphs, usually with weight estimation and edges count.

Building scenario based on the graph analysis

Ontology consideration using structural approach is most relevant to the task, for graph presentation of the structure allows measuring its properties with a selected metric, determining its quality and suggesting recommendations for its future improvement.

Such structured approach makes it possible to evaluate the ontology model building effectiveness due to graph, describing ontology search cost reduction against the usual manual method of building scenario for the same ontology.

To estimate ontology describing graph search cost reduction one can estimate productivity using assumptions, based on some formal assessments for different types of graphs, for which it is quite simple and accurate to estimate the values selected for comparison characteristics. These estimates can be based on the difference between full and partial flow over the graph, which is a graphical representation of the analyzed ontology describing the scenario.

As such procedures basis we can select previous information about graph flow during other queries, which partially coincide with the current that is using information obtained during previously made graph analysis. This information is stored in corresponding knowledge base for each distinct ontology, and so for the corresponding graph. In this case to make approach productivity estimation in whole, without reducing the degree of generality, it is possible to consider some specific graph types used for ontology description.

As it was shown in Miller's article, ontology estimations suggest, that number of connections for a definite concept in the fully connected graph, describing ontology, must not overcome 9 [4, 5]. Thus we can assume that in most of the real cases the number of all ingoing and outgoing edges of the directed graph will not exceed 9.

Using this assumption the maximal effect from usage of previously obtained information for the way of possible scenario building in this case can achieve

$$E_{max} = 9n / (9n - 9(n\text{-}m)). \quad (2)$$

To estimate the productivity of the ontology model building, methods based on the shortest route search in the graph, representing the ontology, can be used.

Generally we can write, that selecting routs to achieve target node T in the graph, using the query:

$$T = \{x \mid A(x)\}, \quad (3)$$

where x- all possible routes in the graph,

$A(x)$ – characteristic property, representing the essence of the specific query, will give the needed result.

Then

$$\forall a \ \ \exists x \ \ \forall c \ \ (c \in x \Leftrightarrow c \le a) ,$$

where

c – all routes in the graph, leading to the target node,

a – minimal length route to the target in the graph.

Thus there always exist minimal length route in the graph, which corresponds to target. So the result of the query also exists

$$T_{min} = min(a). \quad (4)$$

The search task of the single source shortest path (SSSP) [6] is defined based on the general ontology model graph description.

The process flow supporting the analyst work in sample domain of budget monitoring can be represented as follows:

1. Building operational OWL model in the subject area, using Protégé software. (Any other software compliant with OWL2 standard can be used as well).
2. Converting OWL model to the GraphML format using specialized converter.
3. Edit obtained graph, for the purpose of scenario creation.
4. Typical scenario corresponding to selected criteria creation.
5. The most effective current scenario search using selected criteria.

Resulting from the expert in subject area work the corresponding activities model is created and further developed to be used in analytic activity, fixed and stored for definite criteria and tasks as the typical scenario. The specialized software converts the ontology OWL file into the tree of the possible analysts operations saved in the form of the GraphML file. This file represents the supporting operations list for the future analyst work, used for selection of operations subsystem, needed to build scenario and set connections between successive steps with selected metrics of the nodes and edges of the graph [7]. Such metrics as execution time and cost are defined for the nodes, while probability of transition is the metric of edges. Each of the nodes can be simple or compound.

Compound node can hold a graph of possible simple or compound steps, connected with transition edges. The input logical function is stored in any node, defining the node with several incoming edges activation method. In all output edges of the node the normalized transition probabilities to the next nodes are stored thus forming the weighted directed graph.

While practical executing the scenario the updated values of metrics and probabilities can be set in the input structure. The simple nodes hold the input metrics values. Compound node metrics are calculated from the corresponding metrics of the lower level nodes. If the graph can have the cyclic nodes, then corresponding node metrics of such node can have several values, depending on the cycle number. The shortest route search method allows defining simple weight coefficients for nodes, as well as complex, which need metrics for calculation [9]. E.g. if the scenario nodes have time and cost parameters, then shortest route can be calculated using the criteria of the minimal general execution time, minimal cost, or some linear combination of both parameters.

While optimizing scenario built on the graph the standard methods of shortest route search, such as Dijkstra algorithm, are using weighted edges to be run. If we need to select shortest route for the parameters of the nodes the most reliable way is to create supporting edges weights defined as the average of the corresponding nodes parameters, placed on the both ends of the edge except beginning and final node, whose parameter values are not divided by 2. It is simple to show, that total route weight, defined on such edge metrics will be the same, as calculated on the nodes parameter values.

In this work we shall not consider multilevel graph because any multilevel graph can be represented with equivalent single level one by substituting contents of the compound node instead of the node itself (Fig. 1). Such graph will be complex enough, but single layer.

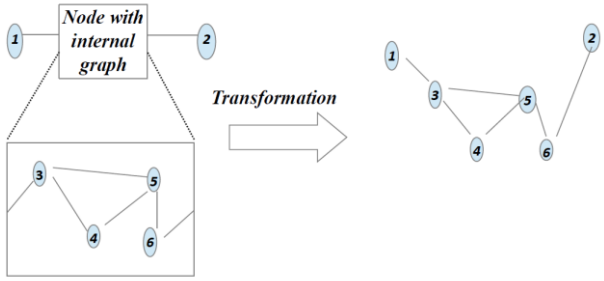


Fig. 1 Multilayer graph transformation to the single layer one.

Let us consider some sample directed weighted graph with defined start node a_s. The edges sequence from node a_s to node a_F ns called route.

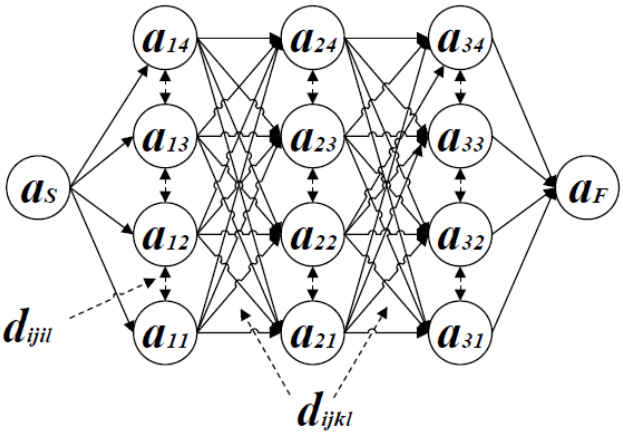


Fig. 2 The base graph structure, used to build a scenario ($n=4$, $m=3$).

The task is for any node v accessible from source-node $a_{S,}$ to find route, having least total weight: $P^*(a_S,a_F)$.

$$f^*(a_S)=f(P^*(a_S,a_F)=min\ f(P(a_S, a_F)). \quad (5)$$

Such problem can be set for both graph type, directed and undirected [9]. The problem solution satisfies the optimality principle: i.e. route in consideration is a part of the shortest route, e.g. from source a_S to finish node a_F . That means that to store the shortest path for each node, it is enough to store its last edges instead of the whole route.

Without decreasing the generality let us revue the task of enhanced rout search on graph, which defines optimized actions sequence in scenario based on the typical scenario of analytical work. Here the typical scenario is already existent and formalized scenario, which usually was received as a result of previous activity of analysts and experts solving such problems.

Let's consider graphical representation of scenario construction on the graph (Fig. 1), which corresponds to this problem formulation with the following assumptions:

1. Every level of graph hierarchy

$$A=\{a_{ij}\} \text{ для } i=1,,,n \text{ та } j=1,,,m \quad (6)$$

corresponds to a full activity set, having analogous by quality results, but differ in activity indicators (execution time or cost).

2. Time consumed by transition between activities in one level is less then transition time from any level to the lower one.

If d_{ijkl} – is an edge between nodes a_{ij} and a_{kl}, and d_{ijil} – is an edge between nodes a_{ij} and a_{il}, then

$$d_{ijil} << d_{ijkl} \quad (7)$$

for i,k = 1,,,n

j,l = 1,,,m

and $i \neq k$.

3. For each graph level, representing subject area ontology there is typical scenario, the defines edges set

$d^{T}_{ii+1.}$ for $i = 1,,,n-1$

4. Scenario can be defined as optimized (enhanced), if total evaluation of scenario edges, defining the newly formed in optimizing process new scenario edge chain d_{iji+1l} is less then total evaluation of edges chain in typical scenario $d^{T}_{ii+1.}$

The successive scenario quality indicators enhancement algorithm, representing the starting ontology of subject area using the structure representation as ordered by activity types graph, can be constructed as the following sequence of steps.

1. For each scenario level

$i = 1,,,n-1$ and $j,l = 1,,,m$

value of d_{iji+1l} is compared to d^{T}_{ii+1}.

2. If values are such that

$d_{iji+1l} < d^{T}_{ii+1}$ then new value for current scenario is defined as

$$d^{TN}_{ii+1} = d_{iji+1l} \quad (8)$$

3. New optimized scenario activities sequence from the node a_{i+1l} sequentially through all subsequent levels to level m.

4. If total sum d^{TN}_{ii+1} for the new chain is less then for d^{T}_{ii+1} of starting typical scenario, then this new scenario is accepted as typical.

5. Where the conditions of paragraph 4 are not fulfilled, the process can be repeated from the last node of the typical scenario, from which a new direction of scenario construction was begun, in another direction up till the last graph level.

Thus, when there is a chain of action better over a chosen criterion (for example, the time of receiving and processing information) compared to a typical scenario, it will be found and the typical scenario will be replaced with a new one.

As a result of running the described algorithm, we obtain a scenario that meets the given conditions, in accordance with the ontology described by the graph. The algorithm presented here shows a sequential process of detecting a sequence of arcs connecting the nodes of the graph from the top level to the bottom, taking into account the matching parameters.

When repeatedly retrieving information by close form of query, we use an existing scenario model, which significantly shortens the construction time by reducing the number of nodes under consideration. This does not exclude the possible need to revise the ontology and rebuild it.

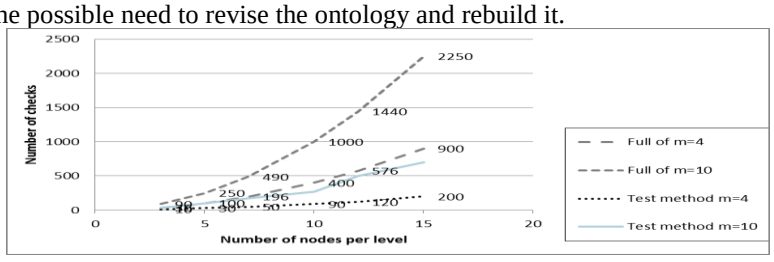


Fig. 3. The successive scenario enhancement algorithm vs full graph analysis approach

Fig.3 shows the scenario optimizing result for test examples with averaging results for two tests. The test examples with 4 and 10 levels and nodes number 3,5,7,10 та 15 for each level where considered. The testing results show, that such approach for scenario optimizing, based on graph representation can significantly reduce the complexity and expenses of building new more effective by defined criteria scenarios. This makes it possible to significantly simplify the analytical work using typical scenarios.

Conclusions

In this article the problems of gathering flow information scenario optimization on the branched network are considered as an example of the construction and further optimization problem of scenarios built on the branched network for the budget monitoring. The structure approach in ontology analysis, as the most appropriate for the considered task, using the graph representation of the structure is suggested. The estimation of the effect of the previously obtained partial information about gathering information in the network scenario construction, described by the graph, for the further clarification of information. The formal description of multilayer hierarchical system structure is provided. The example of ontology elements interaction structure for the problem is presented.

The complex approach based on the shortest route search on the graph and ontology model graph representation is suggested. This makes possible to use algorithms, based on graph nodes in hierarchical levels (layers) traversing.

The approach for modified search in width algorithm building is presented, which significantly decreases routes search for the gathering flow information scenario construction time in the branched network. Described in the article approach for the gathering flow information scenario optimization on the branched network was tested for the pilot system of regional budgets financial analysis project development. Algorithm suggested here makes it possible to develop the software complex, which enables sufficiently full and complete solution to the problems of scenario optimization for scenarios of search and collection of streaming information on an extensive network. One of the perspective directions in the algorithm application is to use its possibilities for information-analytic systems building. This will significantly reduce the time and improve the quality of searching the necessary streaming information on a extensive network.

Literature

1. Alex Guazzelli, Michael Zeller, Wen-Ching Lin and Graham Williams PMML: An Open Standard for Sharing Models: available at: https://journal.r-project.org/archive/2009-1/RJournal_2009-1_ Guazzelli + et+al.pdf
2. Chernov, V.A. The Economic Theory analysis: Textbook // V.A. Chernov .- Moscow: Prospect, 2017 .- 384 p. - ISBN 978-5-392-24867-4
3. Christopher J. Manning, Prabhakar Rahavan, Heinrich Schütz. Introduction Consumer Information Search (trans. With Eng.) - M .: OOO ‘Y.D. Williams’ 2011 – p.504.
4. O. Koval, V. Kuzminykh, S. Otrokh and V. Kravchenko, "Optimization of Scenarios for Collecting Information Streaming Wide-Area Network," 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT), Lviv, Ukraine, 2019, pp. 213-215. doi: 10.1109/AIACT.2019.8847832.
5. Miller G. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review, 1956. 63: pp.81-97.
6. Thorup, Mikkel. “Undirected Single-Source Shortest Paths with Positive Integer Weights in Linear Time.” Journal of the ACM 46, no. 3 (May 1, 1999): pp.362–394
7. Moore, Edward F. “The Shortest Path Through a Maze”. International Symposium on the Theory of Switching, pp.285–292, 1959.
8. Lee, C Y. “An Algorithm for Path Connections and Its Applications.” IEEE Transactions on Electronic Computers 10, no. 3 (September 1961): pp. 346–365.

ВИЗНАЧЕННЯ НАПРЯМКІВ ЗВ’ЯЗКІВ У МЕРЕЖІ ТЕРМІНІВ

Д.В. Ланде [1,2,3][0000-0003-3945-1178], О.О. Дмитренко [1][0000-0001-8501-5313], О.Г. Радзієвська [3][0000-0003-3813-3987], А.В. Бойченко[1]

[1] Інститут проблем реєстрації інформації НАН України, Київ, Україна
[2] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
[3] Науково-дослідний інститут інформатики і права НАПрН України, Київ, Україна

dwlande@gmail.com, dmytrenko.o@gmail.com, radeoksa@gmail.com

Анотація. Розвиток комп’ютерних технологій та, зокрема, мережі Інтернет, яка є джерелом інформаційних ресурсів, які, в свою чергу, є динамічним джерелом текстів, відкриває нові можливості для розробки й застосування покращених методів їх дослідження. Існують різні методи, методики та техніки комп'ютеризованої обробки та аналізу тексту. Проте зважаючи на невпинний розвиток, сучасне програмне забезпечення все більше потребує готових рішень для вдосконалення своїх систем. Адже досі

відкритою та до кінця невирішеною є задача вибору окремих термінів з корпусу текстових документів й автоматизація такого відбору. У зв'язку зі складністю природньої мови, також не менш складною й відкритою проблемою є встановлення синтаксичних зв’язків між термінами у тексті, та визначення напрямків таких зв’язків. В даній роботі розглядаються та аналізуються підходи до побудови мереж термінів, як онтологічної моделі предметної області. Зокрема, представлені нові підходи до визначення напрямків зв’язків між вузлами ненаправленої мережі, побудованої зі слів та словосполучень (окремих уніграм, біграм та триграм) тематичного текстового масиву. Також для побудови направленої мережі термінів розглядається й застосовується один із методів створення термінологічних онтологій – алгоритм формування мереж природніх ієрархій термінів на основі аналізу текстового корпусу. Для створення програмної реалізації запропонованих та розглянутих підходів і методів використовується мова програмування Python, а також окремі функції спеціалізованої надбудови – модуля NLTK (Natural Language Toolkit open source library). Використовуючи засоби програмного забезпечення для моделювання та візуалізації графів – Gephi, побудовані направлені мережі термінів були візуалізовані з метою кращого наочного сприйняття.

Ключові слова: Предметна Область, Онтологія, Мережа Термінів, Граф Горизонтальної Видимості, Мережа Природної Ієрархії Термінів, Синтаксичні та Семантичні Зв’язки, Ненаправлена Мережа, Направлена Мережа.

1. Постановка проблеми

Дуже важливим етапом у комплексних дослідженнях певної предметної області (Subject Domain) є детальне формалізоване представлення її знань, придатне для автоматизованої обробки – побудова онтологічної моделі предметної області. В якості моделі предметної області може розглядатися мережа термінів (Network of Terms), вузли якої відповідають окремим словам та словосполученням у тексті, а ребра – зв’язкам між ними. Процес побудови великих тематичних онтологій зазвичай є складним та ресурсозатратним та становить нерозв’язану досі науково-практичну проблему [1]. Окремий крок такої формалізації – це визначення базових об’єктів. У випадку побудови мереж термінів – це створення словникових номенклатур, тезаурусів та предметних словників термінів, визначених на основі корпусу текстових документів. Ефективний вибір окремих термінів й, тим більше, автоматизація такого відбору з тематичного текстового масиву – актуальна й до кінця невирішена задача [2, 3].

У зв'язку зі складністю природної мови, не менш складною й відкритою проблемою концептуалізації є встановлення семантико-синтаксичних зв’язків між вузлами мережі, що відповідають термінам, та визначення напрямків таких зв’язків.

Метою даної роботи є представити нові підходи до визначення напрямків зв’язків між вузлами ненаправленої мережі, побудованої зі слів та словосполучень тематичного текстового масиву.

2. Метод побудови ненаправлених мереж термінів

Для побудови мереж термінів існує декілька підходів і способів інтерпретації вузлів та зв’язків [4, 5], що призводить до різних видів представлення таких мереж [6].

Для побудови термінологічних онтологій предметних областей в даній роботі використовується алгоритм компактифікованого графа горизонтальної видимості для ключових термінів: окремих уніграм, біграм та триграм.

2.1. Компактифікований граф горизонтальної видимості

Граф горизонтальної видимості [7, 8, 9] є модифікацією стандартного алгоритму графа видимості [10].

Для того, щоб побудувати ненаправлену мережу термінів з використанням аглоритму горизонтальної видимості у роботі [11] пропонується здійснити наступні кроки. Перший етап полягає у тому, що на горизонтальній осі відмічається ряд вузлів, кожен з яких відповідає термінам у тому порядку, в якому вони з’являються в тексті; а по вертикальній осі відкладаються вагові значення – числові оцінки x_i. На другому етапі будується граф горизонтальної видимості. Вважається, що два вузли t_i та t_j, які відповідають елементам часового ряду x_i і x_j, знаходяться у горизонтальній видимості тоді й тільки тоді, коли $x_k < \min(x_i; x_j)$ для всіх t_k таких, що $t_i < t_k < t_j$.

Третій етап полягає в тому, що отримана на попередніх етапах мережа компактифікується: вузли, що відповідають однаковим термінам, об'єднуються в один вузол. Отримана ненаправлена мережа термінів буде називатись компактифікованим графом горизонтальної видимості (рис. 1).

Таким чином, алгоритм компактифікованого графа горизонтальної видимості дозволяє будувати ненаправлені мережеві структури на основі текстів у випадку, коли окремим словам або словосполученням поставлені у відповідність числові вагові значення.

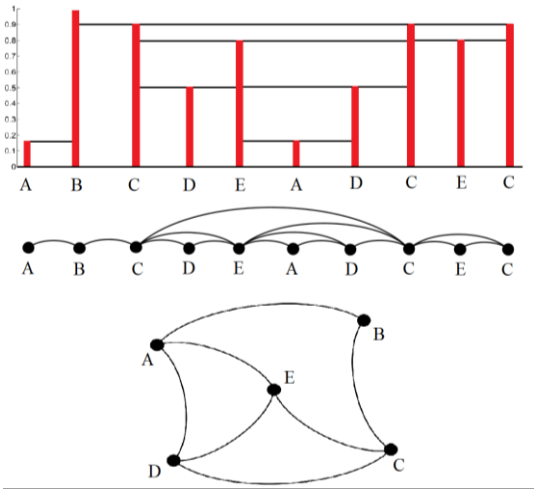


Рис. 12. Етапи побудови компактифікованого графу горизонтальної видимості [11].

2.2. Обробка текстового корпусу

Усна та писемна форми мови складаються зі слів, які часто мають спільне походження.

Мова, що містить різні форми слова, похідні від іншого слова, які використовуються для вираження різного змісту, називається переплетеною мовою (Inflected Language). Зрозуміло, що такі словоформи мають спільну основу.

В даній секції представлені основні етапи комп'ютеризованого процесу обробки текстових документів, які були використані в роботі, такі як: токенізація, розбір на частини мови, лематизація, вилучення стоп-слів, стемінг та зважування термінів.

Токенізація та лематизація

Для проведення попереднього лексичного аналізу здійснюється процес розбиття тексту на елементарні одиниці (токени, лексеми) – токенізація та подальша їх лематизація. Лематизація – процес приведення словоформи до леми – її нормальної (словникової) форми.

У цій роботі була використана бібліотека Python – NLTK, а саме “WordNet Lemmatizer”

Токенізація та лематизація є зазвичай початковими етапами обробки текстів, адже дають змогу працювати зі словом як з окремою сутністю, при цьому знаючи його контекст [12].

Розбір на частини мови

Розбір на частини мови є одним з перших етапів комп'ютерного аналізу тексту.

Для того, щоб кожному слову та його словоформі була надана правильна початкова форма у результаті лематизації, необхідно спочатку надати їм контекст, здійснивши розбір на частини мови [13].

У корпусній лінгвістиці, розбір на частини мови (англ. part-of-speech tagging, POS tagging) – це процес віднесення слова в тексті (корпусі) до певної частини мови, заснований як на його визначенні, так і на його контексті — тобто, на його зв'язку з суміжними і спорідненими словами у фразі, реченні, або абзаці.

Алгоритми розбору на частини мови поділяються на дві відмінні групи: на основі правил і на стохастичні. Метод розбору Е. Брілла [14], який використовує алгоритми на основі правил, є першим й найбільш використовуваним методом розбору англомовних текстів.

Вилучення стоп-слів

Також після етапу обробки текстових документів та виокремлення ключових термів в даному дослідженні пропонується вилучити стоп-слова, які не мають ніякого смислового навантаження, тобто є інформаційно-неважливими, а також біграми, які містять принаймні одне стоп-слово, та триграми, які починаються, або закінчуються стоп-словом. Стоп-словник, який використовувався в межах даної роботи, був сформований на основі різних стоп-словників, які доступні за посиланнями:

https://code.google.com/archive/p/stop-words/downloads/;

http://www.textfixer.com/tutorials/common-english-words.php.

Для створення програмної реалізації додатково було використано “SnowballStemmer” – функцію модуля NLTK (Natural Language Toolkit) – спеціалізованої надбудови мови програмування Python.

Також сформований стоп-словник доповнювався іншими стоп-словами, які були виявлені експертами в межах досліджуваної області.

Стемінг

Після етапів, описаних вище, щоб об’єднати слова, які мають спільний корінь, з метою нормалізації тексту, пропонується здійснити процес стематизації – скорочення слова до основи шляхом відкидання афіксів (закінченнєвих змінних морфем-афіксів, постфіксів, суфіксів та префіксів), що формують похідні форми слова [15]. Результати стемінгу іноді дуже схожі на визначення кореня слова, але його алгоритми базуються на інших принципах [16]. Тому слово після стематизації (обробки алгоритмом стемінгу) може відрізнятися від морфологічного кореня слова.

Варто відзначити, що лематизація та стемінг мають на меті звести словозмінну форму слова, або інколи його похідну споріднену форму, до загальної основної форми.

Проте ці два процеси відрізняються тим, що стемінг зазвичай згортає похідні форми споріднених слів, скорочуючи їх до основи; тоді як лематизація зазвичай лише руйнує різні похідні словоформи, приводячи до нормальної (словникової) форми – леми.

Якщо розглядати англійське слово "saw", стемінг може повернути лише "s", хоча лематизація намагатиметься повернути або "see", або "saw" залежно від того, чи використовувалась лексема (токен) як дієслово чи як іменник.

Тому, щоб уникнути ситуації, описаної вище, в даній роботі процес лематизації проводиться перед стематизацією.

Існує декілька типів алгоритмів стемінгу, які розрізняються відносно продуктивності, точності та відносно того, як долаються проблеми стемінгу [17].

В даній роботі, для створення програмної реалізації, було використано стример “PorterStemmer” (окрема функція спеціалізованої надбудови – модуля NLTK (Natural Language Toolkit)), що розроблений на мові програмування Python та реалізує алгоритм стемінгу Мартіна Портера [18, 19]. Алгоритм Портера є де-факто стандартним алгоритмом стемінгу для англійської мови.

Описані вище кроки дають змогу нормалізувати текст корпусу.

2.3. Зважування та виокремлення ключових термінів

Після проведення попередніх етапів обробки текстового корпусу здійснюється процес зважування та виокремлення ключових термінів. В якості вагових значень термінів, для формування часового ряду в якості функції, яка ставить у відповідність терміну число, в даній роботі використовується модифікація класичного статистичного вагового показника важливості терміна TF-IDF (з англ. Term Frequency – частота терміна, Inverse Document Frequency – обернена частота документа) [20, 21], а саме – GTF (Global Term Frequency – глобальна частота терміна) [22].

Цей підхід дозволяє інформаційно-важливим в глобальному контексті елементам тексту мати високий статистичний показник важливості.

3. Визначення напрямків зв’язків

Нехай G – ненаправлена мережа термінів побудована за принципом, що описаний вище: $G:=(V,T)$ де V — множина вузлів, T — множина невпорядкованих пар вузлів з V, які відповідають причинно-наслідковим зв’язкам між вузлами.

Вважається, що $\forall_{i,j}:(t_i,t_j)\in T$ причинно-наслідковий зв’язок існує у напрямку від вузла t_i до t_j якщо:

1) числове значення, що відповідає: а) степеню [23, 24] б) показнику, що обчислюється за алгоритмом HITS [25] в) показнику, що обчислюється за алгоритмом PageRank [26]) вузла t_i більше за числове значення вузла t_j;

2) у реченні термін, якому відповідає вузол t_i зустрічається раніше терміна, якому відповідає вузол t_j;

3) термін, якому відповідає t_i коротший за термін, якому відповідає t_j.

Для побудови направленої мережі зі слів та словосполучень (уніграм, біграм та триграм) за третім правилом використовується один із методів створення термінологічних онтологій – алгоритм формування мереж природніх ієрархій термінів. Як зазначено у роботі [27] алгоритм створення мереж природніх ієрархій термінів передбачає побудову компактифікованого графу горизонтальної

видимості, та встановлення направлених зв'язків між ключовими термінами, де напрямок зв'язку визначається за принципом входження слова у двочленне або тричленне словосполучення, та входження двочленного у тричленне.

4. Результати дослідження запропонованого підходу

Запропонований підхід для визначення напрямків зв'язків у ненаправлених мережах термінів був апробований на прикладі англомовного тексту, а саме – відомої казки "The story of Little Red Riding Hood".

Відповідно до вищеописаного методу було здійснено обробку обраного текстового документу й виокремлено ключові терміни: уніграми, біграми та триграми (Таблиця 1).

Таблиця 2. Топ-26 ключових уніграм для тексту "The story of Little Red Riding Hood" і їх степінь, HITS та PageRank.

Уніграми	GTF	Степінь	HITS	PageRank
grandmoth	0.065	49	0.444	0.0545
red	0.059	32	0.3099	0.036
hood	0.053	22	0.256	0.0252
ride	0.053	2	0.051	0.0029
wolf	0.031	30	0.301	0.0327
wood	0.025	17	0.204	0.0169
bed	0.019	18	0.191	0.0186
open	0.016	13	0.19	0.0148
time	0.016	14	0.199	0.0162
beauti	0.016	15	0.155	0.0154
big	0.012	9	0.088	0.0119
flower	0.012	8	0.126	0.0093
door	0.012	10	0.136	0.0125
cap	0.012	12	0.162	0.0141
cake	0.012	9	0.101	0.0116
wine	0.012	7	0.099	0.0094
cut	0.009	10	0.095	0.0127
mother	0.009	7	0.118	0.0092
strang	0.009	13	0.116	0.0167
jump	0.009	9	0.101	0.0116
weak	0.009	6	0.083	0.0083
ate	0.009	7	0.134	0.0081
live	0.009	6	0.097	0.0073
pull	0.009	7	0.12	0.0083
sick	0.009	4	0.05	0.0059
huntsman	0.009	11	0.113	0.0148

Таблиця 3. Топ-22 ключові біграми для тексту "The story of Little Red Riding Hood" і їх степінь, HITS та PageRank.

Біграми	GTF	Степінь	HITS	PageRank
ride_hood	0.053	26	0.465	0.0277
red_ride	0.053	28	0.494	0.0303
grandmoth_big	0.009	11	0.177	0.0095
hood_wolf	0.006	5	0.179	0.0054
tasti_bite	0.006	8	0.192	0.0085
bed_pull	0.053	3	0.021	0.0038
hood_grandmoth	0.053	4	0.107	0.0047
leav_path	0.009	7	0.162	0.0084

snore_loudli	0.006	8	0.055	0.0068
grandmoth_live	0.006	6	0.164	0.0071
wolf_bodi	0.006	7	0.160	0.0080
pull_curtain	0.006	5	0.106	0.0056
grandmoth_bed	0.006	5	0.04	0.0060
straight_grandmoth	0.006	7	0.133	0.0076
sick_weak	0.006	5	0.068	0.0057
beauti_wood	0.006	6	0.133	0.0070
cake_wine	0.006	8	0.172	0.0084
beauti_flower	0.006	6	0.105	0.0074
wood_wolf	0.006	7	0.193	0.0076
press_latch	0.006	8	0.047	0.0064
grandmoth_sick	0.006	5	0.096	0.0058
door_open	0.006	8	0.109	0.0085

Таблиця 4. Топ-26 ключових триграм для тексту “The story of Little Red Riding Hood” і їх степінь, HITS та PageRank.

Триграми	GTF	Степінь	HITS	PageRank
red_ride_hood	0.1429	36	0.67	0.1042
grandmoth_what_big	0.0252	6	0.145	0.0114
press_the_latch	0.0168	6	0.059	0.0111
leav_the_path	0.0168	4	0.153	0.0126
sick_and_weak	0.0168	6	0.182	0.0179
bed_and_pull	0.0168	7	0.162	0.0188
cake_and_wine	0.0168	5	0.160	0.0133
hear_how_beauti	0.0084	2	0.111	0.0076
look_so_strang	0.0084	2	0.03	0.0071
hood_and_ate	0.0084	2	0.111	0.0079
listen_littl_red	0.0084	2	0.129	0.007
bite_he_climb	0.0084	2	0.001	0.0101
obey_her_mother	0.0084	2	0.021	0.0084
lay_the_wolf	0.0084	2	0	0.0105
mind_your_manner	0.0084	2	0.054	0.0068
open_hi_belli	0.0084	2	0.003	0.0096
cake_and_drank	0.0084	2	0.018	0.0090
snore_veri_loudli	0.0084	2	0	0.0105
strang_oh_grandmoth	0.0084	2	0.028	0.0059
bird_are_sing	0.0084	2	0.021	0.0084
bed_fell_asleep	0.0084	2	0	0.0103
ride_hood_enter	0.0084	2	0.114	0.0074
larg_heavi_stone	0.0084	2	0.004	0.0093
woman_wa_snore	0.0084	2	0.111	0.0076
loudli_a_huntsman	0.0084	2	0	0.0105
red_ride_hood	0.0084	2	0	0.1042

Побудувавши направлену мережу за першим правилом було отримано наступні результати для різних числових показників вузлів мережі (а) для степеня вузла – рис. 2; б) для HITS – рис. 3; в) для PageRank – рис.4). Візуалізація побудованих мереж термінів здійснювалась за допомогою засобів програмного забезпечення для моделювання та візуалізації графів – Gephi (https://gephi.org).

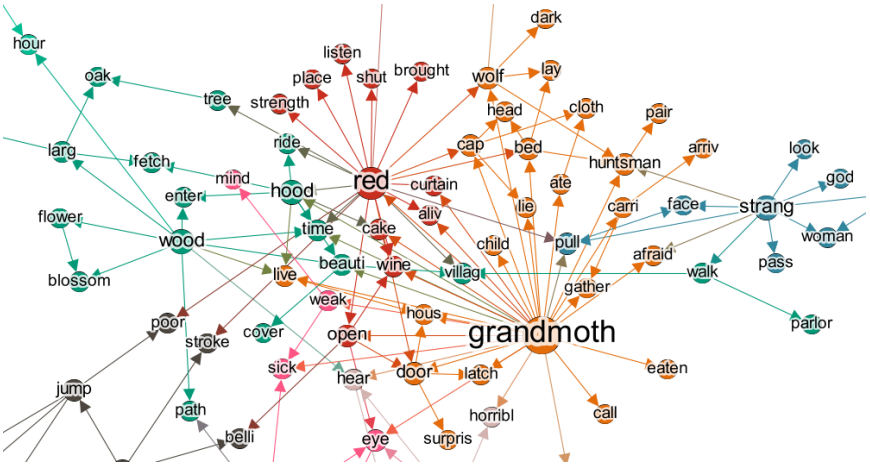


Рис. 13. Фрагмент направленої мережі побудованої за першим правилом для а) степеня вузла.

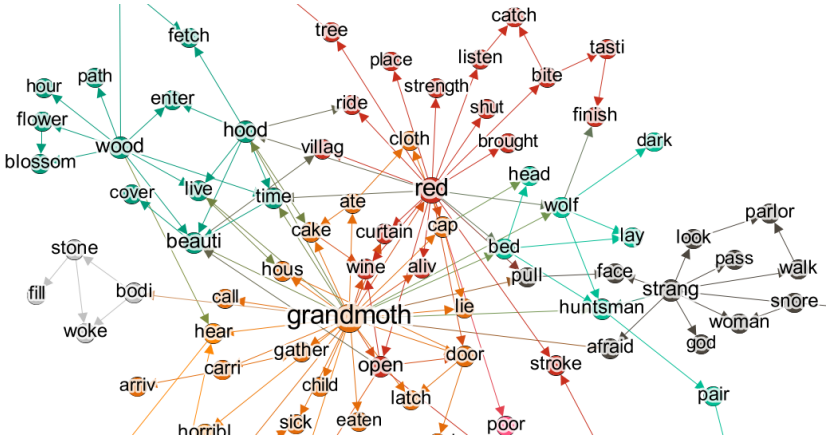


Рис. 14. Фрагмент направленої мережі побудованої за першим правилом для б) HITS.

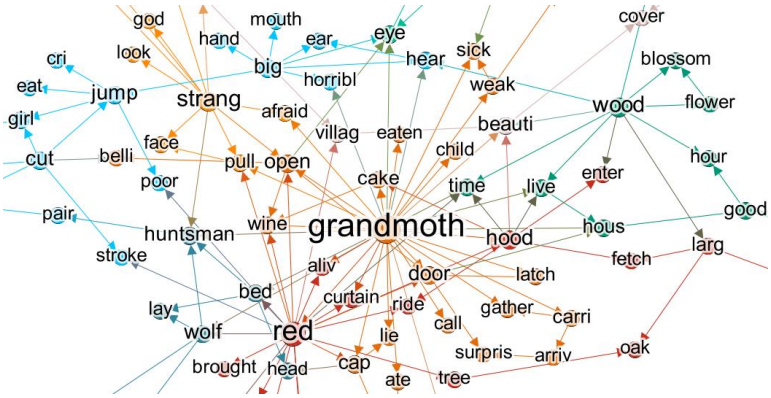


Рис. 15. Фрагмент направленої мережі побудованої за першим правилом для в) PageRank.

На рис. 5 представлена направлена мережа термінів, що побудована за другим правилом.

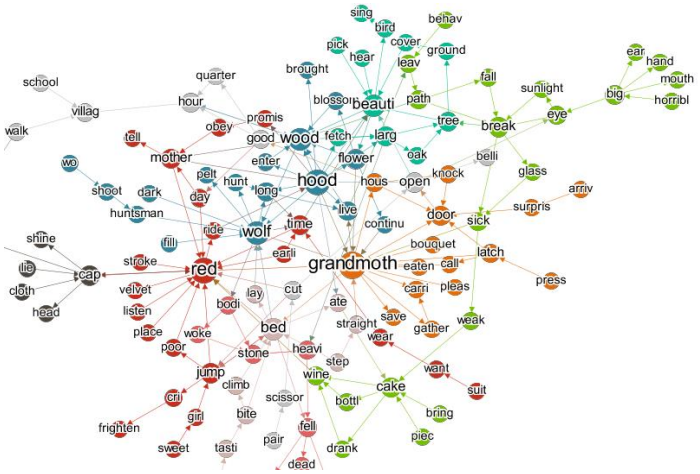


Рис. 16. Фрагмент направленої мережі побудованої за другим правилом.

На рис. 6 представлена мережа природніх ієрархій термінів, що побудована за третім правилом.

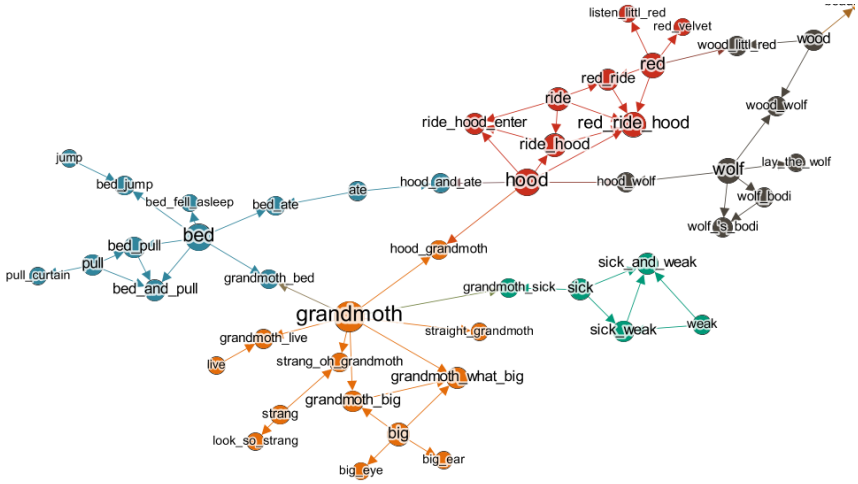


Рис. 17. Фрагмент мережі природніх ієрархій термінів.

Провівши аналіз отриманих результатів, було встановлено, що направлена мережа побудована за другим правилом більш точно відображає напрямок зв'язків, які існують між термінами у розглянутому тексті, ніж мережа, побудована за першим правилом. Мережа природніх ієрархій термінів має свої особливості та переваги, тому її важко порівняти із мережами, що побудовані за першими двома правилами. Зважаючи на природність зв'язків, які встановлюються у такій мережі, можна говорити про їх синтаксичну адекватність.

Якщо розглянути, наприклад, напрямки зв'язків, що визначені для ключових термінів, то видно, що за першим правилом зв'язок між "wolf"-"grandmother"-"red" виглядає наступним чином (див рис. 2,3,4): для степеня вузла, HITS та PageRank – "grandmother" впливає на "wolf" та "red", а "red" на "wolf", що не відповідає реальним напрямкам зв'язків у тексті з точки зору змістовного аналізу; в той час, як за другим правилом – "wolf" впливає на "grandmother", а "grandmother" на "red", що відповідає змісту тексту, який розглядався.

Тож правило визначення напрямків зв'язків, коли у реченні термін, якому відповідає вузол t_i зустрічається раніше терміна, якому відповідає вузол t_j $(t_i, t_j) \in T$, є більш змістовним серед перших двох запропонованих, оскільки зв'язки, визначені за цим правилом більш точно відповідають змісту тексту на думку експертів.

Висновки

Дослідивши запропоновані правила визначення напрямків зв'язків у ненаправлених мережах термінів було встановлено, що друге запропоноване правило більш змістовно відображає напрямок зв'язків, які існують між термінами у тексті. На прикладі загальновідомого тексту – казки "The story

of Little Red Riding Hood" було побудовано ненаправлену мережу термінів, метод побудови якої був представлений у даній роботі. Застосовуючи запропоновані правила визначення напрямків зв'язків, із ненаправленої мережі термінів було отримано направлені мережі. Змістовність зв'язків для мережі побудованої за другим запропонованим правилом є найвищою серед інших двох правил на думку експертів. Зважаючи на природність зв'язків, які встановлюються у мережі природніх ієрархій термінів, можна говорити про їх синтаксичну адекватність.

Направлені мережі зі слів та словосполучень, побудовані за допомогою запропонованого підходу, можна використовувати як основу для автоматизованої побудови онтологічних моделей предметних областей. Також результати роботи можуть бути використані під час створення персональних пошукових інтерфейсів для користувачів інформаційно-пошукових систем, а також у системах навігації у базах даних. Це повинно допомогти користувачам таких систем спростити процес пошуку релевантної інформації.

Оскільки задача підвищення точності визначення напрямків зв'язків у ненаправлених мережах зі слів та словосполучень залишається актуальною, планується продовжити роботу в цьому напрямку, розвиваючи нові та модифікуючи існуючі підходи.

Літературні джерела

1. Ландэ, Д.В., Снарский, А.А.: Подход к созданию терминологических онтологий. Онтология проектирования 2(12). 83-91 (2014).
2. Лукашевич, Н.В., Добров, Б., Чуйко, Д.С.: Отбор словосочетаний для словаря си-стемы автоматической обработки текстов. Компьютерная лингвистика и интеллекту-альные технологии: Труды международной конференции «Диалог–2008», С. 339–344. М. (2008).
3. Филиппович, Ю.Н., Прохоров, А.В.: Семантика информационных технологий: Опыты словарно-тезаурусного описания. М., МГУП (2002).
4. Ferrer-i-Cancho, R., & Solé, R.: The Small World of Human Language. in Proc. of the Royal Society of London, pp. 2261-2265. London (2001).
5. doi: 10.1098/rspb.2001.1800.
6. Caldeira, S. M. G., Petit Lobao, T. C., Andrade, R. F. S., Neme, A., & Miranda, J. G. V.: The network of concepts in written texts. The European Physical Journal B-Condensed Matter and Complex Systems 49(4), 523-529 (2005).
7. Ferrer-i-Cancho, R. F., Solé, R. V., & Köhler, R.: Patterns in syntactic dependency networks. Physical Review E 69(5), (2004).
8. doi: 10.1103/PhysRevE.69.051915
9. Luque, B., Lacasa, L., Ballesteros, F., & Luque, J.: Horizontal visibility graphs: Exact results for random time series. Physical Review E, 80(4), (2009).
10. doi: 10.1103/PhysRevE.80.046103.
11. Gutin, G., Mansour, T., & Severini, S.: A characterization of horizontal visibility graphs and combinatorics on words. Physica A: Statistical Mechanics and its Applications, 390(12), 2421-2428 (2011).
12. doi: 10.1016/j.physa.2011.02.031.
13. Bezsudnov, I. V., & Snarskii, A. A.: From the time series to the complex networks: The parametric natural visibility graph. Physica A: Statistical Mechanics and its Applications, 414, 53-60 (2014).
14. doi: 10.1016/j.physa.2014.07.002.
15. Lacasa, L., Luque, B., Ballesteros, F., Luque, J., & Nuno, J. C.: From time series to complex networks: The visibility graph. Proceedings of the National Academy of Sciences, 105(13), 4972-4975 (2008).
16. doi: 10.1073/pnas.0709247105
17. Lande, D. V., Snarskii, A. A., Yagunova, E. V., & Pronoza, E. V.: The use of horizontal visibility graphs to identify the words that define the informational structure of a text. In: 2013 12th Mexican International Conference on Artificial Intelligence, pp. 209-215 (2013).
18. Manning, C. D., Raghavan, P., & Schütze, H.: An Introduction to Information Retrieval. Cambridge University Press, 22–36 (2009).
19. Schmid, H.: Probabilistic Part-of-Speech Tagging Using Decision Trees. In: Proceedings of International Conference on New Methods in Language Processing, pp. 1–9. Manchester, UK (1994).
20. Brill, E.: A simple rule-based part of speech tagger. In: Proceedings of the third conference on Applied natural language processing (ANLC '92). Association for Computational Linguistics, pp. 152-155. Stroudsburg. PA. USA (1992).
21. doi:10.3115/974499.974526

22. Jongejan, B., & Dalianis, H.: Automatic training of lemmatization rules that handle morphological changes in pre-, in-and suffixes alike. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, pp. 145-153. Association for Computational Linguistics, Singapore (2009)
23. Lovins, J. B.: Development of a stemming algorithm. Mech. Translat. & Comp. Linguistics 11(1-2), 22-31 (1968).
24. Baeza-Yates, R., & Ribeiro-Neto, B.: Modern information retrieval. New York: ACM Press, Harlow. England: Addison-Wesle. (2011).
25. Porter, M. F.: An algorithm for suffix stripping. Program 14(3), 130-137 (1980).
26. doi: 10.1108/eb046814
27. Willett, P.: The Porter stemming algorithm: then and now. Program 40(3), 219-223 (2006).
28. doi: 10.1108/00330330610681295.
29. Salton, G., & Buckley, C.: Term-weighting approaches in automatic text retrieval. Information processing & management 24(5), 513-523 (1988).
30. doi:10.1016/0306-4573(88)90021-0
31. Rajaraman, A., & Ullman, J. D. Mining of massive datasets. Cambridge University Press (2011).
32. Lande, D.V., Dmytrenko, O.O., & Snarskii A.A.: Transformation texts into complex network with applying visability graphs algorithms. In: CEUR Workshop Proceedings (ceur-ws.org). Vol-2318 urn:nbn:de:0074-2318-4. Selected Papers of the XVIIIII International Scientific and Practical Conference on Information Technologies and Security (ITS 2018). vol. 2318. pp. 95-106. (2018).
33. Bondy, J. A., & Murty, U. S. R.: Graph theory with applications. vol. 290. Macmillan, London (1976).
34. Godsil, C., & Royle, G.: Algebraic Graph Theory. Graduate Texts in Mathematics 207. Springer, New York (2001).
35. doi: 10.1007/978-1-4613-0163-9
36. Kleinberg, J. M.: Authoritative sources in a hyperlinked environment. 26. In Processing of ACM-SIAM Symposium on Discrete Algorithms, 46(5), pp. 604–632 (1998).
37. Brin, S., & Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems, 30(1-7), 107-117 (1998).
38. doi:10.1016/S0169-7552(98)00110-X
39. Lande, D.V.: Building of networks of natural hierarchies of terms based on analysis of texts corpora. arXiv preprint arXiv:1405.6068 (2014).

PROBABILISTIC CRITERION OF INFORMATION SECURITY MANAGEMENT SYSTEM DEVELOPMENT

Volodymyr Mokhor[1][0000-0001-5419-9332],
Oleksandr Bakalynskyi[2][0000-0001-9712-2036],
Vasyl Tsurkan[3][0000-0003-1352-042X]

[1] *Pukhov institute for modeling in energy engineering of National academy of sciences of Ukraine, Kyiv-164, Ukraine*
[2] *Department of formation and implementation of state policy on cyber protection of Administration of State serves of special communication and information protection of Ukraine, Kyiv-110, Ukraine*
[3] *Institute of special communication and information protection National technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv-056, Ukraine*
v.mokhor@gmail.com, baov@meta.ua, v.v.tsurkan@gmail.com

Abstract. The risk assessment presentation features of information security by the risk maps are considered. The attention is paid to special aspects of such presentation. Established limit for discreteness and unevenness of changing the step the value of a risk. Using of continuous risk maps is proposed to overcome the limit. This is due to income the information security events with a continuous flow to consider the analogy information security management system with the queuing system. Therefore the justification of continuous risk map is led for probability of occurrence events with risk acceptance. For this the concept and methods of geometric probability are used. Through this received "unit square" as a reflection of probabilistic geometry which is corresponding of the value normalized quantity of information security risk. The admissibility limit is represented by a hyperbola. With this in mind, use the continuous risk maps is justified.

Keywords: information security, risk, risk assessment, continuous risk map.

Introduction

Systemic, visual representation of information security risk assessments is carried out by using maps. Traditionally, they are represented by the coordinate plane whose axes are the risk parameters. For the most part, such parameters are probability of the threat and the amount of losses [1] - [6].

However, in practice, the use of maps is limited to their discreteness and unevenness of changing the step the values of risk [4], [5].

1. Justification of choosing a continuous risk map

Using of continuous risk maps is proposed to overcome the limit [4]. Using of continuous risk maps is proposed to overcome the limit. They are obtained by increasing the multiplicity of discrete risk maps. This transition is due to income the information security events with a continuous flow to consider the analogy information security management system with the queueing system. That's why this process is the best to display by continuous risk maps. That's why effectiveness is confirmed by examples of applications in medicine [4], [6].

2. Establish analogies information security management system with the queuing system

In order to identify the most common analogies between the Information Security Management System (ISMS) and the well-known formal systems, consider the basic characteristics of the ISMS [7]. According to [8], the information security management system is "that part of the overall management system of an organization that is based on risk assessment. It, as part of the overall management system, creates implements, operates, monitors, revises, maintains and improves information security". From this definition it follows that all and any ISMS can consider as a class of systems which is appropriate to repeatedly solve problems of the same type, in a certain sense.

This interpretation suggests an analogy between the ISMS and the queuing system (QS), in which the requirements for the work performed in the form of information security events.

We should also note that in general the sequence of service requirements which have the type of information security events/risks is occasional both in terms of the occurrence of events/risks and the type of such events/risks. The occurrence of the sequence of events/risks served by the ISMS is another aspect of the analogy between the ISMS and the QS.

According to ISO/IEC 27001:2013, all information security events could be divided into separate groups depending on which clauses of Appendix A of the standard [8] they are implemented. In particular, this may be, for example, events in the infrastructure of IT-organization, facts of unauthorized crossing of the security perimeter, personnel problems, non-compliance with certain legislative norms, emergency situations. Information security events fall into each of such separate groups are usually handled by specially trained teams of specialists, and sometimes by external organizations, including law enforcement agencies. Each of these certain teams can be considered as a certain channel for servicing information security events/risks, specializing in events/risks of a certain group, but, in principle, able to serve events/risks which related to other groups. Thus, the presence of processing channels is traced, and this is the essence of another analogy between the ISMS and the QS.

Obviously those information security events involve consequence represented by damage of a certain size H, the occurrence of which is associated with a certain probability p. At the same time, we've known that the combination of the amount of damage and the probability of its occurrence is a risk which is determined in the simplest case by the ratio

$$R = H \cdot p. \qquad (1)$$

In this case, we can say that the ISMS is a QS, in which the requirements for the work performed in the form of information security risks, and the essence of the work performed is the maintenance of these risks in accordance with the recommendations of the ISO/IEC 27k series of standards.

The service is understood in the sense that the ISMS assesses the level of resulting risks and processes such of them which assessment turns out to be higher than the specified threshold. All other events are documented, but the ISMS does not run into the processing state. In other words, the ISMS simply ignores such events. If in a case of separated event, let's say a zero security, we also consider the absence of any information security events, then it's obvious that such a zero, event will be ingored in other words. Thus, we can state that the mechanism of risk management in the ISMS should handle all incoming risks, but it involves two non-overlapping classes of service states: processing and ignoring.

The specific number of analogies between the ISMS and the QS is enough to establish the possibility of interpreting the ISMS as the QS. Fundamentally ISMS may assesses any risk and another analogy between the ISMS and the QS is appeared [7].

3. Evaluation of occurrence probability for acceptable risks

On the basis on relation (1) we can form a trivial risk ranking criterion. But, besides, we can assume that based on the relation (1) and the concept of acceptable risk $R = R_0$, one can determine the probabilistic criterion and its value, set as a design requirement in the construction of the ISMS. For this, using the idea of an approach that called "risk maps" which allow "risk-owners" to set acceptable risk levels $R = R_0$ and divide all risks into acceptable and unacceptable by having the appropriate lines on the "risk maps". Such approach is set out in [1], [3], where the risk map is presented as a two-dimensional table, whose cells at the intersections of the corresponding rows and columns contain the corresponding risk values. At the same time, the risk values are evaluation, for example, on a scale from 0 to 8.

However, it should be noted that the "risk maps" operate with single manifestations of events and do not take into account their possible repeated (multiple) manifestations. The accumulation of the consequences of a set of events, each of which falls into the zone of tolerable, can lead to damage higher than that associated with each of the components of a given level of risk, even without taking into account such phenomena as provocation by one risk of occurrence other. All this leads to the realization that the level of acceptable risk of a single event cannot be used as a correct design requirement for the construction of an ISMS. In other words, the currently existing methods for constructing an ISMS do not have the ability to transform the acceptable risk level set by the owner into the correct formal requirements for building the ISMS. And even if such requirements are formally put forward, then there is no answer to the question of how to make sure that the ISMS, built on the basis of the requirement to ensure a given level of risk, ensures that this requirement is met.

From the thesis stated in the previous paragraph, the conclusion about the non-constructiveness of the project requirement for an ISMS based on the concept "ensure that the level of risk is not higher R_0". On the basis of established analogies between the ISMS and the QS [5], the correct design requirement should be formulated differently, namely this way [9]: the created ISMS should function as a queuing system that provides the processing of the flow of risk events from levels of risk $R \geq R_0$ and a given probability P_0 of occurrence of such events.

In order to substantiate the correctness of such requirement, it is necessary to show the possibility of determining, by a given acceptable risk $R = R_0$, the magnitude of the probability P_0, with which the events associated with the risks occur $R \geq R_0$.

In other words, it is necessary to show the solvability of the following problem: for a given level of acceptable risk $R = R_0$, it is necessary to estimate the probability p_0 of an event with risks $R \geq R_0$. Dual setting of the same task: for a given level of acceptable risk $R = R_0$, estimate the probability p_1 with which events with risks $R < R_0$ may appear. It is obvious that

$$p_0 + p_1 = 1.$$

Probability estimation p_1 can be performed using the concept and methods of geometric probability [8]. First of all, we introduce a two-dimensional Cartesian coordinate system, along the horizontal axis of which we will postpone the values of probabilities p, and along the vertical axis, the values of damage H. It is obvious that the values of probabilities vary in the range from $p = 0$ to $p = 1$, and the values of damage in the range from $H = 0$ to some $H = H_{\max}$. For uniformity of the range of change in the magnitude of damage with the range of variation of probabilities, we introduce into consideration the normalized amount of damage

$$h = \frac{H}{H_{\max}}.$$

Then the normalized damage value will vary in the range from $h = 0$ at ($H = 0$) to $h = 1$ at $H = H_{\max}$.

In Cartesian coordinates ($h0p$) we define the "unit square" $OACE$ as the locus of points corresponding to any possible values of the normalized risk r:

$$r = h \cdot p, \qquad (2)$$

where r is subject to the condition $0 \leq r \leq 1$ due to the fulfillment of the conditions $0 \leq h \leq 1$, and $0 \leq p \leq 1$.

Since the length of each side of the square $OACE$ is equal to one, then the square S_{OACE} of the square $OACE$ is equal to one. We set the level of acceptable normalized risk

$$r = r_0.$$

Then from the relation (2) obviously follows the functional dependence $OACE$

$$h = r_0 \cdot \frac{1}{p}, \qquad (3)$$

the geometrical place of the points of the set of all risks is divided into two subsets (see fig. 1): the figure $OABDE$ determines the geometric location of the points of the set of risk values for which the expression $r < r_0$ holds, and the figure BCD defines the geometric location of the points of the set of risk values for which the ratio is satisfied $r \geq r_0$.

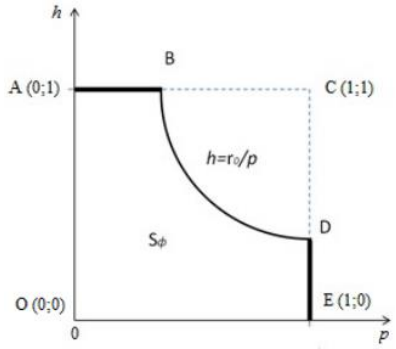


Fig. 1. *Geometrical location of points of a set of risk values, divided by hyperbole* $h = (1/p)$

In this case, the probability p_1 that the value of an arbitrary normalized risk r will not exceed the value of a given standardized risk level $r = r_0$ is determined by the ratio of the figure area $OABDE$ to the area of the "unit square"

$$p_1 = \frac{S_{OABDE}}{S_{OACE}}, \qquad (4)$$

where S_{OABDE} is the area of the figure $OABDE$, and S_{OACE} is the area of the "unit square".

Since it was previously shown that $S_{OACE} = 1$, then relation (4) takes the form:

$$p_1 = S_{OABDE}. \qquad (5)$$

Thus, the probability p_1 that for an arbitrary risk the condition $R > R_0$ will be equal to the area $OABDE$ of the figure. It remains to calculate the area of this figure *(see Fig. 2)*

$$S_{OABDE} = S_1 + S_1 \qquad (6)$$

The area S_1 is calculated as the area of the rectangle with the sides OA and AB. The length of the side OA, as previously stipulated, is equal to 1. And the length of the side AB is determined by the numerical value of the probability coordinate of the point B. The point B is the point of intersection of the line $b = 1$ with the hyperbola, defined by the relation (3). Then the numerical value of the probabilistic coordinate of a point B can be determined by substituting the value $h = 1$ in the left side of relation (3):

$$1 = r_0 \cdot \frac{1}{p}$$

From this relation it follows that the numerical value of the probability coordinate $p = p_0$ of a point B is $p_0 = r_0$.

Then the area S_1 can be expressed by the following relationship:

$$S_1 = 1 \cdot r_0 = r_0. \qquad (7)$$

The area S_2 of the second figure $GBDE$, which is formed by a hyperbola, defined by the relation (3) and three straight lines: $h = 0$, $p = p_0 = r_0$ and $p = 1$, is calculated as a definite integral using the following formula:

$$S_2 = \int_{r_0}^{1} \frac{r_0}{p} dp = r_0 \int_{r_0}^{1} \frac{1}{p} dp = r_0 \ln p \Big|_{r_0}^{1} = r_0 (\ln 1 - \ln r_0).$$

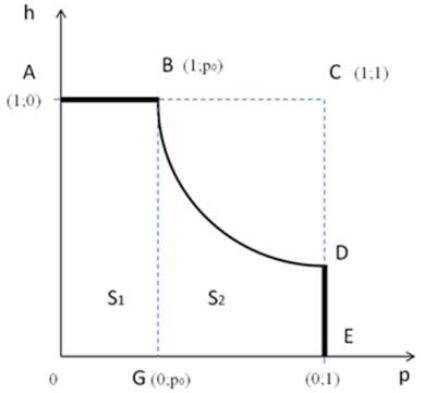


Fig.2. Splitting a figure *OABDE* into two figures: a rectangle *OABG* and a figure *GBDE*

Since $\ln 1 = 0$, the formula for calculating the area S_2 takes the following form:

$$S_2 = r_0 (\ln 1 - \ln r_0) = -r_0 \ln r_0. \quad (8)$$

Then, to calculate the area of the figure *OABDE*, we substitute the values (7) and (8) into (6) and get:

$$S_{OABDE} = S_1 + S_2 = r_0 - r_0 \ln r_0 = r_0 (1 - \ln r_0). \quad (9)$$

So, taking into account (5), a formula is obtained for estimating the probability p_1 that the normalized values of the magnitude of possible risks will not exceed the specified magnitude of the acceptable risk r_0 (see Fig. 3):

$$p_1 = r_0 (1 - \ln r_0)$$

$$p_1 = r_0 \left(1 + \ln r_0^{-1}\right). \quad (10)$$

Looking at the price, there is a list of non-valid cards for the presentation of a number of risks and information.

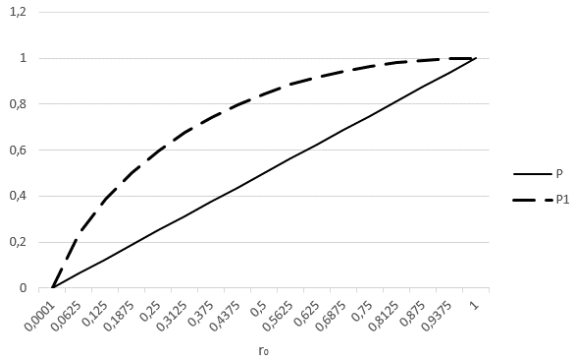


Fig.3. Position of $p_1 = r_0 \left(1 + \ln\left(r_0^{-1}\right)\right)$ function graph relative to the $p = r_0$ function graph

4. Using continuous risk map in information security

Information security events are sent to the ISMS as an continuous stream to according to the analogy of ISMS and QS. That's why this process is the best to image with using continuous risk map. Due to this was choose to use the Cartesian coordinate system with probability values and losses in the ranges from 0 to 1 to reflect such an idea. As a result the risk values are within the "unit square" (see fig. 1), which is created after postponing the maximum value of probability and losses.

When we are using the usual discrete risk maps, an attempt to reduce the uncertainty of the "distances" between two neighboring risks leads to an increase in the multiplicity of the risk map. Increasing the multiplicity of a discrete card leads to a big amount of unacceptable information security risks when it is

developing ISMS. It is difficult to take into account the discreteness of information security risk assessments. Therefore, this restriction is overcome by the use of continuous cards.

For example, continuous risk maps of information security with linguistic criteria are considered. For this, it needs to set these criteria, but each linguistic term still has to find its reflection with a clear classical assessment on the map, for example: 0-20 – very low level; 20-40 – low level; 40-60 – middle level; 60-80 – high level; 80-100 – very high level. The use of linguistic criteria without their "binding" to specific classical estimates when it is applying continuous risk maps is meaningless. Due to the fact that we can find specific numerical values of the probability of realizing the risk and causing losses any risk value on such a map. Of course, using the linguistic values is possible for the division of information security risks into different groups. This action will have a more classificatory character than an applied value.

5. Conclusion

The justification of continuous risk map is led for probability of occurrence events with risk acceptance. For this, firstly we establish analogies information security management system with the mass service system; secondly: the concept and methods of geometric probability are used. Particularly, it is introduced a two-dimensional Cartesian coordinate system on which the value of probability of threat implementation is plotted on the horizontal axis, and the value of the loss is on vertical. The consistency of the change in the value of both ranges is achieved by normalizing the value of the losses. Thanks to this, in the Cartesian coordinate system a "unit square" is obtained. The geometric locus is imaged by this figure which is correspond to the probable value of the normalized value of the risk of information security. By assigning an acceptable value to it the division of the set of risks into a subject of accepted and unacceptable. The admissibility limit is represented by a hyperbola.

It is allowed to income the limit for discreteness and unevenness the value of a risk on map and as a consequence, justifies to use of continuous risk information security maps.

References

1. International Organization for Standardization. (2013, Oktober 01). ISO/IEC 27001. Information technology. Security techniques. Information security risk management. Requirements, https://www.iso.org/standard/54534.html.
2. International Organization for Standardization. (2013, Oktober 01). ISO/IEC 27002. Information technology. Security techniques. Code of practice for information security controls, https://www.iso.org/ standard/54533.html.
3. International Organization for Standardization. (2011, June 10). ISO/IEC 27005. Information technology. Security techniques. Information security risk management, https://www.iso.org/ standard/56742.html.
4. Mokhor, V., Bakalynskyi, O., Tsurkan, V.: Risk assessment presentation of information security by the risks map. Information Technology and Security, vol. 6, iss. 2, 94–104 (2018), doi: 10.20535/2411-1031.2018.6.2.153494.
5. Astakhov, A.: Art of information risk management [Isskustvo upravleniia informatsionnymi riskami]. DMK Press. Moskow (2010).
6. Mokhor, V., Bogdanov, A., Bakalinskii, A., Tsurkan, V.: The Method of the Design Requirements Formation for Information Security Management System, Selected Papers of the XVI International Scientific and Practical Conference "Information Technologies and Security". Kyiv, 2016, pp. 1-6, http://ceur-ws.org/Vol-1813/paper7.pdf.
7. Mokhor, V., Bakalynskyi, O., Bohdanov, O., Tsurkan, V.: Descriptive analysis of analogies between information security management and queuing systems. Zahist ìnformacìï. vol. 19, № 2, 119–126 (2017).
8. International Organization for Standardization (2013, Okt. 1), ISO/IEC 27001. Information technology. Security techniques. Information security management systems. Requirements.
9. https://www.iso.org/ru/standard/54534.html.
10. Mokhor, V., Bakalynskyi, O., Tsurkan, V.: A geometric approach to the acceptable risk probabilities estimation of information security. Zahist ìnformacìï. vol. 18, № 3, 210–217 (2016).
11. Mokhor, V., Bakalynskyi, O., Bohdanov, O., Tsurkan, V.: Analyzing of eligibility of complex risks of information security by analytical geometry methods. Information Technology and Security, vol. 4, iss. 1, 100–107 (2016).

USE OF SEMANTIC SIMILARITY ESTIMATES FOR UNSTRUCTURED DATA ANALYSIS

Rogushina Julia Vitaliivna

Institute of Software systems of National Academy of Sciences of Ukraine, Kyiv, Ukraine, ladamandraka2010@gmail.com

Abstract. The paper discusses problems related to unstructured data analysis in order to acquire implicit knowledge from them. Semantic similarity estimations are used as one of instruments for such analysis. We propose some examples demonstrated the use of ontologies and semantic Wiki markup as a base for generation of semantically similar concepts (in domain-specific sense). Their use increases functionality of the portal version of the Great Ukrainian Encyclopedia. Ontological model of this informational resource is considered as a domain knowledge base that can be used in analysis of unstructured data of this domain. Groupping of concepts is based on high-level ontological classes, and semantic properties and their relations are used for construction of space of attributes.

Keywords: Ontology, Unstructured data, Wiki technology, semantic similarity.

1. Unstructured data analysis

Unstructured data analysis becomes the separate scientific and technical area in the early 2000s, when Gartner analysts published information about high time and labor outlays involved in data processing. This routine, non-automated work with content took up to half the work time. Many problems were caused by processing of unstructured natural language (NL) texts in various formats: emails, business memos, news, chats, reports, marketing materials, presentations, etc. Now the largest part of stored information (more than 80% of all stored data, and their number is increasing by an order of magnitude faster than structured data) is represented by unstructured data (USD) [1], so methods and means of their analysis are evolving rapidly. USD are potentially the greatest source of new knowledge but their use causes problems deal with storage and analysis. Automation of such processing needs in transformation of USD into structured information that can be processed automatically in various ways.

USD defined usually as information collected without any predefined data model or data organization. For the most part USD contain a textual information – arbitrary-length sets of natural language (NL) words combined by weakly formalized linguistic rules. Such USD may also contain dates, numbers, and the like. Examples of text USD are NL documents in various formats, information from social networking, data from mobile devices and content of the Web sites.

The term USD is not well defined for several reasons [2]: it is difficult to distinguish structured and unstructured data if data structure is not defined formally; is not useful for the purposes of data processing or is not applicable for automated processing. One of the criteria for structured data determining is a possibility to create a parser for the data element. Thus, we consider data as USD in cases where information about their structure cannot make data analysis more efficient.

Unstructured information can be stored in the form of objects (files or documents) that have their own structure. For example, the body of email and email attachment are USD, but location of attachment into the mail is determined by structure. The combination of structured and unstructured data are considered also as USD.

2. Properties of unstructured data

USD in contrast to structured data are often created directly by humans, and therefore systems oriented on USD analysis take into account the "human factor".

USD properties:

- *Heterogeneity*. USD can be generated by different ways, in various formats, from various information sources, and these data cannot be structured and placed in any DBMS for various reasons;

- *Ambiguity*. The word for word equal phrases of different persons may have different in kind meanings that depend on their personal experience, views, etc. (for example, an expert phrase “I don’t understand this article” means the poor quality of the article, and the same statement of student means the inadequate education), and the same idea may be expressed in different words .

- *Contextual dependency*. The same word or name is interpreted differently in different contexts (for example, “model” in technology and mathematics has different meanings).

- *Dynamics of value.* Words can change their meaning very quickly, for example, the name of a previously unknown settlement may become publicly known and gain additional meaning because of the widely known events.

Technologies such as Data Mining, Natural Language Processing, and Text Mining provide different methods for finding structure in USD. Common text structuring methods usually include manual tagging with metadata for further structuring. The Unstructured Information Management Architecture (UIMA) standard provides a general framework for processing this information to make sense and create structured data.

3. Text Mining as a basis for analyzing unstructured text information

Text Mining as a separate scientific direction appeared in the late 1990s of the 20th century [3]. Early approaches regarded text as a "bag of words" such as abbreviations, plurals and combination of words, as well as multiple word terms known as n-grams. Basic lexical analysis can take into account the frequency of words and terms to perform such tasks as classification of documents by topic. But this approach doesn't consider the semantics of the documents. Now Text Mining is looking for hidden relations and other complex structures into the sets of text data .

Text data analysis as a technology is based on linguistics and Data Mining. Initially it was oriented on recognition of personal and geographical names, dates, phone numbers and email addresses in the text. Now more sophisticated methods made it possible to find concepts and relations between them and even emotions.

The urgency of the complex scientific problem of adding structure to the USD has increased with the spread of Big Data [4]. In the most general form, the solution of this problem is connected with the construction of a marked graph that corresponds to the contents of the USD, and with the matching of such graphs. Another aspect of this problem is related to the finding of relevant knowledge which can be used as a base for the USD markup.

Text Mining should provide a transition from USD to structured information. Most often, this process ignores a lot of the specific features of the NL that are used only at the previous stage of text parsing, and the following phases of analysis use the "bag of words" model where the order of words is not important [5].

4. Elements for structuring of text documents

Such USD as NL text document from some points of view distinct from automated analysis can be considered as a structured object. For example, from a linguistic point of view each document contains a large amount of semantic and syntactic structure that is hidden in the text. In addition, markup elements (punctuation signs, capital letters, numbers and special characters, etc.) and formatting elements (tables, columns, paragraphs, etc.) can be considered as "soft markup" language to help identify important document subcomponents, such as title, names of authors, subdivisions, etc. Word sequence can also be a structurally significant characteristic to a document. In addition, some text documents may contain embedded metadata in the form of formatted markup tags that are automatically generated by text editors.

Documents that have relatively few such structuring elements (for example, scientific publications and business reports) are called *free-format* or *weakly structured*. Documents with relatively more structuring elements (such as e-mail, HTML web pages) are called *semistructured*.

Pre-processing operations allow to take into account for Text Mining a lot of various elements contained in a NL document to convert it from an USD with implicit structuring into explicitly structured data. However, the potentially great number of words, phrases, sentences and formatting elements contained even in a small document (not even considering the potentially large number of different meanings of these elements in different contexts) causes the need in identification of a simplified subset of document properties (features). Such set of features is called a *representative model* of a document: individual documents are characterized by the sets of features contained in their representative models. But each individual document in the collection has an extremely large number of properties even in the most effective representative models. Therefore, problems associated with high dimensionality of characteristics (ie, the size and scale of possible combinations of feature values for data) are usually much more significant in Text Mining systems than in classic Data Mining systems.

Structured representations of NL documents have a much larger number of potentially representative features – and therefore a greater number of possible combinations of their meanings – than in relational or hierarchical databases. For example, relatively small collection of 10-15,000 documents contains more than 25,000 non-trivial words. The number of attributes in a relational databases analyzed in Data Mining tasks are usually much smaller. The high dimensionality of potentially representative properties leads to pre-processing of the text aimed at creating of simplified models of representation .

Another feature of NL documents is *feature sparsity* – only a small subset of the properties available for document collection as a whole appear in each individual document, and thus, when a document is presented as a binary feature vector, almost all vector values are zero.

5. Properties of individual NL document

Symbols, words, terms and concepts define properties of individual NL document. Text Mining algorithms process document representation through the set of properties, not directly the documents themselves, and therefor we need in compromise between two important goals.

The first goal is to achieve a correct classification of the volume and semantic level of the properties to accurately represent the document during the pre-processing operation. The second goal is to select the property definition that is most computationally efficient and practical for pattern detection. This choice can be supported by validation, normalization, or use properties from controlled vocabularies or external sources of knowledge, such as dictionaries, thesauruses, ontologies, or knowledge bases, to create smaller sets of properties with greater semantic significance.

Although many potential properties can be used to present NL documents, the following four types are most commonly used:

- *Symbols*. Letters, numbers, special characters and spaces are the building blocks of higher-level semantic features such as words, terms and concepts. Symbol-level views may include a complete set of all characters for document or some filtered subset. Character-based representations without position information ("bag-of-characters" approaches) are usually have very limited utility for Text Mining. Views that include some level of positional information (such as bigrams or trigrams) are more useful.

- *Words*. Specific words selected directly from the NL document are the baseline for semantics. Every word-level property has at most one linguistic marker.

- *Phrases* and multiword expressions do not constitute separate properties at the word level. Word-level document representation includes features for each word in that document, that is, the text of the document is represented by complete set of word-level properties. Therefor some representations of word-level document collections contain the great number of unique words in their feature space. However, most document submissions at this level have at least some minimal optimization and are therefore composed of subsets of representative properties that are filtered from such elements as stop words, symbols and meaningless numbers.

- *Terms*. Terms are individual words and multiword phrases selected directly from the source NL document body using the term extraction methodology. Term-based document submission consists of a subset of terms in that document. Various methodologies for extracting terms that convert the raw text of a document into a sequence of normalized terms (tokenized and lemmatized word forms) tagged with the relevant parts of the language can be used. Sometimes, an external vocabulary is also used to normalize terms to provide a controlled vocabulary. Term extraction techniques use different approaches to generate and filter a list of the most relevant document terms from this set of normalized terms.

- *Concept*. Concepts are properties created for NL document using various categorization techniques. Concept-level properties can be created manually, but now more commonly they are retrieved from documents through complex pre-processing procedures that identify individual words, multiword expressions, entire sentences or even larger syntax units, which then relate to specific concept identifiers.

Terms and concepts levels represent properties more significant for semantics. Term-level representations are easier to generate automatically from text than concept-level ones. However, concept-level representation is much more useful for processing synonymy and polysemy.

Many categorization methodologies include the referencing to external knowledge source. For example, some statistical methods can use as externals source an annotated collection of documents. For manual and rule-based categorization, cross-referencing and validation of perspective properties at the concept level typically involve interaction with external databases, such as domain ontology, vocabulary or formal hierarchy. In contrast to word-level and term-level properties, concept-level document properties may consist of words not contained in this document. Concept-based representations allow using very complex concept hierarchies and diverse domain knowledge provided by ontologies and knowledge bases. But concept-level representations have several potential drawbacks: 1) the relative complexity of using heuristics in pre-processing operations, 2) the dependence of concepts on the domain specifics.

6. Using background knowledge in Text Mining

Domain knowledge (another common name is background knowledge) can be used for pre-processing to improve the acquisition of concepts. Domain in Text Mining is a specialized area of interest represented by special ontologies, lexicons and taxonomies. Text Mining systems can use information from formalized

external knowledge sources for these domain to improve document pre-processing and knowledge discovery. Concepts used in Text Mining systems are connected not only to the descriptive attributes of a particular document, but also to domains.

Access to background knowledge – while not absolutely necessary for creating concept hierarchies in the context of a single document or document collection – can play an important role in developing more meaningful, consistent and normalized concept hierarchies.

Text Mining uses background knowledge more than Data Mining: properties of USD are not just elements in a flat set, as is often the case with structured data, because they are linked through lexicons and ontologies to support advanced queries.

Although Text Mining pre-processing operations play an important role in transforming unstructured content of raw document collection into more handy concept-level data representation, the core functionality of such systems is oriented on analysis of concept co-occurrence models in documents collection. Text Mining uses algorithmic and heuristics to consider distributions, frequent sets and various associations of concepts on inter-documentary level to identify the nature and relations of concepts represented by this collection.

For example, if collection of news contains many articles about both event X and company Y at the same time, as well as articles that deal with company Y and product Z at the same time, then Text Mining analysis indicates the relation between X and Z, notwithstanding this relation is not present in any document.

In classic Data Mining, background knowledge from external sources is used to limit search. Text Mining systems can use information from external sources of knowledge in pre-processing and concept testing operations. In addition, access to background knowledge can play an important role in developing meaningful, consistent, and normalized concept hierarchies.

Domain knowledge can also be used by other components of the text extraction system. For example, one of the most important applications of background knowledge is the construction of significant constraints on knowledge discovery operations. Similarly, background knowledge can also be used to formulate constraints that allow users to increase the flexibility of viewing large result sets or formatting data for presentation.

Text Mining systems can utilize background knowledge, represented as domain ontology, describing the set of all important facts, classes and relations between these classes. Domain ontology can be used as a vocabulary designed to be both human-readable and machine-readable.

Well-known ontology used in Text Mining is WordNet developed by Princeton University for NL modeling.

7. Formulation of the problem

Due to the fact that many information resources are unstructured textual data, the problem of acquisition of knowledge from them is very actual. The use of traditional Text Mining is not effective enough to process Big Data, and it causes the need in intellectualization of USD analysis tools that are based on background domain knowledge and specialized ontologies for semantic markup of natural-language texts. We propose to use Wiki technologies and their semantic extension as a basis for structuring natural-language texts.

8. Estimations of semantic similarity

It is advisable to apply the domain ontological knowledge for estimation of the semantic similarity of domain concepts. The sets of semantically similar concepts can be used as a base for structuring of USD by linking of data fragments with ontological elements. Such knowledge makes it possible to quantify the substantive similarity of both the domain concepts and the NL words and phrases corresponding to these concepts.

The values of similarity estimations depend either on estimation methods or on the choice of the domain ontology and on ontology pertinence to user conception about domain. Various ontologies that represent different perspectives on the same domain can be used for this purpose. Such ontologies formalize the contexts of the user task but their use necessitates the integration and reconciliation of these ontologies. It should be noted that the integration of independently created ontologies is a non-trivial problem that cannot be fully automated and requires the participation of domain experts to establish correct relations between ontological concepts.

The definition of semantic closeness between domain concepts is quite closely related to the problem of displaying independently constructed ontologies of this domain.

The task of domain ontologies mapping consist of two separate sub-tasks: *local* representation of concepts that require matching between classes and instances of these ontologies; *global* concept mapping – an analysis of the entire set of local ontology element mappings. Global mapping provides additional information about pairs of different ontology concepts from information about their relation with other elements of ontologies.

Similarity analysis of hierarchical (taxonomic) relations is probabilistic. The assessment of the similarity of concepts from different ontologies may be based on the positions of these concepts in the hierarchy of classes for which similarity has already been determined: if the superclasses and subclasses of these concepts are similar, then the same concepts may also be similar.

The similarity of two entities depends on similarity estimation of: direct superclasses of these concepts; all superclasses of these concepts; subclasses of these concepts; and instances of these concepts.

One of the tasks of mathematical semantics is the measurement of semantic distances between the words of a natural language. Estimation of the semantic distance allows us to assess the density of semantic and associative-semantic relations between words and concepts of the dictionary, between units of text and, in the framework of more complex tasks, between fragments of text.

The value of semantic distance plays an important role in determining the meaning with taking into account the context of several sentences. The semantic meanings of words in a sentence should create semantic unity, therefore, the meanings of the concepts (and sems) of words that stand side by side in a sentence should be in the optimal range of semantic proximity.

Determining the coefficients of semantic connectivity of relations between language units makes it possible to assess the correspondence of NL fragment and its phrases to the points of a multidimensional space of a potentially generated ordered set – classification of natural language phrases. Estimation of the distance between language units is also applicable in other tasks: constructing computer thesauruses where computer automatically constructs a coherent and meaningful text, the work of expert systems, determination of the text subject etc. Semantic similarity of NL fragments A and B can be calculated taking into account the frequency of words typical for A and B. Various types of linguistic relations between words in a language, such as homonyms, synonyms, hyperonyms, antonyms, equonyms, have to receive an accurate numerical estimation based on a uniform scalar value.

A special case of ontologies is taxonomies. They are a fairly common and convenient source of knowledge for analyzing the semantic closeness of NL concepts and words.

a. Using taxonomies to assess semantic similarity

Evaluations of semantic similarity with use of network representations of domain knowledge has long history that was started with the spread of the activation approach [6, 7]. Some researchers consider assessment of similarity in semantic networks using one taxonomic relations “is-a” and exclude other types of relations [8]; other analyze also relations “part-of-part” [9].

Although many similarity criteria have been identified in the literature but they are rarely accompanied by an independent characteristic of the phenomenon that they measure, in particular when the criterion is used in software (for example, similarity of documents in information retrieval, similarity of cases in case-based considerations). On the contrary, the value of such measures depend on their usefulness for particular tasks.

A common and long-known way of evaluation of semantic similarity in taxonomy lies in measuring the distance between net nodes that correspond to the elements being compared – the shorter path from one node to another means their higher similar. If elements are connected by multiple paths between then the shortest path length is used [10, 11].

However, this approach is compounded by the notion that all connections in the taxonomy represent homogeneous distances. Unfortunately, it is difficult to define uniformity of taxonomy distance because real taxonomies have great variability of distances covered by a single taxonomic relation, especially if some taxonomy subsets are much denser than others. For example, WordNet [12] contains a lot of direct links between either fairly similar concepts or relatively distant ones. An alternative way of evaluating semantic similarity in a taxonomy, based on the concept of informational content, which is also not sensitive to the different sizes of distances between relations is offered in [13].

b. Similarity and informative content

One of the key factors in the similarity of taxonomy concepts is the degree of their information sharing that defines by the number of highly specific terms that applies to both of these concepts. The edge-counting method takes it into account indirectly, because if the long minimum connection path by “is-a” relations

between two nodes means that it is necessary to ascend more in taxonomy to general abstract concepts to find the smallest upper bound – a concept to which both concepts under review relate.

Following the standard argumentation of information theory the probability of the concept increases then it's information content decreases. This quantitative characterization of information provides a new way of semantic similarity measuring. The more information is shared by two concepts, the more similar they are, and the information shared by two concepts is determined by the informational content of the concepts in taxonomy.

In practice, we often need to measure the similarity of words, not concepts. Such measure can be based on representation of words through the set of taxonomy concepts that represent meanings (contents) of these words.

Although there is no standard way of evaluating computational measures of semantic similarity, it is appropriate to use estimates that are consistent with human similarity estimates for this purpose. We can use computational similarity measure of word pairs and compare them with human-constructed similarity ratings of the same pairs. In [14] similarity measures are based on maximum depth of the taxonomy and the length of the shortest path in the taxonomy between the concepts. Another points of view for comparing the similarity of concepts is based on the use of the probability of concepts rather than information content. Probability-based similarity estimations consider the word occurrence frequency more important than information content.

9. Wiki technology as a means of information structuring

Wiki technology is a technology for building of the Web resource that allows visitors without using special software to edit its content, specify explicitly links between individual pages through hyperlinks and define their categories [15].

The Wiki page format uses simplified markup language to distinguish various structural and visual elements. Now a large number of Wiki engines and information resources on their base are realized. The largest and most well-known of them is Wikipedia. The main elements of Wiki markup are hyperlinks and categories. Their use makes it easy to convert USD into partially structured data. In addition, the analysis of the structure of Wiki resources at the level of words and concepts allows acquiring knowledge for structuring other USD.

Wiki resources can be used as an external source of features for text categorization and determining the semantic relatedness between NL texts [16].

a. Semanticization of Wiki resources

Semantic MediaWiki (SMW) is an add-on to the MediaWiki [17]. The advantages of SMW are semantic processing of information, the availability of group knowledge management tools, relatively high expressive power and reliable implementation. SMW allows to integrate information from different Wiki pages by searching at the knowledge level and to generate ontological structures on Wiki Pages that can be used by other intelligent software.

In addition to categories, SMW structures information by semantic properties. They allow to link Wiki pages semantically with each other and with other data. Each semantic property has a type, a name and a value, as well as own Wiki page in a special namespace that allows it to determine its place in the property hierarchy and to document how that property should be used.

In terms of ontological analysis, each Wiki page is an ontological element, that is, an element of one of the RDF [18] classes – Thing, Class, ObjectProperty, DatatypeProperty, AnnotationProperty. In addition, each page has its own URI. Usually, Wiki pages are instances of OWL [19] ontology classes, Wiki categories are classes, and Wiki semantic properties are object properties and data properties of ontology.

Therefore, an appropriate OWL/RDF file can be generated on request for any SMW page or the set of pages. Semantic Wiki resources can be used as a basis for automated generation of distributed knowledge bases in RDF format. Exporting to OWL / RDF is a means of ensuring external reuse of data from Wiki, but only practical application of this feature can show the quality of the RDF generated. To this end, system developers have used a number of Semantic Web tools to issue RDFs.

SMW is compatible with the OWL DL knowledge model, therefor external ontologies can be used in Wiki resources. There are two ways to do this: ontology importing allows to create and modify Wiki pages to

represent the relations specified in an existing OWL DL document; and reusing the dictionary allows users to match Wiki pages with elements of existing ontologies.

b. Use of semantic similarity estimations in online version of the Great Ukrainian Encyclopedia

Theoretical principles for development of semantic search and navigation means are implemented into e-VUE – the portal version of the Great Ukrainian Encyclopedia (vue.gov.ua). This resource is based on ontological representation of knowledge base [20]. To use a semantic Wiki resource as a distributed knowledge base we develop knowledge model of this resource represented by Wiki ontology [21]. Using this model for semantic markup provides the formation and software implementation of an appropriate set of hierarchically related categories, templates for typical IOs, their semantic properties and the queries that use them [22].

Application of semantic similarity estimation for this ontology provides the functional extension of Encyclopedia by new ways of content access and analysis on the semantic level.

An ontological model of the structure of the e-VUE is used to support semantic navigation on the portal. One of the significant advantages of e-VUE as a semantic portal is the ability to find semantically similar concepts (SSC). This search is based on the following assumptions: 1. concepts that correspond to Wiki pages that belong to the same set of categories are semantically closer to each other than other concepts on the portal; 2. concepts that correspond to Wiki pages that have the same or similar meanings of semantic properties, are semantically closer to each other than concepts that correspond to Wiki pages with different values of semantic properties or those ones with not defined values of semantic properties; 3| concepts defined as semantically similar by the both preceding criteria are more semantically similar than concepts similar by one of criteria.

For e-VUE, user need to locate the SSP if he (she) is unable to select correctly the knowledge field of concept or enters it with errors. In such cases, user can find similar concepts and then go to desired concept. For example, user wants to find information about a writer or artist whose last name he does not remember accurately, and is not able to accurately determine the style of his (her) works of art, but may indicate the name of more famous person who worked in the same sphere. In some cases, the problem of SSP search is solved by search of the semantically similar words into the NL definition of concept.

In order to extend e-VUE functionality related to search and navigation we propose means of retrieval of semantically similar close IOs – both globally similar (by the full set of features – either categories and values of semantic properties) and locally similar (only by some subset of these features). Concepts of e-VUE are matched with the current Wiki page.

To demonstrate the capabilities of the described above approach we propose the following examples of local SSPs retrived by: 1. the fixed subset of categories of current page; 2. the values of the fixed subset of semantic properties of current page; 3. the combination of categories and values of semantic properties of current page.

The semantic closeness of the search terms is determined relative to the characteristics of current Wiki page that the user is viewing, that is, the categories and properties of this page are analyzed as parameters of such calculation.

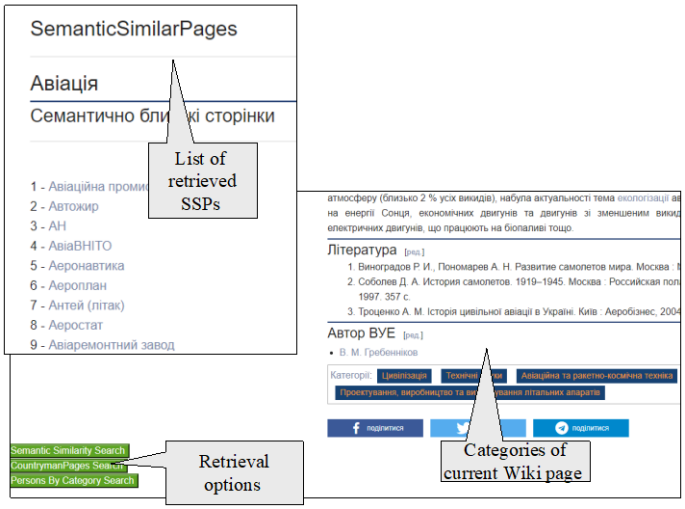


Fig. 1. Search for SSPs for e-VUE page "Aviation".

In the first case, for the current Wiki page you need to find the concepts o e-VUE that are assigned to the categories of the current page. Now this retrieval considers categories and sub-categories of knowledge fields and categories of typical IOs (for example, “Countries”, “Rivers”) IS, and don’t take into account service categories related to the form of publication (e.g. “VUE, Volume 1”) (see Fig. 1).

It should be noted that all these cases of SSP search (locally and globally) cannot be performed automatically by Semantic MediaWiki's built-in tools. . User can achieve the same result with Semantic MediaWiki language of semantic queries by manual entering of all matching parameters – categories and semantic properties copies from selected Wiki page. Therefore these searches are provides by special software code for analyzing the page content.

The second case deals with the search of e-VUE pages corresponding to fixed typical IOs – personalities, cities, countries, etc. Such searches take into account the values of some semantic properties specific to this IO, and match them with values of these properties for the current page. For example, for a typical IO “Person” it is possible to search for persons born in the same place, work in the same field, contemporaries, etc. The third case search option allows to search SSP of the selected category (or the set of categories) with a set of semantic properties defined for selected page. For example, you can find persons (pages from category "Person") who specialize in the fields relates to current page (categories of this page) (see Fig. 2).

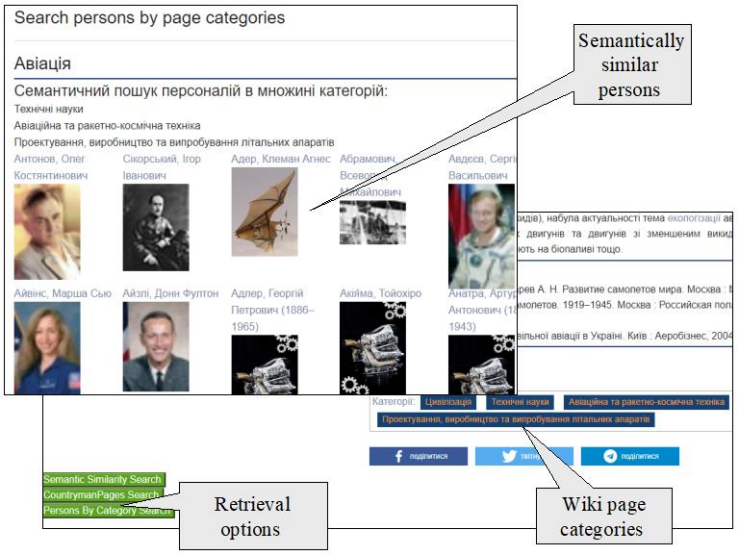


Fig. 2. Search for specialists (by set of current page category) for e-VUE pages.

SSP search generate the Wiki pages with the sets of locally similar concepts. These SSPs can be used in analysis of textual part of USD.

10. Conclusion

We propose to use semantic markup of the Wiki-resource as a source for generation of domain ontologies and groups of semantically similar concepts from these domains. The evaluation of semantic similarity can be based on such characteristics of pages as categories (and their taxonomies), semantic properties and their values, the natural language content and links with other pages. Then these SSC sets can be used as a base for analysis of NL unstructured data. The implementation of proposed approach needs in creation of relevant Wiki resources with appropriate domain knowledge. The use of semantic Wiki-technologies for development of distributed information resources simplifies the process of natural text structuring by users and also generates the source of background knowledge for the analysis of arbitrary texts of the corresponding domains. The models and methods proposed in the work allow to improve this process.

References

1. Grimes S. Unstructured Data and the 80 Percent Rule, Clarabridge, Bridgepoints, (2008), http://breakthroughanalysis.com/ 2008/08/01/unstructured-data-and-the-80-percent-rule/.
2. Unstructured_data, https://en.wikipedia.org/wiki/ Unstructured_ data.
3. Grimes S.: A Brief History of Text Analytics, http://www.b-eye-network.com/view/6311.
4. Buneman P., Davidson S., Fernandez M., Suciu D.: Adding structure to unstructured data. In: International Conference on Database Theory, pp. 336-350. (1997).

5. Feldman R., Sanger, J.: The text mining handbook: advanced approaches in analyzing unstructured data (2007), https://wtlab.um.ac.ir/images/e-library/text_mining/The%20Text%20Mining%20HandBook.pdf.
6. Quillian M. R.: Semantic memory. In: Minsky, M. (Ed.), Semantic Information Processing. MIT Press, Cambridge, MA, (1968)
7. Collins, A., Loftus, E.: A spreading activation theory of semantic processing. In: Psychological Review, 82, pp.407-428, (1975) .
8. Rada R., Mili H., Bicknel E., Blettner M. (1989) Development and application of a metric on semantic nets. IEEE Transaction on Systems, Man, and Cybernetics, 19(1), P.17-30.
9. Richardson R., Smeaton A. F., Murphy J.: Using WordNet as a knowledge base for measuring semantic similarity between words. In: Working paper CA-1294, Dublin City University, School of Computer AppUcations, Dublin, (1994).
10. Lee J. H., Kim M. H., Lee Y. J.: Information retrieval based on conceptual distance in IS-A hierarchies. In: Journal of Documentation, 49(2), pp.188-207, (1993).
11. Rada R., Bicknell E.: Ranking documents with a thesaurus. In: JASIS, V.10(5), pp.304-310, (1989).
12. Fellbaum C.: WordNet. In: Theory and applications of ontology: computer applications, pp. 231-243). Springer, Dordrecht, (2010).
13. Resnik P. (1999) Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language // Journal of Artificial Intelligence Research 11, P.95-130.
14. Miller G. A., Charles, W. G.: Contextual correlates of semantic similarity. In: Language and cognitive processes, 6(1), pp.1-28, (1991).
15. Wagner C.: Wiki: A technology for conversational knowledge management and group collaboration. In: The Communications of the Association for Information Systems, V. 13(1), pp. 264-289 (2004).
16. Banerjee S., Ramanathan K., Gupta, A.: Clustering short texts using wikipedia. In: Proc.of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 787-788, ACM, (2007).
17. MediaWiki, https://www.mediawiki.org/wiki/MediaWiki.
18. Broekstra J., Klein M., Decker S., Fensel D., Van Harmelen F., Horrocks I..: Enabling knowledge representation on the web by extending RDF schema. In: Computer networks, 39(5), pp. 609-634, (2002).
19. McGuinness D. L., Van Harmelen F.: OWL web ontology language overview. In: W3C recommendation, 10(10), (2004).
20. Rogushina J.V.: Use of semantic properties of the Wiki resources for expansion of functional posibilities of “Great Ukrainian Encyclopedia”. In: Encyclopaedias in the modern information space/ Ed. Kirillon A.M., Kyiv, pp.104-115, (2017) [in Ukrainian]
21. Rogushina J.: Analysis of Automated Matching of the Semantic Wiki Resources with Elements of Domain Ontologies. In: International Journal of Mathematical Sciences and Computing (IJMSC), Vol.3, No.3, 2017, pp.50-58.
22. Rogushina J.V.: The Use of Ontological Knowledge for Semantic Search of Complex Information Objects. In: Proc. of OSTIS-2017, pp.127-132, (2017).

ONTOLOGY-BASED APPROACH TO VALIDATION OF LEARNING OUTCOMES FOR INFORMATION SECURITY DOMAIN

Julia Rogushina[1], Anatoly Gladun[2], Serhii Pryima[3], Oksana Strokan[4]

[1] *Institute of Software systems of National Academy of Sciences of Ukraine, Kyiv, Ukraine, ladamandraka2010@gmail.com*

[2] *International Research and Training Center of Information Technologies and Systems of National Academy of Sciences of Ukraine and Ministry of Education and Science of Ukraine, Kyiv, Ukraine, glanat@yahoo.com*

[3] *Dmytro Motornyi Tavria State Agrotechnological University, Melitopol, Ukraine, pryima.serhii@gmail.com*

[4] *Dmytro Motornyi Tavria State Agrotechnological University, Melitopol, Ukraine, oksana.strokan@tsatu.edu.ua*

Abstract. Validation of the results of non-formal and informal learning is an effective way to solve a number of socio-economic problems in various spheres. Recognition of learning outcomes for dynamic domains that requires the use of background knowledge from open information resources. It causes the interest to ontological representation of learning domain knowledge and ontology-based methods of their analysis. Now a lot of useful information is represented by unstructured natural language texts. The large volume of such data necessitates scalable means of analysis, and more efficient and rapid processing can be ensured through the use of task-

specific thesauri based on domain ontologies. The approach proposed by the authors to recognize non-formal and informal learning outcomes is exemplified by information security domain. The referencing to this domain is determined by the heterogeneity and dynamics of its information sources, the complex hierarchy of knowledge as well as the growing need for information security specialists in the context of the digitalization of society.

Keywords: Ontology, Information Security, Validation of Learning Outcomes, Non-Formal Learning, In-Formal Learning, Unstructured Data, Big Data.

1. Unstructured data analysis

The rapid development and integration of various domains requires constant updating of information about relevant professional knowledge and skills in condition of labor mobility. Therefore we need in analysis of their pertinence to new types of tasks that arise at all stages of life activity of individuals and functioning of organizations. In turn, such dynamism of knowledge complicates the determining of learning outcomes in formal, non-formal and informal education. Recognition of the non-formal and informal learning outcomes is increasingly seen as a way of improving life-long education for employment and improving of life quality.

Most European countries emphasize the importance of learning visualization that is organizes outside of official education institutions and proceeds at home, at work or leisure.

This problem is very actual for those domains that are rapidly changing and integrating different directions of activities. To work effectively in these areas, professionals need lifelong learning, gaining experience and new learning outcomes. These features apply equally to most of the information technologies. On conditions of digitalization it is especially important for information security (IS) domain, since practically all employees need now to have basic knowledge in IS for processing of digital data without their loss and disclosure.

Individuals and organizations alike depend on the cyber world. As the world grows more connected through digital environment, a highly skilled information security workforce is required to secure, protect, and defend global, national, commercial and personal information. The IS strategies and development plans require the improvement of the know-how of the citizens and actors of the economic life and public administration. But a lot of people who need now in modern IS skills and knowledge can receive them only from self-education.

Therefor IS competence is not just another professional field of expertise. Rather, it ranges from civic skills all the way to international-level professions and should be included in different educational and institutional levels.

Validation of learning outcomes is a process that recognizes various learning outcomes with the help of such steps as identification, documentation, assessment and certification. Successful validation of learning outcomes is based on use of formalized knowledge of the relevant subject domain: such knowledge is used for matching of learning results obtained by individual with domain references.

Now the problem of validation of non-formal and informal learning outcomes is complicated by the unstructured and unclear nature of many domains, such as IS. It is difficult to clearly define the boundaries of each domain, because many professions are located at the intersection of several industries or have unformalized domain-independent requirements (such as response rate, stress resistance) and require external knowledge from other fields. The problem solution requires the use of modern methods for representation and analysis of knowledge that ensure the information processing at the semantic level.

In this paper we consider IS as an example of complex and dynamic domain where the need to validate the results of non-formal and informal learning is very actual but faces many theoretical and practical obstacles.

2. Formulation of the problem

Digitization of all spheres of life of modern society [1] increases the need for modern information security means and awareness in this area for a wide variety of users and developers of information technology. Therefore, a significant number of people have gained knowledge in the process of self-study, i.e., by non-formal and informal education in this field. Informality of this process and diversity of terminology used for describing of such outcomes complicates to obtain formal definition of learning results for employers for matching with clear criteria and requirements for job responsibilities. This necessitates a reconciliation of domain terminology with knowledge structures pertinent to current IS standards. But now the big number of standards and examples deals with various IS aspects, new versions and editions appear frequently. Therefor Big Data methods can be applied for tracking of all changes in this sphere.

We suggest to harmonize a hierarchy of IS ontologies to define the terms of this domain and construct thesaurus of IS. Each term of this thesaurus is associated with one or more elements of these IS ontologies (concepts, relations etc.). Use of this thesaurus allows to describe various learning outcomes (by natural

language (NL) texts or by structured data) obtained by non-formal and informal learning, and provides their comparison with formalized (or semi-formalized) descriptions of jobs and tasks that arise in this field. Such IS thesaurus helps to take into consider the most task-relevant domain knowledge. Time required for processing and construction of thesaurus is much less than the time of matching of domain ontologies with unstructured NL text, which is usually used to describe learning outcomes and vacancies. Thus, the approach proposed in the work should ensure the transition from processing of unstructured large-scale data to analysis of much more compactly represented knowledge obtained from appropriate Big Data and from domain ontologies.

3. Information Security Domain Specificity

The current tendencies regarding the storage, exchange and processing of information are characterized by the intensive implementation of information technologies, the spread of local, corporate and global networks in all spheres of life of the state and the individuals. This situation creates new opportunities for information exchange and requires that the information becomes secure and safe at the same time.

The actuality of this problem is determined by the following factors:

- the exponential growth of the number of personal computers, mobile devices and other means of information exchange;
- rapid expansion of the circle of users with direct access to digital networks and information resources;
- an increase in the amount of information that is accumulated, stored and processed in digital form;
- spread of hardware, software and information technologies that do not satisfy the current requirements of safety;
- discrepancy between the rapid development of information processing facilities, the national laws of information security and the development of international standards and legal norms that will provide the necessary level of information protection;
- information war in and around the country caused by external aggression;
- cybercrime, sometimes of national importance, etc.

IS deals with the protection of information systems, both hardware and software, from theft, unauthorized access and disclosure, as well as from intentional or accidental harm. IS has to protects all segments related to the Internet (from the networks themselves to the information transmitted over the network) and stored in databases, to the various applications and devices that control the operation of the equipment through network connections.

Today, the issue of IS becomes a cornerstone in the activities of every organization or individual. IS refers to the security of entire organization data against deliberate or accidental actions that cause damage to its owners or users. IS domain is a very dynamic sphere that requires constant monitoring of information coming from both internal and external information resources.

Information resources in IS are generated usually by different tools, sensors and systems expressed using different standards and formats, published by different sources and is often scattered as isolated pieces of information. Furthermore, cybersecurity data are available in structured, semi-structured and unstructured forms from both, internal sources i.e. within the organization, and external sources i.e. outside the organization. Unifying such scattered information provides better visibility and situational awareness to cybersecurity analysts.

Now new terms in IS domain appears frequently, and old ones change the meanings making it difficult to agree on different approaches to knowledge representation about IS. In addition, documents (standards, protocols, descriptions) from the IS are characterized by large volume and not structured form. Therefore, it is advisable to use Big Data analytics methods – a new technology that enables the collection, storage, processing and visualization of these vast amounts of data. Such analytics takes into account the basic characteristics of big data [2].

Specialized analytical methods extract useful information from distributed data to obtain the information about the specialists` learning outcomes relevant to the challenges in the IS field. Analytics Big Data can be used to make decisions in such IS tasks [3].

Big Data Analytics applied to IS data can provides insights useful to improve current security practices of network operators and administrators. To use Big Data as an external source of information, we first need to filter out multiple datasets by pertinence to task of analytics. Preliminary cleaning of the data and preparation for analysis can be performed on the basis of metadata analysis (and quite often the analysis of the NL part of it, for example, annotations) using the knowledge of the domain of processing.

It is advisable to treat some information as Big Data if it has one or more characteristics from the "5V": Volume, Variety, Velocity, Value, Veracity [4]. Big Data, also referred to as Data Intensive Technologies, now are becoming a new technology trend in science, industry and business [5] . Big Data can be defined as a new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data by enabling high velocity capture, discovery, and/or analysis.

Now Big Data technology is applied in different human activity domains empowered by significant growth of the computer power, ubiquitous availability of computing and storage resources, increase of digital

content production, mobility and security. Each stage of the Big Data lifecycle provides the data transformation or changes in processing of the dataset content. In many cases original data and processed data are linked by referral integrity. This motivates such Big Data features as dynamism and variability that reflect the fact that data are in constant change and may have a definite state, besides commonly defined as data in move, in rest, or being processed. Supporting these data properly will require scalable provenance models and tools incorporating also data integrity and confidentiality.

Problems of big data use for analytic in IS often are caused by the lack of descriptions of metadata datasets. Re-use of data requires as much information as possible about the origin of the data and about a complete history of the methods used to collect, maintain and analyze. The availability of metadata increases the likelihood that the datasets are reused correctly.

Widespread Big Data used as a source of background IS knowledge need in the scalability of methods that can be used to structuring and validation of learning outcomes in this domain.

4. Unstructured data in the IS domain

Today, the major portion of cybersecurity data is represented by unstructured NL data. Unstructured data (USD) is information that does not have a predefined model of organization. This causes problems related to storage (traditional databases are not designed for such uncertainty) and analysis. USD contain potentially the greatest value as sources of new knowledge, and the total amount of such data influences on accuracy of the analysis results.

In many cases, USD are represented by textual information in electronic form by sets of natural language words of arbitrary length, combined according to weakly formalized linguistic rules. Such textual information contains the most useful information for further use but USD may also contain dates, numbers, special symbols etc.

Sometimes it is difficult to distinguish between structured and unstructured data. One of the criteria for determining whether data are structured is to create a parser for the data element. USD as a term is not well defined for several reasons [6]:

- data contain the structure without it’s formal definition;
- structure of data is not useful for the specific purposes of their processing ;
- structure of data information cannot be processed automatically without further clarification.

Data are considered as USD in cases where information about their structure cannot make the analysis of these data more efficient. Therefore, for example, standards for IS (both domestic and international) in terms of data analysis are considered as USD, although they have certain structural elements.

A lot of important cybersecurity information can be extracted from USD. For example, various cybersecurity vulnerabilities are typically identified and accessible publicly on the Web but at first not in official documents. It causes a strong need for systems that can automatically analyze unstructured text and extract vulnerability entities and concepts from various non-traditional unstructured data sources such as cybersecurity blogs, security bulletins and hackers forums. Now there is no widely used automatic mechanism to understand and process such USD with non-formalized terminological base.

Technologies such as Data Mining, Natural Language Processing, and Text Mining provide different methods for structure acquisition for USD. Common text structuring methods usually include manual tagging with metadata or other specific tags for further structuring. The Unstructured Information Management Architecture (UIMA) standard provides a general framework for processing this information to acquire semantics and create structured data.

The earliest Business Intelligence studies focused on unstructured text data, not numerical data [7]. The development of Big Data in the late 2000s caused increased interest in the use of USD analysis.

Separating unstructured data analysis (UDA) into the separate scientific and technical area dates back to the early 2000s, when Gartner analysts published information about high time and labor involved in processing data because such routine, non-automated content processing took up half the working time.

Now data without formalized structure are the largest share of stored digitized information (more than 80% of all stored data, and their number is increasing by an order of magnitude faster than structured data), so methods and means of their use are evolving rapidly. These methods are intended to transform USD into structured information that can be used in various ways.

Adding a structure to USD is a complex scientific problem that has been paying attention to scientists for a long time [8]. In the most general form, the solution of the problem is related to the construction of a spaced graph corresponding to the USD content and to the comparison of such graphs generated for different USD sets. Another aspect of this problem is related to the finding of relevant knowledge for USD structuring tags and means of representation of such knowledge.

Text Mining can be defined as the process of acquiring knowledge from collections of USD documents using various toolkits for their analyzing [9]. Text Mining can be considered as a separate case of Data Mining. Processing of NL data should provide a transition from USD to structured ones with subsequent analysis. Most often, this process ignores most of the specific features of the NL, which are used only at the previous stage of parsing the texts, and in the following uses the "bag of words" model, in which the word

order is not difficult. Similar to Data Mining, Text Mining tools seek to retrieve information relevant to user's activity. In the case of Text Mining such templates that can be applied for user tasks should be found not in formalized database records, but in non-structured text data in the documents of these collections.

An analysis of current research in this area demonstrates that the most effective approaches to the analysis of NL data are based on the use of the background knowledge of the corresponding domains formalized by ontology [10].

5. Ontologies of IS domain

Cybersecurity information needs to be machine-readable to enable efficient information exchange, retrieval and operation automation deal with validation of learning in this domain. Today, software development in the IS field is characterized by intellectualization, which allows to anticipate the threats that can lead to APT (advanced persistent threat) attacks, phishing and redemption software. Ontological representation of knowledge is widely used now in distributed intelligent application oriented on processing into the Web open environment to support knowledge sharing and reutilization. Ontologies provide an explicit specification of a conceptualization. Therefore creation of IS domain ontologies covered a set of concepts and categories of this domain, their various properties and relationships between them is of great interest to developers and users of modern IS means.

Ontologies in IS have now evolved into an active provider of data element links that can use machine learning and artificial intelligence algorithms to adapt to changes in the environment [11]. Development of ontological model for entire IS domain requires integration of existing ontologies and their refinement.

The Unified Cyber Security Ontology (UCO) [12] is designed to support the knowledge integration into IS systems and should unify the most commonly used information security standards. This ontology integrates heterogeneous data and knowledge schemas from different IS subsystems and the most commonly used IS standards for data sharing. UCO can serve as the core of knowledge for the IS domain. The UCO ontology has also been mapped to a number of existing cybersecurity ontologies as well as concepts in the Linked Open Data cloud.

UCO is an ontology developed for integration of the unifying information from heterogeneous sources. This ontology supports reasoning and inferring new information from existing information. UCO also provides capturing specialized knowledge of a cybersecurity analyst expressed by ontology classes and individuals as well as rules. UCO ontology provides a common understanding of cybersecurity domain and unifies most commonly used cybersecurity standards.

The most important top-level IS concepts represented by UCO classes:

- Means: various methods of attack executing
- Consequences: possible outcomes of attacks.
- Attack: cyber threat attacks.
- Attacker: identifications and characterization of the adversary.
- AttackPattern: descriptions of common methods for exploiting software that provide the attackers perspective and guidance on ways to mitigate their effect.
- Exploit: descriptions of individual exploits.
- Exploit Target: weaknesses and vulnerabilities of software, systems, networks or configurations that are targeted for exploitation by the TTP (cyber threat adversary Tactic, Technique or Procedure).

UCO ontology serves as the core for cybersecurity Linked Open Data (LOD) cloud and integrates knowledge from some international sources but general ontological model of IS has take into account various national IS standards and laws.

A detailed description of knowledge about the IS domain requires the development of an ontology hierarchy, ranging from the top level to the bottom one. Top-level ontology (fundamental or basic) includes basic domain concepts that have already been previously defined in ontologies on this topic, an low-level ontologies describes the specific aspects of problem with more exact detailing of features. An interesting example of multilevel ontological system of cyber security is proposed in [13] This ontological framework CRATELO consists of three layers:

- top-level ontology designed on the basis of SPRAY (simplified version of DOLCE top ontology),
- Security Core Ontology (SECCO) – a set of security concepts based on DOLCE-SPRAY primitives,
- security-related middle-level ontologies of secure operations (OSCO).

In [14] authors propose a reference ontology for cybersecurity operational information. IS in this approach is considered as the preservation of information confidentiality, integrity and availability. This work takes into account a lot of ontologies in the area of cybersecurity that formalize organizing knowledge of IS risk management, vulnerabilities, attackers, security metrics, countermeasures and other relevant concepts [15]. Many middle-level ontologies have already been created in the field for reflecting the various individual aspects of IS domain. For example, researchers have developed application ontologies to identify and classify network attacks:

- an ontology for distinguishing the state of network security [16];
- an ontology of intrusion detection [17];.
- an ontology for automated classification of network attacks [18].
- an ontology of prediction of potential network attacks [19].

Other IS ontologies provide and enrich an adaptive vocabulary that can improve behavioral analysis and help stop the spread of threats. As well terms for IS ontologies can be obtained from the open sources, such as the dictionary of cyber security terms [21] and the various standards of this domain. But processing of USD with NL text requires special software and expert participation. Acquisition of knowledge from NL documents that contain semantic markup (such as semantic Wiki resources) is easier. Wiki pages correspond to domain concepts, and links between them explicitly defined in content can be used to build an ontological relations. For example, we use the "Security systems" category pages of the portal of the Great Ukrainian Encyclopedia (e-VUE) [22] and Ukrainian Wikipedia as information sources of IS terms (Fig.2).

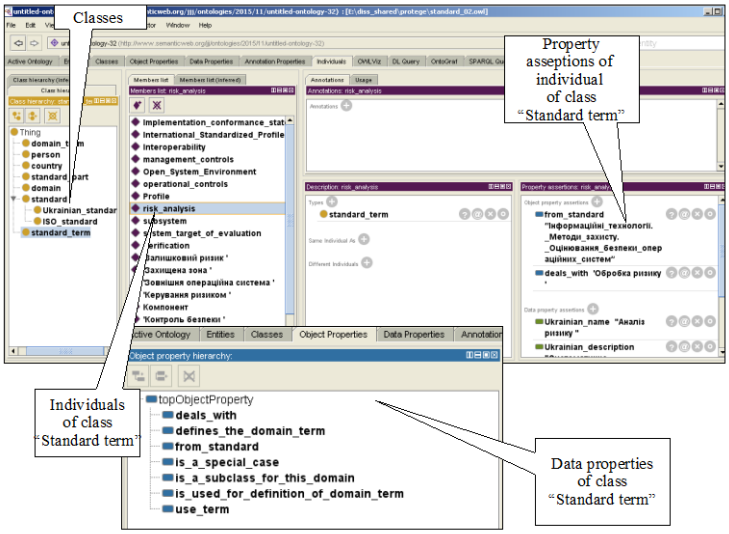


Fig. 1. Information security ontology (fragment)

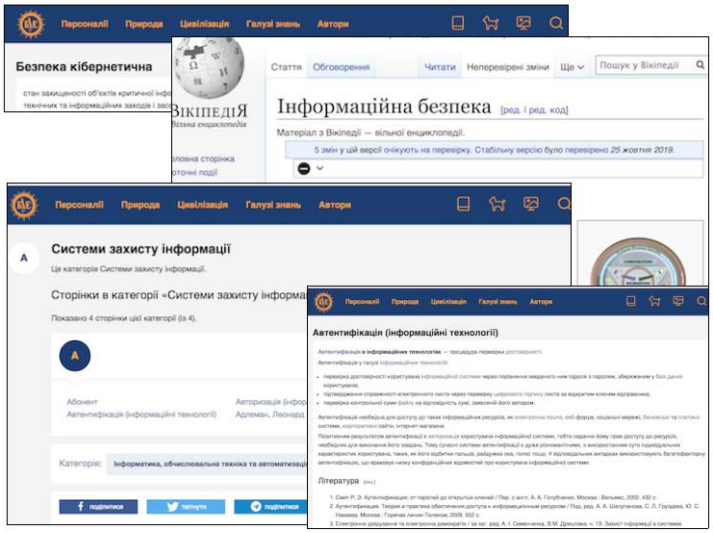


Fig. 2. Wiki pages with semantic markup on e-VUE and Wikipedia.

Wiki markup can be based on existing IS ontologies. It is advisable to create such Wiki pages for each IS standard in whole as well as for their separate subdivisions and definitions. Thus, the semantic markup of NL information resources enables the automated generation of IS ontologies.

6. Use of IS ontological models in validation of learning outcomes

The effective development of IS demands the determination of competence areas and their contents, so as to make it possible to define learning outcomes for each level and type of education [23]. Validation of

learning outcomes can be realized by ontology-based matching of results of informal learning with vacancies and task in relevant domain.

Terms and relations of domain ontology provide the semantic matching of NL texts. At the level of ontological models, it is possible to correlate learning outcomes, for example, the competencies of IS professionals who have acquired their knowledge and skills in non-formal and informal learning. These people can describe their learning outcomes in the same terms used by employers to describe the jobs and tasks. The set of competencies is defined by domain specifics.

We proposed to divide each learning result into the unique set of atomic learning results [24], and to establish the relation and hierarchy of learning outcomes and qualifications through an appropriate ontology. Atomic learning outcomes (atomic competencies) have the following properties:

- $a \in C$, where C is the set of information objects of the class "Learning Outcomes", and C_{atomic} - the set of atomic learning outcomes, $C_{atom} \subseteq C$;

- each learning outcome c is an unique non-empty set of atomic learning results $\forall c \in C \exists a_i \in C_{atomic}, i = \overline{1,n}, k = \bigcup_{i=1}^{n} a_i$;

- atomic learning outcomes are not a subset of other learning outcomes $\forall a, b \in C, a \subseteq b \Rightarrow b \notin C_{atomic}$.

In order to relate the learning results to the domain terminology we define for each $a \in C$ the terms $x_i \in X, i = \overline{1,n}, n \geq 1$ and relations $x_{i_a} = r_j(x_{i_b})$ from the domain ontology (in this case, from one of the ontologies included into hierarchy of IS ontological model). Learning outcomes without such ontological correspondences cannot used (if IS ontologies does not contain an appropriate terms for important learning outcomes then we have to expand IS ontology by appropriate terms).

Thus, the matching of individual learning outcomes with task requirements is executed on semantic level by comparison of finite sets of atomic learning outcomes defined in domain terms. Therefore the time of comparison is linear proportional to the number of atomic learning outcomes, and each atomic learning outcome is a finite unique nonempty set of domain terms.

Use of such atomic learning outcomes requires in generation of domain thesauri that are simpler then ontologies but contain all domain terms deal with some specific task (for example, some job definition).

7. Generation of IS thesauri based on domain ontologies

Thesaurus approach is widely used for analysis of domain term systems. Domain models or thesauri can be used to generate domain representations assisted by computers with the help of integrated combination of different kinds of techniques from computing, statistics and artificial intelligence [25]. Some methods of thesauri generation use domain ontologies as a source of knowledge and integrate it with current task description.

In [26] researchers propose to use a thesaurus approach for formalization of domain terminology in the field of IS. IS thesaurus displays a wide range of essential properties, attributes and relationships that are inherent in this specific type of security. An other example of IS thesaurus is described in [27]. This thesaurus integrates the concepts of cybersecurity, data security, application security, network security, Web security, and critical information infrastructure security.

Application security is determined by the application software as well as by the software resources and processes involved in their lifecycle. Network security is related to the design, implementation and use of networks within the organization, between organizations, between organizations and users. Web security refers to Web services and related information and communication systems and networks. Security of critical information infrastructure is characterized by protection against relevant threats, in particular information security.

Researchers [28] distinguish three groups of IS terms:

1. Terms defining the scientific basis of IS. This group includes terms used in many fields of knowledge that are unambiguous, semantically unified, and stylistically neutral. Examples: information, communication, conflict, influence, threat, danger, security, system. The concepts in this group meet the requirements of uniqueness and stability, i.e. they are uniquely applied in one area of knowledge and retain content in any other. This group of terms corresponds with the top-level ontologies.

2. Terms defining the subject basis of IS. This group of terms refers to concepts and their relation to other concepts within information security as a specific area of knowledge. This group of terms corresponds with the middle-level ontologies of IS.

3. Terms defining the nature of IS activities. This group covers terms denoting objects, phenomena, processes, their properties and relationships that are characteristic of this field. This group of terms corresponds with the low-level ontologies of special subspheres of IS.

IS thesaurus oriented on processing NL texts deal with learning outcomes in IS has to contain terms from all categories if they are used into the task description (here task is a NL description of work or problem that employer proposes to specialists with appropriate competencies).

We define task thesaurus as a special case of domain ontology oriented on analyses of NL texts. Thesaurus contains only ontological terms (classes and instances) but does not describe the semantics of relations between them. It can be automatically generated on base of the domain ontology and the NL description of the problem [29].

A simple task thesaurus is a task thesaurus based on the terms of a single ontology. A composite task thesaurus is a task thesaurus that is based on the terms of two or more domain ontologies. It should be noted that IS domain is very dynamic, and therefore there is a problem in keeping the terminological base of the current state of research works and their integration with the modern grammar of the Ukrainian language.

8. An algorithm for constructing of task thesaurus

A simple task thesaurus is constructed according to the user-selected domain ontology and task – the description of the current problem (Fig. 3). This algorithm is described in detail in [30].

The task description can be represented via NL text that contains elements related to the ontology elements, or through conditions for domain terms relevant to the task.

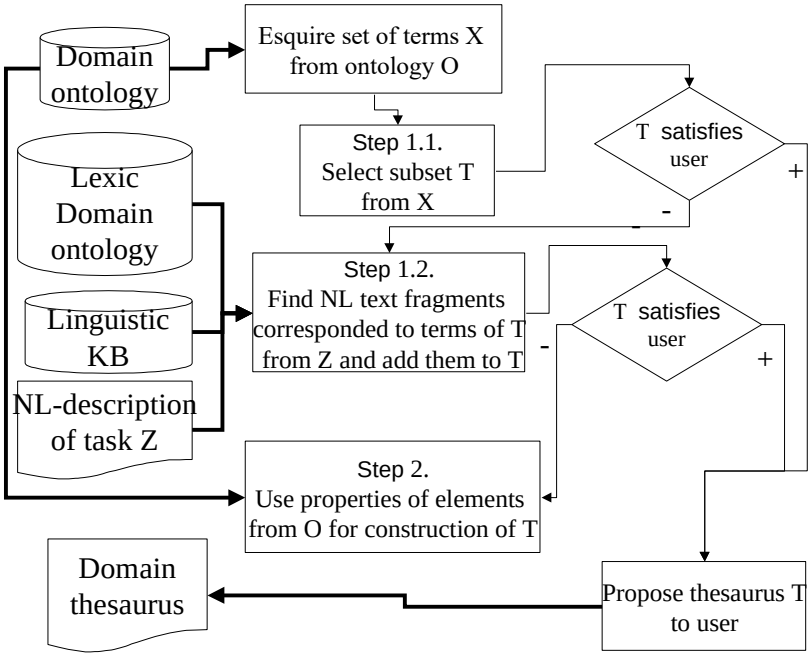


Fig. 3. Main steps for constructing of task thesaurus

Thesaurus constructing consists from two main steps:

Step 1. Automated generation of a simple task thesaurus by task description;

Step 2. Expansion of a simple thesaurus by other ontology concepts by the set of conditions that use elements of the O ontology different from instances and classes.

We divide step 1 into two substeps. On Step 1.1 user explicitly and manually selects task-pertinent terms from the automatically generated list of classes and instance X. In the simplest cases, the construction of the thesaurus may be completed in this step, but it requires more efforts from the user.

Stage 1.2 uses a variety of methods for processing a natural-language task description (linguistic analysis, statistical processing, semantic markup analysis) that allow to detect NL fragments related to terms from O.

Those ontological terms that correspond to some fragments of task description are added to the simple task thesaurus.

Linguistic knowledge bases (KB) can be used to construct a thesaurus. We can use specific domain-oriented linguistic KBs if a large amount of lexical information has already accumulated. Such information is not universal and depends either from domain and natural language used in task definition.

Therefore we cannot use Text Mining systems – they often are oriented on processing only English texts. We apply direct updating of the domain lexical ontology by users and export linguistic knowledge from relevant vocabularies and knowledge bases, as well as from semantically marked Ukrainian texts.

In many cases, information about other elements of the ontology is appropriate in thesaurus constructing. This information allows taking into account the properties of individual terms and their relations with other terms.

Such actions are applied on step 2 for refining the initially formed thesaurus in accordance with explicitly formulated user conditions. These conditions depend from the specific nature of the task, but are not derived from its description. Such conditions can be considered as a set of meta-rules to describe the information the user is seeking.

Stage 2 can be represented as a function that transforms the O ontology into simple task thesaurus according to proposed meta-rules deal with the finite set of conditions that the user formulates for classes and instances of domain ontology.

Complex task thesauri are generated from the built earlier task thesauri (simple or complex) with the help of set theory operations such as sum of sets, intersection of sets etc.

9. Use of task thesauri for validation of learning outcomes

Task thesauri can be used in various intelligent tasks. For example, such thesauruses can be very useful to validate learning outcomes that are represented by NL unstructured text. The main requirement of such approach is an ontological model of learning domain. As we describe above, complex and dynamic IS domain has a lot of ontological representations of its various aspects. Therefore we can apply this approach to problem of learning outcomes validation for IS.

Use of task thesauri for validation of learning outcomes contain such main steps:

Step 1. A set of IS ontologies and a NL job description from employers (vacancy description, set of requirements for a specific position or qualification etc.) are input to the algorithm for thesaurus constructing. Only the concepts used in this description are retrieved from the ontological model of IS.

Step 2. We retrieve in the NL descriptions of the learning outcomes not all IS domain concepts but only those that are included in the task thesaurus. This greatly speeds up the comparison because ontological model of IS in general contains a great number of concepts and their NL representations.

Descriptions are sufficient for to matching procedure if they contain all the necessary concepts of ontology. It is important that thesauri allow us to group synonyms or words represented by different NL, thus matching all semantically similar terms.

Step 3. If any learning outcome completely matches with job description then we try to find descriptions of the learning outcomes that are most similar to the task description.

10. Acknowledgement

We use in this work scientific results received from project of Information programs of the NASU «Development of dataware, functional support and software of electronic version of encyclopedic editions of Ukraine» (2019) in Institute of Software Systems of NASU and from the work prepared in accordance with the thematic plan of scientific researches of the Dmytro Motornyi Tavria State Agrotechnological University (project of applied research at the expense of the state budget expenditures «Theoretical substantiation and development of information system of semantic identification, documentation and processing of results of non-formal and informal learning»).

Conclusions

The information security industry is a dynamic field that is rapidly changing and improving its methods and tools. Therefore, cybersecurity-related specialists require a steady flow of knowledge, both from internal and external sources. This enhances the importance of non-formal and informal learning in this field for the effective accomplishment of various tasks.

The complex hierarchical structure of domain knowledge necessitate the use of ontological analysis to the processing of data related to IS. The unstructured and large-volume information causes the application of Big Data analytic techniques in preliminary processing of data sources used for generation of IS ontological model.

Such ontological model of domain contains its conceptualization, integrates heterogeneous data structures and knowledge from different information security subsystems, as well as national and international standards. Ontologies provide formalization of the relations between data elements.

The authors propose to use task thesauri based on appropriate ontologies as a tool for matching the informalized results of learning outcomes of IS specialists with task descriptions. These thesauri provide a terminological basis for the integration of IS ontologies of different levels that represent the knowledge structure of this domain. Use of thesauri reduces the computational complexity of the comparing problem of heterogeneous knowledge structures.

References

1. Parviainen, P., Tihinen, M., Kääriäinen, J., Teppola, S.: Tackling the digitalization challenge: How to benefit from digitization in practice. In: International Journal of Information Systems and Project Management, 5 (1), pp.63-77.(2017).
2. Grimes S.: Unstructured Data and the 80 Percent Rule, Clarabridge, Bridgepoints, (2008), http://breakthroughanalysis.com/2008/08/01 / unstructured-data-and-the -80-percent-rule /.
3. Savas O., Deng J.: Big Data Analytics in Cybersecurity In: CRC Press, .- 353p. (2018).
4. Zhang, Y., Ren, J., Liu, J., Xu, C., Guo, H., Liu, Y.: A survey on emerging computing paradigms for big data. In: Chinese Journal of Electronics, 26 (1), pp.1-12, (2017).
5. Demchenko, Y., De Laat, C., Membrey, P. : Defining architecture components of the Big Data Ecosystem. In: 2014 International Conference on Collaboration Technologies and Systems (CTS), pp. 104-112, (2014).
6. Unstructured_data. - https://en.wikipedia.org/ wiki / Unstructured_data.
7. Grimes S.: A Brief History of Text Analytics. B Eye Network, (2016), http://www.b-eye-network.com/view/6311.
8. Buneman P., Davidson S., Fernandez M., Suciu D. Adding structure to unstructured data. In: International Conference on Database Theory, pp. 336-350, (1997).
9. Feldman R., Sanger, J.: The text mining handbook: advanced approaches in analyzing unstructured data. In: Cambridge university press, (2007), https://wtlab.um.ac.ir/images/e-library/text_mining/The%20Text%20Mining%20HandBook.pdf.
10. Rogushina J.: Means and methods of unstructured data analysis. In: Problems in Programming, N 1, pp.57-77, (2019), http://pp.isofts.kiev.ua/ojs1/article/view/348/346. (in Ukrainian).
11. Obrst, L., Chase, P., Markeloff, R.: Developing an Ontology of the Cyber Security Domain. In: STIDS, pp. 49-56, (2012).
12. Syed, Z., Padia, A., Finin, T., Mathews, L., Joshi, A.: UCO: A unified cybersecurity ontology. In: The Workshops at the Thirtieth AAAI Conference on Artificial Intelligence. (2016), https://www.aaai.org/ocs/index.php/WS/AAAIW16/paper/download/12574/12365.
13. Oltramari, A., Cranor, L. F., Walls, R. J., McDaniel, P. D.: Building an Ontology of Cyber Security. In: STIDS, pp. 54-61, (2014), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.664.3593&rep=rep1&type=pdf.
14. Takahashi, T., & Kadobayashi, Y.: Reference ontology for cybersecurity operational information. In: The Computer Journal, 58(10), pp.2297-2312, (2015), https://ieeexplore.ieee.org/stamp/ stamp.jsp?arnumber=8205615
15. Wang, J.A. and Guo, M.: OVM: An Ontology for Vulnerability Management. In: Proc. 5th Annual Workshop on Cyber Security and Information Intelligence Research, pp. 1-4, (2009).
16. Bhandari, P. and Guiral, M.S.: Ontology Based Approach for Perception of Network Security State. In: Recent Advances in Engineering and Computational Sciences (RAECS), pp.1-6, (2014).
17. Ye, D., Bai, Q., Zhang, M., Ye, Z.: P2P distributed intrusion detections by using mobile agents. In: Seventh IEEE/ACIS International Conference on Computer and Information Science (ICIS 08), pp.259-265, (2008).
18. van Heerden, R., Leenen, L., Irwin, B.: Automated classification of computer network attacks. In: International Conference on Adaptive Science and Technology (ICAST 2013), pp.157-163, (2013).
19. Salahi, A. Ansarinia, M.: Predicting Network Attacks Using Ontology-Driven Inference, (2011), http://arxiv.org/ftp/arxiv/ papers/1304/1304.0913.pdf.
20. Motik, B., Patel-Schneider, P. F., Parsia, B., Bock, C., Fokoue, A., Haase, P., Smith, M.: OWL 2 web ontology language: Structural specification and functional-style syntax. In: W3C recommendation, (2009), www.w3.org/2007/OWL/draft/ED-owl2-syntax-20090914/
21. Gladun A.Y., Puchkov O.O., Subach I.Y., Khala K.O.: English-Ukrainian Dictionary of Information Technology and Cybersecurity Terms. In: Igor Sikorsky KPI, P. 376, (2018).
22. Rogushina J.: Processing of Wiki Resource Semantics on the Base of Ontological Analysis. In: Proc.of OSTIS-2018, pp.159-162, (2018), https://libeldoc.bsuir.by/bitstream/123456789/30389/1/Rogushina_ Processing.PDF.
23. Lehto, M.: Cyber security competencies: cyber security education and research in Finnish universities. In: ECCWS2015-Proceedings of the 14th European Conference on Cyber Warfare & Security: ECCWS 2015, pp. 179-88, (2015).
24. Rogushina J., Priyma S.: Use of competence ontological model for matching qualifications. In: Chemistry: Bulgarian Journal of Science Education, Volume 26, Number 2, pp.216- 228, (2017).
25. Lloréns, J., Velasco, M., de Amescua, A., Moreiro, J. A., Martínez, V.: Automatic generation of domain representations using thesaurus structures. In: Journal of the American Society for Information Science and Technology, 55(10), pp.846-858, (2004).
26. Sachuk Yu.E.: Training of cybersecurity and information security professionals: thesaurus and ontology. In: Problems of Engineering and Pedagogical Education, N 59, pp.35-40, (2018).

27. Kalichensky A.: The Concept of Creating the National Information and Communication Infrastructure of Ukraine. In: Regional Forum M SE, (2012), https: //www.itu.int/ITU-D/tech/events/2012/ Spectrum_ CIS_Kiev_Sept12/Presentatio ns / Session2 / A_Kalichensky _a.pdf.
28. Diorditsa I.V.: The presentation of the terminology of cybersecurity policy in the texts of regulatory acts of Ukraine. In: Scientific Bulletin of the International Humanities University. Jurisprudence Series, N 29 (1), pp. 64-67, (2017).
29. Gladun A., Rogushina J.: Use of Semantic Web Technologies and Multilinguistic Thesauri for Knowledge-Based Access to Biomedical Resources. In: International Journal of Intelligent Systems and Applications, №1, P.11-20, (2012), http://www.mecs-press.org/ ijisa/ijisa-v4-n1/IJISA-V4-N1-2.pdf.
30. Rogushina Yu.V. Use of Thesauri for Search of Complex Information Objects in the Web on Base of Ontologies. In: Problems in Programming, No. 4, pp.11-27. (2019) (in Ukrainian).

МЕТОДИ І ЗАСОБИ ІНФОРМАЦІЙНО-АНАЛІТИЧНОЇ ПІДТРИМКИ ПРОТИДІЇ ГІБРИДНИМ ЗАГРОЗАМ ДЕРЖАВИ

Шнурко-Табакова Е.В.[1], Ланде Д.В. [1,2]

[1]ГО «Рада інформбезепеки та кіберзахисту»,

[2]ІПРІ НАН України

На цей час в Україні до протидії гібридним загрозам державі залучаються всі шари держави і суспільства, зокрема, силові структури держави, інші міністерства та відомства, недержавні організації, бізнес, громадянські об'єднання. З огляду на те, що ворог поряд із економічними, енергетичними, гуманітарними та іншими важелями, проти нашої країни широко застосовується інформаційна зброя, проводяться інформаційні операції, одним з першочергових завдань стає створення високоефективної системи інформаційно-аналітичної протидії гібридним загрозам держави.

Саме застосування таких підходів може забезпечити швидке оперативне реагування на загрози, виявлення інформаційних кампаній, атак, операцій [1], зокрема, виявлення мереж ворожих інформаційних ботів, прогнозування розвитку подій, явищ, процесів тощо.

Зокрема, потребує інформаційно-аналітичної підтримки протидія інформаційним операціям в мережі Інтернет, які умовно поділяються на «пропагандистські» (інформаційні впливи мають переважно пропагандистський характер), «дезінформаційні» (основною метою є дезінформація шляхом створення «фейків»), «маніпулятивні» (маніпулювання, модифікація установок людей), оборонні («контроперації» – нейтралізація інформаційного впливу супротивника).

Для рішення завдань інформаційно-аналітичної підтримки протидії загрозам такого роду мають застосовуватися найсучасніші методи і засоби аналітичної роботи, що базуються на використанні таких сучасних концепцій, як Data Science (наука про дані), Big Data (великі дані), Text/Data Mining (глибинний аналіз текстів/даних), методи нелінійного (кореляційного, фрактального) аналізу, Complex Networks (складні мережі), OSINT (розвідка за відкритими джерелами) тощо.

Для здійснення інформаційно-аналітичної підтримки мають застосовуватися методи і засоби, що дозволяють:

- виявляти кількісну динаміку, притаманну процесу чи явищу, наприклад, кількість подій або повідомлень щодо події в одиницю часу;
- визначати критичні, порогові точки, що відповідають кількісної динаміці явища;
- визначати прояви подій, процесів, об'єктів в критичних точках, наприклад, виявлення основних сюжетів повідомлень щодо обраного процесу або явища;
- ранжирування цих проявів і дослідження динаміки їх розвитку до та після певних критичних точок;
- здійснення статистичного, кореляційного і фрактального аналізу загальної динаміки і динаміки окремих проявів, на основі яких має здійснюватися прогнозування розвитку події, процесу і окремих його проявів.

Для дослідження взаємозв'язку реальних подій і публікацій про них в мережі Інтернет, зокрема, авторами використовується інформаційна OSINT-системах [2] InfoStream (http://online.infostream.ua), що забезпечує інтеграцію і моніторинг мережевих інформаційних ресурсів, а також аналітична система Attack Index (http://attackindex.com), що дозволяє визначати наявність і рівень інформаційних операцій на базі аналізу відкритих джерел. Використання статистичних методів аналізу інформаційних потоків, методів нелінійного, зокрема, фрактального аналізу, дозволяє прогнозувати розвиток подій або керованих інформаційних процесів. Шляхом застосування цих систем, даних, що отримуються від них, здійснюється як прогнозування реальних процесів, так і інтеграція із системами прийняття рішень.

Окремим питанням є методологія декомпозиції гібридних загроз до семантичного ядра, що має складати основу ключових запитів до інформаційно-аналітичних систем. Сучасні виклики вимагають створення системи рейтенгування загроз та регламентів відстеження ситуації в інформаційному просторі з безперервним супроводженням та реагування.

Сьогодні складні завдання інформаційно-аналітичної протидії, з одного боку, стимулюють розвиток систем керування знаннями, глибинного аналізу даних і текстів, а з іншого – найбільш розвинені із цих систем у явному вигляді містять адаптовані й готові до використання аналітичні блоки.

Для прийняття ґрунтовних рішень у галузі національної безпеки держави, зокрема щодо протидії гібридним загрозам, необхідне використання комплексних інформаційно-аналітичних систем, які дозволяють збирати, обробляти та узагальнювати інформацію, отриману з різних джерел із застосуванням різних технологічних рішень. Тому вже є широкий вибір засобів автоматизації інформаційно-аналітичної діяльності. Причому рівні функціональності таких систем можуть бути дуже різноманітним – від простих засобів контент-моніторингу та інформаційно-пошукових модулів, необхідних на етапі становлення аналітичних систем, до дорогих ресурсномістких систем керування знаннями та глибинного аналізу даних і текстів.

Зокрема, при аналізі тематичних інформаційних потоків із різних мережевих джерел, що аналізуються в системі контент-моніторингу, вирішується проблема формування і дослідження зв’язків інформаційних джерел, які розповсюджують маніпулятивну інформацію. При цьому підставою для можливого зв'язку між двома джерелами може служити той факт, що вони часто публікують документи, що збігаються або близькі за темою. За допомогою такої мережі можна визначити, які з інформаційних джерел з даної тематики є основними, найбільш впливовими, які схильні до певних інформаційних впливів. Підхід базується на тому факті, що джерела інформації ранжовані за обсягами публікацій виходячи з досвіду спостереження за ними протягом тривалого часу, тобто їм вже приписані деякі вагові значення.

Таким чином встановлюється зв'язок джерела інформації з іншим, більш рейтинговим, якщо він існує і опублікував інформацію раніше. Побудована таким чином мережу (Рис. 1) відображає зв'язок джерел по заданій тематиці, дозволяє визначати лідерів серед них, робити припущення щодо першоджерел інформації. Наприклад, на Рис. 5 показана мережа зв'язків джерел інформації із соціальних мереж, що відображають тематику «Формула Штанмайера» у жовтні-листопаді 2019 р за даними системи контент-моніторингу.

Сучасні системи інформаційно-аналітичної протидії гібридним загрозам забезпечують вирішення цілого комплексу проблем, серед яких збір інформації про об'єкти, визначення зв'язків об'єктів, виявлення тенденцій, прогнозування. Не слід вважати, що такі системи є цілком автоматичними, навпаки, у таких системах широко використовується людський досвід, знання експертів. Функціональні можливості таких систем мають виконувати діагностику, прогнозування розвитку ситуацій. Поряд з цим, нині очевидно, що реальний прорив у сфері і інформаційно-аналітичної роботи можливий лише в результаті агрегування усіх наведених напрямків.

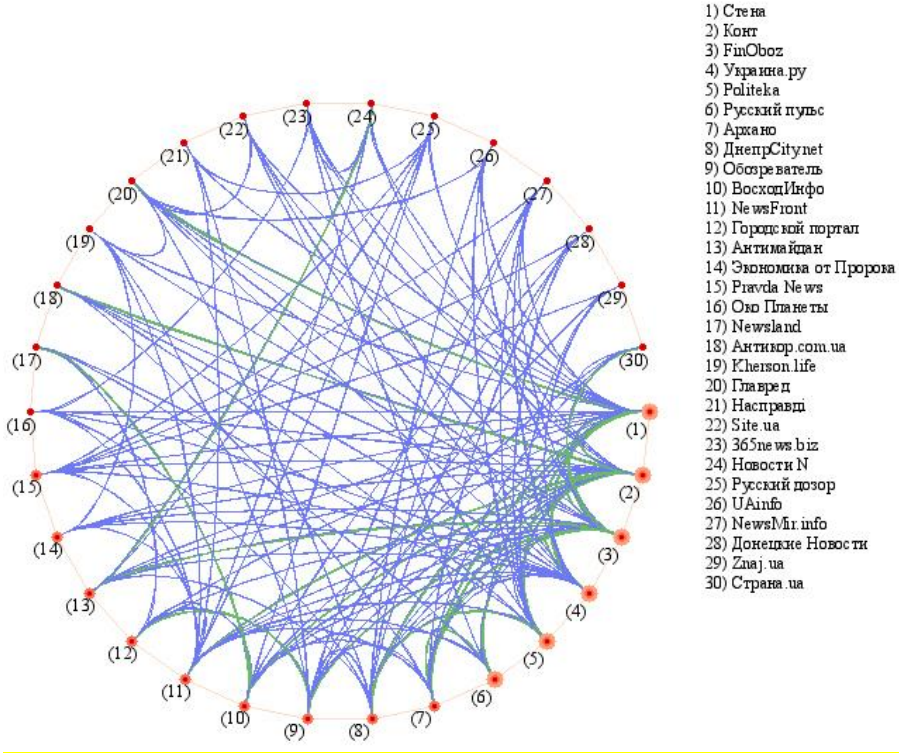


Рисунок 1 – Мережа джерел інформації за тематикою «Формула Штанмайера»

Література

1. Information Operations Recognition. From Nonlinear Analysis to Decision-Making / A. Dodonov, D. Lande, V. Tsyganok, O. Andriichuk, S. Kadenko, A. Graivoronskaya – LAP Lambert Academic Publishing, 2019. - 292 p.
2. Dmytro Lande, Ellina Shnurko-Tabakova. OSINT as a part of cyber defense system // Theoretical and Applied Cybersecurity, 2019. - N. 1. - pp. 103-108.
3. Ландэ Д.В., Снарский А.А. Применение графов горизонтальной видимости в информационной аналитике // CEUR Workshop Proceedings. Selected Papers of the XVII International Scientific and Practical Conference on Information Technologies and Security (ITS 2017) . – С. 86-91.

ПОКАЗНИК РЕЛАКСАЦІЇ В СКЛАДНИХ МЕРЕЖАХ

А.О. Снарський[1,2][0000-0002-4468-4542], Д.В. Ланде[1,2][0000-0003-3945-1178], О.О. Дмитренко [1][0000-0001-8501-5313]

[1]Інститут проблем реєстрації інформації НАН України, Київ, Україна
[2] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
dwlande@gmail.com asnarskii@gmail.com dmytrenko.o@gmail.com

Анотація. В роботі досліджуються нові характеристики вузлів мережевих структур – показник релаксації мережі та індивідуальний показник релаксації вузла. Для отримання показника релаксації застосовуються так звані уповільнені ітераційні алгоритми для HITS та PageRank. Встановлено, що на відновлення традиційних показників мережі, після збурення окремих вузлів, впливає її топологія. Як приклад, показник релаксації мережі та індивідуальний показник релаксації вузла були використані для дослідження структури мережі термінів, побудованої для предметної області “Кібербезпека”. Завдяки застосуванню показників релаксації вдалося визначити найбільш важливі змістовні компоненти мережі та ранжувати їх за введеними показниками. Отримане ранжування у порівнянні з ранжуванням за показникамиHITSта PageRank показує унікальністьзапропонованих показника релаксації мережі та індивідуального показника релаксації вузла.

Ключові слова:складна мережа, показник релаксації,індивідуальний показник релаксації,HITS, PageRank, предметна область, мережа термінів.

Вступ

Складні мережі широко поширені у природі. Більшість об'єктів у природі і суспільстві мають бінарні зв'язки, які можна представити у вигляді мережі. Топологічні властивості мереж, що розглядаються абстрактно від їх фізичної природи, але істотно визначають функціонування мереж, становлять предмет дослідження комплексних мереж. У багатьох прикладних галузях та сферах науки і техніки задачі аналізу топології мережі та дослідження особливостей її вузлівмають досить важливе значення.

Вивченням характеристик складних мереж займається область дискретної математики, що має назву теорія складних мереж (від англ. – ComplexNetworks) [1, 2], потужний математичний апарат якої дозволяє досліджувати, зокрема, поведінку окремих об’єктів такихмереж.

Наукові роботи вітчизняних та зарубіжних вчених В. М. Глушкова, В. М. Томашевського, П. Ердоша, А. Рені, М. Е. Дж. Ньюмана, Р. Альберт, А.-Л. Барабаші, Д. Дж. Ваттса, С. Г. Строґаца та інших дослідників внесли суттєвий вклад у розвиток теоретичних основ і практичних рішень для створення методів і засобів дослідження та проектування складних мереж. Пропонуються також нові методи до вирішення обчислювально складних задач, характерних для сучасних мережевих структур [1-3]. Незважаючи на наявність уже існуючих традиційних підходів, дослідження статистичних властивостей, які характеризують поведінку мереж; створення моделі мереж; прогнозування поведінки мереж при зміні структурних властивостей абопід час різних зовнішніх впливах – актуальні завдання теорії складних мереж.

У прикладних дослідженнях зазвичай застосовують типові для мережевого аналізу характеристики вузлів мережі, які описують її певну визначену властивість, найважливішими серед яких на цей час вважають степінь вузла та показники, що відповідають алгоритмам HITS та PageRank. Поруч із вже традиційними показниками мережі таких як HITS та PageRank в даній роботі були запропоновані та досліджені такі нові характеристики як: показник релаксації мережі та індивідуальний показник релаксації вузла.

Метою даної роботи є ввести нові характеристики вузлів складної мережі, визначити їх “фізичний зміст” і показати унікальність серед інших характеристик, а також навести приклади застосування, зокрема у комп’ютерній лінгвістиці.

1. Ітераційні алгоритми HITS та PageRank

1.1. HITS

Алгоритм ранжування *HITS* (HyperlinkInducedTopicSearch), що був запропонований та розроблений в 1998 році Дж. Клейнбергом (J. M. Kleinberg) [4] забезпечує вибір із інформаційного масиву кращих «авторів» (першоджерел, на які посилаються інші документи) та «посередників» (документів, які посилаються на ці першоджерела). Документ буде вважатися хорошим «автором», якщо на нього посилаються хороші «посередники». В свою чергу, хорошими «посередниками» вважаються ті, які містять посилання на цінні першоджерела.

Для кожного документа j обчислюється його важливість як «автора» $a(j)$ і як «посередника» $h(j)$ відповідно до формул:

$$a(j) = \sum_{i \to j} h(i)\,,\ h(j) = \sum_{j \to i} a(i) \qquad (1)$$

В ітераційному представлені наведений вище алгоритм можна записати наступним чином. Нехай **E**–множина всіх направлених ребер у графі, де e_{ij} –направлене ребро з вершини *i*у вершину *j*. Також задані початкові значення важливості документа як «автора» $a_i^{(0)}$ та «посередника» $h_i^{(0)}$. Далі ітераційно обчислюються значення:

$$a_i^{(k)} = \sum_{j:e_{ij} \in E} h_j^{(k-1)},\quad h_i^{(k)} = \sum_{j:e_{ij} \in E} a_j^{(k-1)},\quad k = 1,2,3,... \qquad (2)$$

В матричному вигляді ці рівняння можна записати за допомогою матриці суміжності *L* направленого графа.

$$\mathbf{L} = \begin{cases} 1, & \text{якщо існує ребро з вершини } i \text{ у вершину } j, \\ 0, & \text{в іншому випадку.} \end{cases} \qquad (3)$$

Отримуємо:

$$a^{(k)} = \mathbf{L}^T h^{(k-1)},\quad h^{(k)} = \mathbf{L}a^{(k)}, \qquad (4)$$

де $a^{(k)}$ та $h^{(k)}$–вектори значень важливості як «автора» та «посередника» на кожному ітераційному кроці.

2.2. PageRank

PageRank (Пейдж-ранк) – один з алгоритмів оцінки важливості та ранжирування веб-сторінок за гіперпосиланнями, був створений в Стенфордському університеті ЛарріПейджем і Сергієм Бріном в 1996 році в рамках науково-дослідного проекту про новий вид інформаційно-пошукової системи [5] й вперше використаний в Google.

Для складної мережі, що задається матрицею суміжності $\hat{\mathbf{H}}$, обчислюється $\hat{\mathbf{G}}$:

$$\hat{\mathbf{G}} = \alpha\hat{\mathbf{H}} + [\alpha\bar{\mathbf{a}} + (1-\alpha)\bar{\mathbf{e}}]\frac{1}{n}\bar{\mathbf{e}}^T, \qquad (4)$$

де $a_i = 1$якщо з *i*-го вузла не виходить жодна зв'язок,

та $a_i = 1$– в протилежному випадку;

n – кількість вузлів в мережі;

α – коефіцієнт загасання (зазвичай $\alpha = 0.85$).

Ліві власні значення $\hat{\mathbf{G}}$ і є PageRank мережі.

Незважаючи на відмінності HITS і PageRank, в цих алгоритмах спільним є те, що “авторитетність” (вага) вузла як «автора» залежить від ваги інших вузлів, а “авторитетність” «посередника» залежить від того, наскільки “авторитетними” є вузли, на які він посилається [6, 7].

3. Показник релаксації мережі

В даній роботі пропонуються новіхарактеристикивузлів складної мережі – показник релаксації мережі та індивідуальний показник релаксації вузла, які дозвлдятьранжувати відповідні вузли складної мережі.

Показник релаксації є аналогом часу релаксації Максвелла [8], яка відіграє важливу роль у фізиці твердого тіла.

Час релаксації τ – характерний час, за який "розсмоктується" електричний заряд у середовищі з питомою електричною провідністю σ та діелектричною проникністю ε. В однорідному нескінченному середовищі неоднорідність розподілу електричного заряду нестійка (систему можна вивести з рівноважного стану), з часом заряд "розсмоктується", розподіляється рівномірно в середовищі та уходить на скінченність. Час релаксації Максвелла – τ і є характерним часом переходу середовища в рівноважний стан, де зменшення щільності заряду ρ з часом t має вигляд $\rho(t) \sim e^{-t/\tau}$, де $\tau = \varepsilon/\sigma$.

По аналогії, введемо в складній мережі час релаксації k-го вузла – τ_k. Спочатку визначимо рівноважний стан складної мережі як набір значень вузлів s_k^0 (у векторному виді – $\vec{\mathbf{s}}^0$), які визначаються за певним правилом, наприклад за їх значенням HITS чи PageRank, чи будь-яким іншим [7].

Обчислення $\vec{\mathbf{s}}^0$ відповідно до вибраного правила (ітераційного алгоритму) завжди може бути записане в ітераційному виді:

$$\vec{\mathbf{s}}(n+1) = \vec{\mathbf{s}}(n) + \hat{\mathbf{L}}\vec{\mathbf{s}}(n), \quad n = 0,1,... \qquad (5)$$

де номери компонент вектора $\vec{\mathbf{s}}$ – номери вузлів, $\hat{\mathbf{L}}$ – оператор відповідного ітераційного алгоритму (в нашій роботі розглянуті ітераційні алгоритми, що відповідають HITSта PageRank), $\vec{\mathbf{s}}(0)$ – задані початкові значення вузлів.

$$\vec{\mathbf{s}}^0 = \lim_{n\to\infty} \vec{s}(n), \quad n = 0,1,... \qquad (6)$$

і, звичайно,

$$\hat{\mathbf{L}}\vec{\mathbf{s}}^0 = 0 \ . \qquad (7)$$

Візьмемо тепер величину початкових значень вузлів $\vec{\mathbf{s}}(0)$ у вигляді розв'язку – $\vec{\mathbf{s}}^0$ (будемо вважати ці значення рівноважними) й відхилимо значення, наприклад, m-го вузла:

$$\vec{\mathbf{s}}_i(0) = \vec{\mathbf{s}}_i^0 + \alpha\delta_{im}\vec{\mathbf{s}}_i^0, \qquad (8)$$

де α – величина відхилення (збурення) m-ї компоненти, δ_{ik} – символ Кронекера.

У векторному вигляді – ми відхиляємо від рівноважного стану одну із компонент (проекцій) вектора $\vec{\mathbf{s}}^0$.

Відхилення вектора $\vec{\mathbf{s}}^0$, за рахунок зсуву компоненти, виводить систему з рівноваги.

Тепер (1) для n=0 має вигляд:

$$s_i = s_i(0) + \sum_k L_{ik} s_i(0), \qquad (9)$$

що з урахуванням (4) дає:

$$s_i(1) = s_i(0) + \sum_k L_{ik} s_i^{\,0} + \alpha \sum_k L_{ik}\delta_{km} s_k^0 = s_i^{\,0} + \alpha q_i^{(m)}, \qquad (10)$$

де вектор $q_i^{(m)} = \sum_k L_{ik}\delta_{km} s_k^0$, для кращого наочного сприйняття, запишемо у вигляді

$$q_i^{(m)} = \begin{pmatrix} L_{11} & L_{12} & \dots & L_{1m} & \dots & L_{1N} \\ L_{21} & \dots & \dots & L_{2m} & \dots & L_{2N} \\ \vdots & & & \vdots & & \\ L_{m1} & & & \vdots & & \\ \vdots & & & \vdots & & \\ L_{N1} & & & L_{Nm} & & L_{NN} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ s_m^0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} L_{1m} \\ L_{2m} \\ \vdots \\ \vdots \\ \vdots \\ L_{1m} \end{pmatrix} s_m^0. \qquad (11)$$

Для початкової умови $s_i(1)$ (6) $s_i(n)$ при збільшенні nвідповідно до (2) збігається до рівноважного розв'язку $\vec{\mathbf{s}}^0$.

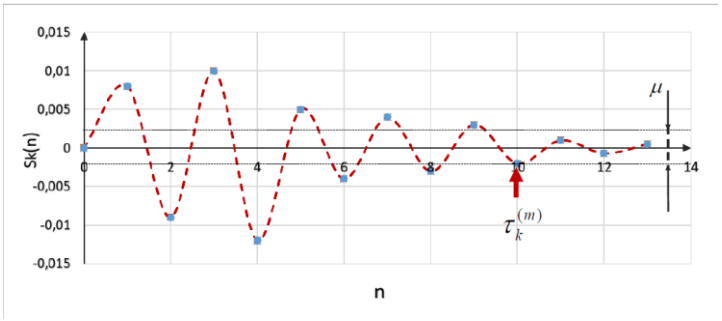


Рис. 1. Схематичне зображення збіжності *k*-го вузла

Вибравши випадок $k \neq m$, початкові значення $s_k(0) = s_k^0$. Починаючи з деякого $n \geq 10$ значення $s_k(n \geq 10)$ стає меншим μ – наперед заданого значення, яке визначає точність збіжності (рис. 1).
Те значення $\tau_k^{(m)}$, при якому для *k*-го вузла виконується умова (при заданому значенні μ)

$$\left| s_k(n \geq \tau_k^{(m)}) \right| \leq \mu, \tag{10}$$

і є показником релаксації *k*-го вузла у випадку збурення *m*-го вузла.
Загалом нас буде цікавити показник релаксації мережі [9] для *m*-го вузла $\max_k(\tau_k^{(m)})$ – найбільше значення $\tau_k^{(m)}$ серед $\forall_k$ при збуренні *m*-го вузла.

Також у роботі паралельно досліджується індивідуальний показник релаксації $\tau_m^{(m)}$, тобто показник релаксації вузла, який і був виведений зі стану рівноваги. В цьому випадку далі будемо користуватись записом τ_m, опустивши верхній символ у $\tau_m^{(m)}$.

4. Дослідження показника релаксації для мережі термінів

Як приклад, показник релаксації мережі та індивідуальний показник релаксації вузла були використані для дослідження структури мережі термінів, побудованої для предметної області "Кібербезпека" (рис. 2).

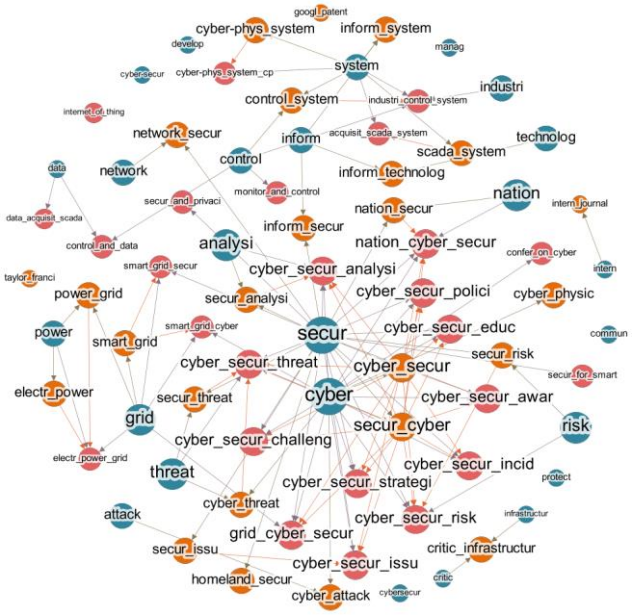


Рис. 2. Мережа термінів, що представляє предметну область "Кібербезпека"

У таблицях 1, 2 представлені топ-20 значень показника релаксаціїмережі та індивідуального показника релаксаціїдля вузлів мережі, отриманих для уповільненого алгоритму HITS та PageRank (коефіцієнт уповільнення рівний 0.9) відповідно (див. Додаток А). Терміни у таблицях відсортовані за спаданням значень показника релаксації мережі. Значення μ було підібрано таким чином, щоб значення показника релаксації мережі для кожного вузла відрізнялись одне від одного якомога

більше: для уповільненого алгоритму HITS було обрано значення 0.001, а для уповільненого алгоритму PageRank – 0.00001.

Таблиця 1.Топ-31 вузол мережі та їх показник релаксації для уповільненого та звичайного алгоритму HITS

Термін	Показник релаксації мережі для звичайного HITS	Показник релаксації мережі для уповільненого HITS	Індивідуальний показник релаксації вузла	HITS
secur	8	107	107	0.0869
cyber	8	104	102	0.0765
cyber_secur	7	99	94	0.0725
secur_cyber	7	99	94	0.0725
grid_cyber_secur	7	94	77	0.0373
cyber_secur_analysi	7	93	76	0.039
cyber_secur_awar	7	93	75	0.039
cyber_secur_challeng	7	93	75	0.039
cyber_secur_educ	7	93	75	0.036
cyber_secur_incid	7	93	75	0.036
cyber_secur_polici	7	93	75	0.036
cyber_secur_risk	7	93	76	0.036
cyber_secur_strategi	7	93	75	0.036
nation_cyber_secur	7	93	76	0.041
cyber_secur_issu	7	92	75	0.036
cyber_secur_threat	7	92	76	0.038
smart_grid_secur	6	78	71	0.0115
inform_secur	6	75	69	0.01042
network_secur	6	75	69	0.01041
smart_grid_cyber	6	75	70	0.0103
grid	6	74	74	0.0076
homeland_secur	6	74	68	0.0102
secur_and_privaci	6	74	68	0.0102
secur_for_smart	6	74	68	0.0102
system	4	74	74	0.0001
confer_on_cyber	6	71	68	0.009
control	4	71	71	0.0001
cyber_attack	6	71	69	0.0091
cyber_physic	6	71	68	0.009
electr_power_grid	3	71	71	0.001
industri_control_system	3	71	71	0.0001

Таблиця 2.Топ-32 вузлів мережі та їх показник релаксації для уповільненого та звичайного алгоритму PageRank

Термін	Показник релаксації мережі для звичайного PageRank	Показник релаксації мережі для уповільненого PageRank	Індивідуальний показник релаксації вузла	PageRank
power	4	75	48	0.0094
threat	4	75	48	0.0094
analysi	4	72	48	0.0094
nation	4	72	48	0.0094
risk	4	72	48	0.0094

control_system	3	66	49	0.01256
cyber_threat	3	66	49	0.01251
electr_power	3	66	49	0.012
industri	3	66	49	0.0094
power_grid	3	66	49	0.0134
secur_threat	3	66	49	0.0124
attack	3	65	49	0.0094
critic	3	65	49	0.0094
cyber-phys_system	3	65	49	0.0105
infrastructur	3	65	49	0.0094
intern	3	65	49	0.0094
nation_secur	3	65	49	0.0137
network	3	65	49	0.0094
scada_system	3	65	49	0.0105
secur_analysi	3	65	49	0.0137
secur_issu	3	65	49	0.0097
secur_risk	3	65	49	0.0137
technolog	3	65	49	0.0094
control	4	63	48	0
grid	4	57	48	0
data	3	56	49	0
smart_grid	3	56	49	0.0512
system	4	54	48	0
cyber_secur_threat	2	51	51	1
electr_power_grid	2	51	51	0.9868
industri_control_system	2	51	51	0.8387

З таблиці 1 та таблиці 2 видно, що по-перше, ранжування вузлів за показником релаксації мережі, яке отримане для звичайних алгоритмів та уповільнених алгоритмів HITS та PageRank,відрізняється:на думку експертів, застосування уповільнених алгоритму HITS та PageRankдає краще ранжування вузлів за показником релаксації мережі. По-друге, ранжування вузлів за показником релаксації мережі в порівнянні з ранжуванням за показником, відповідно,HITS та PageRankзначно відрізняється. А отже, запропонований показник релаксації мережі є унікальною числовою характеристикою вузлів мережі. По-третє, розглянувши числові значення індивідуального показника релаксації вузлів, можна зробити висновок, що він є окремою характеристикою вузлів, яка не подібна до показника релаксації мережі.Відповідно до таблиць 1 та 2 на рис. 2 та рис. 3 зображені порівняльні стовбчасті діаграми, де представленінормовані значення показника релаксації, отриманого для уповільнених алгоритмів HITS та PageRank, та відповідні нормовані показники HITS та PageRank для кожного вузла мережі (вузли відсортовані у порядку зростання їх показника релаксації).

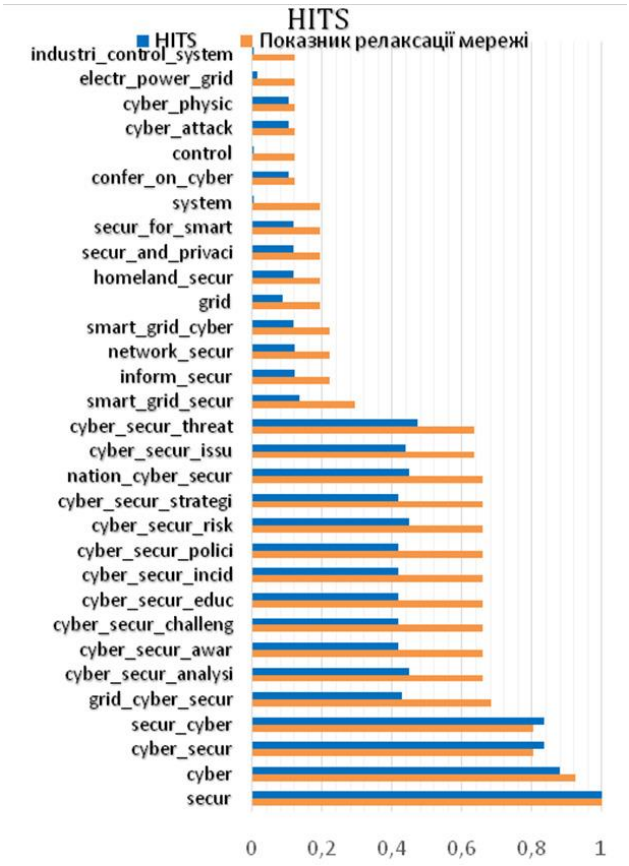


Рис. 2. Стовбчаста діаграманормованогопоказника релаксації, отриманого для уповільненого алгоритму HITS

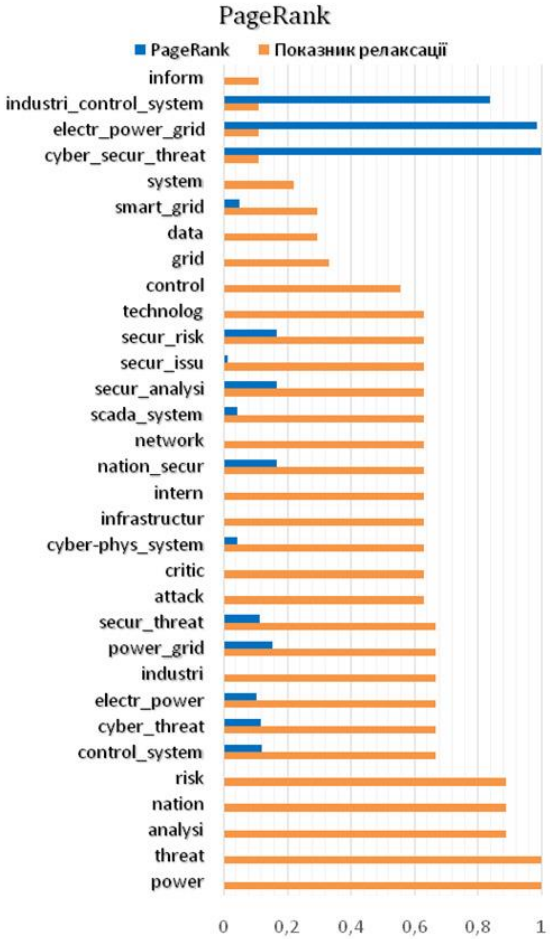


Рис. 3. Стовбчаста діаграма нормованогопоказника релаксації, отриманого для уповільненого алгоритму PageRank

Висновки

В роботі було досліджено нові характеристики вузлів мережевих структур – показник релаксації мережі та індивідуальний показник релаксації вузла.

Було показано, що процес уповільнення алгоритмів впливає на результат обчислення показника релаксації. При цьому ранжування вузлів за показником релаксації у випадку застосування уповільнених алгоритмів HITS та PageRank, на думку експертів, є кращим. Також важливе значення має порогове значення, яке є головною умовою зупинки алгоритму, та від якого залежить показник релаксації. Порогове значення вибиралося таким чином, щоб значення показника релаксації мережі для кожного вузла відрізнялись одне від одного якомога більше, й щоб надалі можна було ранжувативузли за цим показником.

Також було встановлено, що релаксація числових значень деяких вузлів відбувається швидше, ніж релаксація всієї мережі у результаті застосування одного із уповільнених ітераційних алгоритмів (HITS або PageRank). В результаті цього було досліджено так званий індивідуальний показник релаксації вузла.

Показник релаксації мережі та індивідуальний показник релаксації вузла були використані для дослідження структури мережі термінів, побудованої для предметної області “Кібербезпека”. Застосування показників релаксації дало змогуранжувати та визначити найбільш важливі змістовні компоненти мережі.

Отже, запропоновані числові характеристики вузлів мережі можуть бути використані під час дослідження та аналізу структури мережі, даючи змогу виявити найбільш важливі змістовні елементи.

Список використаних джерел

1. Newman, M. E. J.: The structure and function of complex networks. SIAM Review, vol. 45. pp. 167–256. (2003).
2. doi:10.1137/S003614450342480
3. Snarskii, A.,Lande, D.: Modeling of complex networks: tutorial. K.: Engineering, (2015).
4. Dorogovtsev,S.N., Mendes, J.F.: Evolution of networks: from biological networks to the Internet and WWW. Oxford University Press, Oxford (2013).
5. Kleinberg, J. M.: Authoritative sources in a hyperlinked environment. In Processing of ACM-SIAM Symposium on Discrete Algorithms, 46(5), pp. 604–632 (1998).
6. Brin, S., & Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems, 30(1-7), 107-117 (1998).
7. doi:10.1016/S0169-7552(98)00110-X
8. Rajaraman, A., Ullman, J. D.: Mining of massive datasets. Cambridge University Press, Cambridge(2011).
9. doi:10.1017/cbo9781139058452
10. Langville, A. N., Meyer, C. D.: Google's PageRank and beyond: The science of search engine rankings. Princeton University Press, Princeton(2011).
11. Christensen, R.: Theory of viscoelasticity: an introduction. Elsevier(2012).
12. Lande, D.V., Dmytrenko, O.O., Snarskii, A.A.: Research of network relaxation time as characteristics of network nodes. Data Registering, Storing and Processing 21(1), 83-94 (2019) (in Ukrainian).

Додаток А (Уповільнення ітераційних алгоритмів)

Розглянутий вище процес пошуку показника релаксації мережі може бути успішно застосований для розріджених матриць. Проте у випадку, коли вузли мережі мають велику кількість взаємозв’язків, ітераційний процес перерахунку значень вузлів після їх збурення буде швидким. Достатньо велика кількість вхідних та вихідних посилань на вузли спричиняє швидку релаксацію числових значень вузлів мережі. Як наслідок, при дослідженні показника релаксації мережі достатньо лише декілька ітераційних кроків, аби числові значення вузлів повернулись до рівноважного стану після збурення. Для того, щоб уповільнити процес збіжності до рівноважного розв’язку числових значень вузлів після їх збурення, у даній роботі пропонується здійснити уповільнення алгоритмів. Після надання збурення

одному із вузлів мережі, як і у випадку описаному вище, застосовується відповідний ітераційний алгоритм HITS або PageRank з уповільненням:

$$h_0(i) \to h_1(i) \to h_2(i) \to .... \to h_{рівноважне}(i)\ \hat{\mathbf{A}}h_1(i) = h_2(i)$$
$$\hat{\mathbf{A}}h_n(i) = h_{n+1}(i)$$
$$h_{n+1}(i) \longleftarrow\!\!\!\!- \beta h_{n+1}(i)\,,$$

де $0 < \beta < 1-$ коефіцієнт уповільнення.

Таким чином, зменшуючи або збільшуючи числове значення показника “авторитетності” або PageRank вузлів на відповідному ітераційному кроці (уповільнюючи алгоритм), вдається досягти уповільнення релаксаційчислових значень вузлів мережі.

METHODOLOGY OF RATIONAL CHOICE OF SECURITY INCIDENT MANAGEMENT SYSTEM FOR BUILDING OPERATIONAL SECURITY CENTER

Igor Y. Subach, Volodymyr O. Kubrak, Artem V. Mykytiuk

Institute of Special Communications and Information Protection of the National Technical University of Ukraine "Igor Sikorsky Kiev Polytechnic Institute", Ukraine

igor_subach@ukr.net

Abstract. This article discusses the purpose, tasks and composition of the Operational Security Center (SOC). The basic technological tools which should include modern effective SOC are indicated. The focus is on the key role of the Information Security Incident Management System (SIEM) in the SOC. The purpose of SIEM and the main tasks that it should solve are reviewed. The peculiarities of solving the problem of choosing of information security incident management system (SIEM) are analyzed. The groups of indicators that characterize the degree of fulfillment of the requirements to SIEM are highlighted. The application of fuzzy set theory for processing expert information on qualitative indicators characterizing SIEM is proposed. The formulation of the SIEM selection problem is done and the main stages of its solution are proposed: preparation of initial data; choosing the method of solving the multicriteria problem; algorithm development. The method of normalization of SIEM quantitative indicators and the method of paired comparison based on the rank estimates for processing of SIEM qualitative indicators are proposed. It is proposed to use the 9-point Saaty scale to derive functions of SIEM qualitative values based on the processing of expert assessments. The algorithm of the considered method is implemented. Methods for solving multicriteria problems are analyzed and the use of a lexographic method is proposed for solving the SIEM solution for the Security Center (SOC). An algorithm for its implementation has been developed. To illustrate the operation of the proposed algorithm, we give an example of how to apply it to choose a rational SIEM option. Recommendations for application of the results obtained are offered.

Key words: cybersecurity, SIEM, SOC, lexographic method, fuzzy sets theory.

1. Introduction

It is impossible to counteract the modern cyber threats without the use of modern cybersecurity technologies that enable monitoring, collection, collation and processing of information in order to identify existing and predict future threats. Important role is given to the special units that deal with information and cyber security issues at the organizational and technical level – the Security Operation Centers (SOC).

Modern SOC solves the following tasks [1]:

taking immediate actions to protect against cyberattacks and minimize their damage;

identification of system security vulnerabilities and taking actions to eliminate them;

centralized security management of various devices in the system;

continuous monitoring of system threats status;

technical support for cyber security of the system and others.

Structurally, the SOC has three main components: personnel – skilled professionals using modern cybersecurity technologies with teamwork and management competencies; processes – business processes,

technological processes, operational and analytical processes; technologies – tools for detecting, counteracting and preventing cyber threats.

Effective SOC should include the following modern technological tools to ensure cyber security [2]: Next Generation Firewall, Intrusion Prevention System (IPS), Web Application Firewall (WAF), Database Protection, Email Security, Endpoint Detection and Response, Vulnerability Scanners, Data Loss Prevention, Forensics, Network Access Control and others.

However, the basis for building an effective SOC is the use of the SIEM system (Security Information and Event Management) – a system for managing information and security events. The use of SIEM in protection system enables proactive management of cyber incidents. That is, to predict future events that will occur in the system by applying automated mechanisms that use information about events that have already occurred in the system, as well as to adapt the protection settings of the system to its current state, thereby implementing preventive measures even before the situation in the system becomes critical [2]. In accordance with this, SIEM system should solve a range of tasks which include [3]:

collection, processing and analysis of security events coming from a variety of heterogeneous distributed sources;

detection of real-time or close cyber attacks and violations of security policies;

investigation of cyber incidents;

developing effective solutions for cyber security;

generation of reporting documents and visualization of system status and others.

In order to solve these problems, the SIEM-system, on the basis of the initial data collected from the log files which accumulate information about the events that occur in the system, selects those events that may be a sign of cyber attacks or other undesirable actions in the system.

The main feature of the solution to the problem of choosing a SIEM-system for building SOC is a large number of indicators that characterize the degree of fulfillment the requirements for systems of this type which can be both quantitative and qualitative. Qualitative indicators, first of all, include those that characterize how effectively the SIEM system can be used to solve the functional tasks entrusted to it by the SOC; what will be the cost of purchasing and using the system; how reliable it is and easy to operate, etc.

Analysis of recent publications [4-11] showed that these figures can be represented as follows:

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}\},$$

where x_1 – event source support;

x_2 – event collection;

x_3 – correlation;

x_4 – search and analytics;

x_5 – visualization and reporting;

x_6 – prioritization and notification;

x_7 – general settings and installed;

x_8 – scalability, fault tolerance, storing;

x_9 – system component monitoring and internal audit;

x_{10} – ease of use;

x_{11} – availability of state certificates of conformity;

x_{12} – additional system modules;

x_{13} – cost.

Therefore, the problem of rational selection of SIEM-system for building the SOC is characterized by multicriteria and the need to consider a large number of qualitative and quantitative indicators.

In its turn, the first characteristic requires the use of an effective method of solving multicriteria problems, and the second – the application of fuzzy set theory for the processing of expert information on qualitative indicators [12, 13].

2 The problem of rational choice of SIEM

The general statement of the problem of rational choice of SIEM-system can be described as follows.

It is necessary to find

$$S_0 = \arg\underset{s\in S}{opt} W\left(\overline{X}(s)\right), \qquad (1)$$

where W − some generalized indicator of system quality;

S − a set of possible system choices;

$\overline{X}(s) = \left|x_1(s), x_2(s), \ldots, x_k(s), x_{k+1}(s), \ldots, x_n(s)\right|$ − vector of SIEM quality indicators, besides first k ($i = \overline{1,k}$) requirements are quantitative, and the other n-k $k = \overline{k+1,n}$ – qualitative.

The value of the partial indicator i, which characterizes the degree of fulfillment the SIEM requirement i, is determined by its approximation to the optimal value.

The main stages in solving the task (1) are: preparing initial data; choosing a method for solving a multicriteria problem; algorithm development.

3 The method of solving the problem

It is advisable to use normalized values to estimate the degree of proximity of the quantitative indicator i to the optimal value for j variant of the SIEM. $x_{ij}, i = \overline{1,k};\ j = \overline{1,k};\ 0 \le x_{ij} \le 1.$

Normalization of the value of a quantitative indicator can be made as follows:

$$x_{ij} = \frac{x_{ij} - x_{ij}^{*}}{x_{ij}^{**} - x_{ij}^{*}}, \qquad (2)$$

where x_{ij} – the value of indicator i for j variant of the system;

x_{ij}^{*}, x_{ij}^{**} – the worst and the best indicator value.

Accordingly, the degree of proximity of the quality indicator i to the optimal value for the j variant of the SIEM can be determined using the membership function $\mu_S(x_i)$. To build a membership function $\mu_S(x_i)$ it is advisable to use a rank-based method or pairwise ranking method [14, 15].

In this case, the rank of an element $x_i \in X$ refers to a number $r_s(x_i)$ that characterizes its importance in the formation of the SIEM property which is described by a fuzzy term S. Suppose that the greater the rank of an indicator, the greater the value of its membership function.

If you introduce the following figures

$$r_S(x_i) = r_i, \mu_S(x_i) = \mu_i; i = \overline{1,n},$$

then the distribution of membership degrees can be represented as follows:

$$\frac{\mu_1}{r_1} = \frac{\mu_2}{r_2} = \ldots \frac{\mu_n}{r_n}, \qquad (3)$$

in case of normalization:

$$\mu_1 + \mu_2 + \ldots + \mu_n = 1. \qquad (4)$$

On the basis of (3), the membership degree of all elements of the set is determined by the membership degree of the so-called supporting member.

For supporting member $x_1 \in X$ that has a membership function μ_1:

$$\mu_2 = \frac{r_2}{r_1} \cdot \mu_1;\ \mu_3 = \frac{r_3}{r_1} \cdot \mu_1;\ \ldots;\ \mu_n = \frac{r_n}{r_1} \cdot \mu_1; \qquad (5)$$

For supporting member $x_2 \in X$ that has a membership function μ_2:

$$\mu_2 = \frac{r_1}{r_2} \cdot \mu_2;\ \mu_3 = \frac{r_3}{r_2} \cdot \mu_2;\ \ldots;\ \mu_n = \frac{r_n}{r_2} \cdot \mu_2; \qquad (6)$$

Accordingly, for supporting member $x_n \in X$, that has a membership function μ_n:

$$\mu_n = \frac{r_1}{r_n} \cdot \mu_n;\ \mu_2 = \frac{r_2}{r_n} \cdot \mu_n;\ \ldots;\ \mu_{n-1} = \frac{r_{n-1}}{r_n} \cdot \mu_n; \qquad (7)$$

From (5-7) and in case of normalization (4) we obtain:

$$\begin{cases} \mu_1 = \left(1 + \frac{r_2}{r_1} + \frac{r_3}{r_1} + \ldots + \frac{r_n}{r_1}\right)^{-1} \\ \mu_2 = \left(\frac{r_1}{r_2} + 1 + \frac{r_3}{r_2} + \ldots + \frac{r_n}{r_2}\right)^{-1} \\ \ldots \\ \mu_n = \left(\frac{r_1}{r_n} + \frac{r_2}{r_n} + \frac{r_3}{r_n} + \ldots + 1\right)^{-1} \end{cases} \qquad (8)$$

On the basis of (8), it is possible to calculate the membership degrees $\mu_s(x_i)$ on the relative estimates of the ranks $\frac{r_i}{r_j} = \xi_{ij}, i, j = \overline{1,n},$ which create the following matrix:

$$\Xi = \begin{bmatrix} 1 & \frac{r_2}{r_1} & \frac{r_3}{r_1} & \ldots & \frac{r_n}{r_1} \\ \frac{r_1}{r_2} & 1 & \frac{r_3}{r_2} & \ldots & \frac{r_n}{r_2} \\ \ldots & \ldots & \ldots & \ldots & \vdots \\ \frac{r_1}{r_n} & \frac{r_2}{r_n} & \frac{r_3}{r_n} & \ldots & 1 \end{bmatrix} \qquad (9)$$

It is easy to see that the properties of the matrix (9) are the following: it is diagonal, transitive, and elements of the matrix that are symmetric about the main diagonal are connected by dependence relation: $\xi_{ij} = 1/\xi_{ji}$.

Since matrix (9) is a matrix of paired comparison of the element ranks, a 9-point Saaty scale can be used for expert evaluation of its elements: $\xi_{ij} = r_i / r_j$ (Table 1).

Table 1. Relative Importance Scale

Intensity of relative importance	Definition
1	Equal importance of compared requirements
3	Weak importance of one over another
5	Strong importance
7	Demonstrated importance
9	Absolute importance
2,4,6,8	Intermediate values between the two adjacent judgments

Thus, using (8), the expert data on element ranks (their paired comparison) are transformed into a fuzzy term membership function.

The algorithm for constructing the membership function includes the following steps.

1. Set a linguistic variable (qualitative characteristic of SIEM).
2. Determine the universal set on which the linguistic variable is set (the value of the qualitative characteristic of SIEM).

3. Set a variety of fuzzy terms $\{S_1, S_2, \ldots, S_n\}$ that are used to evaluate the variable set in the first step.
4. Form a matrix (9) for each term S_j, $j = \overline{1,m}$.
5. Using the formulas (8) calculate the membership functions of the elements (SIEM characteristics) for each fuzzy term.
6. The procedure for the normalization of the received membership functions should be carried out by dividing them by the largest value of the membership function.

The most common methods for solving a multicriteria problem (1) are the following [16]: the principal indicator method, generalized additive/multiplicative indicator method, generalized minimax indicator method and lexographic method. The analysis shows that they all have their pros and cons, and the choice of a method is largely determined by the completeness and credibility of the expert knowledge of the importance and degree of interrelation of partial quality indicators. Since lexographic method is the least demanding for expert information about the degree of preference for partial indicators, it is advisable to choose the lexographic method in order to solve the problem of rational choice of SIEM-system for building the SOC. The essence of use of this method is the following.

At the previous stage of solving the task it is possible to find a set of "good solutions" (Pareto-optimal solutions) by consistently comparing possible SIEM options for all quality indicators. [17, 18, 19].

Further, all the partial indicators are ordered by importance. Then the set of alternatives with the best score by the most important indicator is outlined. When such an alternative is the only one it is considered to be the best. Otherwise, when several alternatives are obtained, they are distinguished by those that have a better rating on another indicator and so on. Thus, the algorithm for implementing the lexographic method for solving the problem of rational choice of the system consists of the following steps.

1. Partial quality indicators are ranked by importance:

$$x_1(s) > x_2(s) > \ldots > x_n(s)$$

2. For each indicator the value of permissible concession is determined $\Delta x_i, i = \overline{1,n},$ within which the compared SIEM variants are considered to be equivalent;

3. For the first indicator $x_1(s)$ a set Ψ_1 of equivalent SIEM variants is formed which meets the following condition:

$$\max\left(x_{1j} - x_{1k}\right) \le \Delta x_1,\ j = \overline{1,m};\ k = \overline{1,m};\ k \ne j. \quad (10)$$

4. If the set contains only one variant, it is considered to be the best. Otherwise, when it contains more than one alternative, you need consider all variants of the set by indicator $x_2(s)$.

5. For the second indicator $x_2(s)$, from a set of variants Ψ_1, a set of variants Ψ_2 is formed which meet the condition:

$$\max\left(x_{2j} - x_{2k}\right) \le \Delta x_2, i \in \Psi_1;\ k \in \Psi_1;\ k \ne j. \quad (11)$$

6. If the set Ψ_2 contains one variant, it is considered to be the best. Otherwise, found variants are considered by indicator $x_3(s)$ and so on.

7. In the case where all indicators are consistently reviewed and a set $\Psi = \Psi_1 \times \Psi_2 \times \ldots \times \Psi_n,$ containing more than one alternative is obtained, there are two options: reduce the value of the permissible concession $\Delta x_i, i = \overline{1,n},$ from the first most important indicator and repeat the algorithm from the beginning or allow the decision maker to choose the best option.

To illustrate the proposed algorithm in work we give an example of its application to the selection of a rational variant of the SIEM system.

To select a SIEM system, we use four partial indicators: $x_1(s)$ – cost and $x_2(s)$ – event source support which are quantitative indicators, as well as $\mu_{x_3}(s)$ – scalability and $\mu_{x_4}(s)$ – ease to use which are qualitative indicators.

Five options for choosing a SIEM system s_j, $j = \overline{1,5}$ have been selected for consideration.

As a result of the calculations and expert assessments, the following data were obtained characterizing the degree of SIEM compliance with the specified requirements:

$$x_1 = \left\{ \frac{0,8}{s_1}; \frac{0,8}{s_2}; \frac{0,7}{s_3}; \frac{0,5}{s_4}; \frac{0,6}{s_5} \right\};$$

$$x_2 = \left\{ \frac{0,7}{s_1}; \frac{0,8}{s_2}; \frac{0,6}{s_3}; \frac{0,7}{s_4}; \frac{0,8}{s_5} \right\};$$

$$x_3 = \left\{ \frac{0,4}{s_1}; \frac{0,6}{s_2}; \frac{0,7}{s_3}; \frac{0,8}{s_4}; \frac{0,7}{s_5} \right\};$$

$$x_4 = \left\{ \frac{0,5}{s_1}; \frac{0,6}{s_2}; \frac{0,5}{s_3}; \frac{0,6}{s_4}; \frac{0,3}{s_5} \right\}.$$

1. Indicators are ranked by importance as follows:

$$x_1 > x_2 > \mu_S(x_3) > \mu_s(x_4).$$

2. The value of permissible concession $\Delta x_i = 0,1,\ i = \overline{1,4}.$

3. With the maximum value of the first indicator $x_1 = 0,8$ and the value of permissible concession $\Delta x_1 = 0,1$ to the set Ψ_1 of equal variants for SIEM, which meet the condition (2) the following variants are included:

$$\Psi_1 = \{S_1, S_2, S_3\}.$$

4. Of the set Ψ_1, by the second indicator x_2 meeting the condition (3): $x_2 = 0,8$ and $\Delta x_2 = 0,1$ to the set Ψ_2 the following variants are included:

$$\Psi_2 = \{S_1, S_2\}.$$

5. Of the set of variants: $\Psi = \Psi_1 \times \Psi_2$ for the third indicator x_3 meeting the condition (3) $x_3 = 0,6$ and $\Delta x_3 = 0,1$ to the set Ψ_3 the following variants are included:

$$\Psi_3 = \{S_2\}.$$

A rational choice of SIEM for building a SOC is the second option.

4 Conclusion

The conducted research shows that the lexographic method is an effective method for solving the multicriteria problem of SIEM selection for SOC. Groups of quantitative and qualitative indicators characterizing the requirements for SIEM in the SOC are formulated. Methods of processing quantitative and qualitative indicators of SIEM are offered. The expedience of applying the procedure for rationing quantitative indicators of SIEM and applying the method of paired comparison based on rank evaluations for processing its qualitative indicators is justified. The formulation of the SIEM selection problem is done and the main stages of its solution are outlined. An algorithm for the implementation of the lexographic method is developed and brought to practical implementation.
The results obtained can be used in practice to solve the problems of creating SOC and rational choice of its software such as SIEM.

References

1. KG, Vogel Business Media GmbH & Co. Was ist ein Security Operations Center (SOC)? – https://www.security-insider.de/was-ist-ein-security-operations-center-soc-a-617980/, 22.06.17.
2. Laskin S. How to build a competent, scalable and effective information security management center / S. Laskin // LAN Network Solutions Magazine. – https:// https://www.osp.ru/lan/2017/04/13051902/, 28.03.2017.
3. Kotenko I. Proactive security mechanisms against network worms: approach, implementation and results of the experiments / I.V. Kotenko, V.V. Voroncov, A.A. Chechulin, A.V. Ulanov // Information Technology. – 2009. – Vol. 1. – pp. 37 - 42.

4. Kotenko I. Application of security information and event management technology for information security in critical infrastructures / I. Kotenko, I. Saenko, O. Polubelova, A. Chechulin // SPIIRAS Proceeding, ISSN: 2078-9181. – 2012. – Issue 1(20). – pp. 27–56.
5. Paley L. Comparison of SIEM systems. Part 1. – https://www.anti-alware.ru/compare/ SIEM-systems, 07/04/2018.
6. Paley L. Comparison of SIEM systems. Part 2. – https://www.anti-alware.ru/compare/ SIEM-systems-part2, 1/16, 08/05/2019.
7. Niyazov T. Comparison of SIEM solutions for building SOC, www.jetinfo.ru/stati/ sravnenie-siem-reshenij-dlya-postroeniya-soc, 08/05/2019.
8. Comparison of SIEM systems, www.siem.su/compare_SIEM_systems.php, 08/05/2019.
9. Donskoy K.A. SIEM overview on the Russian market // Donskoy K.A., Levin L.S., Trushin V.A. // Collection of scientific works of NSTU. – 2017. – No. 3 (89). – pp. 124–132.]
10. SIEM product comparison – https://comminity.softwaregrp.com/dcvta86296, 10/18/2019.
11. SIEM competitive comparision – https://www.securonix.com/products/ competitive-comparison, 10/18/2019.
12. Boehm B. Characteristics of software quality / [Boehm B., Brown J., Caspar H. et al.: Translated from English]. – Moscow: Mir, 1981. – 208 p.
13. Borisov A.N. Decision making based on fuzzy models: examples of use / [Borisov A.N., Krumberg O.A., Fedorov I.P.] / Riga: Zinatne, 1990. – 184 p.
14. Zaichenko Y.P. Operations Research: Fuzzy Optimization. – Kiev: High school, 1991. – 191 p.
15. Rothstein A.P. Intelligent identification technologies: fuzzy sets, genetic algorithms, neural networks. – Vinnytsia: UNIVERSUM, 1999. – 320 p.
16. Gerasimov B.M. Decision support systems: design, application, performance evaluation / [Gerasimov B.M., Divizinyuk M.M., Subach I.Y.] / Monograph. – Sevastopol: Publishing Center SNIYE and P, 2004. – 320 p.
17. Tutkin L.S. Optimization of electronic devices according to a set of quality indicators / Tutkin L.S. Moscow: Radio and communications, 1975. – 367 p.
18. Podinovsky V.V. Optimization according to successively applied criteria / [Podinovsky V.V., Gavrilov V.M.]. – Moscow: Soviet Radio, 1975. – 234 p.
19. Podinovsky V.V. Pareto-optimal solutions to multicriteria problems / [Podinovsky V.V., Nogin V.D.]. – Moscow: Science, Main edition of the physical and mathematical literature, 1982. – 256 p.

SYSTEM OF INTELLECTUAL UKRAINIAN LANGUAGE PROCESSING

Tmienova Nataliia[1] and Sus' Bogdan B.[2]

[1]Faculty of Information Technology
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[2]Institute of High Technologies
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
tmyenovox@gmail.com,bnsuse@gmail.com

Abstract. The problem of high-quality automatic natural language processing is one of the most important tasks in computational linguistics. Automatic natural language processing is used in information retrieval, in tasks of text generation and text recognition, in machine translation, in sentiment analysis and so on. All of these areas require specialized linguistic and mathematical models to represent the morphology, syntax, and semantics of text in a form that is convenient for automatic processing.

The article describes a developed system that implements specific linguistic tasks related to the processing of the Ukrainian language, that is text preprocessing, morphological and lexical analyzes of text. In order to create such a system, an analysis of the available natural language text-processing tools was carried out and the possibility of using them for text processing of Ukrainian language was examined. Also, the most appropriate text processing tools in the Ukrainian language were selected. The basic stages of text preprocessing are considered in detail and algorithms of their program implementation are given. In addition, the results of the developed system were demonstrated.

Keywords: Natural language processing, NLP, Text Mining, Ukrainian language processing, language analysis, text corpus

Introduction

Natural language processing (NLP) is a subfield of computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data [1].

NLP is a component of text mining that performs a special kind of linguistic analysis that essentially helps a machine “read” text**.** NLP uses a variety of methodologies to decipher the ambiguities in human language**,** including the following: automatic summarization, part-of-speech tagging, disambiguation, entity extraction and relations extraction, as well as disambiguation and natural language understanding and recognition.

NLP is used in information retrieval, in tasks of text generation and text recognition, in machine translation, in sentiment analysis and so on [2]. All of these areas require specialized linguistic and mathematical models to represent the morphology, syntax, and semantics of text in a form that is convenient for automatic processing.

To work, any natural language processing software needs a consistent knowledge base such as a detailed thesaurus, a lexicon of words, a data set for linguistic and grammatical rules, an ontology and up-to-date entities, text corpora.

It is now quite difficult to imagine linguistic researches, language learning and the translation process without the use of corpora [3, 4, 5]. Corpora are widely used in lexicographic and grammatical studies, semantics and stylistics. The term "linguistic corpus" means the electronic collection of natural language texts, which is organized, organized and designed in a certain way and is intended for the scientific and practical study of language [6].

Data collected in the corpus can be very different in quality and quantity depending on the project of the investigation [7]. So, the formation of a corpus of texts of any language should begin with the clarification of the research and educational tasks that are supposed to be solved on the basis of the content of the created corpus. The pupose of the corpus determines the features of its structure. Corpus can be consiered as an important means of verifying a variety of linguistic theories. Therefore, the corpus must be representative and balanced. Accordingly, linguistic tools for corpus work aredeveloped with aim at performing a particular task.

The major problem of natural language processing is the ambiguity, so most natural language processing problems can be considered as findig proper interpretation. According to the structure of natural language, we can distinguish the following basic stages of linguistic analysis, each of which is ambiguous in its own way: previous, morphological, syntactic and semantic. These stages of analysis are performed squentially and each subsequent stage uses the results of the previous ones. Similarly, mistakes in the previous stages of analysis are affected by the results of the following stages.

The purpose of the article

There are many different types of systems that implement the steps described above in English and Russian (AOT, Rosette text analytics, Polyglot). Since there are not enough tools for processing Ukrainian Development of systems for processing Ukrainian-language texts is a rather urgent problem, namely: libraries for programming languages, marked up corpora, dictionaries, thesauruses, etc.

The purpose of development a system for the intellectual processing of Ukrainian-language texts was a creation of tools for the processing of natural-language texts in Ukrainian at preliminary, morphological and lexical levels of analysis.

The main objectives of the preliminary analysis include the following:

• tokenization is the process of splitting the text into units called tokens (tokens can be words, numbers, punctuation marks, etc.);

• sentence boundary detection;

It is also desirable at this stage to solve the following minor problems:

• removing of non-text elements (tags, meta-information);

• removing formatting (italics, underline, bold);

• selection of e-mail addresses;

• selection of file names;

• assembly of words written in the discharge;

• removing of stop words;

• named entity recognition.

The tasks of morphological analysis include the following:

• definition of the grammatical attributes of the word (finding the part of the language and the grammatical categories that are inherent in the corresponding part of the language - number, gender, etc.);

• stemming is the process of reducing words consisting of several morphemes to their stem, i.e. to the fixed basis;

• Lemmatization is the process of reducing words to their vocabulary form.

Among the problems of lexical analysis there are the following:

• defining unique words;

• determining the frequency of words;

• calculating the lexical diversity.

In the analysis of literature of such a complex system, which would cover the processing of Ukrainian texts at several levels, was not found. Only some separate language processing tools have been found. After testing, it turned out that some of them produce incorrect results. An attempt was also made to adapt the stemming algorithm for the Russian language [8] to the Ukrainian language, but the results of this algorithm were not sufficient.

The NLTK library tools were selected for preliminary analysis [9]. On this basis, tools were developed to address the following problems of preliminary analysis of texts submitted in Ukrainian: tokenization; sentence boundary detection; removing of non-text elements (tags, meta-information); e-mail highlighting; selection of file names; assembly of words written in the discharge; removing of stop words; named entity recognition.

The pymorphy2 library was selected as the basis of the morphological analyzer [10]. Pymorphy2 uses a large electronic dictionary of Ukrainian language () [11] converted to the OpenCorpora format [12]. On the basis of the tools of this library were created tools for realization of the following tasks of morphological analysis: morphological analysis of a word; stemming; lemmatization.

Simple lexical analysis was developed using Python tools. The matplotlib library was used to display the lexical information on the graph. For lexical analysis of texts in Ukrainian, the system contains the following tools: calculating word frequencies and displaying them on a graph; calculating lexical variety of text.

.

Available NLP tools review

Modern tools for text analysis can be divided into two major categories:

- Specialized tools - tools for language-specific analysis (morphological analyzers, syntax parsers, etc.);
- Integrated packages are software instruments that provide features for analyzing text at different levels.

Let's describe several platforms and libraries for word processing in natural languages. The following software products provide tools for analyzing texts in natural language, both at one level and at many levels.

The Rosette text analytics platform develops software [13] to extract unstructured text knowledge for use by search engines and data analytics applications.

For basic text analysis, the platform provides specific language tools for tokenization, selection of parts of the language, lemmatization, classification, locating named objects, and relations between named objects. For text preprocessing and morphological analysis, the platform provides the following features:

1) Definition of the boundaries of sentences.

• Includes ambiguous punctuation marks for abbreviations, file names, email addresses, etc.

• Uses machine learning and statistical analysis.

• Supports Ukrainian.

2) Tokenization.

• Uses statistical modeling.

Determines a part of the ambiguous word language by the means of statistical modeling for lemmatization.

• Decomposes difficult words.

• Uses Stemming lemmatization.

• Does not support Ukrainian.

Polyglot [14] is a library on Python for natural language processing. For text preprocessing and morphological analysis it contains the following modules:

1) Tokenization module. In addition to the tokenization itself, it has tools for splitting text into sentences.

• Supports Ukrainian.

• Sentencing does not take into account Ukrainian cuts.

1) Definition of a part of the word language.

• Does not support Ukrainian language.

3) Divide words into morphemes.

• Incorrectly divides Ukrainian words.

AOT system - Automatic word processing [15] is a set of packages designed for word processing in Russian. It contains separate components – processors for text preprocessing and morphological analysis.

The Aot.ru working group develops software in the field of automatic word processing, which mainly includes analysis of the Russian language.

The components that form up the language model are the linguistic processors that process the input text by conveyor method. The input of one processor is the output of another. The following components are distinguished:

- Preliminary analysis;
- Morphological analysis;
- parsing;

• Semantic analysis. The grapheme processor contains an algorithm for splitting into tokens and sentences.

The morphological processor uses a special Russian morphological dictionary based on the A.A. Zaliznyak grammar dictionary. It includes 161 thousand lemmas.

In case of lemmatization, a lot of morphological interpretations of the following form are given for each word of the input:

1) lemma (always in capital letters);
2) the morphological part of the language;
3) a set of common grammars (which apply to all word forms of the word paradigm);
4) multiple sets of gramems.

Natural Language Toolkit Library or NLTK [9] is a software package for symbolic and statistical processing of natural language in Python. It contains graphical representations and examples of data. It is accompanied by extensive documentation, including a book explaining the basic concepts behind the natural language processing tasks that can be accomplished with this package [16].

NLTK is optimal software platform for prototyping and development of linguistic research systems. NLTK is free software.

NLTK supports classification, tokenization, language part definition, syntactic analysis

Key features:

• Lexical analysis: a tokenizer of words and texts.
• n-grams and word combinations.
• Defining parts of languages.
• Lemmatization
• Stemming
• Named Object Recognition.

NLTK is a powerful library for English. The morphological and syntactic analyzers do not support Ukrainian, but the pymorphy2 module is available for morphological analysis of the Russian and Ukrainian languages.

Pymorphy2 Morphological Analyzer [10] is an open-source Python programming library for morphological analysis of Russian and Ukrainian words. Main Functionality:

• Provides information on basic grammatical categories;
• cancels words;
• puts the word in its original form.

In pymorphy2, the MorphAnalyzer class is used for morphological analysis of words.

System structure, methods used to the development of system, description of implementation

The structure of the system can be represented as the following flowchart:

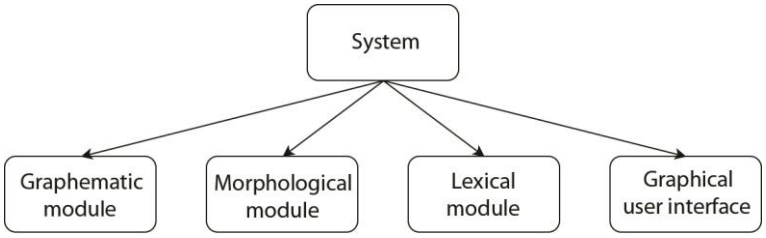


Fig.18**.** Structure of system.

Regular expressions are used to solve a large number of sub-tasks. A regular expression is a special text string that describes or matches multiple lines according to a set of special syntax rules (that is, a search pattern). They are used in many text editors and auxiliary tools to find and modify text based on specified templates.

The NLTK library was selected for the tokenization and solving of sentence boundary problems. Because abbreviations are required to determine word and sentence boundaries, a tokenizer of the NLTK library that works with regular expression patterns was used and a list of abbreviations made using DSTU 3582-97 [17]. Regular expression was used for tokenization:

[^,.;:!?\s]+(?:\.[^ЙЦУКЕНГШЩЗХЇҐФІВАПРОЛДЖЄЯЧСМИТЬБЮAZ\s])*[^,;.:!?\s]*|[,;.:!]

The boundaries of sentences are determined by regular expression

.*?(?:\S{2,}|\s)[.!?](?![йцукенгшщзххїґфівапролджєячсмитьба-z])

Regular expressions were used to highlight non-text items, e-mail addresses, and filenames.

The removing of stop words was performed using a stop word list.

A recursive algorithm that uses a dictionary of words was developed to minimize the quantity of words written with expanded spacing.

The pymorphy2 bibliography was used to obtain the morphological information about the word. Based on information about the lemma and word form, stemming is performed.

The search for named objects in the text is based on a list of new word names based on the NER annotation of the Ukrainian corpus [18].

Simple lexical analysis was written on Python. The matplotlib library was used to display the lexical information on the graph.

The Python was chosen to implement the software system. The program's GUI is written using the PyQt GUI, a Python for the Qt library. A matplotlib library was used to build the graphs - a Python library to build 2D graphs that creates shapes in a variety of print formats and interactive environments across platforms.

Demonstration of resaults

The GUI of the system consists of four areas, which are shown on fig. 2.

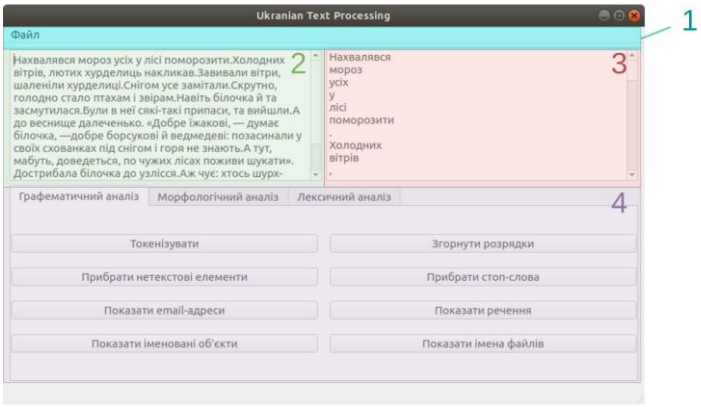


Fig. 2. 1 - menu area, 2 - input area, 3 - token list area, 4 - language feature set processing area

The results of tokenization and partitioning of the input text are shown in the fig. 3 and 4 accordingly.

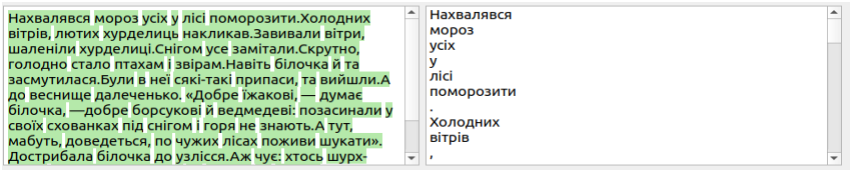


Fig. 3. The result of tokenization

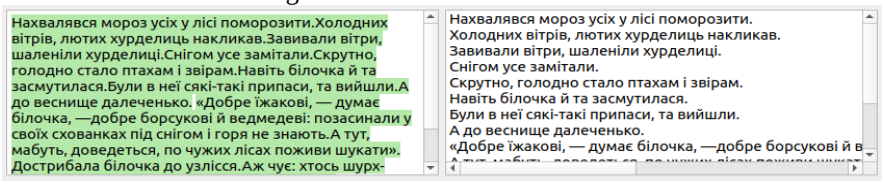


Fig. 4. The result of the sentence splitting

The functions and results of text preprocessing are shown on fig. 5-10

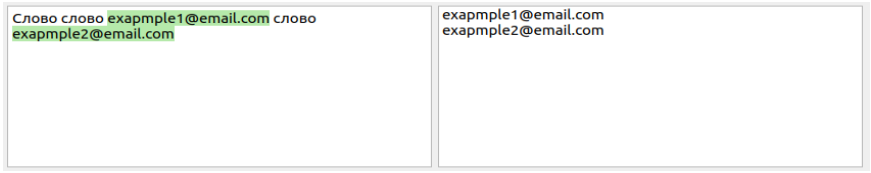


Fig. 5. Selection of e-mails

Fig. 6. Selection of file names

Fig. 7. Selection of named entity

Fig. 8. Selection of non-text elements

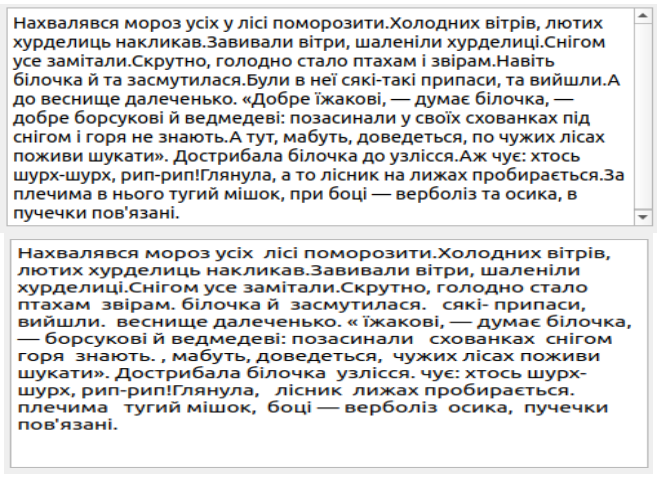


Fig 9. Removing of stop-words.

For morphological analysis, a tokenization must be carried out in advance. To extract morphological information for a word from a text, it is necessary to click on the desired word in the text or in the list on the right.

The morphological analysis of the selected word from the text is shown in Fig. 10.

Morphological analysis is shown in a table where rows are possible parsing words and columns are grammatical categories. The lexical analysis tab and the lexical characteristics of a news article are presented on Fig. 11.

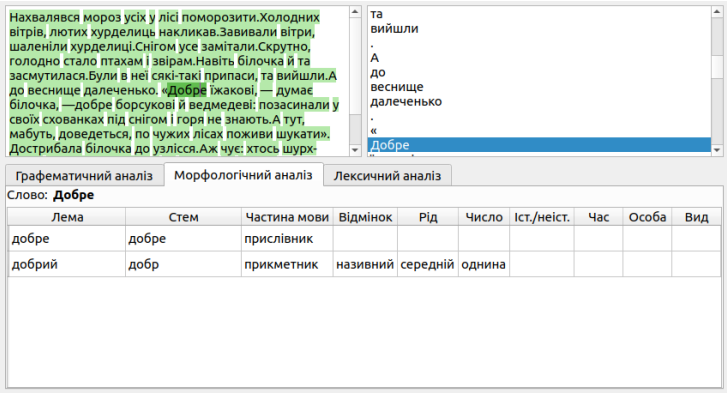


Fig 10. The window tab of morphological analysis

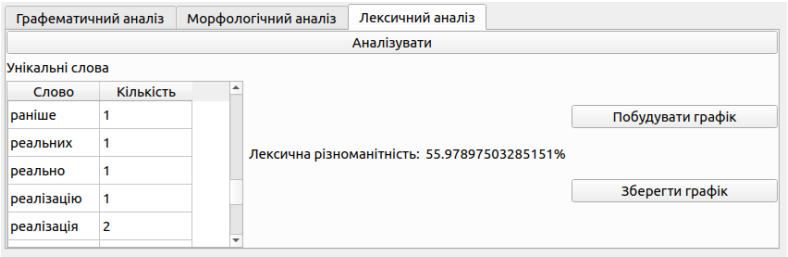


Fig 11. The result of lexical analysis of news article

A graphical representation of the lexical variety of the text is shown in Figure 13. To save the received graphs, it is necessary to click on the button "Save graph"

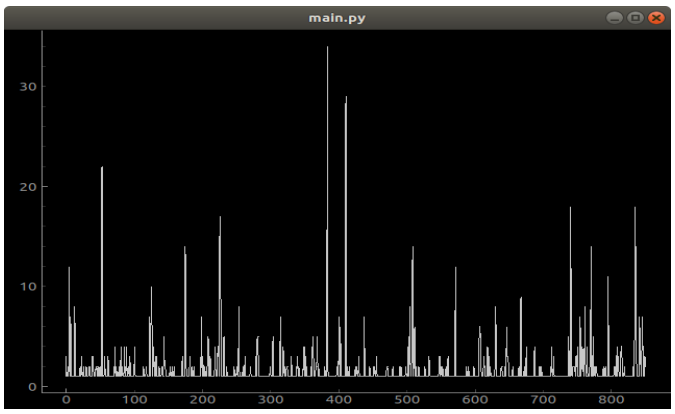


Fig 12. News article

Conclusions

During research, several natural language processing tools have been identified and reviewed. The advantages of these systems are that they perform natural language processing at several levels: text preprocessing, morphological, syntactic and semantic. The disadvantage of the systems that is not all of them provide the means of word processing in the Ukrainian language at several levels of analysis, and the systems that have tools for word processing in the Ukrainian language produce incorrect results, in particular at the text preprocessing and morphological levels of the analysis. As a result, the system that implements specific linguistic tasks related to the processing of the Ukrainian language, namely, text preprocessing, morphological and lexical analyzes of text was developed. The basic stages of text preprocessing are considered in detail and algorithms of their program implementation were presented. In addition, the results of the developed system were demonstrated.

The advantage of the developed system is that it provides the features for complex processing of Ukrainian texts in at three levels of analysis. The disadvantages of the developed system include the following: suboptimal algorithm of the collection of words written in a word (the algorithm uses a dictionary to identify words written in a word; to reduce the volume of word search, the length limit is introduced, so words, larger limit values could be convoluted with mistakes); incorrect parsing of non-vocabulary words,

which is, in fact, a disadvantage of pymorphy2, which is taken as the basis of the morphological module of the system; simple lexical analysis.

References

1. Handbook of Natural Language Processing, (Chapman & Hall/CRC Machine Learning & Pattern Recognition), 2nd Edition – 2010. – 667 p.
2. Jurafsky and Martin. Speech and Language Processing, second edition. – 2009. – 950 p.
3. Brown Corpus [Electronic Resource]. Mode of access:: http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM
4. The Brown Corpus of Ukrainian Language [Electronic Resource]. Mode of access: https://github.com/brown-uk/corpus
5. Corpus of the Ukrainian Language (Ukrainian) [Electronic Resource]. Mode of access: http://www.mova.info/ corpus.aspx?l1=209
6. S. Kübler and H. Zinsmeister, Corpus Linguistics and Linguistically Annotated Corpora (English), New York; London: Bloomsbury Academic, 2015, pp. 21-156. http://dx.doi.org/10.5040/9781472593573
7. L. Anthony, A critical look at software tools in corpus linguistics, (2013) Linguitic Research, 30 (2), pp. 141-161. https://doi.org/10.17250/khisli.30.2.201308.0012
8. Russian stemming algorithm. [Electronic Resource]. Mode of access: http://snowball.tartarus.org/ algorithms/russian/stemmer.html
9. Natural Language Toolkit [Electronic Resource]. Mode of access: https://ru.wikipedia.org/ wiki/Natural_Language_Toolkit
10. Morphological analyzer pymorphy2 [Electronic Resource]. Mode of access: http://pymorphy2.readthedocs.io
11. brown-uk/dict_uk: Project to generate POS tag dictionary for Ukrainian language GitHub [Electronic Resource]. Mode of access: Режим доступу: https://github.com/brown-uk/dict_uk
12. LT2OpenCorpora – GitHub [Electronic Resource]. Mode of access: https://github.com/ dchaplinsky/LT2OpenCorpora
13. Rosette Text Analytics - AI for Human Language [Electronic Resource]. Mode of access: https://www.rosette.com /
14. polyglot 16.07.04 documentation [Electronic Resource]. Mode of access: http://polyglot.readthedocs.io/
15. AOT [Electronic Resource]. Mode of access: http://www.aot.ru/
16. Steven Bird. Natural Language Processing with Python/ Steven Bird, Ewan Klein, and Edward Loper. – O'Reilly Media. – 2009. – 479 c.
17. Shortening of words in the Ukrainian language in the bibliographic description. DSTU 3582-97. [Electronic Resource]. Mode of access: http://www.library.ukma.edu.ua/fileadmin/documents/Bibliography/26_DCTU3582-97.pdf
18. GitHub - lang-uk/ner-uk: Ukranian NER annotation project [Electronic Resource]. Mode of access: https://github.com/lang-uk/ner-uk

IMPLEMENTATION OF COMBINED METHOD IN CONSTRUCTING A TRAJECTORY FOR STRUCTURE RECONFIGURATIOIN OF A COMPUTER SYSTEM WITH RECONSTRUCTIBLE STRUCTURE AND PROGRAMMBLE LOGIC

Volodymyr Tokariev, Vitalii Tkachov, Iryna Ilina, Stanislav Partyka

Kharkiv National University of Radio Electronics,
61166, Ukraine, Kharkiv, Nauky Ave, 14,
tkachov@ieee.org, tokarev@ieee.org, ilirvi_71@ukr.net, stanislav.partyka@nure.ua

Abstract. In modern conditions providing the continuity of technological processes as well as increasing the reliability and survivability of operation of computer systems with configurable structure and programmable logic is one of the main strategic directions of modern social-economic and technical complexes. This is conditioned by the need to maintain the resilience and stability in the operation of computer systems with a reconstructible structure and programmable logic under various conditions of adverse effects of external and internal factors, both natural and technogenic, as well as those human-induced. It has been found that neutralizing threats and minimizing losses caused by abnormal, critical, emergency and catastrophic situations leading to an avalanche-like increase in degradation processes and destruction of computer systems with a reconstructible structure and programmable logic, requires the development of new principles, approaches, ways and methods of operational monitoring, analysis and forecasting of situations, development of options for control decisions, procedures for their selection and implementation in the framework of theory of structural

dynamics control. To solve the optimization problem of constructing scenarios for the structural reconfiguration of computer systems with a reconstructible structure and programmable logic, a method and an algorithm that implements this method are proposed. The novelty of the method is the combined use of the random guided search method and the method of cutting -off unpromising variants of structural reconfiguration of computer systems with a reconstructible structure and programmable logic such as « branch and bound». The proposed approach allows to solve the optimization problems of constructing scenarios for the structural reconfiguration of computer systems with a reconstructible structure and programmable logic, both unconditional and conditional, that can be described using various structural and topological indicators.

Key words: combined method, reconfiguration, functional element, structure, computer system, survivability, graph.

1. Introduction

The analysis of existing and predicted critical and emergency situations that are currently occurring everywhere in various subject areas shows that they stop being sectoral, and develop into accidents and disasters that are already intersectoral in nature. Under these conditions, it is necessary to investigate and solve the problems of increasing the reliability and survivability of computer systems with a reconstructible structure and programmable logic within the framework of the interdisciplinary approach. The causes that lead to the emergence of critical situations, accidents and disasters, which have natural-ecological, technical-industrial or anthropogenic-social causes, are of particular danger to modern computer systems with a reconstructible structure and programmable logic. Moreover, the range of threats to economic, physical and information security, as well as the list of vulnerabilities in the hardware-software and information structures of computer systems with a reconstructible structure and programmable logic is constantly growing [1-4]. Thus, in real life, situations are possible when these threats are combined and lead to an avalanche-like occurrence and development of negative events, ultimately resulting in catastrophic consequences. Under these conditions, ensuring the continuity of technological processes and increasing the reliability and survivability of the corresponding production systems is one of the most important strategic directions in the development of modern socio-economic and technical complexes. This is caused by the need to maintain the resilience and stability in the functioning of computer systems with a reconstructible structure and programmable logic under various conditions of adverse effects of external and internal factors, which can be anthropogenic and/or natural.

1. Analysis of known solutions

In practice, when solving the problems of ensuring survivability, computer systems with a reconstructible structure and programmable logic have received such an option for managing the structures of computer systems as reconfiguration. The standard (classical) technology for reconfiguring computer systems with a reconstructible structure and programmable logic in case of failure of one of its resources includes the following main steps [5-7].

Step 1. Determination and analysis of the time and place of a resource failure; removal of the function (task) performed on this resource from the solution; transfer of the function (task) to another resource (with/without saving the obtained intermediate results).

Step 2. Exclusion of a failed resource from the configuration of computer systems with a reconstructible structure and programmable logic; an attempt to replace it with a backup (of the same type), or a backup of a different type with similar functionality.

Step 3. Exclusion of links to a failed resource, denial of access to it and for the failed resource itself – an attempt to restore it. In case if a high-priority function (task) has been solved on a failed resource, which, when transferred to other resources, begins to conflict with functions (tasks) assigned to this resource, then, depending on the service discipline, the execution of less priority functions (tasks) is interrupted; or simple removal from the solution. Thus, in the framework of the standard reconfiguration in case of failures and violations of the correct functioning of the corresponding node in the computer system in order to maintain the highest priority functions of the specified object or acceptable operating conditions, they “sacrifice” other functions or part of the operable elements. In some cases, this reconfiguration is called a “blind” reconfiguration, since during its implementation, as a rule, the following operations are not carried out: accounting and analysis of the current characteristics of tasks and functions performed in the computer system; analysis and assessment of the current state of the computer system as a whole; operative calculation, evaluation and analysis of the target and information-technical capabilities of the computer system for a reasonable redistribution of the functions of a computer system between its operable elements and subsystems. A prerequisite for the implementation of standard reconfiguration technology is the presence of a multi-level model of the computer system. operability. A possible graphical description of this model for one of the subsystems (structures) is shown in Figure 1. In Figure 1, the vertices of the graph are associated with

the technical states (TSs) of a computer system. During operation, its elements can fall into various types of TSs: a functioning TS, a faulty TS, an operative TS, an inoperative TS, as well as correct and incorrect functioning. Figure 1 shows the main classes of the computer system TSs:

- the class of fully operative and functioning states:

$$S^{(u)} = \{S_{\alpha}^{(u)}\}, \quad \alpha = 1, ..., N^{(u)}.$$

Elements of this class differ from each other in the level of accumulated deviations from the norm of performance parameters in certain elements of the computer system;

- the class of partially operative states with the *B*th level of efficiency and the *y*th number of accumulated deviations. At the same time, it is believed that with an increase in the index number, the *B*th level of the computer system operability will decrease:

$$S^{(p)} = \{S_{By}^{(p)}\}, \quad B = 1, ..., N_1^{(p)}, \; y = 1, ..., N_2^{(p)}.$$

In the graph, operability levels are separated by horizontal dashed lines;

- the class of inoperative (failure) states of the computer system caused by the appearance of the *C*th type with the *N*th amount of losses:

$$S^{(o)} = \{S_{CN}^{(o)}\}, \quad C = 1, ..., N_1^{(O)}, \; N = 1, ..., N_2^{(O)}.$$

In Figure, 1 S_L is the limit state, i.e. the state, in which all elements of the computer system are inoperative and the level of losses is maximum.

In this Figure, the arcs of the graph correspond to transitions of the computer system from one state to another. These transitions, as noted earlier, have a different nature: some are structurally functional, others are not structurally functional. There exists the following classification:

- D-transitions, which are caused by failures of elements of a computer system and its transition into a state with a lower level of performance or a higher amount of losses;
- A-transitions, which are caused by the achievement of the established threshold values of the accumulated deviations number;
- R-transitions, which are caused by partial or complete recovery of the computer system operability;
- C-transitions, which are caused by compensation of the accumulated deviations in the parameters of both the elements and the computer system as a whole.

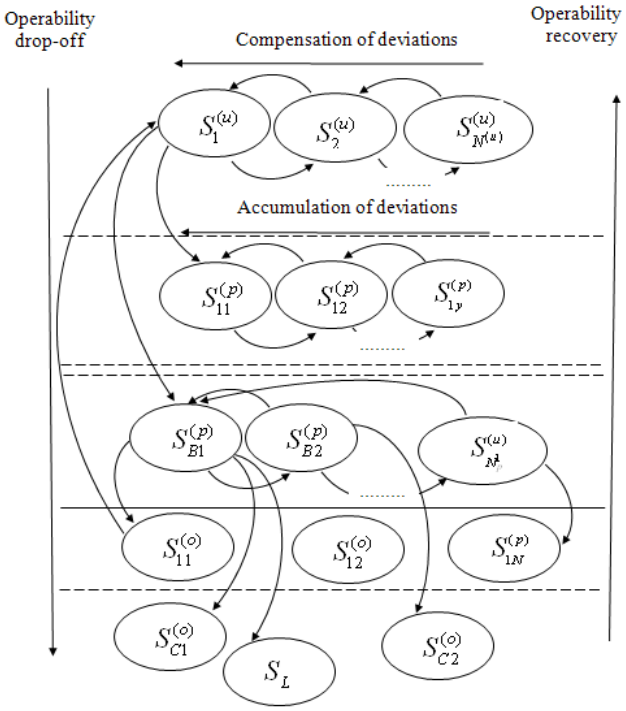


Figure 1 – The graph of evolution of computer system performance with a reconstructible structure and programmable logic

In Figure 1, D-transitions are depicted by arcs directed from top to bottom; A-transitions are given by horizontal arcs going from left to right; R-transitions are shown by arcs going from bottom to top; C-transitions are the arcs from right to left. The dashed line in Figure 1 marks the boundary between the operational and failure states of a computer system. Given the above, the dynamics of the computer system

operability in the general case assumes the simultaneous implementation of the following processes of changing the technical state of both elements of the computer system and the system as a whole:

- degradation processes (D-processes);
- recovery processes (R-processes); - processes of accumulating parameter deviations (A-processes);
- compensation processes of accumulated parameter deviations (C-processes).

In this case, one of the goals of controlling the structural dynamics of computer systems with a reconstructible structure and programmable logic is to provide the highest possible level of operability of a computer system and its elements at every moment in time. This goal is achieved by two complementary processes. First, by influencing (controlling) the D-process (degradation) the possibility (probability) of transitions of the computer system to the least desirable states is eliminated or reduced. Second, it is done by using the organization (management) of the R-process (recovery) and the C-process (compensation). To implement these processes both in the computer system itself and in each of its elements (subsystems), a structural control loop is formed. Due to this loop, the scheduling processes are realized by reconfiguring computer systems with a reconstructible structure and programmable logic. Speaking about the possible classes of tasks for reconfiguring computer systems, one should, first of all, proceed from the so-called functional-structural approach (FSA) to the description of objects of any nature (including computer systems). Characteristic of FSA features are:

- considering the dialectical relationship of functions and structure of objects with the decisive role of the function in relation to the structure;
- a holistic approach to the analysis and synthesis of multi-level systems;
- considering material, energy and information links between elements of the system;
- considering the relationship of the investigated (created) system with the external environment.

The analysis of problems of reconfiguring computer systems taking into account the FSA allowed to identify the basic concepts and definitions that are shown in Figure 2. At the same time, considering various combinations of independent branches of the given morphological tree, we can obtain a kaleidoscopic set of reconfiguration tasks for computer systems with a reconstructible structure and programmable logic.

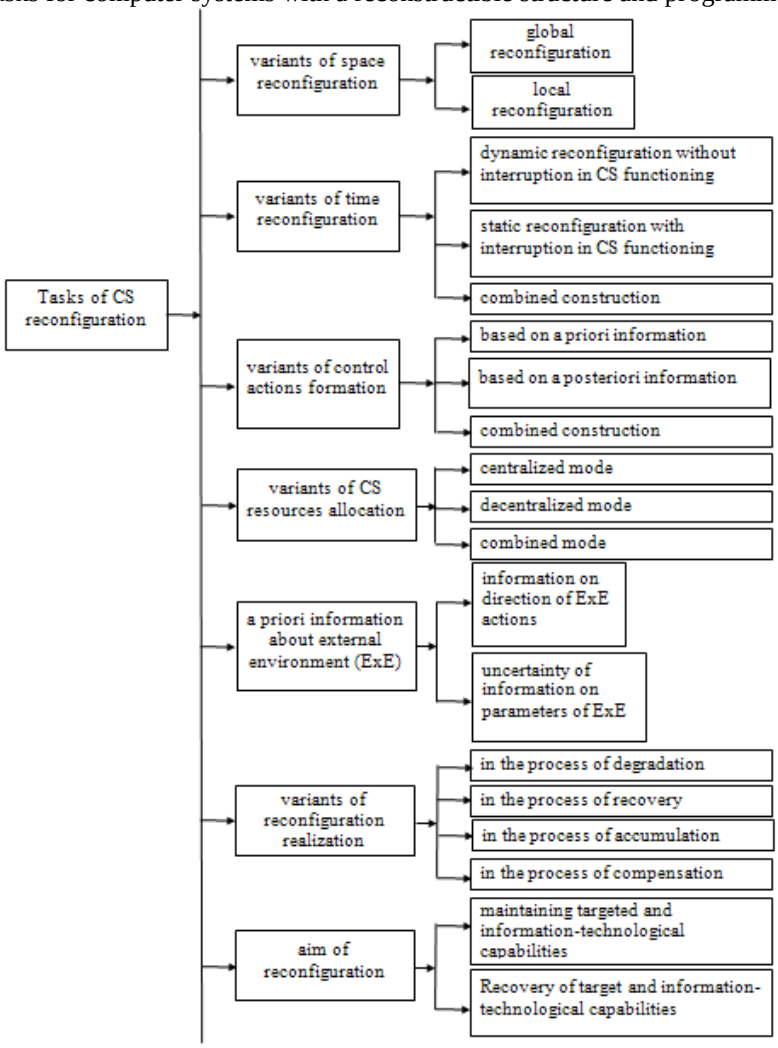


Figure 2 – The classification of reconfiguration tasks for computer systems with a reconstructible structure and programmable logic

These systems operate in conditions of significant non-stationarity and uncertainty associated with changes in the content of goals and tasks, which computer systems with a reconstructible structure and programmable logic face, the influence of disturbing factors from the external environment and factors having a purposeful and/or unfocused nature. Moreover, in real life, disturbing environmental factors can be deterministic or unknown, described using probability-statistical or fuzzy-probable formalisms. In the process of reconfiguring computer systems with a reconstructible structure and programmable logic, the following main goals are pursued: preservation, implementation and restoration of the target and information-technological capabilities of the system. The implementation of reconfiguration can be carried out simultaneously in several directions.

The "vertical" direction implies the implementation of a reconfiguration of a computer system when it affects two complementary processes. The first one is the process of computer system degradation in order to eliminate or reduce the possibility of a computer system transferring to the least desirable states. The second one is the recovery process in order to partially or fully achieve an operable state. The "horizontal" direction is the impact on the process of accumulating deviations of FE parameters (or the computer system as a whole) and the impact on the process of compensating for accumulated deviations. In addition, the following can be considered as the most essential morphological features of the process of a computer system reconfiguration: space, time, resource allocation mode and the formation of control actions. According to the "space" feature, the variants for reconfiguring a computer system can be distinguished into local and global. A local variant of reconfiguration of a computer system deals with individual FEs, sets of FEs and/or individual subsystems of a computer system. In the case of global reconfiguration, the process relates generally to the entire computer system. The "time" classification feature is closely related to the state of the computer system and its objects during reconfiguration. Static reconfiguration is performed after the computer system has been interrupted. The main disadvantages of this type of reconfiguration are: - demand to stop the computer system, which entails not only the shutdown of the computer system, but also the shutdown of the associated service objects; - reconfiguration of the computer system should be carried out by external means since in these conditions the internal resources of the computer system cannot be used. The advantages of the dynamic reconfiguration, which is performed during the operation of a computer system, are the ability to maintain the operability of its service subsystems during the reconfiguration of a computer system and the ability to use its internal resources and reserves [8-10]. The development of control actions during reconfiguration of a computer system can be based on both a priori and a posteriori information about the behavior of individual FEs and subsystems of a computer system. As for the options for reallocating the resources of a computer system in the process of its reconfiguration, the following modes can be distinguished: centralized, decentralized, or a combination of these modes. In the centralized mode, the allocation of resources is carried out using a unified model that includes both the initial state of the computer system and all kinds of (most likely) intermediate states caused by destructive influences. In the decentralized mode, the selection of a resource allocation plan for the initial state of a computer system is carried out in such a way that subsequent reallocation caused by adverse effects will allow the rational use of computer system resources. The determination of optimal control programs for the main elements and subsystems of a computer system can be performed only after the list of functions and algorithms for information processing and control becomes known, which should be implemented in these elements and subsystems [11-16]. In turn, the distribution of functions and algorithms by elements and subsystems of a computer system depends on the structure and parameters of the laws of managing these elements and subsystems. The evolving situation is closely connected with the principle of the duality of the past and the future. This means that you can have knowledge about the past, but you cannot control it, and you can design the future without knowing it.

2. Algorithm for solving the problem of constructing a trajectory for structural reconfiguration of computer systems with a reconstructible structure and programmable logic

The task of constructing a trajectory for the structure reconfiguration during the degradation or recovery of a computer system with a reconstructible structure and programmable logic can be represented as the following optimization problem:

$$\sum_{j=0}^{N} F_{failure}(\vec{x}_{a_j}) \to \min_{\substack{\vec{x}_{a_j} \in X(\vec{x}_{a_{j-1}}) \\ \vec{x}_{a_0}=\vec{x}_0,\ \vec{x}_{aN}=\vec{x}_f, \\ \Psi_l(\vec{x}_{a_j}) \le 0, l=1,2,...L \\ \{Q_{j1}, Q_{j2}, ..., Q_{jN}\}=\hat{Q}}} . \qquad (1)$$

To solve the considered problem of structural reconfiguration of a computer system, we use a combined method, which includes the method of random guided search and cutting off unpromising trajectories for structural reconfiguration of a computer system using the "branch and bound" method. The random guided search in solving the problem of scheduling the structural reconfiguration of a computer system can be implemented as follows. When building a random sequence:

$$\mu_{\zeta}^{(k)} = [\vec{x}_{a_0}, \vec{x}_{a_1}^{(k)}, \vec{x}_{a_2}^{(k)}, ..., \vec{x}_{a_{N-1}}^{(k)}, \vec{x}_{a_N}]. \quad (2)$$

of the intermediate states of the trajectory for structural reconfiguration, we are in some intermediate structural state, which is characterized by the genome $\vec{x}_{a_j}^{(k)}$ from which we can go to one of the states of the next level of the computer system structure degradation through the failure of one of the operable FE sets: $\{Q_{i_1}, Q_{i_2}, ..., Q_{i_p}\} \subseteq \hat{Q}$.

Moreover, the values $a_r = F_{failure}(\vec{x}_{i_r}^{+}), r = 1, ..., p$ characterize the contributions to the structural failure upon failure of these FEs. In the general case, $a_r \in [-1,1]$. To carry out the guided search, the interval [0,1] is divided into subintervals so that the coordinates of the ends of the subintervals form the sequence

$$0, \omega_1 = b_1, \omega_2 = \omega_1 + b_2, ..., \omega_p = \omega_{p-1} + b_p = 1. \quad (3)$$

where $b_r = \frac{1 - a_r}{\Lambda}$; $\Lambda = \sum_{r=1}^{p}(1 - a_r)$. Here, the smaller the contribution a_r to the value of the structural failure index when the FE is removed $Q_{i_r} \in \hat{Q}$, the greater the length of the sub-interval b_r is assigned to this element, i.e. the element Q_{i_r} is "encouraged" by the large sub-interval. For random selection of the FE removed from the structure, a random number ξ is generated that is distributed uniformly in the interval from 0 to 1, which falls into a certain sub-interval $\xi \in (\omega_{r-1}, \omega_r)$ and thereby determines the selection of the FE removed from the structure of the computer system. Thus, in accordance with the constructed probabilistic hypothesis, a search for a global extremum is organized. In addition, in the vicinity of the global extremum, the density of randomly selected trajectories for a reconfigurable computer system will be the highest. To increase the efficiency of random guided search, it is proposed for each test to compare the intermediate values of the total structural failure indicator obtained on this trajectory of the computer system reconfiguration with the record value and stop the started test as soon as the intermediate value of this indicator exceeds the record one. If the specified number of tests has been carried out, then the calculation is considered complete. The calculation can be considered complete even if subsequent tests do not provide an improvement in the result. Using the proposed approach allows to guarantee the reception of the trajectory of the computer system structural reconfiguration in a sufficiently small vicinity of the optimal trajectory. At the same time, the process of finding a solution can be limited both by the time allocated for making a decision, by the productivity of the used computing tools, and by analyzing the nature of the process of taken decisions convergence to the optimum. The proposed approach can be implemented in the form of an algorithm consisting of the following steps.

Step 0. The initial state. We set the number of statistical tests M; the number of tests *d* that does not give an improvement in the result; current test k = 0; the record $F_{failure}^{record} = +\infty$.

Step 1. The beginning of the next test. List of deleted critical FEs $\hat{Q} = \{Q_{i_1}, Q_{i_2}, ..., Q_{i_N}\}$; intermediate total structural failure $F_{failure} = 0$.

Step 2. Random selection of the deleted FE. One of the FEs is randomly selected from the list $\hat{Q}$ in the manner described above, let it be Q_{i_r}. The intermediate structural state of the current trajectory is determined $\vec{x}_{a_j}^{(k)}(\vec{x}_{a_{j-1}}^{(k)} \to \vec{x}_{a_j}^{(k)})$.

Step #3. Analysis of the intermediate structural state $\vec{x}_{a_j}^{(k)}$. The structural failure $F_{failure}(\vec{x}_{a_j}^{(k)})$ and the change in the current total structural failure $F_{failure} = +F_{failure}(\vec{x}_{a_j}^{(k)})$ are calculated. If $-F_{failure} > F_{failure}^{record}$ ($F_{failure}^{record}$ is the current record), then proceed to **Step 5**.

Step 4. Analysis of the list of deleted FEs. FE Q_{i_r} is excluded from the list $\hat{Q}$. If $\hat{Q} = \varnothing$, then the test is over. Moreover, if $F_{failure} < F_{failure}^{record}$, then $F_{failure}^{record} = F_{failure}$. If $\hat{Q} = \varnothing$, go to **Step 2**.

Step 5. Analysis of test results. If the number of consecutive tests with no results $F_{failure} > F_{failure}^{record}$ is greater than *d*, then the procedure ends. If k> M, then the procedure ends. Otherwise, go to **Step 1**. The finiteness of the algorithm constructed according to this scheme follows from the finiteness of the set $\hat{Q} = \{Q_{i_1}, Q_{i_2}, ..., Q_{i_N}\}$, the finiteness of the number of tests *M* and the sequential deletion of the FEs from the set $\hat{Q} = \{Q_{i_1}, Q_{i_2}, ..., Q_{i_N}\}$.

3.

4. **Conclusions and recommendations**

The analysis of the optimization problem features in constructing scenarios for the structural reconfiguration of computer systems with a reconstructible structure and programmable logic was carried out. The results have shown that methods and algorithms of the theory of search engine optimization of evolutionary type should be used to solve the considered problem. To solve the given problem, the method and the algorithm that implements this method are proposed, the novelty of which is the combined use of the random guided search method and the method of cutting off unpromising structural reconfiguration options for computer systems with a reconstructible structure and programmable logic such as “branch and bound”. The proposed approach allows solving the optimization problems of constructing scenarios for the structural reconfiguration of computer systems with a reconstructible structure and programmable logic, both unconditional and conditional, that can be described using various structural and topological indicators. At the same time, modification of the proposed method allows the construction of a series of optimistic and pessimistic scenarios for the structural reconfiguration of computer systems with a reconstructible structure and programmable logic. The research has been carried out on the basis of the Educational and Scientific Laboratory of “Reconfigurable and Mobile Systems” at the Department of Electronic Computers in Kharkov National University of Radio Electronics.

References

1. Churyumov, G., Tkachov, V., Tokariev, V., Diachenko, V.: Method for Ensuring Survivability of Flying Ad-hoc Network Based on Structural and Functional Reconfiguration. In: XVIII International Scientific and Practical Conference «Information Technologies and Security» (ITS 2018), pp. 64-76. Kyiv, Ukraine (2018).
2. Tkachov, V., Tokariev, V., Dukh, Ya., Volotka, V.: Method of Data Collection in Wireless Sensor Networks Using Flying Ad Hoc Network. In 5th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology, pp 197-201, Kharkiv, Ukraine (2018).
3. Dodono,v A., Gorbachyk, O., Kuznietsova, M.: Management organization of mobile technical objects group / In: XVII International Scientific and Practical Conference «Information Technologies and Security» (ITS 2017), pp. 1-7. Kyiv, Ukraine (2017).
4. Dodonov, A., Gorbachyk, O., Kuznietsova, M.: Increasing the survivability of automated systems of organizational management as a way to security of critical infrastructures. In: XVIII International Scientific and Practical Conference «Information Technologies and Security» (ITS 2018), pp. 261-270. Kyiv, Ukraine (2018).
5. Barabash, O., Kravchenko, Y., Mukhin, V., Kornaga, Y., Leshchenko, O.: Optimization of Parameters at SDN Technologie Networks. International Journal of Intelligent Systems and Applications. 9 (9), 1-9 (2017).
6. Lemeshko, O., Yeremenko, O., Tariki, N.: Solution for the Default Gateway Protection within Fault-Tolerant Routing in an IP Network. International Journal of Electrical and Computer Engineering Systems. 8 (1), 19–26 (2017).
7. Bielievtsov, S., Ruban, I., Smelyakov, K., Sumtsov, D.: Network technology for transmission of visual information // In: XVIII International Scientific and Practical Conference «Information Technologies and Security» (ITS 2018), pp. 161-275. Kyiv, Ukraine (2018).
8. Pavlov, A.N.: Metodologicheskiye osnovy resheniya problemy planirovaniya struktur-no-funktsional'noy rekonfiguratsii slozhnykh ob"yektov. Izvestiya Vysshikh Uchebnykh Zavedeniy. Priborostroyeniye, 55 (11), 7-13.
9. Dodonov, A.G., Kuznetsova, M.G., Gorbachik, Ye.S.: Vvedeniye v Teoriyu Zhivuchesti Vychis-litel'nykh Sistem, Nauk. Dumka, Kyiv (1990).
10. Dodonov, A.G., Lande, D.V.: Zhivuchest' Informatsionnykh Sistem. Nauk. Dumka, Kyiv (2011).
11. Dodonov, A.G., Kuznetsova, M.G.: O nekotoroj strategii rekonfiguracii zhivuchih vychislitelnyh sistem. Gibridnye vychislitelnye mashiny i kompleksy, 11, 31-35 (1988).
12. Dodonov, A.G., Kuznetsova, M.G.: Dodonov, A.G., Kuznetsova, M.G.: Raspredelenie resursov v zhivuchih vychislitelnyh sistemah. Gibridnye vychislitelnye mashiny i kompleksy, 7, 20-23 (1984).
13. Westmark, V. R.: A definition for information system survivability. In 37th Annual Hawaii International Conference on System Sciences, p. 10 (2004).
14. Romashkova, O.N., Dedova Ye.V.: Zhivuchest' besprovodnykh setey svyazi v usloviyakh chrezvychaynoy si-tuatsii. T-Comm: Telekommunikatsii i Transport, 6, 40-43 (2014).
15. Lande, D.V.: Metodi pidvishennya zhivuchosti informacijnoyi skladovoyi korporativnih informacijno-analitichnih sistem pidtrimki prijnyattya rishen. Reyestraciya, zberigannya i obrobka danih, 14 (2), 48-58 (2012).
16. Dodonov, A.G., Lande, D.V., Putyatin V.G. Informacijni potoki v globalnih komp'yuternih merezhah. Naukova dumka, Kyiv (2009).

PROTECTION DATA TRANSMISSION SYSTEMS FROM THE INFLUENCE INTERSYMBOL INTERFERENCE SIGNALS

Volodymyr Yartsev[1] and Dmytro Hololobov [2]

[1] *State University of Telecommunications, Kyiv, Ukrain*
jvp57@ukr.net
[2]*Taras Shevchenko national University of Kyiv , Ukrain*
vobd@ukr.net

Abstract. The issues of creating a regenerator of a optical system for transmitting discrete messages in order to protect against the influence of intersymbol interference of signals using methods of the statistical theory of mixtures are considered. A demodulator adaptation model was created and studied that calculates the weighted sum of 4 samples of the input signal and compares the resulting sum with a threshold value. The basic operations for adaptation of the regenerator during demodulation of signals distorted by inter-character interference are determined under specific conditions of transmission, which are formulated from the point of view of statistical mixture theory as a task of estimating some parameters of the probability distribution of the mixture. A block diagram of an optical signal regenerator is proposed, which implements a digital processing algorithm for the mixture. The analysis is performed and the requirements for the parameters of the analog-to-digital converter, which is used in the regenerator, are determined. The results of simulation of digital processing of a discrete test message are presented to verify the process of adaptation of the device to the influence of intersymbol distortions of signals. The structural scheme of the digital processing device is substantiated, in which the fast-acting nodes that perform real-time processing are combined with a microprocessor that implements an algorithm for adapting the regenerator to specific conditions of message transfer conditions.

Keywords: statistical mixture theory, fiber optic communication lines, intersymbol interference signal, demodulator adaptation, optical signal regenerator.

1 Using mixture theory to eliminate reduces the effects of intersymbol interference

1.1 Basic concepts of mixtures

The protection of information systems for the transmission of discrete signals from the effects of inter-symbol interference (IIS) remains one of the pressing problems in connection. IIS causes distortion of the signal corresponding to the i-th transmitted symbol, by signals corresponding to the i-1, i + 1, ... symbol. This makes it difficult for the demodulator to make the right decision about the meaning of each transmitted character. To eliminate the influence of IIS, analog and digital signal correctors are used. Analog correctors of linear distortion are used to neutralize IIS, which compensate for the imperfection of the frequency characteristics of the communication channel. Operations with discrete samples of the received signal are carried out by a demodulator, which actually performs the functions of correction in the time domain.

The transmission of a discrete information often is accompanied by an IIS which can be caused by an imperfection of frequency responses of the channel or is created it artificial. In any case the presence of IIS hampers demodulation of signals and forces to use rather complicated algorithms of handling [1].

The principles of constructing correctors and demodulators to neutralize IIS do not take into account the stochastic nature of the signal, which is due to additional signal distortions due to random noise and the pseudo-random nature of the sequence of transmitted symbols. The concept of ISI neutralization with the help of correctors loses its meaning in situations when interference is artificially created by using signals with a partial response. Solving IIS problems is possible while simultaneously performing the tasks of correction and demodulation with minimization of errors. This is possible when using a method of protection based on a probabilistic interpretation of the phenomena of IIS and using the concepts and methods of the statistical theory of mixtures[2].

The use of methods of the statistical theory of mixtures opens new possibilities in a solution of problems caused IIS.

Let is given a parametric family $\mathbf{F} = \{f(x; A): A \in \mathbf{A}\}$ of probability distributions of random variables. The elements $f(x; A)$ of this family are described by an identical known analytical dependence from x (value of an random variable) and differ only by values of some parameters A, which belong to a set $\mathbf{A}$ of admissible values. In a common case the probability distribution $f_P(x)$ of a infinite mixture from a family $\mathbf{F}$ is defined by expression.

$$f_P(x) = \int_{\mathbf{A}} f(x; A) dP(A), \qquad (1)$$

where $P(A)$ - function of mixing distribution. If the whole mass of distribution $P(A)$ is concentrated in a finite number Q of points $A_1,...,A_Q$, with which there correspond probabilities $p_1,...,p_Q$, the expression (1) will be transformed to a probability distribution of a finite mixture

$$f_P(x; p_1,..., p_Q, A_1,..., A_Q) = \sum_{j=1}^{Q} p_j f(x; A_j). \qquad (2)$$

The given expression statistically describes a mixed sample $\mathbf{X}_P=\mathbf{X}_1\cup...\cup\mathbf{X}_Q$ with volume $N = N_1 +... + N_Q$, in which in an arbitrary sequence the elements of partial samples $\mathbf{X}_1,...,\mathbf{X}_Q$ are mixed. Each partial sample $\mathbf{X}_j$ consist from N_j of random variables, which submit to distribution $f(x;A_j)$ and mix up in a proportion $p_j=N_j/N$ [3].

In a common case the statistical theory of mixtures allows to solve the following tasks by handling concrete sample $\mathbf{X}_P$:

1) to determine number Q of partial samples, which have formed a mixture;

2) to find evaluations of probabilities $p_1,...,p_Q$ and parameters $A_1,...,A_Q$ of partial distributions;

3) to accept a solution about membership of each element x_i (i = 1,...,N) of mixed sample $\mathbf{X}_P$ (that is, to what partial sample from $\mathbf{X}_1,...,\mathbf{X}_Q$ it belongs).

The theory of mixtures defines a population of conditions, which realization ensures a possibility of a solution of tasks 1 and 2 (the task 3 is reduced to a typical task of statistical solutions). A major of them are the conditions of identifiability, which consist in the requirement of linear independence of the elements of a family $\mathbf{F}$ [1]. To these conditions satisfy many families of probability distributions, in particular, normal and uniform distributions [2,4].

It is important to note two circumstances. At first, the property of identifiability is stipulated by a not stochastic nature of function $f(x; A)$, but by a peculiarity of dependence from arguments x and A. Secondly, both appearances - IIS and mixing of samples of random variables - represent an outcome of linear transformations of partial actions. The specified circumstances justify formal an identification of physical signals with probability distributions of random variables.

1.2 Example of statement of a task of demodulation of signals

Let's consider the communication channel with amplitude modulation, on which the sequence of symbols "1", "0" with a velocity $1/T$ bit/sec (where T - duration of a bit interval) is transmitted. Let for "1" the impulse $g(t,\theta)$ of the Gaussian form is formed, where the function $g(t,\theta)$ is identical to analytical exposition of a normal probability distribution with parameters $m = \theta T$, $\sigma = 0.425T$. Thus the breadth of impulse at a level 0.5 is equal T. The values of function $g(t,\theta)$ are truncated up to zero with $t = m \pm T$. Such signal borrows 2 bit intervals and well reproduces influence of dispersing distortions in optical fiber communication systems. Let's designate $G(t)$ - signal, which will be formed of a transmitted sequence "1" and "0" in an outcome of superposition of the appropriate impulses $g(t,\theta)$. The additional distortions are caused by random noise $N(t)$ with zero expectation and standard deviation σ_n.

Let's assume, that the demodulator produces references of a signal

$$S(t) = G(t) + N(t) \qquad (3)$$

in discrete instants t_m with a periodicity $dt = 0.5T$, and that receiver is synchronized with the transmitter of a signal and the references are produced on the boundary and in a middle of bit intervals. Let's designate the references as $s_m = S(t_m)$.

The elementary method of demodulation is *the comparison of values s_m with a threshold value sd* (method CWT). That is, the solution r_i about presence " 1 "or" 0 " on a bit position with number i, should be: $r_i = 1$, if $s_{2i+2} > s_d$, otherwise $r_i = 0$. However thus is not taken into account, that the useful information is contained not in one, but in all references on stretch of an interval $2T$ of existence of a signal $g(t,\theta)$, owing to what reliability of demodulation considerably decreases.

The best outcomes can be received if to accept *maximum likelihood solutions* (method MLS). For a solution about a bit position with number i it is necessary to analyze $M = 4$ references

$$s_{2i}, s_{2i+1}, s_{2i+2}, s_{2i+3}, \qquad (4)$$

which hit in " a sliding window " by duration $2T$, and also to take into account influence of the following i+1-st and previous i-1-st of a symbol. Thus, the values of references in "window" of the analysis depend on a combination "1" and "0" on 3 positions. For all possible $2^3 = 8$ combinations are necessary to generate the

etalons of a signal $e_{h,0},...,e_{h,3}$ (h = 0,...,7) as values of process $G(t_{2i})$, ..., $G(t_{2i+3})$ in expression (3). For normal distributions of noise $N(t)$ the solution accepts such aspect r_i = central bit of 3-digit binary number j

$$j = \arg \min_{h=0,...,7} \sum_{m=0}^{M-1} (s_{2i+m} - e_{h,m})^2 \qquad (5)$$

Shortcoming of such method of demodulation is rather large volume of calculations.

Now we shall consider a solution of a task of demodulation *on the basis of statistical theory of mixtures* (method STM). As are known number of impulses in "window" of the analysis (Q = 3), their form and position, it is enough to solve only task 2, where is necessary to estimate probabilities p_1, p_2, p_3 . In this case sequence of references (4) plays a role of an empirical probability distribution of sample of random variables from a general population, which submits to a discontinuous distribution of probabilities of a mixture

$$f_m(p_1,...,p_Q) = \sum_{j=1}^{Q} p_j f_{m,j}, \qquad (6)$$

where p_j - parameter of presence of a symbol "1" or "0" on position j (accordingly $p_j \neq 0$ or $p_j = 0$), and $f_{m,j}$ - values of a partial discontinuous distribution are equal to values of function $g(m \times 0.5, j\text{-}1)$.

It is necessary to find estimations $p_1^{(i)},...,p_Q^{(i)}$ of parameters $p_1,...,p_Q$ by handling empirical distribution (4). For an estimation by a method of least squares [3] they correspond to a minimum of a criterion

$$D(p_1,...,p_Q) = \sum_{m=0}^{M-1} (s_{2i+m} - f_m(p_1,...,p_Q))^2 = \sum_{m=0}^{M-1} (s_{2i+m} - \sum_{j=1}^{Q} p_j f_{m,j})^2. \qquad (7)$$

If to equate partial derivatives of a criterion (7) on $p_1,...,p_Q$ to zero, then obtain a system of simple equations

$$\sum_{j=1}^{Q} p_j c_{j,h} = b_h^{(i)}, \qquad h=1,...,Q, \qquad (8)$$

where

$$c_{j,h} = \sum_{m=0}^{M-1} f_{m,j} f_{m,h} \qquad b_h^{(i)} = \sum_{m=0}^{M-1} s_{2i+m} f_{m,h} \qquad (9)$$

The solution of a set of equations (8) gives all estimations $p1^{(i)},...,pQ^{(i)}$ however it is expedient to use only one of them $p2^{(i)}$ which it is possible to calculate according to formula

$$p_2^{(i)} = \sum_{j=1}^{Q} ci_{2,j} b_j^{(i)}, \qquad (10)$$

where ci - elements of a matrix, inverse in relation to a matrix $\left\| c_{j,h} \right\|$. These matrixes can be calculated beforehand. The rule of a solution for the given method consists in check of inequalities: if $p_2^{(i)} > pd$, then r_i = 1, otherwise r_i = 0. It is easy to see, that volume of calculations of magnitudes $b_h^{(i)}$, $p_2^{(i)}$ is significantly smaller on a comparison with volume of calculations for a method MLS.

1.3 Outcomes of statistical simulation

For the considered methods of demodulation, the statistical experiment is carried out, which outcomes are submitted on Fig.1. The axis of abscissas corresponds to a ratio SN of a power of a useful signal $G(t)$ to a power of noise $N(t)$ (in decibels). The axis of ordinates corresponds to probability of errors of demodulation for methods CWT, MLS and STM. In experiment is taken into account, that the moments of references of a signal are displaced on $0.2T$ concerning of specified above positions owing to a phase lag of the circuit of a synchronization. The threshold values are equal $sd = pd = 0.5$. As well as it was necessary to expect, the method CWT is least effective, and the method MLS is most effective and can be considered as an optimum method of demodulation. With magnification of a ratio a signal / noise the winning in effectiveness is increased and with SN = 18 dB the method MLS ensures a diminution of errors of demodulation on 3 order on a comparison with a method CWT. The effectiveness of a method STM concedes a little to method MLS, however also considerably exceeds a method CWT.

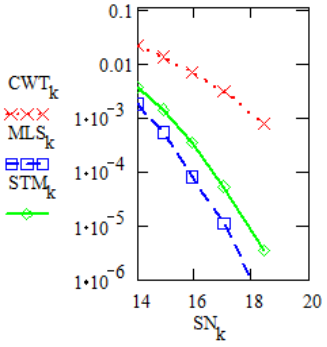


Fig.1. Statistical experiment results

On Fig.2 the diagrams of time of simulation are represented. The lower part of a column of each diagram corresponds to expenditures of time on simulation of signals, upper part - expenditures of time on operation of demodulation. From the diagrams it is visible, that the method STM requires approximately in 8 times of smaller volume of calculations, that is the important argument for the benefit of its choice.

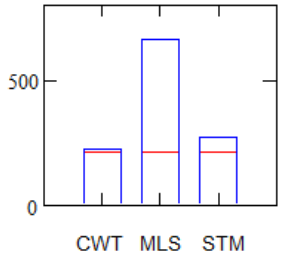


Fig.2. Diagrams of time of simulation

The results of statistical modeling showed that:

- the offered method STM based on use of the statistical theory of mixtures, ensures the almost same effectiveness of demodulation, as well as the method MLS of maximum likelihood demodulation, however requires approximately on an order of smaller volume of calculations;
- the method STM can be spreaded to more complicated cases, for example, when it is required to estimate a position of signals in "window" of the analysis (with a solution of a task 2 it is required to estimate not only probability $p_1,...,p_Q$, but also parameters $A_1,...,A_Q$).

2 Technical solution for eliminating inter-character signal interference in the communication information channel

2.1 Model of a device for reducing the influence of inter-symbol interference in an optical information transmission channel

The traditional means of combating inter-symbol interference of signals are the use of various analog linear distortion correctors that compensate for the imperfection of the amplitude-frequency characteristics of the communication channel. Digital corrections are also used, which directly operate on the signal as a function of time [5].

The general block diagram of an optical signal regenerator with digital processing that uses a conversion algorithm based on a mixture theory is shown in Fig. 3.

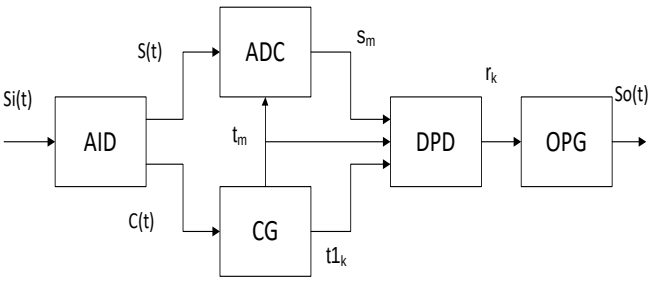


Fig. 3. Block diagram of an optical signal regenerator

The regenerator circuit consists of such elements. The regenerator circuit consists of such elements:

AID - analog input devices that receive Si(t) optical signals, detect them and filter them. Received optical signals Si(t) is converted using a fiber optic converter (OE converter), at the output of which an electric signal S(t) is generated, which contains information about the sequence of transmitted symbols “0” and “1”, as well as a synchronization signal C(t).

CG is a clock generator that generates the clock sequences $t1_k$ (one pulse per bit interval) and tm (nb pulses per bit interval) from the signal C(t).

ADC is an analog-to-digital converter that encodes samples $s_m = S(t_m)$ in the form of an n_r-bit binary code.

DPD is a digital processing device that implements adaptation and demodulation operations; in the demodulation mode, it gives a decision $r_k = 1$ or $r_k = 0$ for each k-th bit interval;

OPG is an optical pulse generator that, in accordance with the solution rk, generates the reconstructed optical pulses So(t) at the output of the regenerator.

Regardless of the technical implementation of the CPU, it is important to determine the time discreteness of the counts of the signal S(t), that is, the number of readings nb during the bit interval and the discreteness of the readings of the amplitude sm, that is, the number of nr encoding bits. The first parameter must satisfy the inequality $n_b = 2$.

However, the increase in n_b necessitates an increase in the speed of ADC and DSP, which complicates the regenerator. Moreover, the increase in nb is justified only if the independence of the random distortion of the signal S(t) by the noise N(t) is maintained for each reference. If the WUA performs a consistent filtering of the signal S(t), then the random components of the counts are found to be correlated more significantly, the greater the n_b. Given these features, it is advisable to take $n_b = 2$ [5].

Increasing the number of n_r encoding bits complicates ADCs and DTCs if the demodulation algorithm has a technical implementation. Due to low-bit coding, rather rough "images" of the signal are formed during the adaptation phase and this reduces the reliability of demodulation. If, as the number of digits increases, the encoding discreteness decreases to values smaller than the rms noise σ_n, then a further increase in n_r becomes useless. Thus, determining the minimum number of encoding bits that provides acceptable demodulation accuracy is one of the important tasks of the technical implementation of the proposed processing algorithm [6,7].

The signal amplitude sampling in the ADC can be described by the following function:

$$\text{kvant}(x) = \begin{cases} \dfrac{Nr-1}{Nr}, & \text{якщо } x > \dfrac{Nr-1}{Nr} \\ \dfrac{[x \cdot Nr]}{Nr}, & \text{якщо } 0 \le x \le \dfrac{Nr-1}{Nr} \\ 0, & \text{якщо } x < 0 \end{cases}$$

where [n] is the integer part of n; $Nr = 2n_r$ is the number of quantization levels. The operational value of argument x is in the range $0 \le x \le 1$.

When processing digital sample values, the required number of bits must exceed n_r by several units. If the CPU is based on microprocessors, then this requirement is met, since modern signal processors have at least 64 bits. To test the effect of the number of coding bits nr, statistical modeling was performed.

The discretization of the readout values is taken into account in the form of replacement of sm by quant (s_m) in the calculation of the array of accumulated readings s_m for adaptation of the regenerator, which is performed according to the formula:

$$sm_m = \frac{1}{Nw} \sum_{j=0}^{Nw-1} s_{6j+m}, \text{ m = 0, ..., 5.} \qquad (11)$$

Quantum (s_m) is also used to calculate the probability estimate $p_2(k)$ that the symbol "1" is in the kth position when demodulating the signals:

$$p_2(k) = \sum_{m=1}^{M} c1_{2,m} \cdot s_{k \cdot n_b + m - 1}., \qquad (12)$$

All other calculations were performed with the maximum accuracy possible in MathCAD14 [8].

To adapt the regenerator, a sequence of repeating triples of symbols "010" (TPS1) was used, which consisted of $N_w = 257$ triples of symbols "010". In each demodulation simulation session, a sampling of signals over a 256-bit interval was generated. For the demodulation error statistics set, the simulation sessions were repeated 400 or 4000 times. Thus, the total number of characters that were modeled in each session group reached 100,000 or 1,000,000. The simulation results are shown in Table. 1.

The first column presents the noise intensity in the form of a value of σ_n. In the second column, the signal-to-noise ratio in dB is calculated as 20lg ($1 / \sigma_n$), given that the maximum amplitude of the useful

signal C(t) is approximately 1. Other columns provide estimates of the probability of demodulation errors for different numbers of n_r encoding bits in the ADC.

Table 1. Results simulation regenerator test operation

σ_n	S/ N	$n_r = 2$	$n_r = 3$	$n_r = 4$	$n_r = 6$	$n_r = 10$	$n_r = 15$
0.2	14	0.0043	0.0030	0.0028	0.0030	0.0030	0.0035
0.18	14.9	0.0019	0.0011	0.0011	0.0011	0.0011	0.0015
0.16	15.9	0.00062	0.00034	0.00024	0.00024	0.00025	0.00041
0.141	17	0.00015	0.00008	0.00002	0.00004	0.00004	0.00008
0.12	18.4	0.00003	0.00005	0.00001	0.00001	0.00001	0.00001

As we can see, with a small number of digits $n_r = 2$, the errors are, as expected, maximal. However, at $n_r = 4$, they reach a minimum. Further increase in the number of digits of the coding does not actually improve the accuracy of demodulation, but on the contrary, worsens it. For example, for the highest possible accuracy of calculations in MathCAD14, which corresponds to 15 decimal places ($n_r = 15$), the results are even worse than for $n_r = 4$.

This can be explained by the fact that quantization makes the "image" of the signal coarser. On the other hand, quantization negates the effect of noise in the range of ΔS values between quantization levels. If S << n, no leveling occurs. At the same time, it is obvious that, as noise levels decrease, it is advisable to reduce ΔS, that is, to increase n_r. Given that the simulation noise level exceeds the actual noise level in the regenerators (since it is impossible to obtain statistically reliable estimates of error probabilities at low noise), 5 digits of coding in the ADC can be considered sufficient.

Digital processing operations associated with adaptation of the regenerator and demodulation of the signal can be performed using modern signal processors.

Such a microprocessor embodiment of digital processing of the received signal is technically the simplest and cheapest. However, in this case, only sequential execution of operations is possible. This limitation is related to the actual DPD speed. However, the most critical in terms of performance are two groups of operations directly related to the processing of s_m counters. These are operations of accumulation of sum_s (11) at the stage of adaptation and operations of calculating the probabilistic estimate p_2 (k) (12).

At the demodulation stage, high speed is required in the decision-making operation r_k, where it is necessary to compare the estimate obtained with the decision threshold p_{por} in terms of:

$$r_k = \begin{cases} 1 & \text{if } p_2(k) > p_{por}, \\ 0 & \text{— else.} \end{cases} \qquad (13)$$

Due to the simplicity of these operations, it is advisable to create specialized nodes for their execution, which provide minimal processing time, in particular due to the organization of parallel calculations. Such a principle and the order of processing can be recommended. Firstly, it is necessary to form a sufficiently long test signal sequence (TPS1) to adapt the regenerator. Moreover, only the initial fragment with a length of 3·*Nw* characters is actually used to obtain information about the properties of the signals. On this fragment, it is necessary to form partial sums (for each m = 0, ..., 5) in expression (11). Corresponding operations must be performed in real time by the accumulation unit (AU). After the initial fragment, the sequence of TPS1 should continue for a time exceeding the time required by the microprocessor calculator (MPC) to perform all subsequent operations, including the calculation of the coefficients $c1_{2.1}$ ..., $c1_{2.4}$, which are used in expression (12). During this time, which may be called the "adaptation period" (PA), the input signals of the regenerator are ignored. To determine the duration of the PA interval need to carry out simulations of MPC. After the end of the PA interval, regenerator enters the operating mode of signal processing in accordance with expressions (11), (12). This processing is performed by a separate demodulation node (DN).

The block diagram of a possible option for constructing such a hardware-software DPD is shown in Fig. 2. According to the results of simulation, we take the number of bits of the ADC coding $n_r = 5$, and the length of the initial fragment of TPS1 is 768 characters ($N_w = 256$).

The accumulation node (AN) consists of the following elements:

- IRg - input 5 - bit register, which receives s_m signals from the ADC;
- ARg1 ..., ARg6 - 13-bit registers of accumulation of private sums (11). The number of digits is defined as 5 + log2 (N_w) = 5 + 8 = 13;
- RgP - 13-bit intermediate register, with the help of which communication with ARg1 ..., ARg6 and MPC is carried out;

- SUM1-13-bit adder adds the next signal from the ADC to the corresponding partial sum.

The received amounts after completion of accumulation are transferred using RgP to microprocessor calculator.

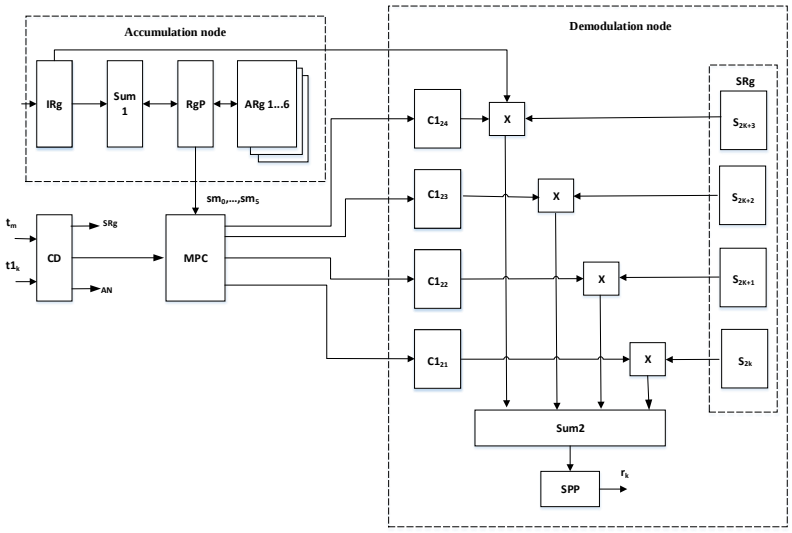


Fig. 2. Block diagram of a digital processing device

The numbers with which this node operates are treated as non-negative integers (with a point fixed after the least significant bit). The result of the calculations of the operations performed by the MPC are the coefficients $c1_{2.1}$ ..., $c1_{2.4,}$ written to the corresponding registers $C1_{21,}$ ..., $C1_{24}$ of the demodulation node (as 16-bit numbers with a fixed or floating point).

The demodulation node (DN) consists of the following elements:

- $C1_{21}$, ..., $C1_{24}$ - registers with weight coefficients, which are determined at the stage of adaptation of the regenerator;
- SRg - shift register from 4 cells to 5 bits, where sm samples are received in the operating mode (each clock pulse tm shifts information by one cell);
- MUX is a 16-bit multiplier that calculates each sum of (12), where s_m is treated as fractional numbers with a point fixed before the highest digit;
- SUM2 is a 16-bit adder that accumulates the sum (12);
- SPP is a comparison scheme with the threshold $p_{por} = 0.5$, which gives the solution r_k in accordance with (13).

The solution is the value of the discharge, which is located to the right of the point separating the integer and fractional part of the number at the output of the adder. The operations of multiplication and summation in the demodulation node are performed once during the bit interval, that is, by the clock cycle $t1_k$.

The control device (CD) synchronizes the interaction of all elements of the circuit using clock pulses t_m, $t1_k$.

According to the logical sequence of signal processing in the DPU, the microprocessor calculator starts working after the accumulation node completes its work. However, it must be taken into account that the MPC and AN start to function simultaneously after turning on the power or the receipt of a special signal to start operation. While AN carries out the accumulation of TPN1, it is advisable to provide for testing MPC. Since the ratio of accumulation time and testing time is unknown, we will proceed from the following interaction principle:

- AN, after completing its work, transfers to the MPC the accumulated amounts sm1, ..., sm6 and the readiness flag, which are recorded in the MPC memory in the interrupt mode;
- After completion of testing, the MPC checks the sign of readiness and, if there is one, proceeds to the processing of information, and in the absence, the signs of readiness are expected.

After completing all the processing associated with the adaptation of the regenerator, the MPC issues the obtained weight coefficients $c1_{21}$ ..., $c1_{24}$ to the demodulation unit and gives the adaptation termination flag to the control device, by which the latter opens the arrival of the s_m signal samples to the shift register in the ID and after 2 bit intervals (when the register is full) allows the operation of a comparison scheme with a threshold that produces a solution r_k.

Conclusions

The basic operations of protecting the regenerator of a signals transmission system that have distortion due to intersymbol interference can be considered from the point of view of the statistical theory of mixtures as a task of estimating some parameters of the probability distribution of the mixture.

Due to a significant decrease in the effective bandwidth of the communication line, as the length of the link increases, a significant pulse expansion occurs and its shape becomes bell-shaped, which causes significant inter-symbol interference of the signals.

It is proposed to adapt the digital corrector to the features of a particular communi-cation line using a special test sequence of characters that is generated at the begin-ning of the communication session and allows you to create a "image" of the distort-ed signal in the communication line.

To adapt the regenerator need to accumulate 6 averaged real-time counts that correspond to three bit intervals (2 counts per bit interval). Then need center and highlight 4 of the 6 averaged counts. Selected readings form an "image" of a real signal in the form of a response of the transmission line to a single character "1", which, as a result of inter-character distortions, takes about 2 bit intervals. Finally - to calculate the values of the weighting coefficients for the obtained "image" of the real signal.

References

1. Steklov, V. Theory of electrical communication. Kiev, Technics (2006).
2. Milenky, A. Classification of signals in conditions of uncertainty. Moscow, Soviet Radio, (1975).
3. Andreev, A., Baushev, S. State of the theory and practice of using signals with a partial response. Foreign Radioelectronics. Radio and communications 9, 57-83, (1992).
4. Teicher, H. Identifiability of finite mixtures. Ann.Math.Stat 34(4),126-129, (1963).
5. Sklyar, B. Digital communication. Theoretical basis and practical application. Moscow, Williams, (2003).
6. Zasov, V. Algorithms and computing devices for separating and reconstructing signals in multidimensional dynamic systems. Samara: SamGUPS, (2012).
7. Sergienko, A. Digital signal processing. St. Petersburg, Peter, (2003).
8. Makarov, E. MathCAD. Training course. SPb, BHV-Petersburg, (2007).
9. Sounduchkov, A., Fadeeva E.A., Chests, A. Transmission rate, interchannel and intersymbol distortions. http://ena.lp.edu.ua last accessed 2019/10/24.
10. Yartsev, V. Use of the statistical theory of mixtures for the fight against inter-sumbol interference of signals in the FOTS. Zv'yazok 3, 71-80, (2018).

МОДИФІКОВАНИЙ МЕТОД ЕКСПОНЕНЦІАЛЬНОГО ЗГЛАДЖУВАННЯ ДЛЯ ФІЛЬТРАЦІЇ КУРСУ РУХОМИХ ОБ'ЄКТІВ ПРИ ЇХ СУПРОВОДЖЕННІ

Володимир Юзефович[0000-0002-6336-9548]

Інститут проблем реєстрації інформації НАН України, Київ, Україна
uzefv71@gmail.com

Abstract. У статті запропоновано модифікований метод експоненціального згладжування із змінним коефіцієнтом для згладжування значень курсу малорухомих об'єктів, розрахованих за рівноточними та нерівноточними вимірюваннями (оцінками) координат таких об'єктів. Запропонований метод дозволяє скоротити налаштування фільтру на дійсне значення курсу у середньому на 5-6 циклів оновлення інформації, при цьому забезпечується додаткове зменшення середньоквадратичної похибки оцінки курсу.

Keywords: супроводження рухомих об'єктів, курс руху, експоненціальне згладжування, рівноточні вимірювання, нерівноточні вимірювання, фільтрація.

Вступ

В рамках багатьох систем безпекового сектору, у тому числі й військового призначення, розгортаються системи моніторингу рухомих об'єктів (РО), що складаються із різнотипних засобів розвідки навколишнього простору та засобів обробки отриманої від них інформації. Метою таких систем моніторингу є спостереження за діями повітряних, наземних, чи надводних об'єктів у певній зоні відповідальності та інформування про них визначених споживачів. Для досягнення зазначеної мети системи спостереження (моніторингу) вирішують задачі виявлення, оцінки параметрів руху, супроводження (формування траєкторії) РО, прогноз їх подальшого руху та інформування споживачів про дії об'єктів у часі.

Існує ряд задач на етапі супроводження РО та прогнозування характеру їх подальшого руху, якісне вирішення яких потребує згладжування поточного курсу об'єктів, який розраховується алгоритмом траєкторної (вторинної) обробки координатної інформації. Така задача, зокрема, виникає у випадках, коли швидкість руху РО є такою, що його лінійне пересування за інтервал спостереження

(оновлення інформації про РО засобами розвідки) менше, ніж лінійні значення похибок вимірювання координат об'єкта. Такі об'єкти в рамках даної роботи умовно названі «малорухомими». Описана ситуація є цілком характерною для супроводження надводних і наземних об'єктів та за певних умов (великої відстані до об'єкта та низької точності вимірювання його координат засобами виявлення) може виникати і при супроводженні повітряних РО. Для прикладу, на рис. 1 пунктирними стрілками показано напрями поточних курсів РО, які визначалися за «сусідніми» оцінками координат об'єкта. Як видно з рисунку, оцінки курсу змінюються у широкому діапазоні значень, в той час, коли дійсний курс об'єкта є постійним. При цьому, найбільш невизначеною ситуація виглядає на початковому етапі супроводження РО (на початку формування (визначення) траєкторії його руху).

На рис. 1 показано результат моделювання процесу супроводження повітряного РО, що рухається прямолінійно та рівномірно, у площинних координатах.

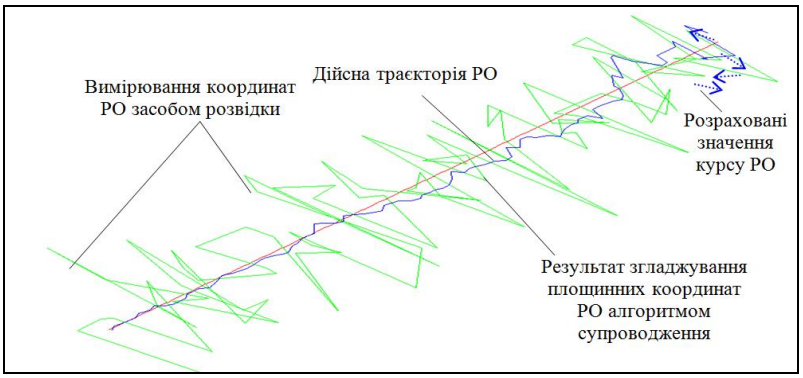


Рис. 19. Результати моделювання руху повітряного об'єкта зі швидкістю 40м/с на відстані 200км від джерела інформації з періодом огляду простору 2,5с (середньоквадратична похибка вимірювання дальності 300м, СКП вимірювання азимуту та кута місця 20'). Моделювання руху РО здійснювалося протягом 100 періодів оновлення інформації (250с).

Основою алгоритму супроводження є фільтр Калмана І-го порядку, що працює у декартовій системі координат і є оптимальним для зазначеного типу руху. При цьому, як видно із результатів моделювання, виконання вищезазначеної умови «малорухомості» РО призводить до майже хаотичної зміни оцінок його поточного площинного курсу алгоритмом супроводження від огляду до огляду протягом досить тривалого часу (якщо курс РО розраховується безпосередньо за результатами роботи фільтру Калмана). Така ситуація призводить до складності аналізу (пролонгації) траєкторії подальшого руху РО та визначення спрямованості його дій. Крім того, це ускладнює вирішення завдань диспетчерського контролю, забезпечення безпеки та управління рухом множини РО. Отже, згладжування поточних оцінок курсу РО «малорухомих» об'єктів є актуальною задачею.

Загалом, задача згладжування формулюється, як задача зменшення дисперсії похибки оцінки деякого параметру, однак оскільки виникнення помітної систематичної похибки такої оцінки в нашому випадку є небажаним - в подальшому будемо наряду с поняттям «згладжування» використовувати терміни «фільтр» та «фільтрація».

Характер руху будь-якого РО за курсом можна поділити на рух без маневру та рух із маневром за цим параметром. Розглянемо найпростіший характер руху РО для його супроводження – без здійснення маневру за курсом.

У [2] для вирішення задач згладжування курсу запропоновано використовувати експоненціальне згладжування, яке є досить простим та ефективним способом згладжування незмінних у часі параметрів [3,4] та на відміну від методів ковзаючого середнього та найменших квадратів починає «згладжувати» параметр з другої його оцінки:

$$\hat{Q}_k = (1-\xi)Q_k + \xi\hat{Q}_{k-1}, \qquad (3)$$

де $\hat{Q}_k$ - оцінка курсу на k-му кроці супроводження РО, отримана у результаті експоненціального згладжування ($\hat{Q}_0$ =0));

Q_k – розраховане значення курсу за даними оцінок параметрів руху РО на *k*-му кроці обробки інформації алгоритмом супроводження;

ξ – коефіцієнт експоненціального згладжування (ξ змінюється від 0 до 1).

Як відомо, при $\xi \to 1$ зростає якість згладжування параметру та точність його визначення після завершення перехідного процесу) однак спостерігається значна інертність відповідного фільтру (зростає час його налаштування на дійсне значення). При зменшенні ξ - ситуація змінюється на протилежну. (На рис. 2 показано результати моделювання згладжування курсу РО, який не маневрує, за допомогою експоненціального фільтру з коефіцієнтами ξ = 0,9, та ξ = 0,7 (суцільна та пунктирна криві, відповідно, помічені чотирикутним маркером), що підтверджує зазначене вище). Отже, при використанні експоненціального згладжування необхідно шукати компроміс між швидкістю

налаштування фільтру на дійсне значення параметру та точністю визначення параметру після завершення перехідного процесу. Необхідно також враховувати, що тривалий час налаштування фільтру ($\xi \to 1$) призведене до неякісного вирішення задачі відслідковування реального курсу РО під час поступового маневру за цим параметром та після його завершення. Логічним вирішенням щодо використання експоненціального згладжування виглядає використання різних значень коефіцієнта ξ на різних кроках налагодження експоненціального фільтру.

Мета роботи

Розробити модифікований експоненціальний фільтр, що дозволяє швидко налаштовуватися на дійсне значення курсу «малорухомого» РО, що розраховується за вимірюваннями (оцінками) його координат, за умови забезпечення високої точності оцінки курсу після завершення перехідного процесу.

Основний зміст

Для забезпечення високої якості згладжування при одночасному зменшенні інертності фільтру пропонується використовувати змінний коефіцієнт згладжування ξ, значення якого однозначно визначається часом (тривалістю) супроводження РО, тобто для дискретного спостереження за об'єктом - $\xi = f(k)$.

Розглянемо обґрунтування залежності $f(k)$ для випадку, коли статистична похибка розрахованого курсу РО постійна у часі і не залежить від k (усі оцінки курсу є рівноточними), а сам курс об'єкта не змінюється.

За такої умови ваговий коефіцієнт ξ може бути визначено на основі рівняння для отримання статистичної оцінки математичного очікування випадкової величини:

$$\hat{Q} = \frac{1}{K}\sum_{k=1}^{K} Q_k \ , \qquad (4)$$

де K – кількість спостережень (оцінок курсу).

Тобто, виходячи з наведеного виразу, вага чергової k-ї оцінки курсу ($K = k$) дорівнює $1/k$, а сумарна вага усіх попередніх оцінок $(k-1)/k$.

Тоді, значення коефіцієнта ξ на k-му кроці згладжування:

$$\xi_k = (k-1)/k \ . \qquad (5)$$

Підставимо вираз (3) у вираз (1):

$$\hat{Q}_k = (1-\xi_k)Q_k + \xi_k \hat{Q}_{k-1} = (1-(k-1)/k)Q_k + (k-1)/k \cdot \hat{Q}_{k-1}$$

і остаточно отримаємо:

$$\hat{Q}_k = (1/k)Q_k + (k-1)/k \cdot \hat{Q}_{k-1} . \qquad (6)$$

Фактично вираз (4) є відомим виразом для ітераційного (рекурентного) розрахунку математичного очікування випадкової величини.

Оскільки при $k \to \infty$ - $\xi \to 1$, за великих значень k, фільтр втратить чутливість до нових значень курсу. Для забезпечення можливості відслідковування даним фільтром незначного маневру по курсу доцільним є обмеження значення вагового коефіцієнта ξ деяким числом k_{max}:

$$\xi_k = \begin{cases} (k-1)/k, & \text{при} \quad k < k_{max}; \\ (k_{max}-1)/k_{max}, & \text{при} \quad k \geq k_{max}. \end{cases} \qquad (7)$$

На рис. 2 суцільною лінією, маркованою зірочкою, показано результат згладжування курсу, де k_{max}=10.

З рисунку видно, що інертність згладжування при використанні виразів (4) та (5) суттєво зменшується і при цьому його якість у стаціонарному режимі приблизно відповідає випадку для $\xi = 0{,}9$.

Фільтр за виразом (4) краще (у середньостатистичному сенсі) ніж фільтр за виразом (1) налаштовується на дійсне значення параметру і в умовах незначного маневру РО за курсом в процесі налаштування (див. рис. 3, перші десять циклів оновлення курсу). Поведінка фільтру (4) у подальшому є загалом аналогічною фільтру (1) та дозволяє відслідковувати незначний і короткий у часі маневр РО.

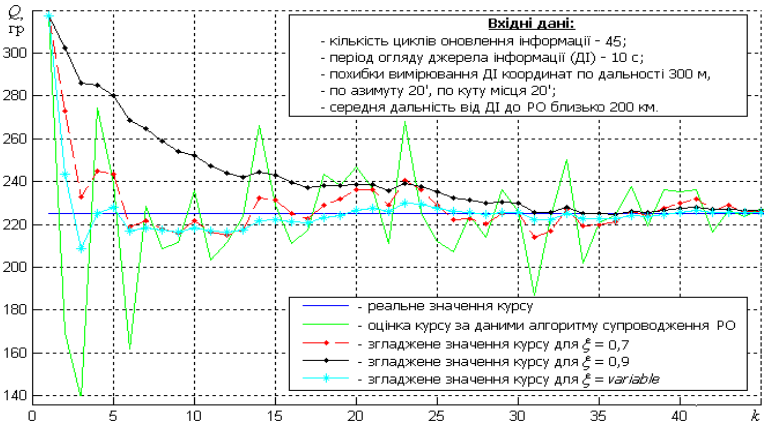


Рис. 20. Результати моделювання оцінки та згладжування курсу РО, що не маневрує по курсу.

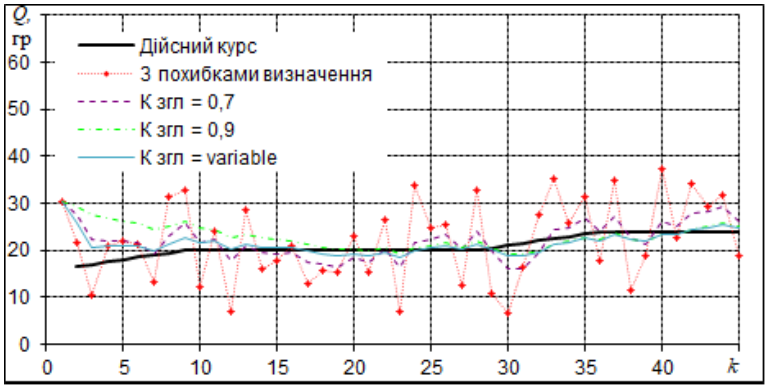


Рис. 21. Результати моделювання згладжування курсу РО, що незначно маневрує за курсом.

Статистичні дослідження показали, що час налаштування модифікованого експоненціального фільтру на дійсне значення курсу у середньому зменшується на 5-6 циклів оновлення інформації і при цьому забезпечується зменшення середньоквадратичної похибки оцінки курсу приблизно у 1,5 рази (до закінчення процесу налаштування фільтрів).

У випадку значного маневру РО за курсом (інформація про такий маневр може надходити від алгоритму формування трас РО засобів розвідки), необхідно здійснювати «скорочення» передісторії фільтру і продовжувати її «накопичення» після завершення маневру. Як відомо, на етапі маневрування, для згладжування параметрів руху РО доцільно використовувати лише дві, три останні їх оцінки [3-5].

Отже, на етапі маневру, для згладжування курсу РО, доцільно використовувати вираз (5), з k_{max} = 3, або 4 ($\xi \approx 0{,}7$ - 0,75). Після завершення маневру РО значення k_{max} обмежується значенням 10-15.

На рис. 4 показано результат згладжування курсу РО для умов в цілому аналогічних рис. 2, однак об’єкт здійснює значений маневр за курсом. З рисунку видно, що динамічне налаштування коефіцієнта ξ за виразом (4) (за умов обмеження k, як описано вище) дозволяє точніше відслідковувати дійсне значення курсу для заданих варіантів руху РО у порівнянні з експоненціальними фільтрами з $\xi = 0{,}7$ та $\xi = 0{,}9$.

Таким чином, модифікований експоненціальний фільтр може ефективно використовуватися для згладжування курсу РО. При цьому, використання перемінного коефіцієнта згладжування дозволяє зменшити час налаштування фільтру на дійсне значення курсу, у тому числі й після завершення маневру РО.

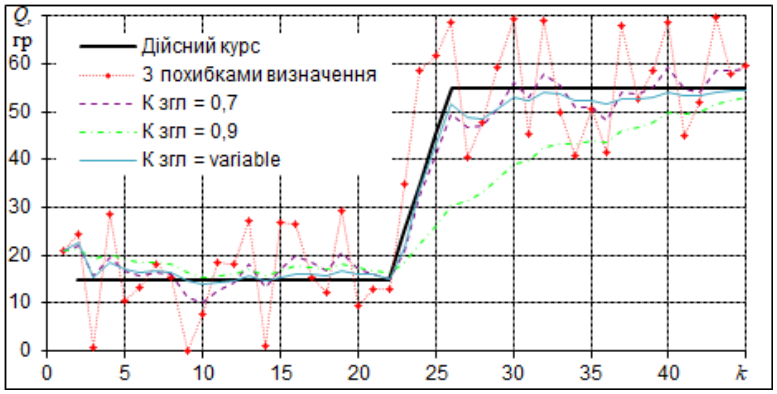


Рис. 22. Результати моделювання оцінки та згладжування курсу РО, що здійснив маневр по курсу

Розглянемо більш загальний випадок, коли оцінки курсу РО здійснюються за нерівноточними оцінками координат об'єкта. Такий випадок є характерним, для так званої третинної, або мультирадарної обробки інформації в системах моніторингу, коли курс об'єкта визначається не за даними одного засобу розвідки, а одразу за даними декількох джерел інформації із різними характеристиками. Тоді коефіцієнт згладжування ξ може бути розрахований, виходячи з виразу для статистичної оцінки математичного очікування випадкової величини за нерівноточними вимірюваннями.

Виходячи з виразу (1) для експоненціального згладжування, нескладно показати, що поточна оцінка курсу для нерівноточних даних буде розраховуватися за виразами:

$$\begin{aligned} \hat{Q}_k &= p_k / P_k \cdot Q_k + P_{k-1} / P_k \, \hat{Q}_{k-1}; \\ P_k &= P_{k-1} + p_k, \end{aligned} \qquad (8)$$

де $p_k = 1/\sigma_{Qk}^2$ – вага k-ї оцінки курсу, отриманої від алгоритму супроводження РО;

σ_{Qk}^2 - середньоквадратична похибка k-ї оцінки курсу за результатами обробки одним джерелом;

P_k - сумарна вага усіх оцінок курсу, що отримані до k-го періоду оновлення інформації включно ($P_1 = p_1$).

При цьому, середньоквадратичну похибку оцінки курсу (σ) на кожному кроці супроводження РО може бути розраховано, виходячи з середньоквадратичних похибок оцінки площинних координат (σ_x та σ_y) РО, на основі яких курс об'єкта власне й розраховується:

$$Q = \begin{cases} \dfrac{\pi}{2}, \text{если} \left(\Delta x = 0\right) \wedge \left(\Delta y \geq 0\right); \\ \dfrac{3}{2}\pi, \text{если} \left(\Delta x = 0\right) \wedge \left(\Delta y < 0\right); \\ arctg \dfrac{\Delta y}{\Delta x}, \text{если} \left(\Delta x > 0\right) \wedge \left(\Delta y \geq 0\right); \\ \pi + arctg \dfrac{\Delta y}{\Delta x}, \text{если} \left(\Delta x < 0\right); \\ 2\pi + arctg \dfrac{\Delta y}{\Delta x}, \text{если} \left(\Delta x > 0\right) \wedge \left(\Delta y < 0\right), \end{cases} \qquad (9)$$

де Δx, Δy - поточне приростання площинних координат об'єкта (x, y) у декартовій системі координат.

Для визначення σ, що є функцією від декількох випадкових аргументів, використаємо метод лінеаризації, який детально описаний, наприклад, у [6].

У відповідності до цього методу:

$$\sigma_Q{}^2 = (dQ/dx)^2 \sigma_{\Delta x}{}^2 + (dQ/dy)^2 \sigma_{\Delta y}{}^2, \qquad (10)$$

де $\sigma_{\Delta x}^2 = \sigma_{x-1}^2 + \sigma_x^2 \approx 2\sigma_x^2$ и $\sigma_{\Delta y}^2 = \sigma_{y-1}^2 + \sigma_y^2 \approx 2\sigma_y^2$.

В кінцевому підсумку, враховуючи (7), для σ^2 можна записати:

$$\sigma_Q^2 = \left(\frac{\Delta x}{\Delta x^2 + \Delta y^2} \right)^2 \sigma_{\Delta y}^2 + \left(-\frac{\Delta y}{\Delta x^2 + \Delta y^2} \right)^2 \sigma_{\Delta x}^2 = \frac{\Delta x^2 \sigma_{\Delta y}^2}{\left(\Delta x^2 + \Delta y^2\right)^2} + \frac{\Delta y^2 \sigma_{\Delta x}^2}{\left(\Delta x^2 + \Delta y^2\right)^2},$$

або остаточно, разом із індексом *k*:

$$\sigma_{Qk}^2 = \frac{2\Delta x_k^2 \sigma_{yk}^2 + 2\Delta y_k^2 \sigma_{xk}^2}{\left(\Delta x_k^2 + \Delta y_k^2\right)^2}. \qquad (11)$$

Вирази (6) та (9) відображають функціональну залежність коефіцієнта згладжування для модифікованого експоненціального фільтра курсу РО), що, на відміну від виразу (4), враховує нерівноточність поточних оцінок курсу.

На рис. 5 показано результати згладжування курсу при використанні експоненціального фільтру із ξ=0,9 та модифікованих експоненціальних фільтрів без врахування та з урахуванням нерівночності оцінок параметрів руху РО різними засобами розвідки. В ході модельного експерименту вважалося, що дані про параметри руху об'єкта надходять послідовно від трьох нерівноточних джерел.

Як видно з рисунку, модифікований експоненціальний фільтр, що враховує нерівноточність оцінок параметру, який згладжується, дозволяє отримати більш якісні результати згладжування у порівнянні з іншими експоненціальними фільтрами. У середньому такий фільтр дозволяє додатково скоротити час налаштування на декілька періодів оновлення значення параметру та додатково зменшити середньоквадратичну похибку визначення параметру (конкретні характеристики такого фільтру прямо пропорційно залежать від числа джерел інформації та їх точності).

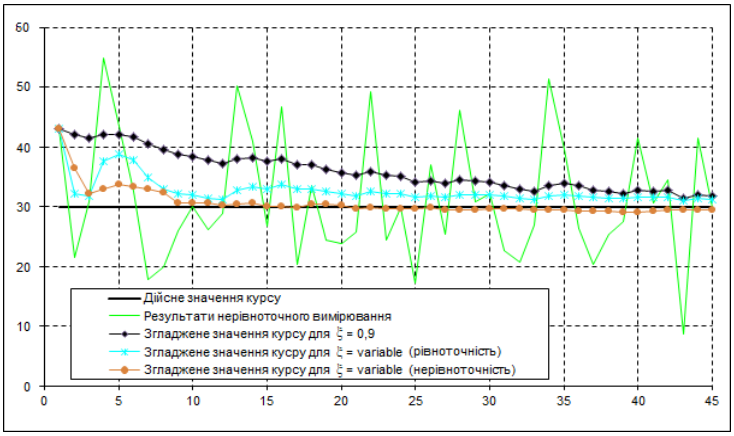


Рис. 23. Результати моделювання оцінки та згладжування курсу РО, що інтенсивно маневрує по курсу на деякій ділянці руху

Висновки

У роботі для згладжування оцінок курсу малорухомих об'єктів в процесі їх супроводження запропоновано використовувати модифікацію відомого способу експоненціального згладжування динамічних даних, яка полягає у використанні замість постійного значення коефіцієнта згладжування його перемінне значення. Таке перемінне значення коефіцієнта розраховується на основі рівнянь для отримання статистичної оцінки математичного очікування випадкової величини.

Задачу визначення коефіцієнта згладжування модифікованого експоненціального фільтру вирішено для часткового випадку рівноточних оцінок курсу та для більш загального випадку – нерівноточних оцінок параметру.

В результаті, час налаштування модифікованого фільтру на дійсне значення курсу у середньому зменшується на 5-8 циклів оновлення інформації і при цьому забезпечується додаткове зменшення середньоквадратичної похибки оцінки курсу у 1,5-2 рази. Це дозволяє отримувати більш точні оцінки курсу малорухомих об'єктів на початковому етапі їх супроводження, коли існує найбільша невизначеність щодо їх дій.

Запропонований модифікований метод експоненціального згладжування може використовуватися для згладжування послідовного ряду рівноточних та нерівноточних вимірювань (оцінок) будь-яких постійних у часі величин для скорочення перехідного процесу налаштування на дійсне значення величини.

References

1. Теоретичні та методологічні основи створення розподілених інформаційно-аналітичних систем (ІАС) моніторингу множини динамічних об'єктів у реальному часі. Звіт про НДР шифр "Контроль", Книга 2 – К.: ІПРІ НАН України. – 2009 – 222с. Держ. реєс. .№ 0107U002592 УДК 004.75:004.9.

2. Кожешкурт В.И., Юзефович В.В. Дослідження схем фільтрації алгоритмів трасової обробки інформації в системах моніторингу динамічних об'єктів // Реєстрація, зберігання і оброб. даних. – 2010. – Т. 12, № 4. – С. 3–12.
3. Кузьмин С.З. Основы теории цифровой обработки радиолокационной информации. – М.: Сов. радио, 1974. – 432 с.
4. Кузьмин С.З. Цифровая радиолокация. Введение в теорию. – Киев: Издательство КВіЦ, 2000. – 428 с.: илл.
5. Фарина А., Студер Ф. Цифровая обработка радиолокационной информации. Сопровождение целей: Пер. с англ. – М.: Радио и связь, 1993. – 320 с.
6. Вентцель Е.К. Теория вероятностей./ Е.К.Вентцель – М.: Наука. – 1969. – 576 с.

МАНІПУЛЮВАННЯ ВИБОРОМ У ЗАДАЧАХ БАГАТОКРИТЕРІАЛЬНОЇ ОПТИМІЗАЦІЇ

Гнатієнко Григорій Миколайович

Київський національний університет імені Тараса Шевченка

G.Gna5@ukr.net

Анотація. Розглянуто проблеми маніпулювання вибором варіантів рішень у ситуаціях експертного оцінювання. Наведено описання задачі експертного оцінювання у вигляді кортежу, на основі якого досліджено можливості маніпулювання вибором. Особливу увагу приділено задачам багатокритеріальної оптимізації (ЗБКО). Описано постановку ЗБКО з акцентом на її складові та евристики, які можуть бути використані для маніпулювання. Запропоновано класифікацію задач маніпулювання вибором в ситуаціях експертного оцінювання, які формалізуються у класі багатокритеріальних оптимізаційних задач.

Ключові слова: експертне оцінювання, прийняття рішень, маніпулювання, евристика, багатокритеріальна оптимізація.

1 Вступ

У сучасному світі проблема маніпулювання вибором стає все більш актуальною. Адже задачі вибору виникають у найрізноманітніших сферах, а можливість вибору є основою демократії. В той же час відомо, що маніпулювання – це вид прихованого управління. Тому дослідження можливостей маніпулювання для різних моделей прийняття рішень на сьогодні є актуальним і перспективним напрямком досліджень. Вивчення проблем маніпулювання є тим більш важливим тому, що нерідко застосування шкал при вимірюванні деяких явищ перетворюється на тривіальне присвоєння цифр, а не на задекларованe дослідниками вимірювання суб'єктивних складових задачі прийняття рішень.

Маніпулювання в задачах вибору добре досліджене і для таких задач розроблена теорія агенди [1] – маніпулювання «порядком денним», послідовністю застосування процедур вибору. Відомо також [2], що правило вибору є захищеним від маніпулювання тоді і лише тоді, коли воно є диктаторським.

У цій роботі будемо розглядати задачі маніпулювання вибором, які виникають при формалізації практичних проблем у класі задач багатокритеріальної оптимізації. Критеріальний підхід є розповсюдженим підходом при розв'язанні задач прийняття рішень. Цей підхід до описання задач прийняття рішень пов'язаний з припущенням, що кожен об'єкт можна оцінити конкретним числом, яке є значенням критерія, тому порівняння об'єктів зводиться до порівняння відповідних їм чисел.

У багатьох ситуаціях експертного оцінювання є доцільним задавати не одну скалярну оцінку якості об'єктів, а сукупність показників якості і розглядати задачу у векторній постановці. При цьому виникає задача багатокритеріальної оптимізації. Часто багатокритеріальність є способом збільшення адекватності описання цілі. В той же час, як буде показано нижче, багатокритеріальність відкриває широкі можливості для маніпулювання вибором.

2 Задачі експертного оцінювання

Для описання задач експертного оцінювання використовується її символьне представлення у вигляді кортежу [3] <A, S, R, E, C, P>, де A – множина об'єктів (варіантів, альтернатив, наборів параметрів), S – множина обмежень, R – множина критеріїв, шкал вимірювання по критеріях, згорток критеріїв, відображення множини допустимих альтернатив у множину критеріальних оцінок, E –

множина формальних характеристик експертів, С – множина цілей, що стоять перед дослідниками, Р – система переваг вирішуючого елемента: особи, що приймає рішення (ОПР), або групи експертів.

Множина $A,$ що складається з n об'єктів, які порівнюються між собою:

$$a_i \in A,\ i \in \{1,...,n\} = I.$$

Об'єкти описуються множиною m параметрів, тобто кожен об'єкт є точкою деякого параметричного простору Ω^m

$$a_i = \left(a_i^1,...,a_i^m\right),\ a_i \in A,\ i \in I,\ A \subset \Omega^m,$$

де $a_i^j,\ i \in I,\ j \in \{1,...,m\} = J,$ – значення j-го параметра i-го об`єкта.

З множини A ОПР вибирається найкращий за встановленими критеріями об'єкт та обґрунтовується його вибір. На етапі обґрунтування підходів до вибору та власне самої процедури вибору виникають можливості для маніпулювання результатами розв'язання задачі.

Для опису методів та алгоритмів розв'язання ЗЕО наведемо основні поняття, які будуть використовуватися у подальшому викладенні. Під ЗЕО розуміють пару $\langle C, A\rangle,$ де $C-$ принцип оптимальності, $A-$ задана множина об'єктів. Розв'язком ЗЕО $\langle C, A\rangle$ є підмножина об'єктів $A_0 \subset A,$ одержана з використанням принципу оптимальності: $A_0 = C(A).$ В теорії дослідження операцій [4] пара $\langle C, A\rangle$ називається:

– задачею прийняття рішень, якщо C та A не є апріорно заданими і можуть варіюватися;

– задачею вибору, якщо множина об'єктів A є заданою, а принцип оптимальності $C-$ таким, що варіюється;

– загальною задачею оптимізації, коли C та A є заданими.

Зрозуміло, що у ситуаціях, де допускаються варіації принципу оптимальності або множини об'єктів, виникає можливість для маніпулювання. Але зосередимося на останній ситуації – задачах оптимізації. Причому будемо досліджувати проблематику задач багатокритеріальної оптимізації.

3 Постановка та особливості визначення розв'язків задачі багатокритеріальної оптимізації

Задача багатокритеріальної оптимізації формалізується у такій постановці [3, 5]:

$$y_i = f_i(x) \to \max,\ i \in L_1, \quad (1)$$

$$y_i = f_i(x) \to \min,\ i \in L_2, \quad (2)$$

$$x \in A,\ A \subseteq E^m, \quad (3)$$

де $A-$ множина об'єктів, які характеризуються m параметрами, тобто належать простору E^m; $y(x) = (f_1(x),...,f_k(x))$ – вектор оцінок об'єктів або критеріїв, який задається відображенням $f: A \to E^k,$ $L = \{1,...,k\}-$ множина індексів критеріїв, $L_1 = \{1,...,k_1\},\ L_2 = \{k_1+1,...,k\}$ – множини індексів критеріальних функцій, які, відповідно, максимізуються та мінімізуються, $L_1 \cup L_2 = L.$ У деяких задачах множина об'єктів виділяється із ширшої множини $A \subseteq D \subset E^m$ за допомогою обмежень, які найчастіше задаються системою нерівностей.

Вектори $y = (y_1,...,y_k)$ та $y' = (y'_1,...,y'_k),\ y \neq y',$ знаходяться у відношенні ">", якщо $y_i \geq y'_i,\ \forall i \in L,$ і хоча б одна нерівність є строгою.

Оцінка об'єкта $y^0 = f(a_0),\ y^0 = (y_1^0,...,y_k^0),$ називається ефективною (оптимальною за Парето, Парето-оптимальною, непокращуваною, мажорантною, недомінованою), за відношенням ">", якщо не існує іншої оцінки $y,$ яка б строго переважала $y^0,$ тобто виконувалося б відношення $y > y^0.$

Відомо [5], що два ефективних об'єкти є або еквівалентними, або непорівняними між собою за множиною критеріїв.

Проблема визначення області Парето є строго об`єктивною і вирішується без застосування будь-яких евристик [5]. Але область компромісів є множиною точок, з яких у більшості випадків треба вибрати одну. Звуження області ефективних об`єктів і, тим більше, вибір єдиного з них принципово потребує застосування додаткової інформації від експертів, оскільки ефективні набори параметрів не можна порівняти один з одним формально.

Оцінки об'єктів мають різну розмірність, оскільки характеризують різні фізичні властивості об'єктів, тому доцільно розглядати не абсолютні значення оцінок об'єктів, а їх безрозмірні значення. Монотонні перетворення, які переводять оцінки об'єктів до безрозмірного виду, наведено у першому розділі, в якому вони представлені формулами...

Для описання одного із способів розв'язання ЗБКО, вводяться евристики.

Евристика Е1. Вибір виду монотонного перетворення $\omega_i(a),\ i \in L,\ a \in A,$ для переведення значень параметрів об'єктів до безрозмірного виду здійснюється експертом за формулами, які мають задовольняти таким вимогам:

– враховувати необхідність мінімізації відхилень від оптимальних значень по кожній цільовій функції (ЦФ);

– мати загальний початок відліку і однаковий порядок зміни значень на усій множині об'єктів;

– зберігати відношення переваги на множині об'єктів, які порівнюються, за сукупністю ЦФ, і ,таким чином, не змінювати множину ефективних об'єктів.

Розв'язок ЗБКО може виявитися не оптимальним для жодної ЦФ виду (1), (2), але одночасно бути у певному розумінні найкращим розв'язком для усіх критеріальних функцій одночасно.

Евристика Е2. Найкращим об'єктом при розв'язанні ЗБКО слід вважати такий об'єкт, для якого відхилення від найкращих значень по кожній оцінці є мінімальними.

Якщо найменші значення відхилень по кожному критерію не досягаються одночасно на жодному об'єкті, то виникає необхідність порівнювати ці відхилення між собою, що пов'язано з необхідністю залучення додаткових евристик від експертів. Для конструктивного визначення найкращого об'єкта наведемо ще одну теорему.

Відомо, що для кожного об'єкта $a \in A,$ такого, що у просторі перетворених значень ЦФ справедливі нерівності $0 < \omega_i(a) < 1,\ \ \forall i \in L,$ існує вектор $\rho = (\rho_i, i \in L),$ який задовольняє співвідношенням нормованості:

$$\rho_i > 0,\ \ i \in L, \tag{4}$$

$$\sum_{i \in L} \rho_i = 1, \tag{5}$$

і число $k_0,$ такі, що об'єкт $a \in A,$ задовольняє одночасно k рівностям

$$\rho_i \omega_i(a) = k_0,\ \ \forall i \in L. \tag{6}$$

Введемо ще три евристики.

Евристика Е3. Вектор вагових коефіцієнтів ЦФ $\rho = (\rho_i, i \in L),$ який задовольняє співвідношенням (4), (5), будемо інтерпретувати як відношення переваги у кількісному вигляді на множині ЦФ.

Евристика Е4. Під розв'язком ЗБКО для заданого вектора вагових коефіцієнтів $\rho = (\rho_i, i \in L),$ будемо розуміти такий компромісний об'єкт, який належить множині ефективних об'єктів і знаходиться на напрямку, який визначається вектором $\rho = (\rho_i, i \in L).$

Евристика Е5. Компромісний розв'язок $a_0 \in A$ ЗБКО має забезпечувати однакові мінімальні зважені відносні відхилення $\rho_i \omega_i(a_0),\ \ i \in L,$ за всіма критеріями одночасно.

Доведено [5], що якщо $a_0 \in A$ є ефективним об'єктом для вектора $\rho = (\rho_i, i \in L),$ то цьому об'єкту відповідає найменше значення параметра $k_0,$ при якому система (6) виконується одночасно для усіх ЦФ.

Відомо також [5], що для того, щоб об'єкт $a^* \in A,$ такий, що $\omega_i(a^*) > 0,\ \forall i \in L,$ був ефективним при заданому векторі вагових коефіцієнтів $\rho = (\rho_i, i \in L),$ достатньо, щоб $a^* \in A$ був єдиним розв'язком системи нерівностей

$$\rho_i \omega_i(a) \le k_0,\ \ \forall i \in L, \tag{7}$$

для мінімального значення параметра $k_0^*,$ при якому ця система є сумісною.

Таким чином, компромісний розв'язок ЗБКО, визначений евристикою Е5, може бути знайдений як єдиний розв'язок системи нерівностей виду (7) для мінімального значення параметра $k_0,$ при якому ця система ще сумісна. У просторі відносних значень параметрів об'єктів компромісному об'єкту відповідає точка перетину променя, направляючі косинуси якого визначаються заданим вектором відносної важливості ЦФ $\rho = (\rho_i, i \in L),$ з множиною ефективних об'єктів. Якщо такої точки не існує, тобто на промені, що визначається вектором $\rho = (\rho_i, i \in L),$ не лежить жодна точка, яка відповідає ефективному об'єкту, то компромісним об'єктом вважається той, для якого виконується система нерівностей (7) і цьому об'єкту відповідає точка, найближча до заданого променя. Якщо компромісний об'єкт не єдиний, тобто існує деяка підмножина ефективних об'єктів, еквівалентних з точністю до деякого достатньо малого числа за значенням параметра $k_0,$ то вибір компромісного об'єкта здійснюється з допомогою іншого критерію.

Зазначимо, що альтернативами в ЗБКО (1)-(3) можуть бути, зокрема, ранжування [3] або об'єкти чи явища іншої природи.

4 Класіфікація задач маніпулювання вибором при розв'язанні задач багатокритеріальної оптимізації

Відповідно до загального опису задач експертного оцінювання та описаного вище підходу до розв'язання ЗБКО, наведемо розроблену автором класифікацію задач маніпулювання вибором у задачах багатокритеріальної оптимізації.

4.1. Задачі маніпулювання множиною альтернатив

Множина альтернатив може бути змінена різними способами:
− волюнтаристським підходом ОПР;
− імітацією демократичних процедур, які в результаті призводять до зміни початкової множини А;
− введенням додаткових обмежень та умов участі альтернатив у подальшому розв'язанні багатокритеріальної задачі;
− свідомим ігноруванням деяких обмежень з метою розширення початкової множини альтернатив;
− іншими підходами, які призводять до зміни початкової множини альтернатив з метою маніпулювання та «обґрунтованого» вибору потрібної для ОПР альтернативи.

4.1.1. Доповнення початкової множини альтернатив A деякою підмножиною A^*, $A^* \cap A = \varnothing$, яка містить альтернативу $a^* \in A^*$, що має бути вибраною як розв'язок у результаті розв'язання ЗБКО виду (1)-(3). Тоді $a^* \in A^1 = A \cup A^*$.

4.1.2. Вилучення з початкової множини альтернатив A деякої підмножини A^-, $A \setminus A^- = A^1$, яка містить альтернативу $a^* \in A^-$, яка не повинна бути вибраною як розв'язок у результаті розв'язання задачі БКО (…).

4.1.3. Комбінація підходів 4.1.1 та 4.1.2: одночасне доповнення початкової множини допустимих альтернатив A деякою підмножиною A^* і вилучення небажаної для вибору підмножини A^-. Тобто $A^1 = A \cup A^* \setminus A^-$.

4.1.4. Зміна вагомості параметрів, які характеризують об'єкти початкової множини A. Для дослідження цього підходу до маніпулювання вибором автором розроблено групу методів адаптивного визначення вагових коефіцієнтів параметрів об'єктів [3].

Крім цілеспрямованої зміни вагомості параметрів об'єктів можуть бути застосовані також більш радикальні варіації цього підходу.

4.1.5. Доповнення початкової множини параметрів альтернатив A деякою підмножиною A^*, $A^* \cap A = \varnothing$, яка містить альтернативу $a^* \in A^*$, що має бути вибраною як розв'язок у результаті розв'язання ЗБКО виду (1)-(3). Тоді $a^* \in A^1 = A \cup A^*$.

4.1.6. Вилучення з початкової множини альтернатив A деякої підмножини A^-, $A \setminus A^- = A^1$, яка містить альтернативу $a^* \in A^-$, яка не повинна бути вибраною як розв'язок у результаті розв'язання задачі БКО (1)-(3).

4.1.7. Комбінація підходів 4.1.5 та 4.1.6: одночасне доповнення початкової множини допустимих альтернатив A деякою підмножиною A^* і вилучення небажаної для вибору підмножини A^-. Тобто $A^1 = A \cup A^* \setminus A^-$.

4.2. Маніпулювання з обмеженнями ЗБКО та шкалами вимірювання

4.2.1. Доповнення початкової множини обмежень S деякою підмножиною S^*, $S^* \cap S = \varnothing$.

4.2.2. Вилучення з початкової множини обмежень S деякої підмножини S^-, $S \setminus S^- = S^1$.

4.2.3. Комбінація підходів 4.2.1. та 4.2.2: одночасне доповнення обмежень початкової множини обмежень та вилучення небажаної підмножини обмежень.

4.2.4. Маніпулювання шкалами при вимірюванні параметрів альтернатив: заміна шкал вимірювання, застосування неприпустимих операцій при операціях над результатами вимірювання.

4.2.5. Маніпулювання з шкалами вимірювання при визначенні вагових коефіцієнтів параметрів, відносної ваги критеріїв, коефцієнтів компетентності експертів (вагомості джерел інформації [6]).

4.2.5.1. При вимірюванні в ординальних шкалах організатори експертизи можуть примусити експертів визначати свої переваги у просторі строгих відношень переваги, позбавивши таким чином учасників експертної групи можливості виражати ситуацію рівноцінності або байдужості.

4.2.5.2. Можуть бути встановлені обмеження в кардинальних шкалах: наприклад, в діапазоні від 1/9 до 9, що суттєво порушує транзитивність вже на етапі початкової експертизи та апріорно не дозволяє

досягти ситуації надтранзитивності відношень між параметрами, критеріями чи компетентністю експертів.

4.2.5.3. Встановлення вимоги фіксованого задання значень параметрів чи відношень переваги між об'єктами, їх параметрами, критеріями чи компетентністю експертів. Це нерідко призводить до суттєвого порушення адекватності моделювання предметної області та навмисної втрати інформації про явище, яке моделюється.

4.3. Маніпулювання з множиною метрик, що застосовуються, критеріїв та способів згортки критеріїв

Для досягнення мети маніпулювання можна майже непомітно для непідготовленого спостерігача вибрати з початкової множини допустимих альтернатив A практично будь-яку альтернативу, бажану для ОПР.

4.3.1. При розв'язанні задачі в ординальних шкалах існує кілька найпопулярніших метрик – Хемінга, Кука, Евкліда тощо [3]. Як правило, застосування кожної з метрик дає результати, що не збігаються з розв'язками, одержаними в інших метриках. Тому ОПР достатньо лише вибрати, розв'язок за якою метрикою її найбільше влаштовує.

4.3.2. При застосуванні формул для переходу до безрозмірного виду критеріїв суттєвими є границі зміни критеріальних функцій. Вибираючи гіпотетично ці границі, можна підібрати потрібні для маніпулювання розв'язком значення і суттєво вплинути на остаточне рішення.

4.3.3. На етапі підготовки до розв'язання задачі багатокритеральної оптимізації на попередніх етапах можна застосувати один або кілька турів процедур послідовного аналізу варіантів. Тобто попередньо встановити такі критерії, які здійснять відсів небажаних для ОПР альтернатив з початкової множини A, як заздалегідь безперспективних з огляду на «легітимно» встановлені критерії.

4.3.4. Зміна класу задач, в якому слід формалізувати початкову модель багатокритеріальної оптимізації. Це автоматично тягне за собою обґрунтоване та непомітне для фахівця застосування потрібних для досягнення мети маніпулювання методів розв'язання задачі та застосування відповідних видів згортки критеріїв задачі.

4.3.5. Доповнення початкової множини критеріїв деякою новою підмножиною або одним, навіть несуттєвим критерієм. Таким чином можна зробити ефективною альтернативу, яка у попередній моделі була домінованою. Така маніпуляція може суттєво вплинути на множину ефективних рішень початкової ЗБКО. А відтак – і на визначення компромісного розв'язку задачі (1)-(3).

4.3.6. Вилучення з початкової множини критеріїв R деякої підмножини критеріїв або навіть одного з них. Таке втручання у процес може суттєво вплинути та побудовану багатокритеріальну модель досліджуваного явища.

4.3.7. Комбінація підходів 4.3.5 та 4.3.6 може спричинити синергетичний ефект у порівнянні з автономним застосуванням кожного з цих підходів. Зокрема, можна вивести з вибору невигідну для ОПР домінуючу альтернативу під деяким слушним приводом.

4.4. Маніпулювання шляхом впливу на експертну групу

4.4.1. Заміна (відвід) неугодного експерта у разі, коли призначена одноосібна експертиза.

4.4.2. Доповнення початкової множини експертів деякою підмножиною експертів, на яких має вплив ОПР або інші учасники прийняття рішення, зацікавлені у маніпулюванні – для підготовки та обґрунтування потрібного рішення.

4.3. Вилучення з початкової множини експертів тих, які є незручними для ОПР та мають свою особисту думку.

4.4. Комбінація підходів 4.4.2 та 4.4.3 – одночасне доповнення групи експертів новими членами групи та відведення з групи деяких її членів.

4.5. Маніпулювання множиною цілей, що задекларовані при розв'язанні задачі

4.5.1. Збільшення підмножини вибору в ході розв'язання задачі, що дозволить ввести до підмножини переможців бажані для ОПР альтернативи.

4.5.2. Досягнення бажаного для ОПР рішення: з допомогою потужного інструменту, оригінального математичного забезпечення можна досягти як узгодження рішень різних рівнів, так і надати широкі і непомітні можливості для недоброчесного маніпулювання.

4.5.3. У випадку, коли підготовлене та обґрунтоване рішення не задовольняє ОПР, можна досягти визнання вибору недійсним з різних мотивів.

4.6. Маніпулювання системою переваг експертів

4.6.1. Мінімальними змінами у деяких відношеннях переваги, заданих експертами, можна досягти зміни початкового компромісного рішення.

4.6.2. Для досягнення мети маніпулювання може слугувати заміна лише деяких індивідуальних переваг експерта з метою згладжування його переваг або узгодження групового експертного рішення.

4.6.3. Маніпулювання ваговими коефіцієнтами компетентності експертів (вагомості джерел інформації [6]) можна досягнути бажаного для ОПР рішення або, щонайменше, блокування того рішення, яке є найбільш неприйнятним.

4.6.4. Виключення деяких елементів матриці парних порівнянь з розгляду – тобто переведення її у категорію неповних матриць. Таким чином змінюється не тільки інформація, але й методи розв'язання задачі експертного оцінювання. Це може потягнути за собою обчислення потрібного компромісного рішення.

4.6.5. Заміна елемента матриці парних порівнянь, який найсуттєвішим чином впливає на результат – спосіб непрямого впливу на остаточний розв'язок ЗБКО

5 Висновки

Досліджено різні аспекти та можливості маніпулювання вибором у задачах експертного оцінювання у випадку формалізації їх у вигляді моделі багатокритеральної оптимізації.

У загальному випадку розглядаються різні аспекти та джерела можливого маніпулювання у задачах експертного оцінювання:

– виборцем;

– організатором процедури;

– ОПР.

Вибір може бути обґрунтованим або волюнтаристським. При добре організованій маніпуляції ОПР можна підвести до потрібного для маніпулятора вибору. Тому в задачах експертного оцінювання особливо важливим є чітке представлення евристик, на які спираються дослідники при визначенні компромісного розв'язку.

Джерела

1. Лезина З.М. "Манипулирование выбором вариантов (теория агенды)", Автомат. и телемех., 1985, № 4, С.5–22.

2. Волошин О.Ф., Мащенко С.О. Теорія прийняття рішень: Навчальний посібник. – К.: Видавничо-поліграфічний центр "Київський університет", 2006. − 304 с.

3. Гнатієнко Г.М., Снитюк В.Є. Експертні технології прийняття рішень: Монографія. – К.: ТОВ «Маклаут», 2008. – 444с.

4. Макаров И.М., Виноградская Т.М., Рубчинский А.А. и др. Теория выбора и принятия решений: Учебное пособие. − М.:Наука, 1982. − 328 с.

5. Михалевич В.С., Волкович В.Л. Вычислительные методы исследования и проектирования сложных систем. − М.: Наука, 1982. − 286 с.

6. Додонов А.Г., Ландэ Д.В., Цыганок В.В., Андрейчук О.В., Каденко С.В, Грайворонская А.Н. Распознавание информационных операций – К.: ООО «Инжиниринг», 2017. – 282 с.

АПОСТЕРІОРНЕ ВИЗНАЧЕННЯ КОМПЕТЕНТНОСТІ ЕКСПЕРТІВ В УМОВАХ НЕВИЗНАЧЕНОСТІ

Гнатієнко Г.М., Снитюк В.Є.

Київський національний університет
G.Gna5@ukr.net, snytyuk@gmail.com

Анотація. Розглянуто проблему визначення відносної компетентності експертів у задачах експертного оцінювання. Запропоновано класифікацію методів визначення компетентності та обґрунтовано використання об'єктивних підходів до апостеріорного визначення зазначених коефіцієнтів. Наведено постановку задачі визначення результуючого ранжування об'єктів та способи обчислення її розв'язків. На основі аналізу одержаних розв'язків задачі групового ранжування пропонується метод визначення коефіцієнтів відносної компетентності експертів у вигляді функцій належності нечіткій множині.

Ключові слова: експертне оцінювання, результуюче ранжування, компетентність експертів, прийняття рішень, функція належності нечіткій множині.'

1 Вступ

Визначення коефіцієнтів компетентності експертів є важливим елементом розв'язання задач експертного оцінювання. Точність визначення відносної компетентності експертів впливає на результати розв'язання задачі, достовірність розв'язку та довіру до одержаного результату. Урахування компетентності експертів є запорукою якості визначення розв'язку та значною мірою впливає на результат прийняття рішення в задачі експертного оцінювання. Тому дослідження компетентності експертів та аргументоване її урахування є актуальним напрямком досліджень і може значною мірою сприяти підвищенню якості результатів експертного оцінювання та їх обґрунтованості.

2 Класифікація методів визначення компетентності експертів

Визначення компетентності експертів є важливим елементом експертного оцінювання. Цей чинник суттєво впливає на результати прийняття рішення і потребує поглибленого вивчення, дослідження та розвитку. На сьогодні існує кілька підходів до визначення коефіцієнтів відносної компетентності експертів. Але кожен з напрямків не є досконалим:

– документальний нерідко викликає недовіру до способів формалізації та врахування об'єктивних даних про експертів, а також характерною для сучасності тенденцією до девальвації офіційних документів та некритичністю деяких інституційних рішень;

– самооцінка, при проведенні якої нерідко отримується інформація про рівень самовпевненості експерта, а не про його реальну компетентність;

– взаємооцінка, з допомогою якої у деяких задачах можна виявити конфронтацію в експертній комісії та наявності коаліцій її членів, які інколи спотворюють справжню індивідуальну компетентність експертів, яка дійсно має впливати на результати;

– комбіновані методи інколи замість синергетичного ефекту призводить до накопичення сукупної помилки;

– об'єктивні підходи, серед яких є зручні для застосування і достатньою мірою обґрунтовані.

Тому зупинимося на методах апостеріорного визначення компетентності експертів, розроблених авторами. До об'єктивних підходів апостеріорного визначення коефіцієнтів компетентності експертів відносяться:

– дослідження результативності участі експерта у попередніх експертизах;

– обчислененя компетентності за результатами контрольної експертизи;

– дослідження результатів участі експерта у конкретній експертизі.

Останній підхід, дослідження участі у конкретній експертизі, може розглядатися як:

– непрямий аналіз показників без попереднього агрегування результатів;

– співвідношення індивідуальних експертних даних та обчислених інтегральних показників.

Підхід, який полягає в дослідженні співвідношень індивідуальних експертних даних та обчислених інтегральних показників в задачах визначення групового ранжування об'єктів, класифікується таким чином:

– обчислення відстаней до матриць парних порівнянь (МПП) на основі знайденого результуючого ранжування (в різних метриках та за різними критеріями);

– апостеріорне визначення на основі результуючого ранжування об'єктів за заданими індивідуальними експертними ранжуваннями:

= кількість інверсій у експертному ранжуванні по відношенню до результуючого ранжування;

= кількість несуперечливо проранжованих експертом об'єктів;

= сума модулів різниць між рангами об'єктів у результуючому ранжуванніта заданих експертами ранжуваннях;

– визначення індивідуальних відносних оцінок для колективно заданих відносних оцінок:

= сума інтервалів;

= максимальний інтервал;

= «об'єм» інтервалів;

= «площа» інтервалів;

= відхилення МПП від «ідеального» вектора відносних оцінок;

– відстань до групового вектора вагових коефіцієнтів при заданні експертами МПП у кардинальних шкалах;

– відносна відстань експертних ранжувань до результуючого ранжування.

3 Постановка задачі визначення результуючого ранжування при груповому оцінюванні

Методи обробки експертної інформації поділяються на три основні групи [1]:

– статистичні методи;

– методи шкалювання;

– алгебраїчні методи.-

Суть алгебраїчних методів полягає в тому, що на множині допустимих оцінок задається відстань і результуюча оцінка визначається як така, відстань якої до оцінок експертів (за певним критерієм) мінімальна. На основі алгебраїчного підходу будемо визначати коефіцієнти відносної компетентності експертів.

Нехай k експертів з індексами $l \in L = \{1,...,k\}$, $k > 1$, задають свої переваги на множині n об'єктів A у вигляді ранжувань $R^l, l \in L$.

Для визначення відстаней між ранжуваннями об'єктів використовують:

– метрику Кука неспівпадання рангів об'єктів у індивідуальних ранжуваннях

$$d\left(R^j, R^l\right) = \sum_{i \in I} \left| r_i^j - r_i^l \right|, \qquad (1)$$

де r_i^l – ранг *i*-го об'єкта у ранжуванні *l*-го експерта, $R^l, l \in L$, $1 \le r_i^l \le n$,

– метрику Хемінга

$$d\left(B^j, B^l\right) = 0{,}5 \sum_{i \in I} \sum_{s \in I} \left| b_{is}^j - b_{is}^l \right|, \qquad (2)$$

де $B^l = \left(b_{is}^l\right)$, $l \in L$, $i, s \in I$, – МПП, що відповідають ранжуванню $R^l, l \in L$.

Інколи для задач визначення результуючого ранжування використовується евклідова метрика.

Поширеною є така задача експертного оцінювання: знайти результуюче (колективне, групове, компромісне, інтегральне, агреговане, узгоджене тощо) ранжування, яке за деяким критерієм є «найближчим» до усіх експертних ранжувань. Найбільш обґрунтованим методом знаходження результуючого ранжування об'єктів вважається обчислення медіани заданих ранжувань.

Позначимо множину усіх можливих ранжувань n об'єктів через $\Re^{\Omega}$.

Для метрики Кука (неспівпадання рангів об'єктів) (1) при використанні утилітарного критерію обчислюється медіана Кука-Сейфорда:

$$R^{CS} \in Arg \min_{R \in \Re^{\Omega}} \sum_{l \in L} d\left(R, R^l\right). \qquad (3)$$

При використанні егалітарного критерію обчислюється ГВ-медіана (компроміс):

$$R^{ГВ} \in Arg \min_{R \in \Re^{\Omega}} \max_{l \in L} d\left(R, R^l\right). \qquad (4)$$

Інколи визначається також середнє з використанням евклідової метрики:

$$R^{S} \in Arg \min_{R \in \Re^{\Omega}} \sum_{l \in L} d^2\left(R, R^l\right). \qquad (5)$$

Позначимо через $\Re^{\Theta}$ – множину МПП, які відповідають усім можливим ранжуванням n об'єктів.

Для метрики Хемінга (2) при використанні утилітарного критерію обчислюється медіана Кемені-Снелла:

$$R^{KC} \in Arg \min_{B \in \Re^{\Theta}} \sum_{l \in L} d\left(B, B^l\right). \qquad (6)$$

При використанні егалітарного критерію обчислюється ВГ-медіана (компроміс):

$$R^{ВГ} \in Arg \min_{B \in \Re^{\Theta}} \max_{l \in L} d\left(B, B^l\right). \qquad (7)$$

Середнє при використанні метрики Хемінга (2) визначається як

$$R^{C} \in Arg \min_{B \in \Re^{\Theta}} \sum_{l \in L} d^2\left(B, B^l\right). \qquad (8)$$

Методи визначення медіан виду (3)-(8) та їх особливості розглядаються у монографії [2].

На основі обчислення визначених вище медіан (3)-(8) у цій роботі пропонується обчислювати коефіцієнти відносної компетентності експертів у вигляді функції належності нечіткій множині.

4 Спосіб визначення функції належності нечіткій множині шляхом аналізу частотності значень

На сьогодні в напрямку експертного оцінювання розроблено та обґрунтовано значну кількість методів визначення колективного рішення та обчислення відносної компетентності експертів. Усі вони мають право на існування і можуть бути використані для прийняття рішень. Оскільки розроблено різноманітний апарат визначення групових оцінок, ми його використовуємо, але «згортаємо» примножені сутності в одну ФН. Для визначення компетентності експертів при

розв’язанні конкретної задачі експертного оцінювання авторами пропонується не створювати нові методи, а скористатися аналізом існуючих.

У [7] розглянуто одну з найважливіших задач експертного оцінювання, якою є визначення компетентності експертів. В основу запропонованих методів визначення відносної компетентності експертів в умовах різнопланової невизначеності покладено аксіому незміщеності: «Судження більшості компетентні» та її наслідок про те, що «найбільш компетентним є той експерт, судження якого у більшості випадків співпали з висновками більшості експертів».

Наведемо фактори, які впливають на кількість розв’язків задачі визначення результуючого ранжування:

– множина метрик, що застосовуються;

– множина критеріїв визначення близькості між ранжуваннями;

– неєдиний розв’язок задач визначення медіани;

– застосування для визначення результуючого ранжування множини класичних правило вибору: Кондорсе, Борда, Сімпсона, Копленда, Нансона, альтернативних голосів, відносної більшості тощо [1].

Особливістю неузгоджених МПП та множини ранжувань об’єктів є велика кількість ефективних розв’язків у задачі визначення групового ранжування об’єктів. Зокрема, кожна з обчислених медіан (3)-(8) може бути неєдиною.

У задачах ранжування мірою компетентності слугують відстані від експертних ранжувань до обчислених медіан. Оскільки варіантів визначення результуючого ранжування можуть бути десятки, коефіцієнти компетентності можуть бути обчислені у вигляді ФН.

5 Висновки

Запропоновано класифікацію методів визначення компетентності експертів в умовах різнопланової невизначеності. Обґрунтовано перспективність застосування апостеріорного підходу до обчислення коефіцієнтів відносної компетентності експертів. На основі застосування методів, розроблених для такого підходу, та аналізу їх з використанням методу визначення ФН на основі частотності значень, запропоновано підхід до побудови вагових коефіцієнтів відносної компетентності експертів у вигляді ФН.

Джерела

1. Волошин О.Ф., Мащенко С.О. Теорія прийняття рішень: Навчальний посібник. – К.: Видавничо-поліграфічний центр "Київський університет", 2006. − 304 с.

2. Гнатієнко Г.М., Снитюк В.Є. Експертні технології прийняття рішень: Монографія. – К.: ТОВ «Маклаут», 2008. – 444с.

3. Гнатієнко Г.М. Алгоритми визначення функції належності шляхом аналізу частотності значень//Праці ІІІ-ї міжнародної школи-семінару “Теорія прийняття рішень”, Ужгород, УжНу, 2006.- С.32-34.

4. Снитюк В.Є., Гнатієнко Г.М. Оптимізація процесу оцінювання в умовах невизначеності на основі структуризації предметної області та аксіоми незміщеності // Искусственный интеллект. – 2008. – № 3. – С. 217–222.

5. Снитюк В.Є. Спрямована оптимізація і особливості еволюційної генерації потенційних розв’язків // VI міжнародна школа-семінар «Теорія прийняття рішень», Ужгород, 1 – 6 жовтня 2012 р. – Ужгород : «Інвізор», 2012 – С. 182-183.

6. Гнатієнко Г.М. Визначення вагових коефіцієнтів критеріїв задачі багатокритеріальної оптимізації у вигляді функцій належності нечіткій множині // Матеріали доповідей V Міжнародної науково-практичної конференції «Інформаційні технології та взаємодії» (IT&I – 2018). – К. ВПЦ «Київський університет», 2018. – 343 с. С.29-30.

7. Снитюк В.Е., Рифат Мохаммед Али, Модели и методы определения компетентности экспертов на базе аксиомы несмещенности // Вісник Черкаського інженерно-технологічного інституту. − 2000. − № 4. – С. 121–126.

ЖИВУЧІСТЬ Й КОМПРОМІС: ФОРМУВАННЯ ПАРЕТО-ОПТИМАЛЬНИХ РІШЕНЬ ОРГАНІЗАЦІЙНОГО УПРАВЛІННЯ ЗАСОБАМИ EXCEL

О.Г. Додонов, А.І. Кузьмичов

Інститут проблем реєстрації інформації Національної академії наук України, м. Київ, Україна
dodonov@ipri.kiev.ua, akuzmychov@gmail.com

Жива практика свідчить, що в умовах конфліктів, які є завжди і всюди, за необхідності якнайкраще врахувати декілька (k, $k > 1$) часто суперечливих і недосяжних «ідеальних» альтернатив, прийняте остаточне і компромісне «реальне» рішення знаходиться десь «посередині» проміж цими альтернативами із врахуванням умов і обмежень необхідних ресурсів, визначених природою, штатним розкладом, запасами чи кошторисом. Характерний і простий приклад – купівля/продаж, де продавець і покупець, маючи протилежні цілі, максимум доходу і мінімум витрат, певними зустрічними поступками знаходять компромісне рішення (*trade-off*), що задовольняє обох.

Пошук компромісу – цілком природне явище і реальна проблема – є вимушеним, вкрай важливим й навіть критичним. Адже коли за кількома суперечливими цілями з різних причин формується недостатньо обґрунтоване і зважене рішення, його наслідком можуть статися неочікувані важкі втрати і витрати в конфліктних ситуаціях, де застосовується поняття «живучість», у першу чергу, в області життєзабезпечення верств суспільства і обороноздатності держави[1].

Прийняте рішення може мати стратегічні наслідки, наприклад, вдала конструкція продукту може означати тривалий успіх і стійку репутацію у бізнесі і навпаки, погані властивості можуть стати фатальними для живучості підприємства в умовах конкурентного середовища.

Розроблений продукт характеризується набором ознак: вартістю виробничого процесу, ціною, модифікацією, технічними і економічними параметрами, розмірами, ефективністю чи зручністю у користуванні тощо. Теоретично, існує нескінченна кількість комбінацій цих ознак, тож ключова проблема виробника – прийняття компромісних рішень, де багато цілей, критеріїв та обмежень, аби ще на етапі проектування визначити, скажімо, "найкращу" композицію ознак продукту та їх значень для забезпечення тривалої живучості бізнесу в умовах реальних конфліктів, внутрішніх і зовнішніх.

Прийняття компромісних рішень (*DM, Decision Making*) – це не лише дія, аби знайти задовільний результат, а складний і неперервний *процес* вибору критеріїв та їх вимірів, визначення альтернатив, способів збору, оцінки та обробки інформації, отримання та оцінки поточних результатів, перегляду, повторення і зміни використаних підходів на основі досягнутих результатів. Вважається, що компроміс не можна однозначно вирішити, його можна розв’язати, бо існує уявлення, що певні зі сторін за різних обставин більш податливі на поступки.

Для пошуку зважених управлінських рішень для складно організованих системних задач визначився напрям компромісної (багатокритеріальної, багатоцільової, векторної) чи Парето-оптимізації, популярний будь-де, зокрема, в управлінні проектами, продукт-менеджменті та бізнес-аналітиці, де в умовах постійних змін набору і значень початкових даних знайденими компромісними рішеннями підтримується живучість державних, соціальних, індустріальних, економічних та організаційних процесів й відповідних об'єктів.

Вперше проблему багатокритеріальної оптимізації розглянув італійський економіст В. Парето (1848-1923) в 1896 році, досліджуючи товарний обмін. За його теорією, якщо суспільство досягло і перебуває в стані загальної економічної рівноваги, тоді здійснюється оптимальний розподіл продукції, товарів і послуг, – при мінімально необхідному використанні ресурсів забезпечується максимально можливе задоволення потреб.

На практиці ж таку рівновагу важно досягти, тож відшукують розв’язанням різноманітних багатокритеріальних оптимізаційних задач компромісний оптимум, що має задовольнити декілька, конфліктуючих між собою, критеріїв. Для цього треба знайти безліч рішень, їх називають Парето-оптимальними, де остаточне рішення приймає ОПР, зважуючи на пріоритети поставлених цілей (скажімо, у вигляді їх вагових коефіцієнтів).

Спрощений підхід цього обчислювального процесу – зведення k-критеріальної ($k > 1$) оптимізаційної задачі до 1-критеріальної: тут одна з цілей визначається *домінуючою* і саме для неї формулюється оптимізаційна задача (із «твердими» обмеженнями), де усі інші k - 1 цілей вимушено задані певними «м’якими» обмеженнями на шукані змінні, роль їхніх сторін пасивна і, зрозуміло, не завжди влаштовує відсторонених від прийняття рішень учасників, усі обмеження визначають відповідну ОДР (область допустимих рішень).

[1] «… Наиболее полной из таких компромиссных мер будет …». Г. Ляхов. Очерки по живучести боевого корабля. Управление ВМС РККА, 1932 г.
«Корабль – это всегда компромисс». А.В. Неменко. Черноморский флот в годы войны. Вече, 2015. – 320 с.

Приклад – класична транспортна задача, її цільова функція – мінімум транспортних витрат для забезпечення попиту, яка відтворює інтереси лише транспортної компанії, але ж ніяк інтереси постачальників (у вигляді очікуваних доходів) й споживачів (у вигляді сервісу), які в цій моделі представлені лише пропозиціями і попитом. Але ж мінімізуючи транспортні витрати, не виключено, що одночасно погіршуються певні показники постачальників і споживачів, тож отриманий локальний оптимум можливо сумнівний щодо врахування інтересів усіх учасників процесу. Бізнес-практика реагує на такі ситуації утворенням альтернативних транспортних служб («Нова пошта», різні перевізники), кожна із власними цілями, ресурсами і бюджетами, вони усі грають на одному полі в межах фіксованої ОДР, аби досягти своїх цілей якнайкраще.

Багатокритеріальна оптимізація (БКО) – аналіз[2] та прийняття[3] багатокритеріальних (багатоцільових, компромісних) рішень – власна область досліджень і неодмінна складова математичного програмування, дослідження операцій та бізнес-аналітики. Типова сфера застосування – міждисциплінарні дослідження (техніко-економічні, економіко-екологічні, хіміко-фізичні) зі сфери організаційного управління системами, представлені інтегрованими науками типу економетрія чи обчислювальна хімія, де спільно вивчаються комплекси об'єктів різної природи побудовою специфічних математичних та оптимізаційних моделей за суперечливими критеріями. Історія БКО розпочалася майже зразу із відкриттям моделей і методів лінійної оптимізації (1951 р.), тоді ж сформувався апарат багатокритеріальної лінійної оптимізації (БКЛО), використаний у наведених прикладах.

Оптимальне компромісне рішення визначається одночасним розв'язуванням глобальної і кількох локальних задач оптимізації, рішення зводиться до мінімізації суми відстаней від кількох знайдених локальних оптимумів до шуканої ідеальної точки (її координати – Парето-оптимальне рішення) із врахуванням пріоритетів цілей.

Існують різні метрики і методи пошуку відстаней між точками в *n*-вимірному просторі, відповідно, визначається лінійна чи нелінійна задача оптимізації. Скажімо, при застосуванні нелінійної (евклідової) метрики формується непроста для реалізації задача нелінійного програмування, тому для практиці бажано знайти метод для його реалізації засобами лінійного програмування (зазвичай, симплекс-методом, який дає змогу отримати розв'язки прямої і двоїстої задач ЛП).

Таким є інтерактивний (за участю ОПР) *зважений min-max метод* (*weighted min-max method*), за яким формується і розв'язується задача лінійної *k*-критеріальної лінійної оптимізації, де: *n* – число невідомих, *m* – число обмежень, *k* – число критеріїв, цільових функцій.

Приклад 1.

Підприємство випускає 2 види продукції (П1, П2), відповідно, має дві «позитивні» цілі (ЦФ1, ЦФ2) – *максимізувати* дохід і якість, використовуючи *3* види економічних (ресурсних) обмежень (Р1, Р2 та Р3) типу «ЛЧ ≤ ПЧ».

Виробничий процес вимушено супроводжується викидами трьох шкідливих речовин, це три «негативні» цілі (ЦФ3, ЦФ4, ЦФ5), які треба *мінімізувати*, обмеженнями для них є 4 екологічні граничні значення (В1, В2, В3 та В4) типу «ЛЧ ≥ ПЧ)[4].

Це багатокритеріальний варіант класичної задачі про оптимальний асортимент продукції (про оптимальний розподіл обмежених ресурсів).

Отже, маємо 5-критеріальну (із 5-ма суперечливими цілями й, відповідно, із 5-ма цільовими функціями) задачу лінійної оптимізації з сумісною системою 7 лінійних обмежень для двох шуканих невідомих (x_1, x_2), результат можна показати на площині $x_1 0 x_2$.

7 обмежень утворюють ОДР (область допустимих розв'язків), в межах якої відшукується і пропонується для аналізу і прийняття керівництвом підприємства економіко-екологічний ***компроміс*** – оптимальний глобальний («реальний») план (шукані координати точки Парето) із відповідною ЦФ, побудований на значеннях 5 оптимальних локальних («ідеальних» і, зазвичай, одночасно недосяжних) планів.

Шуканий компроміс має досягатися взаємними поступками економістів та екологів, за ним дохід і якість зменшаться, а викиди збільшаться (теоретично, в глобальний оптимум може увійти певний локальний оптимум).

На початку пріоритети (вагові коефіцієнти цілей) однакові (1).

Початкові дані:

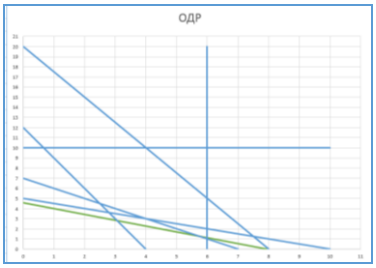
- коефіцієнти цільових функцій (*C*), вагові коефіцієнти (*w*)
- коефіцієнти лівих частин і праві частини обмежень

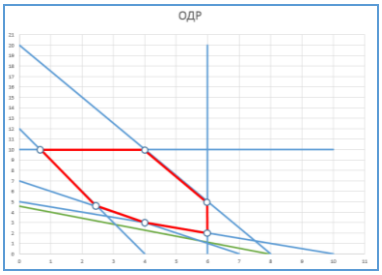
[2] MODA (Multiple Objective Decision Analysis)
[3] MCDM (Multiple Criteria Decision Making)
[4] ЛЧ, ПЧ – ліва, права частина обмеження, вимога Excel: ЛЧ (формула з невідомими), ПЧ (константа)

С	П1	П2	Крит.	Вага
ЦФ1	3	3	max	1
ЦФ2	9	1	max	1
ЦФ3	40	32	min	1
ЦФ4	800	1250	min	1
ЦФ5	0,2	0,45	min	1

Обм.	x1	x2	ПЧ
P1	20	8	**160**
P2	1	0	**6**
P3	0	1	**10**
B4	12	4	**48**
B5	4	4	**28**
B6	10	20	**100**
B7	15	26	**119**

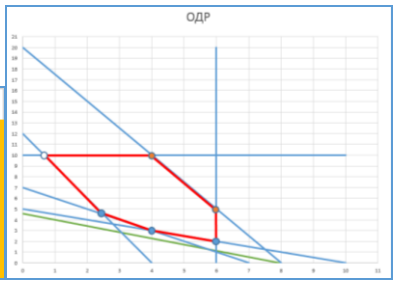
За заданими 7 обмеженнями будується ОДР, обмеження В7 виявилося поза конфлікту цілей, реалізується автоматично, 6 кутових точок – претенденти 5 локальних оптимумів:





Послідовним розв'язком 5 задач ЛП визначено 5 «ідеальних» точок в ОДР, дві зверху (max), три знизу (min):

План	x1	x2	Ідеал
ЦФ1	4	10	42
ЦФ2	6	5	59
ЦФ3	2,5	4,5	244
ЦФ4	4	3	6950
ЦФ5	6	2	2,1



Пошук точки Парето

1) Формування стовпця «реальних» значень ЦФ (Реал.), використовуються коефіцієнти 5 ЦФ (B3:C7) і рядок шуканих координат X^0 (B15:C15)
2) Обчислення відносних відхилень між «ідеальними» (I) та «реальними» (R) значеннями ЦФ: max: (I-R)/I; min: (R-I)/I, стовпець Відх.
3) Формування стовпця зважених відхилень (Відх. × Вага), ZV
4) Формування стовпця ЛЧ(Q) обмежень (I18:I24)
5) шукане значення Q буде розміщено в клітинці D15

Задача лінійної оптимізації

I. Знайти $X^0 = (x_1^0, x_2^0)$ та Q

II. ЦФ $Q \rightarrow$ min

III. за обмежень:

ЛЧ $(Q) \leq \geq$ ПЧ $\quad ZV \leq Q \quad X^0 \geq 0.$

Результат

	A	B	C	D	E	F	G	H	I	J	K
2	С	П1	П2	Крит.							
3	ЦФ1	3	3	max							
4	ЦФ2	9	1	max							
5	ЦФ3	40	32	min							
6	ЦФ4	800	1250	min							
7	ЦФ5	0,2	0,45	min							
8											
9	**План**	x1	x2	Ідеал	Реал.	Відх.	Вага	Зв. відх.	ТЦ		
10	ЦФ1	4	10	42	27,2	0,352	1	**0,352**	-0,68		
11	ЦФ2	6	5	59	48,8	0,173	1	0,173	0		
12	ЦФ3	2,5	4,5	244	329,9	0,352	1	**0,352**	-0,25		
13	ЦФ4	4	3	6950	9101,7	0,310	1	0,310	0		
14	ЦФ5	6	2	2,1	2,8	0,352	1	**0,352**	-0,07		
15	ЦФQ	**4,97**	**4,1**	**0,352**	ЦФQ						
16											
17	Обм.	x1	x2	ЛЧ (1)	ЛЧ (2)	ЛЧ (3)	ЛЧ (4)	ЛЧ (5)	ЛЧ (Q)		ПЧ
18	Р1	20	8	**160**	**160**	86	104	136	132,1	≤	**160**
19	Р2	1	0	4	**6**	2,5	4	**6**	5,0	≤	**6**
20	Р3	0	1	**10**	5	4,5	3	2	4,10	≤	**10**
21	В4	12	4	88	92	**48**	60	80	76,0	≥	**48**
22	В5	4	4	56	44	**28**	**28**	32	36,3	≥	**28**
23	В6	10	20	240	160	115	**100**	**100**	131,7	≥	**100**
24	В7	15	26	320	220	154,5	138	142	181,2	≥	**119**

Пряма задача

Координати точки Парето: $x_1^0 = 4,97; x_2^0 = 4,1$, $Q = 0,352 = 35,2\%$ (це мінімально можливе максимальне відхилення локальних оптимумів від точки Парето, стосується трьох найвіддаленіших точок ЦФ: 1, 3 та 5).

Двоїста задача

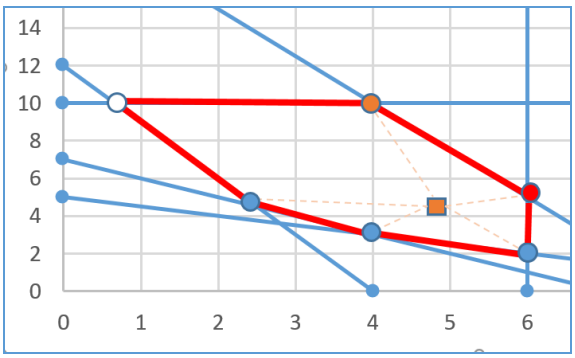
Тіньові ціни (ТЦ) заданих коефіцієнтів глобальної ЦФ – від'ємні чи нульові числа – вказують на можливість зменшення значення Q, це стосується ЦФ1 (-0,68), ЦФ3 (-0,25) та ЦФ5 (-0,07), точки реального оптимуму яких найвіддаленіші від знайденої точки Парето (їх поступки у знайденому компромісі найбільші, 35,2%):

- кутова точка ЦФ1 утворена перетином обмежень Р1 та Р3 (жирний шрифт), зі зменшенням ПЧ цих обмежень (160 та 10), нова кутова точка стане ближчою до нової точки Парето (тоді поступки ЦФ1 будуть меншими із одночасним зростанням поступок інших ЦФ)
- аналогічно, для ЦФ3 та ЦФ5.

Для цього варіанту задачі тіньові ціни заданих значень ПЧ обмежень – нулі (бо точка Парето розташована поза ліній обмежень, ознака – усі обмеження-нерівності.)

Результат

С	П1	П2	Крит.							
ЦФ1	3	3	max							
ЦФ2	9	1	max							
ЦФ3	40	32	min							
ЦФ4	800	1250	min							
ЦФ5	0,2	0,45	min							
План	x1	x2	Ідеал	Реал.	Відх.	Вага	Зв. відх.	ТЦ		
ЦФ1	4	10	42	27,2	0,35	1	0,35	-0,68		
ЦФ2	6	5	59	48,8	0,17	1	0,17	0		
ЦФ3	2,5	4,5	244	329,9	0,35	1	0,35	-0,25		
ЦФ4	4	3	6950	9101,7	0,31	1	0,31	0		
ЦФ5	6	2	2,1	2,8	0,35	1	0,35	-0,07		
ЦФQ	**4,97**	**4,1**	**0,352**	ЦФQ						
Обм.	x1	x2	ЛЧ (1)	ЛЧ (2)	ЛЧ (3)	ЛЧ (4)	ЛЧ (5)	ЛЧ (Q)		ПЧ
Р1	20	8	**160**	**160**	86	104	136	132,1	≤	**160**
Р2	1	0	4	**6**	2,5	4	**6**	4,966	≤	**6**
Р3	0	1	**10**	5	4,5	3	2	4,103	≤	**10**
В4	12	4	88	92	**48**	60	80	76	≥	**48**
В5	4	4	56	44	**28**	**28**	32	36,28	≥	**28**
В6	10	20	240	160	115	**100**	**100**	131,7	≥	**100**



Властивість точки Парето – серед 5 локальних ЦФ немає домінуючих, вони усі рівноправні, з однаковими (так задано) ваговими коефіцієнтами, в житті так буває не часто, бо завжди серед альтернатив є якась важливіша за певних зовнішніх умов. Реальний факт – світова екологічна криза свідчить, що для виробників щоденний дохід важливіший за будь-що інше і екологічні обмеження явно не задовольняються.

Скажімо, у нашому прикладі, якщо таки здійснювати знайдений компроміс із різними інтересами-цілями економістів і екологів, тоді реальний дохід (27,2) підприємства має бути зменшеним майже на третину (42) і з цим виробник, скоріше за все, навряд чи погодиться, значить, він буде намагатися підвищити свою вагу.

Зміна ваги певної цілі хоч на йоту – це руйнування знайденого чутливого і хиткого балансу цілей, отриманого за їхніми заданими ваговими коефіцієнтами, бо з'являється домінуюча ціль, покращення оптимуму якої автоматично погіршує реальні оптимуми інші цілей.

Однак, практика свідчить, що це може бути вимушене і вкрай необхідне рішення на випадок, якщо сталася (чи може статися) критична ситуація: природна, політична, техногенна тощо в разі, скажімо, активного впливу на системний об'єкт неочікуваних факторів.

А послідовно змінюючи вагу тепер вже домінуючої цілі і відслідковуючи наслідки для інших цілей, ми, фактично, здійснюємо прогноз поведінки об'єкту саме у кількісному вимірі за принципом «що - якщо» і тут може бути надзвичайно корисною процедура побудови фронту Парето.

Фронт Парето – множина Парето-оптимальних розв'язків (на площині – точок Парето), кожну ми автоматично обчислюємо за допомогою ефективної надбудови *SolverTable* (у вільному перекладі – таблиця знайдених оптимумів) – зі зміною значень певних початкових даних як аргументів (входів) організується циклічний обчислювальний процес, де в тілі циклу діє стандартна надбудова *Excel Solver* (*Поиск решения*, *Пошук рішення*), яка на кожному кроці обчислює поточні значення функцій (виходів), які наприкінці представлені таблицями і графіками.

Оскільки локальні ЦФ задані своїми попередньо знайденими оптимумами, мова йде про задачу пошуку оптимуму глобальної ЦФ, яка обчислює значення *Q*, на що впливають лише вагові коефіцієнти ЦФ, тож саме вони можуть застосовуватися для побудови фронту в якості аргументів, одного чи двох одночасно.

Приклад 2.

ОПР (аналітик, експерт) хоче побудувати прогноз: реальних значень усіх ЦФ, шуканого плану (X^0) і значення *Q* збільшенням ваги *w*1 (ціль – gbvмаксимум доходу) від 1 до 10, розуміючи, що «реальні» значення ЦФ, окрім ЦФ1, при цьому будуть погіршуватися.

Отже, *w*1 – аргумент (вхід), функцій-«виходів» 8: «реальні» оптимуми ЦФ1 ÷ ЦФ5, шукані значення x_1, x_2, *Q*, фактично, це таблиця розміром 10 × 8, отримана розв'язанням 10 задач оптимізації.

Процедура і результат

w1	x1	x2	Q	Реал 1	Реал 2	Реал 3	Реал 4	Реал 5
1	4,97	4,10	0,35	27,2	48,8	329,9	9101,7	2,8
2	5,63	4,65	0,53	30,8	55,3	373,8	10312,5	3,2
3	6,00	5,00	0,64	33,0	59,0	400,0	11050,0	3,5
4	5,75	5,62	0,75	34,1	57,4	409,9	11623,3	3,7
5	5,56	6,10	0,84	35,0	56,1	417,6	12070,7	3,9
6	5,41	6,48	0,90	35,7	55,1	423,7	12429,7	4,0
7	5,28	6,80	0,96	36,2	54,3	428,8	12724,0	4,1
8	5,17	7,06	1,01	36,7	53,6	433,0	12969,7	4,2
9	5,08	7,29	1,05	37,1	53,1	436,6	13178,0	4,3
10	5,01	7,48	1,08	37,5	52,6	439,7	13356,7	4,4

Рис. 1. Таблиця результатів

Реал 1	Реал 2	Реал 3	Реал 4	Реал 5
27,2	48,8	329,9	9101,7	2,8
30,8	55,3	373,8	10312,5	3,2
33,0	59,0	400,0	11050,0	3,5
34,1	57,4	409,9	11623,3	3,7
35,0	56,1	417,6	12070,7	3,9
35,7	55,1	423,7	12429,7	4,0
36,2	54,3	428,8	12724,0	4,1
36,7	53,6	433,0	12969,7	4,2
37,1	53,1	436,6	13178,0	4,3
37,5	52,6	439,7	13356,7	4,4

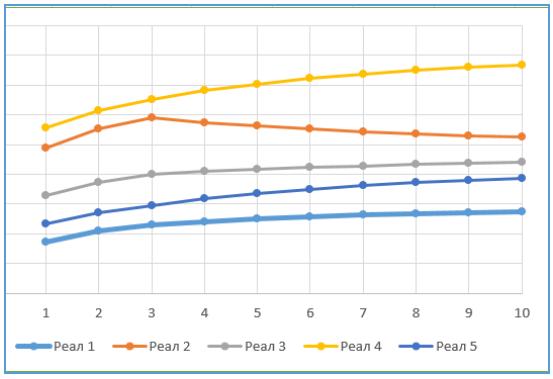
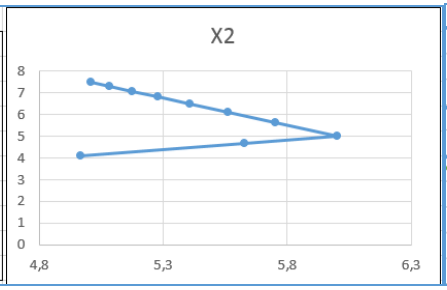


Рис. 2. Зміна «реальних» значень ЦФ (на діаграмі масштаб змінено)

W1	X1	X2
1	4,97	4,10
2	5,63	4,65
3	6,00	5,00
4	5,75	5,62
5	5,56	6,10
6	5,41	6,48
7	5,28	6,80
8	5,17	7,06
9	5,08	7,29
10	5,01	7,48



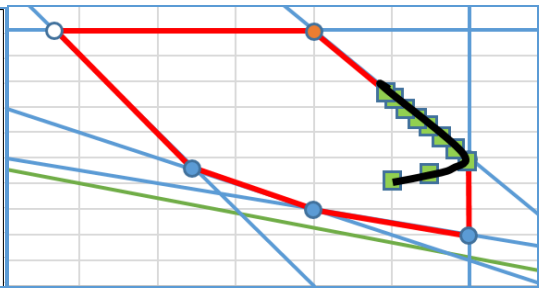
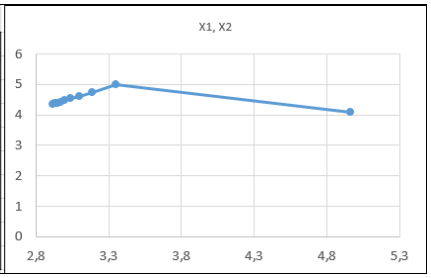


Рис. 4. Фронт Парето: ЦФ1 (max). *w*1 := 1, 10, 1

W3	X1	X2
1	4,97	4,10
2	3,35	4,98
3	3,19	4,74
4	3,10	4,61
5	3,04	4,53
6	3,00	4,47
7	2,97	4,42
8	2,95	4,39
9	2,93	4,37
10	2,92	4,34



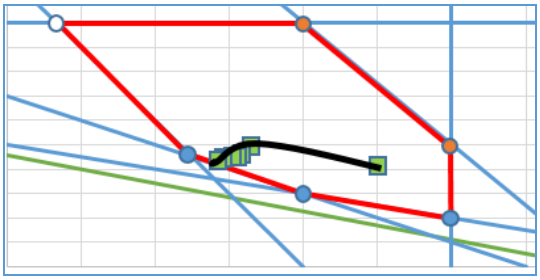


Рис. 4. Фронт Парето: ЦФ3 (min). *w*3 := 1, 10, 1

Аналіз результату

З отриманих результатів видно, що зі зростанням ваги ЦФ1, усі інші показники погіршуються.

На точку Парето постійно діють сили тяжіння *k* локальних оптимумів (наче гумові нитки), пропорційні їхнім ваговим коефіцієнтам, поступове зростання ваги домінуючої ЦФ на кожному кроці змінює баланс сил тяжіння (і смисл компромісу), надаючи можливість наблизитись якнайшвидше до визначеної ідеальної точки, так формується фронт Парето:

- ЦФ1 (max), рис. 3. Зі зміною значення *w*1 від 1 до 10 з кроком 1 поточна точка Парето на 3-му кроці вийшла на пряму між точками (6, 5) і (4, 10), що з'єднує дві точки максимуму ЦФ2 та ЦФ1, і далі рухається «максимальною» прямою до «ідеальної» точки (4, 10). Тобто, фронт Парето складається з двох дільниць:
  - (7,97; 4,10) – (6,00; 5,00), рух на лінію обмеження, та
  - (6,00; 5,00) – (5,03; 7,38) – рух цією лінією,

  аби якнайшвидше послабити вплив трьох ЦФ (min).
- ЦФ3 (min), рис. 4. Зі зростанням ваги *w*3, точка Парето діє аналогічно: спочатку реагує на вплив двох ЦФ (max), а вже на 3-му кроці круто рухається на «мінімальну» пряму і далі нею до визначеної ідеальної точки оптимуму (2,5; 4,5), формуючи фронт.

Отже, якщо в задачі БКО усі ЦФ мають критерій оптимізації мінімум, точка Парето зразу розташовується на лініях, що з'єднують точки мінімумів, де формується й фронт; аналогічно, якщо критерії усіх ЦФ максимум, точка Парето розташовується на лініях, що з'єднують точки максимумів, і там формується фронт. Якщо ж критерії ЦФ мінімум/максимум, точка Парето розташовується десь «посередині» ОДР, а вже потім поступово зростаюча сила тяжіння рухає її на лінію обмеження, де розташована кутова точка визначеної ЦФ.

Приклад 3.

За постановкою задача аналогічна задачі із Прикладу 1, але тут випуск стосується 10 видів продуктів (П1 … П10), відповідно, відшукується оптимальний план $X^0 = (x_1^0,...,x_{10}^0)$ за тими ж ЦФ, критеріями, 6 обмеженнями та тим же методом.

Таблична модель, результат та його аналіз

C	П1	П2	П3	П4	П5	П6	П7	П8	П9	П10									
ЦФ1	40	32	45	100	65	74	15	3	46	26	min								
ЦФ2	45	8	14	11	39	79	12	95	61	93	min								
ЦФ3	0,2	0,5	54	90	68	52	24	2	5	47	min								
ЦФ4	3	3	3	7	40	20	51	1	34	40	max								
ЦФ5	9	1	84	63	77	14	4	86	4	2	max								
План	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	Ідеал	Реал.	Відх.	Вага	Зв.	ТЦ			
ЦФ1	**1,8**	**3,9**	0	0	0	0	**0,4**	0	0	0	**204,8**	**232,7**	0,14	1	0,14	0			
ЦФ2	**0,7**	**10,0**	0	0	0	0	0	0	0	0	**110,0**	**163,6**	0,49	1	0,49	**-0,11**			
ЦФ3	**6,0**	**2,0**	0	0	0	0	0	0	0	0	**2,1**	**3,1**	0,49	1	0,49	**-0,14**			
ЦФ4	**4,5**	**1,8**	0	0	0	0	0	0	0	**0,7**	**48,7**	**25,0**	0,49	1	0,49	**-0,57**			
ЦФ5	**4,6**	**8,2**	**0,1**	0	0	0	0	0	0	0	**61,3**	**31,4**	0,49	1	0,49	**-0,17**			
ЦФQ	**2,68**	**3,67**	0	0	0	0	0	**0,03**	**0,17**	0	**0,49**	ЦФQ							
ПС	0	0	3,3	5,8	3,9	3,28	1,0003	0	0	2,75									
Обмеження											ЛЧ (1)	ЛЧ (2)	ЛЧ (3)	ЛЧ (4)	ЛЧ (5)	ЛЧ (Q)		ПЧ	ТЦ
Рес1	12	4	19	13	1	9	24	2	24	1	**48**	**48**	80	62	90	51	≥	**48**	0
Рес2	4	4	11	7	12	16	24	24	25	20	33	42,67	32	40	53	30,56	≥	**28**	0
Рес3	10	20	20	4	24	2	8	8	24	25	**100**	206,7	**100**	**100**	212	105	≥	**100**	0
Рес4	20	8	22	5	14	19	5	3	23	24	70	93,33	136	123	**160**	86,99	≤	**160**	0
Рес5	1	0	10	16	21	12	3	21	15	2	3	0,667	**6**	**6**	**6**	**6**	≤	**6**	**-0,0011**
Рес6	0	1	13	15	20	15	14	18	25	11	**10**	**10**	2	**10**	**10**	9	≤	**10**	0

Коли фронт Парето не можна зобразити на площині траєкторією точки Парето, на площині будують лінію парних компромісів (*trade-off curve*), де кожна точка – це компроміс, пара координат двох ЦФ із протилежними цілями типу (Що, Якщо), тобто, (дохід, витрати).

Приклад 4.

Значення вагових коефіцієнтів ЦФ4, ЦФ5 (max) *w*4 та w5 одночасно зростають у діапазоні 1 ÷ 10, треба побудувати дві групи парних компромісів з координатами:

- (ЦФ1, ЦФ4), (ЦФ2, ЦФ4) та (ЦФ3, ЦФ4)
- (ЦФ1, ЦФ5), (ЦФ2, ЦФ5) та (ЦФ3, ЦФ5),

із застосуванням надбудови SolverTable (з двома входами) для розв'язання в одному сеансі 100 задач оптимізації із формуванням 3-ох квадратних матриць розміром 10 × 10 для ЦФ на мінімум і 2-ох таких матриць для ЦФ на максимум.

Результат

ЦФ1	1	2	3	4	5	6	7	8	9	10
1	232,7	236,5	242,9	252,3	259,6	265,5	270,3	274,3	277,6	280,5
2	288,6	299,4	325,1	311,3	293,0	278,7	270,3	274,3	277,6	280,5
3	331,3	331,3	343,8	360,8	367,8	358,2	350,4	344,0	338,7	334,2
4	364,8	364,8	364,8	377,1	388,1	396,0	394,8	390,2	386,3	383,0
5	392,9	392,9	392,9	393,8	406,0	414,4	420,3	416,1	416,0	413,5
6	414,7	414,7	414,7	414,7	419,5	428,5	435,2	434,9	431,6	429,8
7	432,2	432,2	432,2	432,2	432,2	439,5	446,5	449,5	446,3	443,7
8	446,4	446,4	446,4	446,4	446,4	448,3	455,6	461,2	458,1	455,6
9	458,3	458,3	458,3	458,3	458,3	458,3	463,1	468,9	467,7	465,2
10	469,0	469,0	469,0	469,0	469,0	469,0	464,0	457,0	453,8	451,1

ЦФ2	1	2	3	4	5	6	7	8	9	10
1	163,6	174,7	192,9	206,6	217,3	225,9	232,8	238,7	243,6	247,8
2	193,9	195,6	199,7	209,6	219,2	226,7	232,8	238,7	243,6	247,8
3	214,6	214,6	216,9	220,0	223,4	231,5	238,0	243,4	247,8	251,6
4	229,3	229,3	229,3	232,0	234,4	236,1	240,8	246,4	251,1	255,1
5	240,4	240,4	240,4	240,6	243,4	245,4	246,8	249,2	253,2	257,3
6	249,1	249,1	249,1	249,1	250,2	252,3	253,9	255,9	257,9	259,7
7	256,0	256,0	256,0	256,0	256,0	257,7	259,4	261,1	263,3	265,1
8	261,6	261,6	261,6	261,6	261,6	262,1	263,9	265,2	267,6	269,5
9	266,3	266,3	266,3	266,3	266,3	266,3	267,5	268,9	271,1	273,1
10	267,1	267,1	267,1	267,1	267,1	267,1	270,4	275,2	277,9	280,0

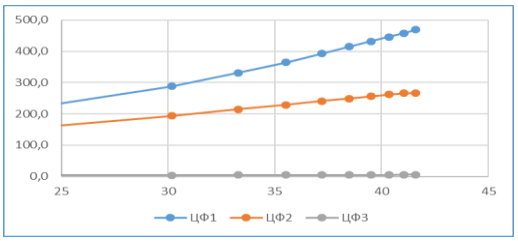
ЦФ3	1	2	3	4	5	6	7	8	9	10
1	3,1	3,3	3,7	3,9	4,1	4,3	4,4	4,6	4,7	4,7
2	3,7	3,7	3,8	4,0	4,2	4,3	4,4	4,6	4,7	4,7
3	4,1	4,1	4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8
4	4,4	4,4	4,4	4,4	4,5	4,5	4,6	4,7	4,8	4,9
5	4,6	4,6	4,6	4,6	4,6	4,7	4,7	4,8	4,8	4,9
6	4,8	4,8	4,8	4,8	4,8	4,8	4,8	4,9	4,9	5,0
7	4,9	4,9	4,9	4,9	4,9	4,9	5,0	5,0	5,0	5,1
8	5,0	5,0	5,0	5,0	5,0	5,0	5,0	5,1	5,1	5,1
9	5,1	5,1	5,1	5,1	5,1	5,1	5,1	5,1	5,2	5,2
10	5,2	5,2	5,2	5,2	5,2	5,2	5,2	5,3	5,3	5,3

ЦФ4	1	2	3	4	5	6	7	8	9	10
1	25,0	20,1	20,2	20,8	21,2	21,6	21,9	22,1	22,3	22,5
2	30,2	29,8	28,9	26,7	24,5	22,9	21,9	22,1	22,3	22,5
3	33,3	33,3	33,0	32,5	32,0	30,8	29,8	29,0	28,4	27,8
4	35,5	35,5	35,5	35,2	35,0	34,8	34,2	33,6	33,1	32,7
5	37,2	37,2	37,2	37,2	36,9	36,7	36,6	36,4	36,1	35,7
6	38,5	38,5	38,5	38,5	38,4	38,2	38,1	38,0	37,8	37,7
7	39,5	39,5	39,5	39,5	39,5	39,4	39,3	39,2	39,0	38,9
8	40,3	40,3	40,3	40,3	40,3	40,3	40,2	40,1	40,0	39,9
9	41,0	41,0	41,0	41,0	41,0	41,0	41,0	40,9	40,8	40,7
10	41,6	41,6	41,6	41,6	41,6	41,6	41,5	41,4	41,3	41,2

ЦФ5	1	2	3	4	5	6	7	8	9	10
1	31,4	43,3	45,9	47,8	49,3	50,5	51,5	52,3	53,0	53,6
2	34,4	37,4	44,6	47,4	49,1	50,5	51,5	52,3	53,0	53,6
3	38,1	38,1	41,5	46,0	48,7	50,0	51,1	52,0	52,8	53,4
4	40,8	40,8	40,8	44,3	47,4	49,6	50,9	51,8	52,6	53,2
5	42,8	42,8	42,8	43,1	46,4	48,7	50,4	51,6	52,4	53,1
6	44,4	44,4	44,4	44,4	45,7	48,1	49,9	51,1	52,1	53,0
7	45,7	45,7	45,7	45,7	45,7	47,6	49,4	50,8	51,8	52,7
8	46,7	46,7	46,7	46,7	46,7	47,2	49,1	50,5	51,5	52,4
9	47,6	47,6	47,6	47,6	47,6	47,6	48,8	50,2	51,3	52,2
10	47,6	47,6	47,6	47,6	47,6	47,6	48,4	49,8	50,9	51,8

За цими даними можна побудувати усі необхідні залежності, як от (ЦФ1, ЦФ4), (ЦФ2, ЦФ4) та (ЦФ3, ЦФ4) при $w4 = 1 \div 10$, $w5 = 1$:

w4	ЦФ4	ЦФ1	ЦФ2	ЦФ3
1	25,0	232,7	163,6	3,1
2	30,2	288,6	193,9	3,7
3	33,3	331,3	214,6	4,1
4	35,5	364,8	229,3	4,4
5	37,2	392,9	240,4	4,6
6	38,5	414,7	249,1	4,8
7	39,5	432,2	256,0	4,9
8	40,3	446,4	261,6	5,0
9	41,0	458,3	266,3	5,1
10	41,6	469,0	267,1	5,2



Література

1. Ragsdale C. Spreadsheet Modeling and Decision Analysis. A Practical Introduction to Business Analytics, 8-ed. – Cengage Learning, 2018. – 869 p.
2. Albright C., Winston W. Business Analytics: Data Analysis and Decision Making, 6-ed. – Cengage Learning, 2017. – 1145 p.
3. Evans G. Multiple Criteria Decision Analysis for Industrial Engineering: Methodology and Applications. CRC Press, 2017. – 468 p.
4. Balakrishnan A. et al. Managerial Decision Modeling. Business Analytics with Spreadsheets, 4-ed. – Deg Press, 2017. – 829 p.
5. Кузьмичов А.І. Оптимізаційне моделювання в Excel. Практикум. – К.: ІПРІ НАНУ, 2017. – 438 с.
6. Додонов О.Г. Комп'ютерне моделювання процесів організаційного управління/Вісник НАН України, 2016, №1. – С. 69-77
7. Badiru A. Industrial and System Engineering, 2-ed. – CRC Press, 2014. – 1239 p.
8. Parnell G. et al. Decision making in systems engineering and management, 2-ed. – Wiley, 2011. – 546 p.
9. Додонов О.Г. та ін. Живучість складних систем: аналіз та моделювання. Навч. пос.. – К.: НТУУ «КПІ», 2009. – 264 с.
10. Badiru A. Handbook of Military Industrial Engineering. – CRC Press, 2009. – 787 p.
11. Ravindran R. Operations Research Methodologies. – CRC Press, 2009. – 478 p.
12. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач, 2-е изд. – М.: Физматлит, 2007. – 256 с.
13. Ehrgott M. Multicriteria Oprimization, 2-ed. – Springer, 2005. – 328 p.
14. Штойер Р. Многокритериальная оптимизация. Теория, вычисления и приложения. Пер. с англ. – М.: Радио исвязь, 1992. – 504 с.
15. Yu P.L. Multiple-Criteria Decision Making. Concepts, Techniques, and Extensions. – Plemum Press, 1985. – 395 p.
16. Кини Р.Л., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения: Пер. с англ. – М.: Радио и связь, 1981. – 560 с.

ПОБУДОВА ФОРМАЛЬНОЇ МОДЕЛІ ІНТЕРНЕТ-МАРШРУТИЗАЦІЇ ДЛЯ ОЦІНКИ ВПЛИВУ АТАК З ПЕРЕХОПЛЕННЯМ МАРШРУТІВ

Віталій Зубок

Інститут проблем моделювання в енергетиці ім. Г.Є.Пухова
Національної академії наук України
Вул. Генерала Наумова, 15, Київ, 03164, Україна
vitaly.zubok@gmail.com

Анотація. Оцінка ризиків перехоплення маршруту потребує кількісного виміру впливу атаки на викривлення маршрутизації, а отже - на збитки від порушення безпеки інформації. Для цього пропонується шлях дослідження топології зв'язків між автономними системами мережі Інтернет для подальшого формулювання задачі обробки ризиків як задачі про топологію. Одним з найважливіших етапів на шляху моделювання впливу атак на маршрутизацію є побудова формальної моделі глобальної Інтернет-маршрутизації. В роботі запропоновано формальний опис об'єктів глобальної маршрутизації та відносин між ними, а також процесу вибору маршруту.

Ключові слова: глобальна маршрутизація, формальна модель глобальної маршрутизації, перехоплення маршрутів, оцінка ризиків, кібербезпека.

1. Актуальність

Автономні системи (Autonomous Systems, AS) використовують протокол прикордонного шлюзу (Border Gateway Protocol, BGP) для обміну маршрутами до своїх префіксів IP та встановлення міждоменних маршрутів в Інтернеті. BGP - це розподілений протокол, в якому бракує валідації маршрутів та автентифікації. Як результат, AS може рекламувати нелегітимні маршрути для IP-префіксів, якими вона не володіє. Ці незаконні анонси маршрутів поширюють та «забруднюють» Інтернет, впливаючи на доступність послуг, цілісність та конфіденційність комунікацій. Це явище, яке називається викраденням префіксу BGP (BGP hijack), може бути спричинене неправильною конфігурацією маршрутизатора або зловмисними атаками.

Ці події мають значний вплив і часто спостерігаються, незважаючи на серйозність такої вразливості інфраструктури Інтернету, і це підкреслює неефективність існуючих контрзаходів. В даний час мережі покладаються на практичні реактивні механізми, щоб спробувати захиститись від викрадення префікса, оскільки запропоновані проактивні механізми (наприклад, RPKI [1]) є повністю ефективними лише при глобальному розгортанні, а оператори неохоче розгортають їх через пов'язані з цим технічні та фінансові витрати. Захист проти викрадення реактивно складається з двох етапів: виявлення та пом'якшення.

Аналіз механізмів проведення атаки в залежності від її цілей та варіанти її реалізації детально описано в [2]. Виявлення в основному забезпечується сторонніми службами, такими як BGPMon, які сповіщають адміністратора мережі про підозрілі події, пов'язані з їх префіксами, на основі інформації про маршрутизацію, відслідковуючи реальні маршрути шляхом трасування та спостерігаючи за оновленнями анонсів маршрутів в BGP. В разі інциденту, постраждалі мережі приступають до пом'якшення наслідків події, наприклад, анонсуючи більш специфічні префікси своїх мереж або звертаючись до інших АС для фільтрації хибних анонсів.

Однак, через поєднання технологічних та практичних проблем розгортання, існуючі реактивні підходи значною мірою недостатні. Зокрема, у найсучасніших технологіях виникають такі основні проблеми:

- різноманітність типів атак на маршрутизацію, комбінування методів призводять до того, що *немає надійної методики виявлення* перехоплення маршруту;
- операторам необхідно *завчасно інформувати про легітимні зміни в своїй політиці маршрутизації* (утворення нових взаємодій між AS, анонсування нового префіксу тощо) аби такі зміни не вважатися підозрілими подіями для систем виявлення атак в умовної третьої сторони. З іншого боку, прийняття менш суворої політики щодо компенсації відсутності оновленої інформації та скорочення кількості false positives, несе небезпеку знехтування реальними подіями і невиявленням інцидентів (false negatives);
- кілька хвилин перенаправлення трафіку можуть спричинити великі фінансові втрати через недоступність послуги або порушення безпеки. А *швидкість реагування на інциденти є повільною* у будь-якому разі, бо за існуючою практикою є необхідність вручну перевірити попередження, що надходять від систем моніторингу та сторонніх сервісів. Тривалість широко відомих інцидентів становила від декількох годин до місяців [3].

В процесі оцінювання ризику особливі вимоги висуваються до якості інформації (максимально можливий рівень повноти, точності і відповідності на момент її отримання) та якості її

джерел [3,4]. Результат ідентифікації ризику повинен бути структурованим та охоплювати чотири елементи: джерела виникнення ризику, безпосередні події в результаті реалізації загроз, причини цих подій та очікувані наслідки. Побудова формальної моделі Інтернет-маршрутизації сприятиме рішенню задачі прогнозування наслідків подій.

2. Графовий підхід до моделювання мережі Інтернет

Телекомунікаційна мережа є невід'ємною частиною інформаційно-телекомунікаційних систем і характеризується різними за формою зв'язками, а також різними видами взаємодії. Часто математичною моделлю таких мереж може служити граф. Граф можна уявити як набір точок, званих вершинами чи вузлами, з'єднаних собою лініями, які називаються дугами чи зв'язками. Кожному зв'язку і вузлу графа може відповідати певна кількість параметрів, що характеризують природні обмеження. Наприклад, мережа може бути представлена в вигляді графа, в якому дуги відповідають каналам зв'язку, а вершини – вузлам комутації. Важливі параметри включені в модель у вигляді чисел або ваг, приписаних до дуг і вершин графа. Ці ваги можуть бути фіксованими і випадковими. Так, для мережі типова вершина, що представляє вузол комутації, може мати такі ваги: максимальну пропускну здатність, обсяг запам'ятовуючого пристрою, і т. і. Типова дуга – лінія зв'язку – може мати такі ваги: максимальну пропускну здатність, середню затримку передачі по каналу зв'язку, надійність каналу зв'язку і таке інше.

Доцільність побудови моделі у вигляді графа залежить від фізичної природи досліджуваної мережі. Найбільш очевидна доцільність користування моделлю у вигляді графа при вирішенні задач зв'язності. Так, в загальному випадку, нас може цікавити завдання доставки інформації з будь-якої точки в будь-яку. Це структурна задача, в якій необхідно встановити чи існує принаймні один шлях з будь-якої вершини в будь-яку іншу. Іншим важливим завданням є пошук найкоротшого шляху між двома, кількома, або всіма вершинами графа, а також пошук найкоротшого шляху з урахуванням різних обмежень. За допомогою графів можна також вирішувати завдання синтезу топології мережі, тобто побудови оптимальної системи. Одним з можливих критеріїв оптимальності можуть бути ступінь живучості або надійності телекомунікаційної системи. Оскільки телекомунікаційній мережі властиві пошкодження і відмови (наслідком чого є порушення зв'язку), можливим завданням може бути побудова системи, в якій наслідки порушень роботи є мінімальними при заданих умовах роботи.

Математичний апарат теорії графів, а пізніше – теорії складних мереж, у застосуванні до топології глобальної телекомунікаційної мережі Інтернет дозволяє аналізувати ступінь вузлів, розподіл ступеню, шліх між парою вузлів, середній шлях в мережі, показники кластерності, посередництва тощо. В даний час в багатьох дослідженнях граф використовують для побудови моделі мережі Інтернет на рівні автономних систем [5,6]. Мережа Інтернет в них представляється графом $G := (V, E)$, де V є множиною автономних систем (AS), а E - їхні зв'язки, утворені протоколом маршрутизації BGP-4. При цьому пропускна спроможність зв'язків між вузлами Інтернет не приймається о уваги і не впливає на вагу ребер, таким чином граф є незваженим. Було досліджено стосунки (node relationships) між автономними системами в Інтернеті. Було виділено декілька типів таких стосунків, зокрема «клієнт-провайдер», «партнер-партнер» та інші. У випадку взаємодії двох рівних за статусом операторів, вони анонсують один іншому власні префікси та префікси мереж своїх клієнтів. В разі взаємодії провайдера і клієнта, провайдер анонсує клієнтові всі наявні в нього префікси, а клієнт анонсує префікси власних мереж. Таким чином, зв'язки між автономними системами представляються двосторонніми, отже ребра є ненаправленими, а граф – неорієнтований.

За допомогою математичного апарату теорії графів було запропоновано метричну функцію відстані в Інтернеті для такого графа і доведено її відповідність аксіомам метрики [7].

3. Формальний опис об'єктів глобальної маршрутизації та відносин між ними

Отже, для виявлення так з перехоплення маршрутів, дослідження масштабів впливу на топологію, а також подальшої оцінки ризиків необхідно мати модель мережі Інтернет на рівні глобальної маршрутизації, тобто - з використанням BGP-зв'язків. Першою відмінністю від наведених в попередньому розділі підходів є те, що сам процес маршрутизації невід'ємно пов'язаний з вибором напрямку, отже при дослідженні втручання в маршрутизацію, коли результатом є несанкціонована зміна напрямку, ми не зможемо використовувати в якості моделі незважений граф. Для визначення необхідних якостей нової моделі пропонується формалізувати поняття маршрутизації.

Сформулюємо вихідні дані.

Існує адресний простір мережі Інтернет A – множина унікальних IP-адрес i, які згруповані в IP-префікси p (надалі - просто «префікси»):

$$A = \{a_1, a_2, a_3, ...a_{|A|} : a_i \neq a_j, \{i, j\} \leq |A|\};$$

$$a \in p \subset A.$$

Префікси, в свою чергу, групуються (інколи вживають термін «агрегуються») з більш специфічних в менш специфічні, як це визначено в [8] і наведено на рис. 1. Повний перелік префіксів наведено в документації по CIDR. З ілюстрації зрозуміло, що будь-яка IP-адреса входить до 32 префіксів, які «інкапсулюються» входять один в одного за допомогою зміни мережевої маски. При цьому весь адресний простір *A* може бути описаний одним префіксом з довжиною мережевої маски 0, або об'єднанням двох префіксів з довжиною маски 1, або чотирьох з довжиною маски 2 і так далі:

$$p_3 \subset p_2 \subset p_1 \subset p_0;\ |p_0| = 2|p_1| = 4|p_2| = 8|p_3|.$$

В загальному випадку можемо так виразити відношення між підмножинами IP-адрес, визначених префіксами префіксами, в множині всіх IP-адрес :

$$\begin{cases} p_i = 2^{j-i}\, p_j; \\ i \le j; \\ 0 \le \{i, j\} \le \log_2 |A| \end{cases} \qquad (1)$$

Нотація префіксу	Довжина netmask, біт	Кількість адрес в префіксі	Загальна кількість префіксів IPv4
x.x.x.x/32	32	1	4294967296
x.x.x.x/31	31	2	2147483648
x.x.x.x/30	30	4	1073741824
x.x.x.x/29	29	8	536870912
…..			
x.x.x.0/24	24	256	16777216
x.x.x.0/23	23	512	8388608
…..			
x.x.0.0/17	17	32768	131072
x.x.0.0/16	16	65536	65536
x.x.0.0/15	15	131072	32768
…..			
x.0.0.0/9	9	33554432	512
x.0.0.0/8	8	16777216	256
x.0.0.0/7	7	8388608	128
…..			
x.0.0.0/1	1	2147483648	2
0.0.0.0/0	0	4294967296	1

Рис.1. Агрегація IPv4-префіксів згідно RFC 4632.

Префікси анонсуються автономними системами (в іноземній літературі використовується термін «originating», отже в кожного префіксу є свій «origin» - автономна система, що анонсує його. Приклад взаємодії AS наведено на рис.2.

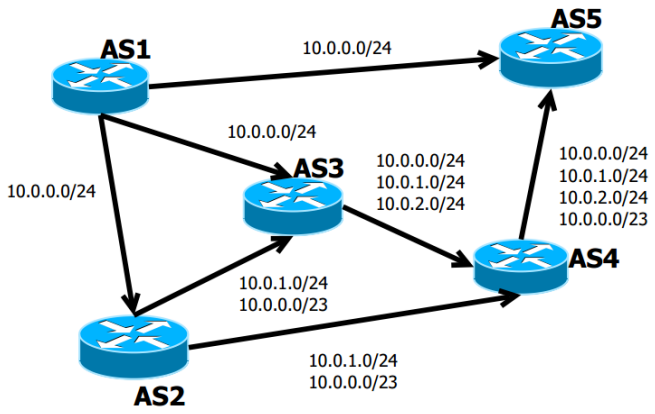


Рис.2. Приклад експорту анонсів IP-префіксів в процесі взаємодії автономних систем

В нормальному стані кожен префікс анонсує лише одна автономна система.

До кожного префіксу існує принаймні один маршрут

$$m(p) := (v_p, e_p), \qquad (2)$$

який складається з направлених ребер e_p, що відповідають анонсам перфіксу між вершинами v_p (автономними системами). До префіксу, який не анонсується, маршрутів нема, і він не може вважатись учасником Інтернету.

Сукупність маршрутів m_p до певного префікса *p* може бути представлена у формі зв'язного орієнтованого графа без кілець

$$M_p := (V_p, E_p), \qquad (3)$$

де V_p – зв'язки між автономними системами, по яких надходять анонси префікса *p*, а E_p – автономні системи, в яких присутні маршрути до префіксу *p*. Вершини графа по вхідних ребрах приймають анонси префіксу і по вихідних ретранслюють їх до інших вершин (чи не ретранслюють, в залежності від політики маршрутизації чи стану каналу). Одна з вершин має тільки віхідні ребра, це – origin. Інші вершини обов'язково мають вхідні ребра (бо приймають анонси) і можуть мати вихідні (якщо ретранслюють анонси).

Сума всіх графів $G_p := (V_p, E_p)$ - це сукупність всіх маршрутів до всіх префіксів. Вона утворює такий граф

$$G := (V, E) : V = \bigcup_p V_p; E = \bigcup_p E_p,$$

який і можна інтерпретувати як мережу Інтернет, представлену на рівні глобальної маршрутизації.

Поглибимось в процес маршрутизації. Якщо префікс p_j є підмножиною префіксу p_i, тобто $p_j \subset p_i$, це не означає, що в них обов'язково однаковий origin, в загальному префікси мають origin незалежно від вкладеності. Приклад: частина великого провайдерського префіксу може санкціоновано анонсуватись одним з клієнтів провайдера в напрямку інших AS. З іншого боку, якщо в певній автономній системі відсутні анонси до префіксу p, це не означає, що через неї цей префікс недосяжний: наявність в якомусь вузлі анонсу до p_i означає також і можливість маршрутизації до p_j (це одностороннє твердження):

$$p_j \subset p_i \Rightarrow m(p_j) \subset m(p_i). \qquad (4)$$

Кардинальним прикладом цього є кінцевий мережевий пристрій, підключений до мережі своїм єдиним мережевим інтерфейсом. Його таблиця маршрутизації може містити виключно маршрут до загального префіксу 0.0.0.0/0 (див. рис.1), який також зветься маршрутом за замовчанням (default route). Слід зауважити, що префікс 0.0.0.0/0 анонсується в BGP виключно для учасників, які використовують обладнання, що не здатне обробити «full view» – повну таблицю маршрутизації (зазвичай це кінцеві споживачі – невеликі мережі автономні системи яких які не є транзитерами).

4. Формальний опис процесу вибору маршруту

Задача знаходження оптимального маршруту є складним завданням. Для вирішення задачі має бути відомою топологія мережі, пропускні спроможності ліній зв'язку, середня довжина повідомлення. Це задача з класу цілочисельного нелінійного програмування, для яких доки не існує алгоритмів навіть з поліноміальною складністю. Відомі лише евристичні алгоритми, яке дозволяють отримувати лише приближене рішення завдання оптимізації. Тому в стеку TCP/IP прийнятий так званий однокроковий підхід до оптимізації маршруту просування пакета (next-hop routing) – кожний маршрутизатор та кінцевий вузол приймають участь виборі тільки одного кроку передачі пакета. Однокроковий підхід означає розподілене обчислення задачі вибору маршруту, і це – перевага, яка лежить в основі умовно безкінечної масштабованості мережі Інтернет.

Першим етапом обрання напрямку доставки є вибір префіксу. В таблиці маршрутизації має бути обраний з урахуванням (1) найбільш специфічний префікс:

$$p(a) = \{\min_j (p_j) \ : \ a \in p \subset A, \ 0 < j \le |A|\}. \qquad (5)$$

Як це пояснено в попередньому параграфі, вибір маршруту властивий не всім учасникам мережі. Суб'єктом вибору маршруту в глобальній маршрутизації є вузол графа, тобто автономна система. Передумовою для початку процесу вибору є наявність більше ніж одного маршруту до одного і того самого префікса *p*. Один з доступних маршрутів може бути визначений як шлях між двома вузлами графа (2), де початковим вузлом є автономна система, в якій приймається рішення, а кінцевим вузлом

– автономна система, яка є origin для префіксу *p*. Якщо не враховувати специфічні локальні атрибути маршруту та ті, що встановлюються адміністративно, єдине, що приймається до уваги, це топологія мережі на момент прийняття рішення. Загальним критерієм обрання шляху є його довжина (best path) – кількість транзитних вузлів між початковим і кінцевим вузлом. З урахуванням (2) та (3), шляхом до префіксу буде такий маршрут:

$$\pi_v(p) = \{\min_v (m_v(p)) : \pi \in M_p, v \in V_p\}, \qquad (6)$$

де *v* - вихідний вузол, в якому приймається рішення.

Отже, виділено та надано формальний опис двом складовим процесу маршрутизації: розрахунку префіксу (5) і вибору шляху (6).

5 Відносини між об'єктами глобальної маршрутизації і різні типи IP-адрес

На сьогоднішній день в Інтернеті використовується два типи IP-адрес, які відрізняються розрядністю. Традиційна IP-адреса складалась з 32 біт. Це наклало обмеження на кількість можливих адрес у 2^{32} = 4294967296. Зростання Інтернету, попри впровадження економного розподілу адресного простору, призвело до нестачі адрес. Сучасна IP-адреса, яка має умовну назву «IPv6-адреса» (на відміну від традиційної, яку стали називати «IPv4-адреса») має довжину 128 біт. Ця різниця впливає на функціонування канального та мережевого рівнів (згідно моделі OSI). Однак принципи CIDR та маршрутизації в цілому не зазнали змін при впровадженні IPv6. З точки зору CIDR, відмінність від IPv4 полягає в наступному:

- максимальна довжина мережевої маски складає 128 біт замість 32;
- кожна адреса входить в 128 IPv6-префіксів, включених один в одного.

Протокол BGP-4 для IPv4 та IPv6 не відрізняється – BGP-спікери використовують одні й ті самі протокольні повідомлення, атрибути шляху, критерії вибору маршруту, як для адресного простору IPv4, так і для адресного простору IPv6. Тому опис моделі глобальної маршрутизації, запропонований в попередньому розділі, зокрема вирази (1), (4) – (6), є незмінним незалежно від типу IP-префіксів.

6 Висновок

Важливим кроком на шляху до оцінки ризику, що спричинений атаками на глобальні маршрутизацію, є прогнозування наслідків атаки, а саме - оцінка масштабу атаки (шляхи розповсюдження, зона впливу, кількість «пошкоджених» маршрутів). Запропоновано формальні описи об'єктів глобальної маршрутизації - автономних систем, IP-префіксів, BGP-анонсів, та відносин між ними. Виділено та надано формальний опис двом складовим процесу маршрутизації: розрахунку IP-префіксу і вибору шляху. Ці етапи є важливим поступом до формулювання задачі поводження з ризиками для певного вузла від перехоплення маршрутів як задачі пошуку для нього найбільш ефективної топології зв'язків.

References

1. G. Huston, G. Michaelson et al.: Resource Public Key Infrastructure (RPKI) Validation Reconsidered. https://tools.ietf.org/html/rfc8360. Дата доступу: 29 жовтня, 2019 р.
2. Зубок, В.: Метричний підхід до оцінки ризику атак на глобальну маршрутизацію : Информационные технологии и безопасность. Мат. XVIII Междунар. науч.-практ. конф. ИТБ-2018. – К.:ООО "Инжиниринг", 2018. - с.43-47.
3. Зубок В., Мохор В.: Дослідження зв'язку між топологією та ризиком внаслідок кібератак на глобальну маршрутизацію. : Моделювання та інформаційні технології, 2018, Вип. 85, с. 23—26.
4. Risk Management – Vocabulary (ISO Guide 73:2009, IDT) : ДСТУ ISO Guide 73:2013. – [Чинний від 2014–07–01] . – Київ : Мінекономрозвитку України, 2014. – 13 с. - (Національні стандарти України).
5. IPv4 and IPv6 AS Core: Visualizing IPv4 and IPv6 Internet Topology at a Macroscopic Scale : http://www.caida.org/research/ topology/as_core_network/ . Дата доступу: 20 жовтня, 2019 р.
6. Pavlos Sermpezis, Vasileios Kotronis et al.: ARTEMIS: Neutralizing BGP Hijacking within a Minute : https://arxiv.org/abs/1801.01085 . Дата доступу: 11 червня, 2019 р.
7. Мохор, В. та Зубок, В.: Формування міжвузлових зв'язків в Інтернет з використанням методів теорії складних мереж. К.:«Прометей», 2017. – 175с.
8. Fuller, V., Li, T.: Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan : https://tools.ietf.org/html/rfc4632 . Дата доступу: 01 жовтня, 2019 р.

ЗАСТОСУВАННЯ ГІПЕРКОМПЛЕКСНИХ ЧИСЛОВИХ СИСТЕМ ДЛЯ ОПИСУ СКЛАДНИХ МЕРЕЖ

Д.В.Ланде[1,2], Ю.Є.Боярінова[2,1], Я.О.Каліновський[1], Т.В.Синькова[1]

[1]Інститут проблем реєстрації інформації НАН України
вул. Шпака, 2, 03113 Київ, Україна
[2] Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
пр. Перемоги, 37, 03056 Київ, Україна

Анотація. Запропоновано для опису складних мереж і різноманітних впливів в цих мережах використовувати гіперкомплексні числові системи. Для спрощення обчислень запропоновано використання ізоморфних гіперкомплексних числових систем, таблиці Келі яких мають діагональний вигляд.

Ключові слова: складні системи, складні мережі, гіперкомплексна числова система, ізоморфізм гіперкомплексних систем, модель впливу

Вступ

В останні роки отримав розвиток напрям дослідження мереж в яких зв'язки у змістовному плані відповідають взаємним впливам вузлів. Дослідженню таких складних мереж (Complex Networks) присвячено багато праць [1-3], але найчастіше вони відносяться до розповсюдження одного виду активності (впливу). В той же час вже з'являються дослідження, що пропонують розглядати декілька характеристик впливу [4-6]. Зокрема, розробляються та досліджуються різноманітні математичні моделі: моделі з порогами, моделі незалежних каскадів, моделі розповсюдження епідемій, моделі марковських процесів та ін.[7-11]

У цій роботі пропонується застосовувати гіперкомплексні числові системи, які є математичним апаратом, що дозволяє моделювати деякі мережеві задачі та вирішувати їх на новому рівні [12].

Постановка задачі

Метою даної роботи є моделювання розповсюдження декількох характеристик впливів на вузли складної мережі із застосуванням апарату гіперкомплексних числових систем (ГЧС).

Терміни

У роботі буде використовуватися така термінологія:
Вузол – елемент мережі, вершина.
Ребро – зв'язок між вузлами, що відповідає впливу одного вузла на інший.
«Фішка» – властивість, яка може впливати на стан вузла.

Мережева модель розповсюдження декількох видів активності

У цій роботі розглядаються мережі, де ребра відповідають взаємним впливам вузлів. Зазвичай розглядаються динамічні мережеві моделі, засновані на розповсюдженні одного виду активності. Запропонована модель, навпаки, базується на мережах, в яких передбачено багато видів активності. Розглядається складна структура вузлів, що зв'язані між собою якимось властивостями. В рамках запропонованої моделі, яка є розширенням моделі, описаної в [4], кожен вузол може віддавати або приймати фішку (змінювати свої властивості). В залежності від стану, вузол по різному реагує на отримання або віддачу різного типу фішки.

В рамках цієї моделі мережа задається графом $G=(V,E)$, де $E\subseteq V\times V$ – множина зважених ребер, крім того задана множина фішок $\{F=1,2,...,n\}$ різного кольору. Вузли мережі з різними наборами представлені на Рис.1

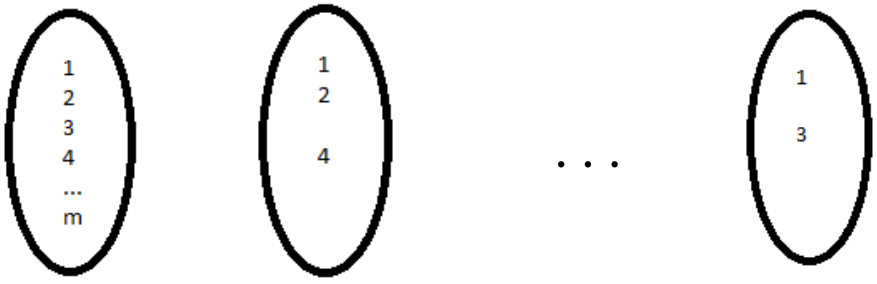


Рисунок 1 – Приклад вузлів з різними наборами фішок

Вузол на кожному такті часу t може мати фішки любого типу або залишатись вільним. Якщо вузол немає визначеної властивості, то відповідної фішки він прийняти не може.

На рис. 2 показано відсутність властивості.

Рисунок 2 – Фішки в вершинах: 1,3 заповнені властивості, 5 – порожня властивість, 2,4 – відсутні властивості

Кількість фішок типу c_i у вузлі v_i відповідно $q_{il}(t)$; стан вузла – вектор $Q_i(t)$ довжини m. Стан мережі – $Q(t)_{nxm}$..

Вузли обмінюються фішками в дискретні моменти часу t по ребрам, що мають різну пропускну здатність. Пропускна здатність – цілі числа. Кожне ребро має m пропускних здатностей. Якщо ребро не може проводити фішки визначеного типу, тоді пропускна здатність дорівнює 0. Приклад мережі представлений на рис.3

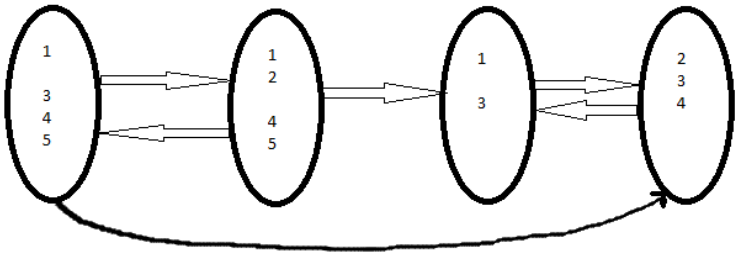


Рисунок 3 – Частина мережі з різними наборами властивостей

Модель є відкритою, тобто фішки в ній з‘являються і зникають. Деякі вузли здатні генерувати фішки визначеної властивості. На кожному такті виникнення фішек типу k відбувається стохастично з деякою ймовірністю p_k.

Ймовірності можуть залежати не тільки від типу фішек, а також від того, в якому вузлі знаходиться та чи інша властивість. Крім того, p_k можуть бути функціями від стану мережі. В мережі фішки можуть не тільки виникати, а також зникати.

Вузли обмінюються фішками в дискретні моменти часу по ребрам (що в запропонованій моделі відповідають таблиці множення відповідної ГЧС).

Якщо кількість фішек певного типу у вузлі v_i в момент часу t не перебільшує визначеної кількості (тобто коефіцієнту при базисному елементі e_i), тоді на даному етапі вузол не активний. Якщо у вузлі на такті t кількість фішек якогось визначеного типу перебільшує визначену кількість – вузол активується.

Нехай $c_{jk}(t)$ – кількість ресурсу (фішек) типу k у вузлі v_i в кінці такту t;

$c_{jk}^{in}(t)$ – кількість фішек типу k, що прийшли до вузла v_i в кінці такту t, тоді

$$c_{jk}(t) = \mu c_{jk}(t-1) + c_{jk}^{in}(t),$$

де $0 < \mu < 1$ – коефіцієнт дисконтування.

Залежність стану властивості k вузла v_i на такті $t+n$ описується формулою:

$$c_{jk}(t+n) = \mu^n c_{jk}(t) + ... + \mu c_{jk}(t+n-1) + c_{jk}^{in}(t).$$

Змінюючи значення μ можна отримати різного вигляду системи з різного роду активністю.

Модель соціальної мережі з агентами впливу

Розглянемо соціальну мережу з множиною агентів $V = \{v_1, v_2, ..., v_n\}$, що задана орієнтованим графом $G = (V, E)$, де $E \subseteq V \times V$ – множина зважених ребер. Агенти в мережі впливають один на другий – наявність ребра e_{ij} від вузла v_i до вузла v_j відповідає впливу *i*-го агенту на *j*-го агента, а його вага $r_{ij} \in N$ – це ступінь впливу. На кожному кроці дискретного часу t вузол обирає одне з трьох дій: два виду активності або бездіяльність. В мережі передаються фішки двох типів – типи 1 та 2 (типи антагоністичної активності, наприклад, створювати або руйнувати тощо). Пропускна здатність по обом видам активності може бути однаковою.

Агенти в мережі, що розглядається, є вузлами, що мають дві властивості – по одній з кожного виду активності. Кількісну характеристику властивості у вузлі v_i позначимо, як d_{j1}, d_{j2}. Крім того в кожному вузлі є два параметра, що показують відношення агенту до кожного з видів активності $s_{jk} \in \{-1, +1\}$, $k = 1, 2$. Позитивне відношення – це +1, а негативне –1. Відношення вузлів двох видів активності можуть бути скомбіновані по різному (табл.1)

Таблиця 1 – Комбінація значень параметрів s_{j1} та s_{j2} для окремої вершини

	s_{j1}	s_{j2}
1	+1	–1
2	–1	+1
3	+1	+1

У випадках 1 та 2 параметри s_{j1} та s_{j2} мають різні знаки, тобто агенти мають різне відношення до якогось виду активності, при цьому один оцінюється як позитивний, а інший, як негативний. Звичайно, кількості властивостей можуть відрізнятись. Кількісне значення властивості відповідає за поріг активації вузла.

Випадок 3 описує ситуацію, коли агент не має чіткої позиції к типам активності. Але він може стати активним, якщо перебільшений поріг активації. При цьому він активується по тому типу, для якого буде наповнена властивість. Такий агент піддається впливу.

Випадок $s_{j1} = s_{j2} = -1$ не розглядається, так як такий агент не буде діяти ні по якому типу.

Кількісні значення d_{j1} та d_{j2} лінійно залежать від вузла v_j. В порогових моделях розповсюдження активності вузол активується, якщо частина активних сусідів перебільшує деякий поріг. В якості порога може бути кількісне значення властивості:

$$d_{jk} = \alpha_{jk} r_j^{in}, \quad k = 1, 2,$$

де r_j^{in} – вплив на вузол v_j. Кількісне значення властивості залежить (з коефіцієнтом α_{jk}) від зваженої кількості вузлів, що впливають на даний, причому вагою служить сила впливу кожного вузла.

Таким чином, кожний вузол визначається параметрами: $\alpha_{j1}, \alpha_{j2}, s_{j1}, s_{j2}$. Кількісне значення властивості введені в моделі для того, щоб параметри α_{jk} задавали порогові значення активності сусідів з врахуванням їх попередньої активності.

Приклади функціонування мережі[4]

І. Розглянемо функціонування мережі, яка представлена графом (Рис. 4) в мережі присутні агент – революціонер (1), агент – конформіст (2) та обережний агент (3). Усі пропускні здатності дорівнюють 1.

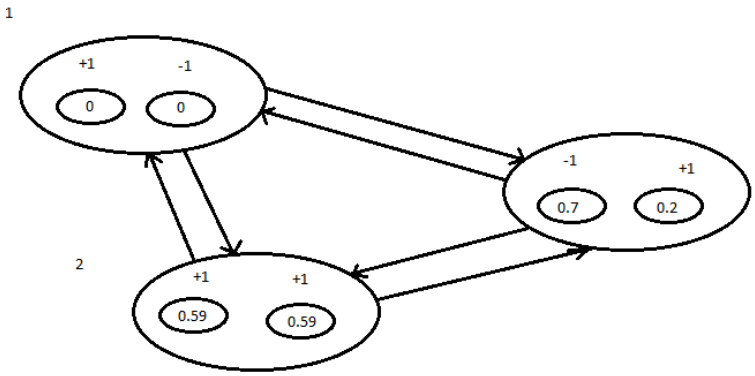


Рисунок 4 – Мережа з трьома агентами

Розглянемо динаміку мережі при $\mu = 0.1$. Типи активності позначимо 1 та 2. Тоді основні характеристики вузлів представлені в табл.2

Таблиця 2 – параметри вершин з трьома агентами

Агенти	r_j^{in}	α_{j1}	α_{j2}	d_{j1}	d_{j2}
v_1	2	0	0	0	0
v_2	2	0.59	0.59	1.18	1.18
v_3	2	0.7	0.2	1.4	0.4

Нехай півтакти позначаються $t.1$ та $t.2$

t = 1.1. Вузол v_1 випускає дві фішки у вузол v_2 та v_3.

t = 1.2. Кількість фішек

$c_{11}(1) = 0$ $\quad$ $c_{12}(1) = 0$

$c_{21}(1) = 1$ $\quad$ $c_{22}(1) = 0$

$c_{31}(1) = 1$ $\quad$ $c_{32}(1) = 0$

Порогові значення вузлів v_2 та v_3 не перевищені.

t = 2.1. Вузли v_2 та v_3 мовчать. Вузол v_1 випускає дві фішки у вузли v_1

t = 2.2. Відповідно:

$c_{11}(2) = 0$ $\quad$ $c_{12}(2) = 0$

$c_{21}(2) = 0.1 * 1 = 1.1$ $\quad$ $c_{22}(2) = 0$

$c_{31}(2) = 0.1 * 1 = 1.1$ $\quad$ $c_{32}(2) = 0$

Порогові значення вузлів v_2 та v_3 не перевищені. На такті t = 3 вони знов мовчать. Можна побачити, що $c_{21}(n) = c_{31}(n) = 1.1 \ldots 1$. Цього значення не вистачає, щоб досягнути кількісного значення в обох вузлах. Вони мовчать завжди. Таким чином вузол v_1 не може активізувати сусідні вузли.

II. Розглянемо динаміку мережі при $\mu = 0.2$.

t = 1.1: Вузол v_1 випускає дві фішки у вузли v_2 та v_3

t = 1.2: Кількість фішек становить:

$c_{11}(1) = 0$ $\quad$ $c_{12}(1) = 0$

$c_{21}(1) = 1$ $\quad$ $c_{22}(1) = 0$

$c_{31}(1) = 1$ $\quad$ $c_{32}(1) = 0$

Порогові значення вузлів v_2 та v_3 не перевищені.

t = 2.1: Вузли v_2 та v_3 мовчать. Вузол v_1 випускає дві фішки у вузол v_2 та v_3

t = 2.2: Кількість фішек:

$c_{11}(2) = 0$ $c_{12}(2) = 0$
$c_{21}(2) = 0.2 * 1 = 1.2$ $c_{22}(2) = 0$
$c_{31}(2) = 0.2 * 1 = 1.2$ $c_{32}(2) = 0$

Ресурс у вузлі v_2 перебільшив порогове значення, а у вузлі v_3 – ні.

t = 3.1: Вузли v_1 та v_2 випускають фішки типу 1. Вузол v_3 мовчить

t = 3.2: Кількість фішек:

$c_{11}(3) = 1$ $c_{12}(3) = 0$
$c_{21}(3) = 1$ $c_{22}(3) = 0$
$c_{31}(3) = 0.2 * 1.2 + 2 = 2.24$ $c_{32}(3) = 0$

t = 4.1: Вузол v_2 мовчить, тому що випустив фішки, він втратив ресурс, а його треба поповнити.

Вузол v_3 випускає фішки типу 2. Вузол v_1 випускає фішки типу 1.

t = 4.2:

$c_{11}(4) = 1$ $c_{12}(4) = 1$
$c_{21}(4) = 0.2 * 1 + 1 = 1.2$ $c_{22}(4) = 1$
$c_{31}(4) = 1$ $c_{32}(4) = 0$

t = 5.1: Вузли v_1 та v_2 випускають фішки типу 1. Вузол v_3 мовчить

t = 5.2:

$c_{11}(5) = 1$ $c_{12}(5) = 0$
$c_{21}(5) = 1$ $c_{22}(5) = 1$
$c_{31}(5) = 0.2 * 1 + 2 = 2.2$ $c_{32}(5) = 0$

t = 6.1 Вузол v_1 випускає фішки типу 1. Вузол v_3 випускає фішки типу 2. Вузол v_2 мовчить.

$c_{11}(6) = 0$ $c_{12}(6) = 1$
$c_{21}(6) = 1.2$ $c_{22}(6) = 1.2$
$c_{31}(6) = 1$ $c_{32}(6) = 0$

t = 7 …3 цього кроку вузол v_2 вносить невизначеність в подальшу динаміку мережі, так як він має однакову кількість ресурсу в двох вузлах і може стати активною по кожному з типів.

Модифікована мережа

Пропонується модель, в якій вершини обмінюються фішками в дискретні моменти часу по ребрам (що відповідають таблиці множення відповідної ГЧС).

Припустимо, що у кожної людини є декілька характеристик (властивостей), що впливають на її стан: робота, сім'я, спілкування, суспільство та ін. Ці відносини можна записати у вигляді таблиці Келі (таблиці множення гіперкомплексної числової системи). Зокрема, у загальному випадку добуток двох базисних елементів ГЧС відносно операції множення має вигляд числа з цієї ж ГЧС (властивість замкнутості ГЧС):

$$e_i e_j = \gamma_{ij}^1 e_1 + \ldots + \gamma_{ij}^n e_n \ . \qquad (1)$$

де e_i – базисні елементи ГЧС; коефіцієнти γ_{ij}^k – структурні константи, є дійсними числами ($\gamma_{ij}^k \in R$);

При такому представленні для повного завдання ГЧС необхідно задати n^3 структурних констант.

При цьому одиничний елемент ГЧС може входити в базис, або ні. Гіперкомплексне число має вигляд:

$$A = a_1 e_1 + a_2 e_2 + \ldots + a_n e_n , \qquad (2)$$

де e_i – базисні елементи ГЧС, a_i – коефіцієнти гіперкомплексного числа *A*.

Відповідно до (1) та (2), при обчисленні добутку двох гіперкомплексних чисел треба виконати n^3 дійсних множень. Для зменшення їх кількості можна використати перехід до ізоморфних ГЧС.

Ізоморфізм гіперкомплексних числових систем

Гіперкомплексні числові системи Γ_1 і Γ_2 називаються ізоморфними, якщо існує таке взаємно однозначне відображення L векторного простору Γ_1 на Γ_2, що виконуються наступні властивості:

1. $L(a+b)=L(a)+L(b)$. (3)
2. $L(a\cdot b)=L(a)\cdot L(b)$. (4)

де $a,b\in\Gamma_1,\ \ L(a),\ \ L(b)\in\Gamma_2$

З (3-4) випливає, що Γ_1 і Γ_2 – лінійні простори з базисами $e^{(1)}=\{e_1^1,e_2^1,...,e_n^1\}$ і $e^{(2)}=\{e_1^2,e_2^2,...,e_n^2\}$ відповідно, а між ними можна встановити взаємно однозначне лінійне перетворення з дійсною матрицею A:

$$\begin{aligned} e_1^1&=\alpha_{11}e_1^1+\alpha_{12}e_2^1+...+\alpha_{1n}e_n^1\\ e_2^1&=\alpha_{21}e_1^1+\alpha_{22}e_2^1+...+\alpha_{2n}e_n^1\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ e_n^1&=\alpha_{n1}e_1^1+\alpha_{n2}e_2^1+...+\alpha_{nn}e_n^1 \end{aligned}\ . \qquad (5)$$

При цьому детермінант матриці A відрізняється від нуля:

$$\|A\|\neq 0\,, \qquad (6)$$

Тоді повинно існувати і зворотне перетворення.

Як випливає з теорії лінійних просторів, для кожної пари лінійно незалежних базисів можна знайти взаємно однозначне лінійне перетворення (5), що переводить один базис в інший і навпаки. Але виконання одночасно вимог (3) та (4)не завжди можливо.

Якщо

й $a=\sum_{i=1}^{n}a_ie_i^1$, $b=\sum_{i=1}^{n}b_ie_i^1$ - гіперкомплексні числа, то їх добуток:

$$ab=\sum_{i=1}^{n}\sum_{i=1}^{n}\sum_{k=1}^{n}a_ib_j\gamma_{ij}^k e_k\,,$$

а лінійне перетворення L:

$$L(ab)=L(\sum_{i=1}^{n}\sum_{i=1}^{n}\sum_{k=1}^{n}a_ib_j\gamma_{ij}^k e_k^1)=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\gamma_{ij}^k a_ib_jL(e_k^1)\,.$$

Але, відповідно до (5):

$$(e_k^1)=\sum_{s=1}^{n}\alpha_{ks}e_s^1\,.$$

Тому:

$$L\,(ab)=\sum_{i=1}^{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{s=1}^{n}\alpha_{ks}\gamma_{ij}^k a_ib_je_s^1\,. \qquad (7)$$

З іншого боку:

$$L(a)\cdot L(b)=\sum_{i=1}^{n}a_iL(e_i^1)\sum_{j=1}^{n}b_jL(e_j^1)=\sum_{i=1}^{n}\sum_{k=1}^{n}\alpha_{ik}a_ie_k^1\sum_{j=1}^{n}\sum_{s=1}^{n}\alpha_{js}b_je_s^1=$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{r=1}^{n}\alpha_{ik}\alpha_{js}\gamma_{ks}^r a_ib_je_r^1\,. \qquad (8)$$

Прирівнюючи в правих частинах виразів (7) і (8) вирази при однакових $a_ib_je_r^1$, отримаємо n^3 нелінійних рівнянь від n^2 невідомих α_{ij}:

$$\sum_{k=1}^{n}\sum_{s=1}^{n}\alpha_{ks}\gamma_{ij}^k=\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{r=1}^{n}\alpha_{ik}\alpha_{js}\gamma_{ks}^r;\ \ i,j,k\in\overline{1,n}\,. \qquad (9)$$

З (9) випливає, що ця система перевизначена. Вона завжди має тривіальне рішення:

$$\alpha_{ij} = 0; \quad i, j = 1...n .$$

Але наявність нетривіальних дійсних рішень з виконанням умови (6) не гарантується. Тому, якщо є хоч одне нетривіальне рішення з виконанням умови (6), тоді ці дві гіперкомплексні числові системи Γ_1 і Γ_2 є ізоморфними. Якщо таких рішень немає, то вони не ізоморфні.

Ізоморфний перехід зберігає нульовий і одиничний елементи системи, тобто нульовий і одиничний елементи системи Γ_1 перетворюються відповідно в нульовий і одиничний елементи системи Γ_2.

Існує декілька способів побудови ізоморфних ГЧС. Один з таких способів – узагальнення процедури подвоєння Грасмана-Кліфорда.

За виглядом таблиць з множення двох деяких ГЧС дуже важко встановити, ізоморфні вони або ні. Регулярним методом встановлення ізоморфізму є розв'язок системи рівнянь ізоморфізму (9) яка є системою з n^3 нелінійних квадратичних рівнянь з n^2 змінними (n – розмір ГЧС), тобто система перевизначена. Якщо для $n < 4$ час вирішення системи рівнянь невеликий, то для $n \geq 4$ він стає неприйнятним навіть для проведення наукових досліджень.

Іншим способом генерації ізоморфних ГЧС є ізоморфне перетворення базису вихідної ГЧС. Цей метод простий в реалізації, але, як правило, отримують неканонічні ГЧС, та складним є вибір оператору ізоморфізму, який привів би до необхідної за структурою ГЧС. Однак цей метод також знаходить застосування.

Як показали дослідження, найбільш підходящий метод генерації ізоморфних ГЧС заснований на процедурі подвоєння Грасмана-Кліфорда методом добутку розмірності.

Процедура подвоєння Грасмана-Кліфорда фактично має на увазі те, що компоненти гіперкомплексного числа для даної ГЧС вже не є дійсними числами, а числами, що належать ГЧС вимірності 2. У загальному випадку, якщо розмір системи не тільки дорівнює 2, вимірність отриманої ГЧС буде вже не подвоюватися по відношенню до вихідної, а множитись на розмірність тієї ГЧС, яка була вихідною. Такі процедури мають назву процедур множення розмірності.

Будемо позначати результат застосування процедури множення розмірності системи $\Gamma_1(e,n)$ системою $\Gamma_2(f,m)$ так: $Д\left(\Gamma_1(e,n),\Gamma_2(f,m)\right)$.

При використанні процедури множення розмірності базис отриманих ГЧС складається з парних добутків базисних елементів вихідної ГЧС. Якщо вихідні ГЧС були комутативні, то і базисні елементи результуючої ГЧС будуть також комутативні. Слід відмітити, що незалежно від порядку розташування ГЧС в операторі множення розмірності, отримані базиси будуть однакові. Таким чином, будуть ідентичними і таблиці множення ГЧС, отриманих при перестановці їх в операторі множення розмірності.

Тому для комутативних ГЧС можна сформулювати наступні принципи множення розмірності:

1. При використанні процедури множення розмірності операнди-ГЧС можна комутувати;
2. Множення розмірності однієї і той же ГЧС ізоморфними ГЧС приводять до ізоморфних ГЧС;
3. При множенні розмірності пари ізоморфних ГЧС іншою парою ізоморфних ГЧС отримується пара ізоморфних ГЧС та існує не вироджене лінійне перетворення базисів отриманої пари ГЧС, яке є добутком лінійних перетворень першої пари ГЧС.

Якщо до однієї і тієї же ГЧС застосовуються процедури множення розмірності різними ГЧС, але розмірності яких однакові,то в результаті отримуємо різні ГЧС однакової розмірності, що, в загальному випадку, неізоморфні між собою. Якщо для множення розмірності застосовуються ізоморфні ГЧС, то і результати після застосування процедури множення розмірності також будуть ізоморфні між собою.

Використовуючи ці властивості, можна побудувати цілій ряд пар ізоморфних ГЧС достатньо великих розмірностей, а також оператори їх ізоморфізму.

В залежності від обраної ГЧС (її вимірності), можна оцінювати декілька відносин впливу одного агенту на іншого (рис.5).

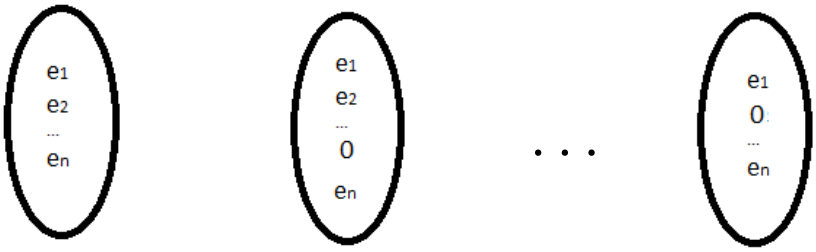


Рисунок 5. Приклади вигляду вершин з різними наборами «фішек» з елементами базисів ГЧС

Розглянемо відносини між 4 особами. Ці відносини можна записати у вигляді таблиці Келі (таблиці множення гіперкомплексної числової системи). Ця таблиця має вигляд:

	e_1	e_2	e_3	e_4
e_1	e_1	e_2	e_3	e_4
e_2	e_2	e_1	e_4	e_3
e_3	e_3	e_4	e_1	e_2
e_4	e_4	e_3	e_2	e_1

За допомогою програмного комплексу гіперкомплексних обчислень [8] можна побудувати ізоморфну систему, яка буде мати вигляд

$W_1^{(2)}$	f_1	f_2	f_3	f_4
f_1	f_1	0	0	0
f_2	0	f_2	0	0
f_3	0	0	f_3	0
f_4	0	0	0	f_4

При цьому будуть отримані такі формули для оператора ізоморфного переходу:

$$\begin{aligned} e_1 &= f_1 + f_2 + f_3 + f_4 \qquad & f_1 &= (e_1 + e_2 + e_3 + e_4)/4 \\ e_2 &= f_1 - f_2 + f_3 - f_4 , \qquad & f_2 &= (e_1 - e_2 + e_3 - e_4)/4 \\ e_3 &= f_1 + f_2 - f_3 - f_4 \qquad & f_3 &= (e_1 + e_2 - e_3 - e_4)/4 \\ e_4 &= f_1 - f_2 - f_3 + f_4 \qquad & f_4 &= (e_1 - e_2 - e_3 + e_4)/4 \end{aligned}$$

Перехід до такого типу системи надає можливість враховувати декілька характеристик впливу одночасно, але таблиця Келі має діагональний вигляд, що спрощує обчислення.

Розглянемо роботу такої системи, яка складається з чотирьох вузлів та взаємовпливів один на інший. При цьому коефіцієнт ($\overline{\mu} = 0.4(e_1 + e_2 + e_3 + e_4)$)

t = 1.1: Нехай вузол v_1 впливає на вузли v_2 і v_3 (за властивостями h_1 та h_3) та на вузол v_4 (за властивістю h_4)

t = 1.2:

$$\begin{aligned} s21 &= s20 + F2(t1) = 1e_1 + 1e_2 + 0e_3 + 1e_4 + 1e_1 + 1e_3 = 2e_1 + 1e_2 + 1e_3 + 1e_4 \\ s31 &= s30 + F3(t1) = 0e_1 + 0e_2 + 1e_3 + 1e_4 + 1e_1 + 1e_3 = 1e_1 + 0e_2 + 2e_3 + 1e_4 \\ s41 &= s40 + F4(t1) = 1e_1 + 1e_2 + 0e_3 + 0e_4 + 1e4 = 1e_1 + 1e_2 + 0e_3 + 1e_4 \end{aligned}$$

t = 2.1: Вузли v_2 і v_3 не відповідають. Вузол v_4 впливає на v_1, v_2 та v_3 (за властивістю h_4)

$$\begin{aligned} s12 &= s11 + F1(t2) = 1e_1 + 1e_2 + 1e_3 + 0e_4 + 1e4 = 1e_1 + 1e_2 + 1e_3 + 1e_4 \\ s22 &= s21 + F2(t2) = 2e_1 + 1e_2 + 2e_3 + 1e_4 + 1e_4 = 2e_1 + 1e_2 + 2e_3 + 2e_4 \\ s32 &= s31 + F3(t2) = 1e_1 + 0e_2 + 2e_3 + 1e_4 + 1e_4 = 1e_1 + 0e_2 + 2e_3 + 2e_4 \end{aligned}$$

t = 3.1: Вузол v_4 впливає на v_2 та v_3 (за властивістю h_4 та h_1)

$$s23 = s22 + F2(t3)\mu = 2e_1 + 1e_2 + 2e_3 + 2e_4 + (1e_1 + 1e_4)\mu = 2.4e_1 + 1e_2 + 2e_3 + 2.4e_4$$

$$s33 = s32 + F3(t3)\mu = 1e_1 + 0e_2 + 2e_3 + 2e_4 + (1e_1 + 1e_4)\mu = 2e_1 + 0e_2 + 2e_3 + 2.4e_4$$

Якщо припустити, що значення коефіцієнтів не може бути більше за значення *H* або норма не повинна перевищувати значення *K*.

Тоді починається зворотній зв'язок і вузол починає впливати на інші вузли, тобто зменшується значення коефіцієнтів до порогового значення.

Наприклад, якщо *H*=2, тоді після останнього кроку вузол v_2 буде вже впливати на інші вузли за властивістями h_2 та h_4, а вузол v_3 буде вже впливати на інші вузли за властивістю h_4.

Висновки

В роботі розглянуто модель впливу декілька характеристик із використанням гіперкомплексних числових систем. Модель, зокрема, може представляти собою взаємодію декількох предметних областей зі складними учасниками та декількома видами активності, якою вони можуть впливати один на іншого з різною ймовірністю та пороговим значенням.

При $\mu = 0$ отримаємо класичну порогову модель розповсюдження активності, при малих значеннях μ стан на минулих тактах дає невеликий вплив на теперішній стан, при $\mu = 1$ отримаємо лінійне накопичення ресурсу.

Таким чином, нами для опису складних мереж і різноманітних впливів в цих мережах запропоновано використовувати гіперкомплексні числові системи. Для спрощення обчислень також запропоновано використання ізоморфних гіперкомплексних числових систем, таблиці Келі яких мають діагональний вигляд.

Така модель може бути застосована в соціальних мережах.

Література

1. BlanchardPh. Random Walks and Diffusions on Graphs and Databases: An Introduction (Springer Series in Syn-ergetics) / Ph. Blanchard, D. Volchenkov. - Berlin–Heidelberg: Springer-Verlag, 2011.
2. Lovasz L. Mixing of Random Walks and Other Diffusions on a Graph / L. Lovasz, P. Winkler // Surveys in Com-binatorics, 1995 (ed. P. Rowlinson), London Math. Soc. Lecture Notes Series 218. - Cambridge Univ. Press. - P. 119–154.
3. Кузнецов О.П. Двусторонние ресурсные сети – новая потоковая модель / О.П. Кузнецов, Л.Ю. Жилякова // Доклады Академии Наук. 2010. Том 433. №5. – С. 609-612.
4. Жилякова, Л.Ю. Сетевая модель распространения нескольких видов активности в среде сложных агентов и ее приложения / Л.Ю. Жилякова // Онтология проектирования, 2015.Т.5, №3(17) - С. 278-295
5. Ланде Д.В. Аналізінформаційнихпотоків у глобальнихкомп.ютерних мережах // Вісник НАН України, 2017. – N 3. – С. 46-54.
6. Ланде Д.В.Обмеження доступу до Інтернету у світі: мережева модель/Д.В., Ланде Ю.О.Ліненко // Інформаційне право: сучаснівиклики і напрмирозвитку: Матеріалипершоїнауково-практичної конференції. 18 жовтня 2018 р., м. Київ. / Упоряд. : В.М. Фурашев, С.Ю. Петряєв. - Київ:НаціональнийтехнічнийуніверситетУкраїни .Київський політехнічний інститут імені Ігоря Сікорського. - 2018. - С. 50-54.
7. Kemple D. maximizing the Spread of Influence through a Social Network/ D. Kemple , J/Kleiberg? E Tardos// Proc. Of 9-th ACM SIGKDD Int.Conf. on Knowledge Discovery and Data Mining, 2003 –P. 137-146
8. Granovetter M. Threshold Models of Collective Behavior/ M.Granovetter // American journal of Sociology, 1978, Vol.83, No.6 – P.1420-1443
9. Watts D.J. A Simple model of global cascade on random networks/ D.J.Watts // Proc.Nat.Acad.Sci.USA,2002, 99(9) – P.5766-5771
10. Goldenberg J. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth/ J. Goldenberg , B.Libai, E.Muller//Marketing Letters, 2001, No.2. – P.11-34
11. Губанов Д.А. Социальные сети: модели информационного влияния, управления и противоборства/Д.А.Губанов, Д.А.Новиков, А.Г. Чхартишвили// М.: Физматлит, 2010 – 228с.
12. Я.А. Калиновский, Ю.Е. Бояринова, А.С. Сукало. Гиперкомплексные числовые системы четвертой размерност - К:ИПРИ НАН Украины, 2017. – 125с.
13. Синьков М.В. Конечномерные гиперкомплексные числовые системы. Основы теории. Применения. / М.В. Синьков, Ю.Е. Бояринова, Я.А. Калиновский. – К.: Инфодрук, 2010. – 388 с.

АДАПТАЦІЯ ХМАРНИХ ОБЧИСЛЕНЬ ЯК ОПТИМІЗАЦІЯ ПРОЦЕСУ НАДАННЯ ПОСЛУГ КОРИСТУВАЧАМ В УМОВАХ ОБМЕЖЕНИХ ОБЧИСЛЮВАЛЬНИХ РЕСУРСІВ

Матов О. Я.

Інститут проблем реєстрації інформації НАН України,

Київ, Україна

Розглянуто інфраструктуру хмарних обчислень (ХО) як об'єкт адаптації і процес адаптації хмарних обчислень як оптимізаційний. Викладено загальну постановку задачі адаптації дисципліни надання обчислювальних ресурсів користувачам ХО. Запропоновано технологію динамічної адаптивної змішаної дисципліни надання обчислювальних ресурсів користувачам ХО. Наведено напрямок вирішення задачі оптимізації динамічної адаптивної змішаної дисципліни. Запропоновано відомий функціонал оптимізації, який базується на припущенні, що резуль-тати використання обчислювальних ресурсів користувачем (вирішення задач користувача) знецінюються пропорційно часу їхнього знаходження в черзі на вирішення та самого вирішення в системі ХО. Можливі й інші функціонали з часовими обмеженнями. Це є актуальним для сучасних глобальних інформаційно-аналітичних систем реального часу із застосуванням технологій хмарних обчислень і може бути критичним при обмежених обчислювальних ресурсах ХО. Наведено, що задача оптимізації вирішується ітераційним методом з використанням відповідних аналітичних моделей функціонування ХО. Наведена характеристика таких моделей.

Ключові слова: хмарні обчислення, дисципліна надання обчислюваль-них ресурсів, абсолютний та відносний пріоритети, адаптація та оптимізація дисциплін обслуговування, ефективність адаптації, змішана дисципліна обслуговування, математична модель.

Вступ

Створення адаптивних інфраструктур хмарних обчислень, які здатні динаміч-но адаптуватися до постійно діючих змін умов функціонування, та розробка відповідних методів організації обчислень є важливим напрямком розвитку сучасних глобальних інформаційно-аналітичних систем із застосуванням технологій хмарних обчислень.

Адаптація, як керування, є вторинною стосовно основного контуру керування. Якщо керування виконує основні цілі, реалізація яких забезпечує функціонування об'єкта, то адаптація забезпечує якість цього функціонування. Тому, коли виникає потреба в підвищенні (чи підтримці на необхідному рівні) якості функціонування об'єкта, завжди виникає необхідність адаптації.

Особливість системи адаптації полягає в тому, що працювати їй приходиться в умовах значної невизначеності зовнішнього середовища і поводження об'єкта.

Невизначеність середовища і об'єкта є характерною рисою, що дозволяє розглядати адаптацію як специфічний вид керування. При цьому ступінь невизначеності визначає важливість рішення задачі адаптації: чим більше невизначеність, тим більше необхідність адаптації.

Хмарні обчислення як об'єкт адаптації

Хмарні обчислення (ХО) є об'єктами з високим рівнем невизначеності процесу функціонування. Тут зовнішню невизначеність потоку запитів на обчислювальні ресурси (ОР) (середовища) доповнює внутрішня невизначеність ХО (об'єк-та), що зв'язана з наявністю чи відсутністю необхідних ОР, випадкових несправностей системи ХО, а також необхідність забезпечення певних часових характеристик для багатьох клієнтів. Саме це визначає необхідність введення адаптації у процес функціонування ХО.

Крім того, введення адаптації у процес функціонування ХО зв'язано з необхідністю підтримки системи в оптимальному, а іноді і просто працездатному стані незалежно від численних факторів зовнішнього та внутрішнього характеру, що виводять ХО з необхідного цільового стану.

Усе сказане однаковою мірою може бути віднесено і до обчислювального процесу як об'єкта адаптації, тому що він розвивається у ХО і є його невід'ємним атрибутом.

У поняття адаптації як активної дії (керування) звичайно вкладають два зміс-ти: пристосування об'єкта до фіксованого середовища (пасивна адаптація) і пошук середовища, адекватного даному об'єкту (активна адаптація) [1]. У першому випадку об'єкт, що адаптується, функціонує так, щоб виконати свою мету в даному середовищі щонайкраще, тобто максимізує свою ефективність у даному середовищі. Активна адаптація, навпаки, має на увазі зміну середовища з метою максимізації ефективності функціонування об'єкта.

Стосовно до ХО, як системи масового обслуговування (СМО), активна адаптація може розглядатись як зміна інтенсивності чи кількості вхідних потоків заявок, а також законів розподілу процесів надходження заявок.

У практиці функціонування ХО найбільше часто використовується пасивна адаптація, при якій адаптуючий вплив може мати різний характер. Воно може привести до зміни або параметрів об'єкта адаптації (параметрична адаптація), або його структури (структурна адаптація). Як керовані параметри ХО можуть розглядатись інтенсивність і закони розподілу процесу обслуговування заявок, обмеження на тривалість очікування (перебування) заявок у черзі (системі), порядок обслуговування заявок (дисципліна обслуговування) тощо. Прикладом структурної адаптації, при якій змінюються число обслуговуючих приладів і зв'язку між ними, є переконфігурація багатосерверного ХО. Структурна адаптація більш радикальна і звичайно супроводжується параметричною, тому що кожна структура має свої параметри.

Залежно від того, є модель об'єкта адаптації чи ні, розрізняють два дуже важ-ливих види адаптації: адаптацію з моделлю і без моделі (пошукову адаптацію), що істотно відрізняються один від одного [1].

За наявності адекватної моделі об'єкта для синтезу впливу, що адаптує, досить виміряти стан середовища, і використовуючи модель, визначити вплив, що повинен перевести об'єкт у необхідний стан.

Однак дуже часто об'єкт адаптації настільки складний, що неможливо побудувати його модель, а адекватну модель — тим більше. При цьому, мабуть, не можна скористатися методом адаптації з моделлю, що змушує звертатися до пошукової адаптації. Цей вид адаптації відрізняється наявністю пошуку — спеціально організованого процесу, що дозволяє визначити необхідний адаптуючий вплив, не маючи моделі об'єкта. Для пошукової адаптації характерні експерименти з об'єктом, у процесі яких одержують інформацію про його властивості. На основі цієї інформації визначається вплив, що адаптує, підвищувальна ефективність об'єкта.

Сам по собі процес пошукової адаптації має послідовний багатоетапний характер — на кожному етапі приймаються заходи для підвищення ефективності об'єкта (на відміну від адаптації з моделлю, за якої можлива адаптація за один етап).

Якщо при адаптації з моделлю стан об'єкта потрібно вимірювати тільки для підстроювання його моделі і не потрібно для самої адаптації, то при пошуковій адаптації стан об'єкта несе основну інформацію для формування впливу, що адаптує.

Труднощі адаптації з моделлю полягають у синтезі моделі об'єкта, а сама адаптація — у рішенні оптимізаційної задачі вибору такого формування впливу, що адаптує, що задовольнило би цілям адаптації. При пошуковій адаптації виникають труднощі іншого роду — потрібно одночасно і експериментувати з об'єк-том, і адаптувати його.

В усіх випадках, коли вдається побудувати адекватну модель об'єкта, питання про вибір виду адаптації зважується однозначно на користь адаптації з моделлю, тому що тільки наявність моделі дозволяє швидко адаптувати об'єкт.

Адаптація хмарних обчислень як оптимізація

Рішення задачі адаптації зводиться до визначення такого керуючого (адаптуючого) впливу, при якому досягається максимальна ефективність роботи об'єк-та у сформованій ситуації.

Сформована ситуація характеризується двома факторами: станом середовища, у якій знаходиться об'єкт, і станом самого об'єкта адаптації.

Для ХО, як системи масового обслуговування, під станом середовища можна розуміти, наприклад, інтенсивність вхідних потоків заявок, а під станом об'єкта (системи) — число чи час очікування (перебування) заявок у черзі (системі), чи справність-несправність обслуговуючого приладу, рівень завантаження системи і т.п.

Залежно від сформованої ситуації повинен бути сформований адаптуючий вплив, що мінімізує середнє число чи середній час чекання (перебування) заявок у черзі (системі), або час входження системи в стаціонарний режим, або сумарну вартість за роботу системи, або вірогідність утрати заявок і т.д. Метою адаптації може бути максимізація доходу від обслуговування заявок, ліквідація перевантаження системи та підтримка її у стаціонарному режимі функціонування.

Таким чином, адаптацію ХО можна розглядати як процес оптимізації роботи у сформованій ситуації.

Загальна постановка задачі адаптації дисципліни обслуговування

Задача адаптації дисципліни обслуговування у ХО виникає в зв'язку з непередбаченими і неконтрольованими змінами в середовищі та системі, що неминуче змінюють оптимальне настроювання дисципліни обслуговування, якщо така була реалізована в системі. Тому систематичне підстроювання (адаптація) дисципліни обслуговування неминуче при бажанні підтримувати систему в оптимальному режимі незалежно від змін, що відбуваються в середовищі та системі.

Сформулюємо в загальному вигляді задачу адаптації дисципліни обслуговування [2].

Нехай X та E — контрольований і неконтрольований стани середовища. Пара $\{X, E\}$ однозначно описує середовище, у якій знаходиться ХО. Наприклад, X — паспортні дані заявок з обслуговування, а E — інтенсивність їхнього надходження.

Аналогічно пара $\{Y, H\}$ описує стан системи. Тут Y та H — відповідно контрольовані і неконтрольовані фактори. Наприклад, Y — довжина черг заявок, а H — інтенсивність їхнього обслуговування чи інтенсивність відмовлень системи.

Показник ефективності системи носить екстремальний характер. Він визначений на контрольованих станах середовища та системи:

$$\acute{Y} = \acute{Y}(X, Y). \quad (1)$$

Як показники ефективності системи можуть виступати середній час перебування (очікування) заявок у системі (черги), середня довжина черги заявок, середня сумарна вартість очікування (перебування) заявок у черзі (системі) і т.п.

Стан системи Y залежить від X, E та H, а також від дисципліни обслуговування S:

$$Y = F(X, E, H, S), \quad (2)$$

де F — оператор системи.

Під дисципліною обслуговування розуміють правило вибору заявок на обслуговування залежно від станів середовища та системи:

$$S = S(X, Y). \quad (3)$$

Оптимальність дисципліни S у більшості випадків зв'язана з екстремізацією показника ефективності функціонування системи (1). Це означає, що для синтезу оптимальної дисципліни S^0 необхідно вирішити наступну оптимізаційну задачу:

$$\acute{Y}[X, F(X, E, H, S)] \to \underset{S \in S^*}{extr} \Rightarrow S^0, \quad (4)$$

де S^* — обмеження, що накладаються на вибір дисципліни обслуговування S.

Ці обмеження можуть бути зв'язані, наприклад, з визначеним набором заздалегідь заданих дисциплін обслуговування і т.п.

Очевидно, що вирішити задачу (4) на стадії проектування обчислювальної системи (ОС) неможливо, тому що апріорі невідомі фактори E та H. Усереднювання за цими факторами вводити не можна, тому що вони можуть мати нестаціо-нарний характер.

Тому задачу синтезу оптимальної дисципліни S^0 варто вирішувати шляхом адаптації ХО, тобто в режимі їхньої експлуатації. Тоді адаптація зводиться до рішення задачі

$$\acute{Y}(S) \to \underset{S \in S^*}{extr} \Rightarrow S^0$$

за локальними спостереженнями оцінок значення показника ефективності при різ-них дисциплінах:

$$\hat{\acute{Y}}_1 = \hat{\acute{Y}}(S_1), \ldots, \hat{\acute{Y}}_\xi = \hat{\acute{Y}}(S_\xi).$$

Алгоритм адаптації повинний указати послідовність переходу від однієї дис-ципліни до іншої: $S_1 \to S_2 \to \cdots \to S_\xi \to \cdots$, яка приводить до рішення S^0, оптимального у сформованій ситуації.

Використання для адаптації технології динамічної адаптивної змішаної дисципліни надання обчислювальних ресурсів користувачам ХО

На даний час відома велика кількість різних дисциплін обслуговування. З них у ХО широко застосовуються дисципліни обслуговування з відносним і абсолютним пріоритетами. Однак ці дисципліни є статичними і внаслідок цього мають ряд істотних недоліків, що знижують ефективність обчислювальних систем (процесів) в умовах невизначеності зовнішнього середовища і поводження самих систем.

При використанні дисципліни з відносним пріоритетом вибір чергової заявки для обслуговування може бути здійснений тільки після завершення поточного обслуговування, навіть

якщо заявка, що обслуговується, має нижчий пріоритет. Унаслідок цього тривалість перебування у ХО деяких найбільш важливих заявок може виявитися неприпустимо великою. Зменшення затримки в обслуговуванні важливих заявок досягається за рахунок переривання, тобто введення для цих заявок абсолютного пріоритету. Однак при цьому зростає тривалість перебування у ХО заявок низьких пріоритетів і в ряді випадків, при інтенсивному надходженні важливих заявок, може відбутися блокування процесу обслуговування заявок низьких пріоритетів, що також знижує ефективність ХО у цілому.

Для компенсації недоліків, які властиві дисциплінам обслуговування з відносним і абсолютним пріоритетами і обліком їхніх переваг, доцільно реалізувати у ХО змішані дисципліни обслуговування, що використовують як відносний, так і абсолютний пріоритети.

Розглянемо одну зі змішаних дисциплін обслуговування. Нехай N вхідних потоків заявок відповідно до їхньої важливості та терміновості в обслуговуванні розбито на M груп, між якими діє абсолютний пріоритет, а всередині — відносний. Це означає, що заявки будь-якого потоку з групи $m(m=\overline{1,M})$ переривають обслуговування заявок, які належать потокам із груп з номерами $\overline{m+1,M}$. У кож-ній групі знаходиться N_m потоків, заявки яких не переривають один одного. Очевидно, що $\sum_{m=1}^{M} N_m = N$. Пріоритет будь-якої заявки в системі з такою дисципліною обслуговування може бути описаний парою чисел m та n, де $n=\overline{1,N_m}$ і визначає номер потоку заявок у групі з номером m.

Описана змішана дисципліна обслуговування дозволяє за рахунок її адаптації більш гнучко реагувати на різні ситуації, що виникають у процесі функціонування ОС. При цьому адаптація дисципліни складається в зміні кількості та положення границь, що розділяють потоки заявок на групи абсолютного пріоритету, тобто в зміні кількості груп і кількості потоків у групах. Варіанти групування будемо називати розбивками. Загальна кількість розбивок Φ визначається числом потоків заявок $N:\hat{O}=2^{N-1}$. Кожна розбивка $\varphi(\varphi\in\hat{O})$ задається сукупністю чисел $\{N_1,N_2,\ldots,N_M\}$.

Така дисципліна правомірно названа динамічною адаптивною змішаною дис-ципліною надання обчислювальних ресурсів користувачам ХО.

Уведена в розгляд змішана дисципліна обслуговування представляє відомий практичний інтерес, оскільки вона при оптимальному виборі розбивки потоків по групах у принципі забезпечує не гірше обслуговування порівняно з «*чистими*» дисциплінами (з відносним і абсолютним пріоритетами). Так, при $M=N$, $N_m=1$ для всіх $m=\overline{1,M}$ виходить дисципліна обслуговування з абсолютним пріоритетом, а при $M=1$, $N_1=N$ — з відносним.

Задачі динамічної адаптивної змішаної дисципліни надання обчислювальних ресурсів з моделлю

Розглянемо дві практичні задачі динамічної адаптивної змішаної дисципліни надання обчислювальних ресурсів (змішаної дисципліни обслуговування) з моделлю.

Одними із основних показників ефективності ХО є показники, що базуються на оцінці часових характеристик цих систем. Такі показники можуть задаватися договором між постачальником і користувачем ОР ХО і здобувають особливе значення для систем, що функціонують у реальному масштабі часу.

Унаслідок випадкового характеру обчислювального процесу виникають додаткові затримки в обробці інформації, порушуються припустимі обмеження на час її перебування в ХО, що негативно позначається на ефективності рішення цільових задач користувачів.

Для забезпечення необхідної ефективності ХО в таких ситуаціях необхідно підтримувати часові характеристики системи на заданому рівні. В умовах дефіциту обчислювальних ресурсів це можливо тільки за рахунок підвищення ефективності обчислювального процесу, зокрема, за рахунок адаптації дисципліни обслуговування. Поряд з цим виникає задача найбільш ефективного використання наявних обчислювальних ресурсів у кожен момент часу функціонування керуючої ХО. Цю задачу також можна вирішити шляхом адаптації дисципліни обслуговування.

У зв’язку з викладеним, як показник ефективності ХО виберемо середню сумарну вартість часу надання (очікування в чергах і часу використання, тобто перебування в ХО як у СМО) ОР за заявками (вимогами) користувачів. Для цього використаємо відомий функціонал [1] — середню сумарну вартість часу надання ОР:

$$C^{(S)}=\sum_{i=1}^{n}\alpha_i\lambda_i v_i^{(s)},$$

з чого маємо

$$C^{(\varphi)} = \sum_{m=1}^{M}\sum_{n=1}^{N}\alpha(m,n)\lambda(m,n)v^{(\varphi)}(m,n)\,, \qquad (5)$$

де α_i — вартість за одиницю часу ОР для *i*-го типу заявок користувачів;

λ_i — інтенсивність *i*-го потоку заявок;

$v_i^{(S)}$ — середній час надання ОР заявок *i*-го потоку;

n — кількість типів заявок;

s — параметр, що характеризує спосіб організації обчислювального процесу;

$v^{(\varphi)}(m,n)(m=\overline{1,M},n=\overline{1,N_m})$ — середній час надання ОР в ХО заявкам (*m, n*)-го потоку;

α (*m, n*) — вартість одиниці часу надання ОР в ХО заявкам (*m, n*)-го потоку;

λ (*m, n*) — інтенсивність (*m, n*)-потоку надання ОР в ХО.

Показник ефективності базується на припущенні, що результати використання ОР користувачем знецінюються пропорційно часу їхнього перебування в системі ХО. Тоді цілями адаптації змішаної дисципліни обслуговування будуть або задоволення вимог вчасного перебування (*m, n*) заявок у системі, що задаються припустимими значеннями цього часу $v_{\ddot{A}}(m,n)$, або мінімізація функціонала (5). Ця мета досягається шляхом відшукання відповідних оптимальних розбивок φ^0, тобто задачі адаптації змішаної дисципліни обслуговування з відносно-абсолютним пріоритетом являють собою оптимізаційні задачі, загальна постановка яких розглянута вище.

Оскільки сформульовані вище цілі адаптації змішаної дисципліни обслуговування можуть бути досягнуті при декількох різних розбивках потоків заявок на групи абсолютного пріоритету, то виникає необхідність уведення додаткового обмеження на вибір розбивки φ.

Наявність у ХО абсолютного пріоритету потребує деяких технологічних втрат ОР, які пропорційні числу груп (рівнів) абсолютного пріоритету. У зв'язку з цим оптимальною необхідно вважати таку розбивку, яка забезпечує досягнення цілей адаптації при мінімальній кількості груп абсолютного пріоритету *M*.

Тоді розглянуті задачі адаптації змішаної дисципліни обслуговування можуть бути формально поставлені в такий спосіб:

$$\begin{gathered} v^{(\varphi)}(m,n) \le v_{\ddot{A}}(m,n) \Rightarrow \varphi^0, \\ \varphi \in \Phi, \\ M = \min. \end{gathered} \qquad (6)$$

$$\begin{gathered} C^{(\varphi)} \to \min \Rightarrow \varphi^0, \\ \varphi \in \Phi, \\ M = \min. \end{gathered} \qquad (7)$$

Рішення задач відшукання оптимальної розбивки (6) і (7) за допомогою відомих аналітичних методів оптимізації не представляється можливим. Єдиний шлях рішення цих задач — евристичний підхід, що не має формального обґрунтування, а спирається лише на специфіку математичних моделей [3…6] і зв'язані з ними розуміння.

З виразів (5)–(7) випливає, що досягнення цілей адаптації змішаної дисципліни обслуговування сполучено з потребою оцінки значення середнього часу перебування в системі заявок (*m, n*)-типу — *v*(*m, n*). Тому виникає необхідність синтезу математичної моделі ХО зі змішаною дисципліною обслуговування [3].

Характеристика аналітичних моделей

Розробляти аналітичні моделі хмарних обчислень пропонується як системи масового обслуговування зі змішаною дисципліною надання ресурсів. Моделі повинні враховувати відмови і різні особливості функціонування та мати довільні закони розподілення для деяких вірогідних процесів.

Загальний опис моделей наступний. На вхід системи ХО , у якій реалізована змішана дисципліни обслуговування (з відносними та абсолютними пріоритетами), надходять N пуасонівських потоків заявок на ресурси з відповідними N пріоритетами. Тривалісті обслуговування заявок різних потоків мають свої довільні закони розподілення. Заявка з відносним пріоритетом, обслуговування якої перерване заявками з абсолютним пріоритетом, повертається в чергу. Розглядаються дві

дисципліни поновлення обслуговування A і B. У межах одного пріоритету заявки обслуговуються в порядку надходження.

ХО виходить з ладу по пуасонівському закону, а відновлюється по довільному закону. У період відновлення використовуваються елементи адаптації до відмов: заявки одних потоків у чергу приймаються і накопичуються, а інші – не приймаються (дисципліна поповнення черги відповідно I і II). Вихід з ладу обслуговуючого приладу може відбутися як під час його вільного стану, так і під час обслуговування заявки. Розглянуто дві дисципліна поновлення обслуговування після відновлення C і D. Перервана заявка дообслуговується з місця її переривання. Сполучення дисциплін поновлення обслуговування і поповнення черги дозволяє розглядати самостійні моделі різних типів систем, що мають відповідне позначення.

Різні особливості функціонування складаються з різних сполучень дисциплін A, B, C, D, I і II. Такі моделі з розкритям особливостей функціонування розглянуті в [3] і для однієї з них наведені фінальні аналітичні вирази.

Висновки

Відшукування оптимальної дисципліни обслуговування не завжди зв'язано з екстремізацією показника ефективності функціонування ХО. Метою адаптації також може бути задоволення обмежень на показник ефективності, що задані у вигляді рівностей чи нерівностей. У будь-якому випадку така постановка задачі адаптації припускає необхідність реалізації у ХО декількох чи однієї змішаної дисципліни обслуговування.

Посилання

1. Матов А.Я., Шпилев В.Н., Комов А.Д. и др. Организация вычислительных процессов в АСУ/под ред. А.Я.Матова. Киев, 1989. 200 с.
2. Матов О.Я. Оптимізація надання послуг обчислювальними ресурсами адаптивної хмарної інфраструктури. Реєстрація, зберігання і обробка даних. 2018. Т.20, №3 С.83-90.
3. Aleksandr Matov. Mathematical models of cloud computing with absolute-relative priorities of providing of computer resources to users in conditions of functioning features and failures // CEUR Workshop Proceedings (ceur-ws.org). Vol-2318 urn:nbn:de:0074-2318-4. Selected Papers of the XVIII International Scientific and Practical Conference on Information Technologies and Security (ITS 2018) PP. 150-159 [http://ceur-ws.org/Vol-2318/paper13.pdf]
4. Mokrov E.V., Samuilov K.E. Cloud computing system model in the form of a queuing system with multiple queues and with a group of requests. https://cyberleninka.ru/article/n/model-sistemy-oblachnyh-vychisleniy-v-vide-sistemy-massovogo-obsluzhivaniya- s-neskolkimi-ocheredyami-i-s-gruppovym-postupleniem-zayavok. Russ.
5. Tsai J.M., Hung S.W. A novel model of technology diffusion:system dynamics perspective for cloud computing // Journal of Engineering and Technology Management. 2014. V. 33. P. 4762. doi: 10.1016/j.jengtecman.2014.02.003
6. Singh P., Dutta M., Aggarwal N. A review of task scheduling based on meta-heuristics approach in cloud computing // Knowledge and Information Systems. 2017. V. 52. N 1. doi: 10.1007/s10115-017-1044

ПІДХІД ДО ДЕЛЕГУВАННЯ ТРАНЗАКЦІЙ У САМОЗАХИСНИХ ДЕЦЕНТРАЛІЗОВАНИХ ПЛАТФОРМАХ ДАНИХ

Савченко М.М., Циганок В.В., Андрійчук О.В.

Інститут проблем реєстрації інформації НАН України,
Київ, Україна

Анотація. Проведено огляд існуючих підходів до делегування транзакцій в самозахисних децентралізованих платформах даних на прикладі найбільш відомих проектів децентралізованої платформи даних Ethereum. Розглянуто чотири найбільш використовуваних підходи, виділено їх переваги та недоліки, показана обмеженість рішень на їх основі. Розроблено універсальний підхід до делегування транзакцій в самозахисних децентралізованих платформах даних, який не залежить від стандарту децентралізованої програми, серверну частину для його підтримки та допоміжну клієнтську частину, які можна застосувати до будь-яких децентралізованих програм. Наведено можливий вектор застосування розробленого підходу до систем підтримки прийняття рішень.

Ключові слова: Технологія розподіленого реєстру, децентралізовані платформи даних, делеговані транзакції, Ethereum, системи підтримки прийняття рішень.

Вступ

Серед сучасних перспективних напрямків розвитку інформаційних технологій, все більшої уваги та важливості набуває розвиток самозахисних децентралізованих платформ даних та технології DLT (Distributed Ledger Technology), яким прогнозується величезна роль у майбутньому майже в будь-яких сферах: фінансових, державних, охорони здоров'я, торгівлі, тощо. Такі децентралізовані платформи на 2019 рік представлені в основному технологією blockchain, а також іншими технологіями, які надають платформам даних на їх основі схожих характеристик.

Самозахисна децентралізована платформа даних — це стандартизована система зберігання, захисту, підтримки, обробки та управління даними, яка підтримується багатьма автоматизованими обчислювальними вузлами, поєднаними в одну цілісну мережу. Основою для таких платформ даних зазвичай виступає технологія Distributed Ledger Technology (DLT, технологія розподіленого реєстру). Основною характерною особливістю таких платформ даних є стійкість до будь-яких атак на дані та до впливу на саму децентралізовану мережу. Це практично унеможливлює підробку, видалення, неправомірну модифікацію або доступність даних. До основних характеристик самозахищених децентралізованих платформ даних відносять їх пропускну спроможність (зазвичай, кількість транзакцій за одиницю часу), масштабованість, ступінь децентралізації (повна або часткова), вартість обслуговування (обчислювальна та грошова) та інші.

Прикладами таких децентралізованих платформ даних є проекти Bitcoin, Ethereum, EOS, IOTA, Hedera Hashgraph, Holochain, Radix, Telegram Open Network (TON) та інші. Кожна платформа підтримує власну окрему мережу за допомогою десятків або сотень тисяч обчислювальних вузлів, розташованих по всій земній кулі. Самозахисні децентралізовані платформи даних уже сьогодні застосовуються для величезної кількості задач: створення програмованих криптовалют, захищених реєстрів даних, таких як, наприклад, кадастровий реєстр, або реєстр публічних документів, проведення непідробних голосувань, реєстрації вантажу або товарів, токенізації предметів реального світу тощо.

Відносно нові, але дедалі більш актуальні технології, що лежать в основі самозахисних децентралізованих платформ даних, стрімко розвиваються, як і самі платформи даних. Найперша самозахисна децентралізована платформа даних Bitcoin, яка дозволяє проводити усього 7 транзакцій за секунду, проіснувала уже більше 10 років і досі не була успішно атакована. Такий рівень захисту і довіри спонукає перенесення усіх класичних програмних систем на самозахисні децентралізовані платформи даних. Але існує ряд недоліків, через які на даний момент це зробити неможливо. По-перше, жодна з тепер наявних децентралізованих платформ не є масштабованою, і тому не може обслуговувати велику кількість транзакцій та запитів. По-друге, дані платформи вимагають певного підходу до роботи з ними, який часто виявляється занадто складним або занадто коштовним для кінцевих користувачів. Нерідко і застосунки, що розробляються на основі певної децентралізованої платформи спрощуються і втрачають властивість захищеності даних, яку їм могла б надати дана платформа.

Отже, однією з найбільш важливих проблем є складність у користуванні такими платформами, як і відсутність загальноприйнятого або універсального підходу до побудови децентралізованих застосунків з метою вирішення проблеми складності для кінцевого користувача. У даній роботі описується універсальний підхід до побудови таких застосунків незалежно від платформи, що використовується. Проведено порівняння переваг і недоліків використання даного підходу в реальних сучасних застосунках та наведено рекомендації щодо його використання у системах підтримки прийняття рішень як приклад використання.

1. Проблема сплати комісій користувачем в самозахисній децентралізованій платформі даних

Жодна з існуючих і теоретично можливих децентралізованих мереж та самозахисних платформ даних, що утворюють такі мережі, не може існувати без їх неперервної підтримки всіма учасниками мережі. Окрім того, кожний учасник має бути достатньо мотивованим щоб залишатися в мережі та виконувати функції забезпечення її цілісності та захищеності [1]. Існує декілька видів стимулів для учасників мережі, які застосовуються в сучасних децентралізованих платформах даних:

- Фінансова нагорода. На даний момент є основним стимулом і використовується майже в будь-яких децентралізованих платформах даних: Bitcoin, Ethereum, EOS. Учасники мережі, що її підтримують, отримують винагороду у вигляді певного активу в мережі, який використовується і для сплати комісії за транзакції користувачами. Також, у алгоритмі децентралізованої платформи може бути закладена передбачувана інфляція, що генерує

певну кількість цього активу (наприклад, додатково 1% на рік) і розподіляє його між учасниками, що підтримують мережу.

- Привілеї у користуванні мережею. Наприклад, можливість виконувати в ній більше транзакцій. Даний стимул використовується у таких платформах, як IOTA. Суть його полягає в тому, що самі користувачі мережі, або компанії, яким необхідно користуватися мережею, зобов'язані виконувати функції підтримки цієї мережі для користування нею.
- Волонтерство або домовленість. Частково принципи волонтерства використовується (або використовувалися, наприклад, під час запуску мережі) у таких децентралізованих мережах як EOS, IOTA. Тобто, мережа підтримується волонтерами, наприклад, з метою зарекомендувати себе у суспільстві (общині), отримати певний рейтинг в мережі тощо. А домовлене волонтерство здебільшого використовується у приватних мережах, але наразі з'являється все більше прикладів використання такого підходу і в публічних мережах, коли декілька інституцій домовляються про підтримку спільної мережі та побудову самозахисної децентралізованої платформи даних на ній. Яскравим прикладом такої платформи можна назвати проект Libra, започаткований Facebook, до якого входять велика кількість таких організацій як Visa, Mastercard, PayPal, Uber та інші.
- Комбінації вище перерахованих стимулів.

Однак, майже кожна сучасна децентралізована платформа даних використовує фінансову винагороду як стимул, так само, як і за виконання транзакції, в ній присутня комісія для її користувачів [2]. Комісія необхідна не тільки для винагороди тих, хто підтримує децентралізовану платформу даних, але і як запобіжний захід проти спам-атак на платформу даних. Тобто, для того, щоб користуватися платформою даних або застосунками на її основі, користувачу необхідно спершу придбати певну кількість криптовалюти, яка використовується для сплати комісій, а лише потім переходити до користування платформою даних або її застосунком.

Придбання криптовалюти для кожної децентралізованої платформи даних є непростою процедурою: окрім того, що користувачам необхідно створювати власний акаунт (гаманець) у рамках децентралізованої платформи, користувачі також вимушені використовувати так звані обмінники для того, щоб отримати криптовалюту, за допомогою якої вони можуть взаємодіяти з децентралізованим застосунком. На сьогодні сервіси інтернет-обміну валют є досить розвиненими, однак, присутні нижчезначні обмеження здійснення такого обміну:

- По-перше, інтернет-обмін дуже залежить від країни, в якій знаходиться користувач, та від її законів. Станом на сьогодні, багато людей досі не мають можливості купувати криптовалюти.
- По-друге, через обмеження обмінних пунктів, користувачі не можуть купляти маленьку суму криптовалюти, необхідну лише для кількох цільових транзакцій в мережі. Тому їм доводиться купувати криптовалюту одразу на значну суму (зазвичай, не меншу від 50 USD).

На жаль, більшість інтернет-користувачів не виявляють бажання використовувати інтернет-сервіси, які мають складну процедуру реєстрації або використання. А у випадку інтернет-сервісу на основі децентралізованих платформ даних використання здебільшого сильно обмежується необхідністю придбання криптовалюти. Нижче на рисунку 1.1 показано початкові етапи, виконання яких необхідне для підготовки до роботи з децентралізованими платформами даних.

Рисунок 1.1 — діаграма етапів через які проходить користувач перед використанням децентралізованих застосунків

Таким чином, використання децентралізованих застосунків на основі самозахисних децентралізованих платформ даних або набуває дуже вузького застосування, або спрощується до систем, у яких використання децентралізованої платформи даних втрачає основну частину своїх важливих функцій через часткову або повну централізацію.

2. Наявні підходи до спрощення рішень на основі самозахисних децентралізованих платформ даних

Використання децентралізованих застосунків, відомих як DApps, або Decentralized Applications потребує спеціального програмного забезпечення, відомого як криптогаманці. Акаунт (гаманець) в децентралізованій платформі даних — це адреса користувача, яка використовується ним для виконання певних дій з платформою даних [3]. Акаунт може бути згенерованим (як, наприклад, у децентралізованих платформах даних Ethereum або Bitcoin), або попередньо зареєстрованим

(наприклад, EOS). У будь-якому випадку, для створення акаунта користувач генерує секрет (приватний ключ), який дозволяє йому, і тільки йому, виконувати певні дії в децентралізованій платформі. Спеціальне програмне забезпечення ("криптогаманець") зберігає секрет (приватний ключ) користувача і виступає допоміжною програмою для взаємодії з децентралізованою платформою. Нерідко за допомогою одного секрету можна згенерувати практично нескінченну кількість акаунтів з використанням алгоритму ієрархічних детермінованих ключів (Hierarchical Deterministic key creation), наприклад, BIP32 [4], якщо це передбачено криптогаманцем.

Криптогаманці, зазвичай, існують у вигляді мобільних додатків, фізичних пристроїв з додатковим програмним забезпеченням, плагінів для інтернет-браузерів або окремих браузерів. Прикладами існуючих криптогаманців є Trezor, Ledger (фізичні пристрої) [5], Trust, Metamask (мобільні додатки), Metamask (додатки для інтернет-браузера), Mist, Brave, Opera (інтернет-браузери з вбудованим криптогаманцем) [6].

Створення гаманця зазвичай не викликає труднощів, але потребує додаткового часу і уваги користувачів. Адже у випадку втрати секрета користувач більше не зможе відновити доступ до його акаунту [7]. Але найбільшою складністю, як було зазначено вище, є поповнення акаунту певною кількістю криптовалюти, що дозволяє користувачу працювати з децентралізованим застосунком. Рішення, що спрощують роботу з децентралізованим застосунком зазвичай застосовуються на етапі купівлі криптовалюти або є такими, що повністю усувають даний етап. Таким чином, із початкових етапів, зображених на рисунку 1.1, залишається тільки один етап — створення криптогаманця і акаунта для децентралізованої платформи даних. Спрощена схема етапів підготовчої роботи користувача показано на рисунку 2.1.

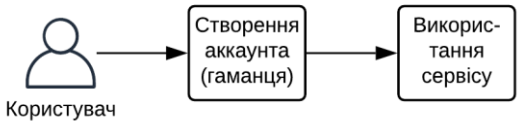


Рисунок 2.1 — схема етапів, через які проходить користувач перед використанням децентралізованих застосунків без необхідності купувати криптовалюту

Дане спрощення зазвичай відбувається за допомогою делегування транзакцій — тобто, делегування права виконувати транзакції в мережі від одного акаунта (користувача) до іншого [8, 9]. Таким чином, комісію за виконання транзакції буде сплачувати делегат (акаунт, що проводить транзакцію), а не акаунт, який має намір виконати транзакцію. На теперішній час використовується ряд підходів до делегування транзакцій, кожен з яких має свої переваги і недоліки. Нижче наведено огляд даних підходів та порівняння їх переваг і недоліків.

2.1. Довірене делегування транзакцій

Яскравим прикладом довіреного делегування транзакцій є надання делегату прав виконувати будь-які не чітко визначені транзакції від імені іншого акаунта. В таких децентралізованих платформах даних як Ethereum навіть розроблено стандарти для написання децентралізованих програм-токенів (наприклад, ERC777 [10]), що регулює делегування транзакцій за допомогою довіреного делегування транзакцій.

Стандарт ERC777, який описує довірене делегування транзакцій, дозволяє акаунтам передавати право розпоряджатися своїми токенами (програмованими криптовалютами) іншим акаунтам, так само як і забирати в них таке право. Таким чином, користування децентралізованим застосунком може бути спрощене, оскільки транзакції, що мав би виконувати користувач, будуть проводитися делегатом, а не самим користувачем. Користувачу залишатиметься лише виконати одну транзакцію, що дозволить довіреному делегату користуватися токенами від імені його акаунта, і будь-які наступні транзакції зможе проводити делегат. Схема роботи довіреного делегування транзакцій показана на рисунку 2.2.

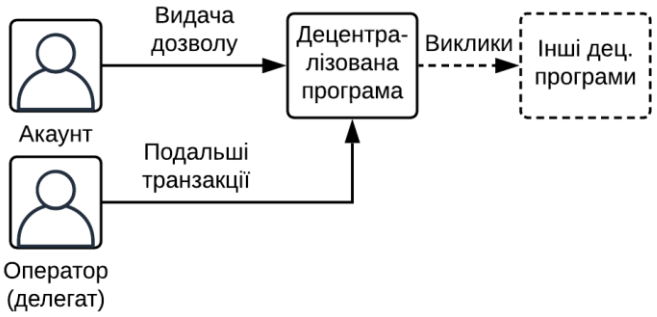


Рисунок 2.2 — схема довіреного делегування транзакцій

Даний підхід спрощує користування децентралізованим застосунком, оскільки всі дії користувача тепер може виконувати довірений делегат, сплачуючи за них комісію. Проте, це можливо лише після виконання користувачем транзакції, що надає делегату такий дозвіл, для виконання якої все ще потрібна криптовалюта. Однак, стандартом ERC777 також передбачено, що у токена може бути присутній делегат за замовчуванням, наприклад, акаунт установи що випустила токен.

Виходячи з вище описаної концепції стандарту ERC777, можна виділити наступні переваги довіреного делегування транзакцій:

- Простий у застосуванні.
- Стандартизований.

Недоліки даного підходу:

- Вимагає проведення першої коштовної транзакції, якщо децентралізованою програмою не передбачений делегат за замовчуванням.
- Вимагає проведення коштовних транзакцій для надання/припинення повноважень делегатів.
- Не обмежує делегата у тому, які транзакції він може виконувати, а тому є недостатньо захищеним (через можливі вектори атаки на делегата).
- Може бути застосований тільки до нових децентралізованих програм, або тих, де даний підхід був передбачений.

2.2. Делегування транзакцій за допомогою допоміжних децентралізованих програм

На відміну від довіреного делегування транзакцій, часто є необхідним більш захищений підхід, який можна також застосувати і до вже існуючих децентралізованих програм (смарт-контрактів). У такому випадку, можна створити проміжну децентралізовану програму, яку прийнято називати контрактом особистості (identity contract). Дана децентралізована программа виступає у якості акаунта, але запрограмованого таким чином, що дозволяє присвоїти їй будь-який акаунт-власник та виконувати захищені делеговані транзакції без необхідності власника платити за транзакції. Акаунт-власник з таким підходом має лише надавати дозвіл на проведення транзакцій від імені його контракта особистості за допомогою електронно-цифрових підписів. Схема делегування транзакцій за допомогою допоміжних децентралізованих програм наведена на рисунку 2.3.

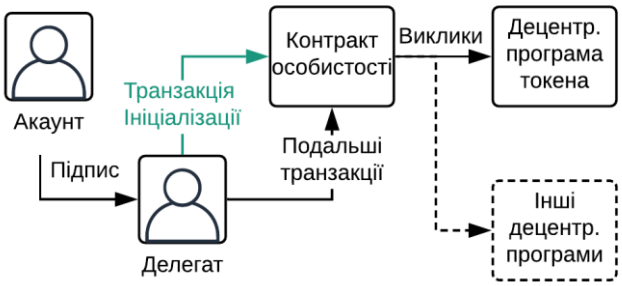


Рисунок 2.3 — схема делегування транзакцій за допомогою допоміжних децентралізованих програм

Серед переваг даного підходу можна виділити наступні:

- Може бути застосований до будь-яких децентралізованих програм.

- Є захищеним, оскільки з таким підходом акаунт-власник контракта особистості може мати повний контроль над діями, які виконуються делегатом за рахунок електронно-цифрових підписів.

Недоліки делегування транзакцій за допомогою допоміжних децентралізованих програм:

- Змінює логіку роботи з децентралізованими програмами і вводить поняття "контракта особистості", таким чином роблячи деякі візуальні застосунки несумісними з ним. Наприклад, у даному підході токенами буде володіти контракт особистості, а не сам акаунт, через що баланс токенів у гаманці власника буде завжди відображатися нульовим (оскільки гаманець нічого не знає про контракт особистості).
- Необхідна попередня ініціалізація контракта особистості, що потенційно відкриває новий вектор для спам-атак на децентралізовані застосунки.
- Складно стандартизувати.

Прикладами стандартів, що описують контракт особистості з можливістю делегування транзакцій за допомогою підписів є ERC1056 або ERC780 [11].

2.3. Делегування транзакцій без використання проміжних децентралізованих програм

Підхід до делегування транзакцій без використання проміжних децентралізованих програм можна класифікувати як комбінацію переваг двох вищеописаних підходів. На відміну від підходу з використанням допоміжних децентралізованих програм, усі властивості допоміжної децентралізованої програми надаються самій децентралізованій програмі, що описує криптовалюту якою сплачується комісія. І, на відміну від довіреного делегування транзакцій, дії, які може виконувати делегат можуть бути обмеженими і повністю контрольованими самим акаунтом за допомогою електронно-цифрових підписів. Схема делегування транзакцій без використання допоміжних децентралізованих програм наведена на рисунку 2.4.

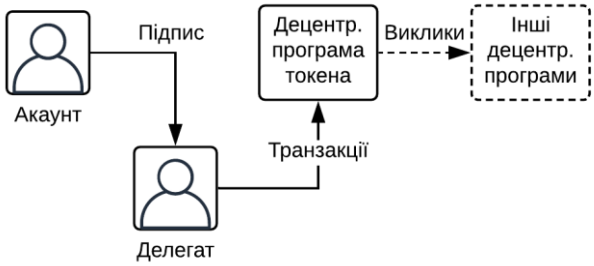


Рисунок 2.4 — схема делегування транзакцій без використання допоміжних децентралізованих програм

Перевагами даного підходу є:

- Немає необхідності в транзакціях ініціалізації.
- Є захищеним, оскільки з таким підходом акаунт-власник контракта особистості може мати повний контроль над діями, які може виконувати делегат за рахунок електронно-цифрових підписів.
- Не змінює логіку роботи децентралізованих застосунків (але має свої особливості).

До недоліків даного підходу можна віднести:

- Складно стандартизувати.
- Може бути застосований тільки до нових децентралізованих програм, або тих, де даний підхід був передбачений.

Прикладами спроб стандартизувати даний підхід є пропозиції стандартів ERC1776, ERC-865, ERC-1003 та ERC-1228 [11, 12], що на 2019 рік залишаються відкритими. Основний недолік даного підходу полягає в тому, що його складно стандартизувати, і це спричинило появу багатьох стандартів токенів.

2.3. Делегування транзакцій на рівні платформи даних

Якщо говорити про самозахисні децентралізовані програми в цілому, існує можливість імплементації делегування транзакцій на рівні самої платформи даних. Однак, досі даний підхід не набув поширення і не імплементований у провідних децентралізованих платформах, таких як Ethereum, TRON, EOS та інших (станом на 2019 рік). Відомо лише, що даний підхід може бути реалізований у децентралізованій платформі Ethereum 2.0, випуск якої планується не раніше ніж в 2022 році [13].

Підхід до делегування транзакцій на рівні самої платформи даних може виглядати як можливість сплати комісії за будь-яку транзакцію будь-яким делегатом, на заздалегідь погоджених умовах. Наприклад, делегат буде проводити тільки такі групи підписаних акаунтом транзакцій, в яких хоча б одна транзакція буде сплачувати відповідну комісію у певному токені, криптовалюты, або проводити вигідну для делегата дію.

Для реалізації даного підходу необхідна підтримка груп підписаних транзакцій на рівні самої платформи даних. Таким чином, необхідно щоб декілька підписаних транзакцій, одна з яких, наприклад, є сплатою комісії в токені, гарантовано проводилися платформою в повному обсязі або були відхилені у разі хоча б однієї невдалої транзакції. Потенційна схема імплементації даного стандарту наведена на рисунку 2.5.

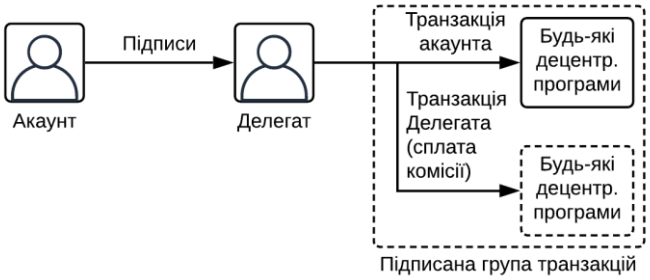


Рисунок 2.5 — схема делегування транзакцій на рівні платформи даних

До переваг даного підходу можна віднести усі переваги вище зазначених підходів, крім наступних недоліків:

- На 2019 рік не підтримується ніякими існуючими розвиненими децентралізованими платформами даних.
- Важко стандартизувати на рівні платформи даних. Однак, стандартизація на рівні децентралізованих застосунків можлива, оскільки делегатами зазвичай виступають самі компанії, що підтримують застосунок і вони мають повний інструментарій для роботи з їх децентралізованими програмами-токенами, включаючи інструменти торгу та обміну криптовалют. Що ж стосується платформи даних, проблема сплати комісій у будь-якій криптовалюті-токені має складну природу обміну, що не дозволяє точно оцінити комісію яку необхідно стягнути з акаунта. Як приклад, сплата комісії в токені X може бути відхилена делегатом A, тому що він просто не довіряє даному токену або не може його обміняти на іншу криптовалюту (немає обмінників, забороняє закон, тощо).

3. Універсальний підхід до делегування транзакцій

Усі вищезазначені методи делегування транзакцій мають свої переваги та недоліки, і не існує такого методу який був би позбавлений будь-яких недоліків. Однак, щоб зрозуміти який саме метод є найбільш застосовуваним і недопрацьованим, було проведено огляд найбільш поширених децентралізованих застосунків, що використовують підходи до делегації транзакцій у 2019 році на основі платформи Ethereum (здебільшого проекти, які успішно провели ICO (Initial Coin Offering): Binance, DreamTeam, Loom Network, Stratis, Maker DAO, OmiseGo, Basic Attention Token, 0x, Golem). З них можна виділити наступні тенденції децентралізованих застосунків, які є актуальними на даний момент:

- Майже кожен децентралізований застосунок використовує свою власну криптовалюту-токен для сплати за користування децентралізованими або частково децентралізованими сервісами [14].
- Децентралізовані застосунки призначені для вузької досвідченої публіки і здебільшого не використовують делеговані транзакції для спрощення користування сервісами (тобто, беруть оплату за послуги в токені і додатково вимагають сплати комісій у валюті децентралізованої

платформи даних) [15]. Непоширені і зазвичай приватні децентралізовані застосунки, які все ж таки використовують делегування транзакцій, мають дуже вузькоспеціалізовані системи для підтримки лише певних децентралізованих програм.

- Як наслідок із попереднього пункту, відсутність необхідності у відкритій екосистемі (ринку) для делегованих транзакцій. Акаунти-делегати, що використовуються в децентралізованих застосунках, як правило підтримуються самим розробником децентралізованого застосунку і не представляють фінансового інтересу для інших делегатів-організацій.
- Делегування транзакцій є сильно централізованим, що випливає з попереднього пункту.
- Відсутність загальноприйнятого стандарту або підходу до побудови децентралізованих застосунків, що використовують делегування транзакцій.

Таким чином, можна зробити висновок про те, що підходи до делегування транзакцій, в яких функції делегування вбудовані в саму децентралізовану програму-токен є найбільш поширеними серед існуючих децентралізованих програм, оскільки кожна організація намагається розробляти свої власні системи делегування транзакцій. Проблема відсутності єдиного стандарту чи підходу до побудови децентралізованих програм або застосунків, що використовують делегування транзакцій все ще залишається не вирішеною.

Розроблений універсальний підхід до делегування транзакцій, що описується в даній статті, пропонує вирішення проблеми стандартизації шляхом використання підходу до написання децентралізованих програм та їх обслуговуючої частини без необхідності у єдиному стандарті на рівні децентралізованих платформ даних. Іншими словами, даний підхід дозволяє використовувати єдину обслуговуючу систему для будь-яких децентралізованих програм з підтримкою делегування транзакцій, в тому числі і уже існуючих децентралізованих программ, незалежно від їх імплементації, що у випадку всіх перерахованих вище підходів до делегування було неможливим або потребувало побудови спеціальних обслуговуючих систем окремо для кожної імплементації.

Універсальний підхід до написання децентралізованих застосунків, описаний в даній статті, складається з наступних частин:

- Підхід до написання децентралізованої програми, в якому робота такої програми не обмежується та може імплементувати будь-який стандарт делегування транзакцій за вибором розробника.
- Розроблена допоміжна серверна частина для підтримки делегування транзакцій є універсальною, і підходить до будь-якого децентралізованого застосунку, що використовує будь-який підхід до делегування транзакцій.
- Розроблена допоміжна клієнтська частина у вигляді вбудовуваного віджету, який також є універсальним і може використовуватися як графічний користувацький інтерфейс для проведення делегованих транзакцій.

3.1. Підхід до написання децентралізованої програми

В основу універсального підходу до делегування транзакцій був покладений підхід без використання проміжних децентралізованих програм, тобто такий, в якому функції делегування вбудовані в децентралізовану програму-токен, який і використовується для можливої сплати комісій. На рисунку 3.1 нижче наведено схему взаємодії користувача із децентралізованою програмою без використання проміжних децентралізованих програм.

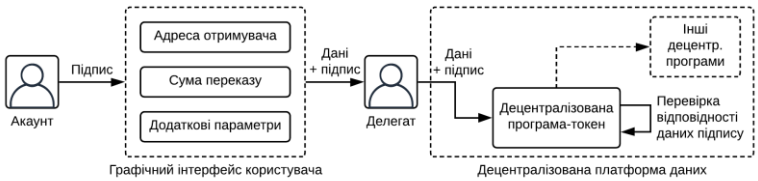


Рисунок 3.1 — взаємодія користувача з децентралізованою програмою з використанням підходу до делегування транзакцій без допоміжних децентралізованих програм

На схемі зображено наступні кроки виконання делегованої транзакції:

1. Користувач отримує підготовлені або вводить нові дані на графічному інтерфейсі користувача.
2. Користувач підписує електронно-цифровим підписом введені дані і додаткові парарметри для виконання делегованої транзакції (такі дані як додаткова комісія, гранична дата виконання, ідентифікатор транзакції, тощо).
3. Користувач передає підпис і дані делегату.

4. Делегат відправляє коштовну транзакцію в децентралізовану платформу даних, сплачуючи комісію.
5. Децентралізована програма-токен перевіряє відповідність підпису та переданих даних, таким чином гарантуючи їх непідробність, і проводить транзакцію.

Децентралізована програма-токен може бути написана з використанням будь-якого існуючого стандарту: наприклад, ERC20 або ERC721 [16]. Але варто зазначити, що для того, щоб до неї можна було застосувати універсальний підхід описаний нижче, вона має задовольняти наступним критеріям:

- Для кожної основної функції децентралізованої програми має бути передбачена аналогічна додаткова "делегована функція", що виконуватиме ту саму дію, але не буде залежати від викликаючого її акаунта (делегата). Іншими словами, якщо, наприклад, у децентралізованій програмі-токені присутня функція transfer, що дозволяє переводити криптовалюту з одного акаунта на інший, має бути додатково передбачена функція transferViaSignature (наприклад), що дозволить акаунту перевести криптовалюту за допомогою делегата. У якості альтернативи, в контракті може бути передбачена єдина функція для виклику всіх інших функцій незалежно від викликаючого акаунта (делегата).
- По можливості, делегована функція має підтримувати декілька існуючих стандартів підписів. Використання декількох стандартів замість одного підвищать шанси безпрецедентної роботи децентралізованого застосунку в майбутньому, адже з плином часу деякі стандарти підписів можуть бути оголошені як застарілі.
- Делегована функція має передбачати ряд необхідних додаткових параметрів для обмеження довільного її використання і надання додаткових гарантій безпеки акаунту, що має намір делегувати транзакцію. Рекомендованими параметрами вважаються: ідентифікатор транзакції; гранична дата виконання транзакції; акаунт, що отримує винагороду. Всі додаткові параметри мають бути включені до електронно-цифрового підпису та перевірені самою децентралізованою програмою.
- Додатково до стандарту децентралізованої програми-токену може бути передбачена функція виклику інших децентралізованих програм, і відповідна їй функція для виклику через делегата.

Роздивимось рекомендований підхід до написання делегованих функцій на прикладі децентралізованої платформи даних Ethereum, а саме на прикладі децентралізованої програми-токена за стандартом ERC20 і функції transfer, передбаченої даним стандартом. Даний підхід спрощує використання децентралізованого застосунку шляхом усунення додаткової криптовалюти, яку необхідно сплачувати для проведення транзакцій майже в будь-якій децентралізованій платформі даних. Варто зазначити, що даний підхід можна застосувати до будь-якої самозахисної децентралізованої платформи даних, якщо самою платформою не передбачено інший варіант проведення делегованих транзакцій.

Функція transfer в стандарті ERC20 призначена для переказу певної кількості токенів (програмованої криптовалюти) з акаунта користувача на певний інший акаунт. Припустимо тепер, що додатково буде імплементована функція transferViaSignature, яка виконуватиме ту саму роботу що і функція transfer, але дозволятиме делегувати її виклик будь-якому іншому акаунту. Тобто, на відміну від функції transfer, викликати функцію transferViaSignature може будь-який акаунт, а її робота буде аналогічна функції transfer по відношенню до іншого акаунта. На рисунку 3.2 показано, як саме відбувається виклик функцій користувачем та переказ токенів у випадках функцій transfer та transferViaSignature. Варто зазначити, що функцією transferViaSignature передбачено сплату комісії акаунтом делегату за його роботу.

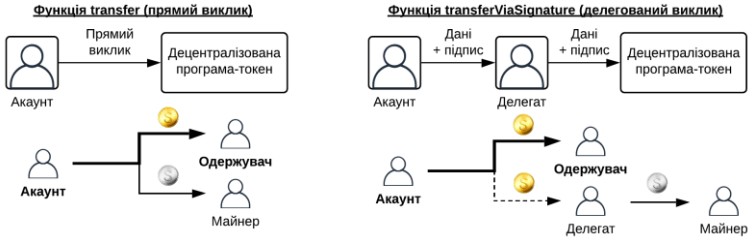


Рисунок 3.2 — схема викликів функцій transfer та transferViaSignature та результатів функції

Як видно зі схеми, зображеної на рисунку 3.2, в делегованому виклику комісія за транзакцію сплачується у токені (криптовалюті) самим акаунтом, який делегував свою транзакцію делегату, а

комісію передбачену самозахищеною децентралізованою платформою даних сплачує делегат. Таким чином усувається необхідність для акаунта тримати декілька криптовалют (у даному випадку, в децентралізованій платформі даних Ethereum сплачується комісія криптовалютою Ether). Варто зазначити, що комісія делегату може і зовсім не сплачуватися, але вона показана на схемі як приклад збалансованої економічної моделі, де делегат отримує певну винагороду за делегування транзакцій.

Функція transferViaSignature приймає на вхід не тільки такі самі параметри, що і функція transfer, а і ряд додаткових параметрів, які гарантують захищене делегування для функції transfer. На рисунку 3.3 показано рекомендовані сигнатури функцій, які необхідно імплементувати в децентралізованій програмі-токені.

```
contract ERC20 {
  function transfer(to, value) external;
  function transferViaSignature(from, to, value, fee, feeRecipient, deadline, sigId, sig, sigStd) external;
}
```

Рисунок 3.3 — рекомендовані сигнатури функцій децентралізованої програми-токену

В той час як функція transfer приймає тільки два параметра, to (одержувач) і value (сума переказу), функція transferViaSignature додатково приймає наступні параметри:

- from — акаунт-відправник (необов'язковий параметр). На відміну від функції transfer, акаунт-відправник не є акаунтом-ініціатором транзакції, а отже його необхідно додатково предати, або отримати із електронно-цифрового підпису у самій децентралізованій програмі.
- fee — сума комісії, яку сплачує акаунт-відправник акаунту-делегату.
- feeRecipient — адреса акаунта делегата-отримувача комісії від акаунта-відправника. Не рекомендується прибирати цей параметр або використовувати в якості адреси акаунта делегата адресу того, хто відправив транзакцію, адже в багатьох платформах транзакцію можна клонувати і виконати швидше, ніж буде виконана оригінальна транзакція (race condition).
- deadline — гранична дата можливості проведення делегованої транзакції. Не рекомендується прибирати цей параметр. За допомогою цього параметру, у випадку втрати підписаних даних або підпису за будь-яких причин, користувач може бути впевнений що його транзакцію ніхто не зможе провести пізніше.
- sigId — додатковий унікальний ідентифікатор транзакції (необов'язковий параметр). Може бути замінений самим підписом як унікальним ідентифікатором.
- sig — електронно-цифровий підпис акаунта-відправника, яким було підписано всі параметри даної функції, тим самим засвідчуючи свою конкретну дію, яку акаунт-відправник має намір виконати.
- sigStd — стандарт підпису у випадку використання декількох стандартів підпису в децентралізованій програмі-токені (рекомендується).

Отже, за допомогою додаткових параметрів, їх підпису і подальшої верифікації в децентралізованій програмі обмежуються можливості делегата будь-яким чином маніпулювати проведенням делегованої транзакції. Таким чином, користувач може бути впевнений у виконанні або невиконанні чітко визначеної делегованої транзакції.

Варто зазначити, що впровадження делегування транзакцій майже завжди є частковою централізацією децентралізованого застосунка, оскільки делегатом виступають лише певні автоматизовані обслуговуючі програми. Але варто роздивлятися дану централізацію виключно з ціллю спростити децентралізований застосунок для користувача. Саме для того, щоб він залишався децентралізованим даним підходом передбачено написання делегованих функцій як додаткових, а не як основних. У випадку коли система делегування транзакцій не працює, користувач все ще може виконати необхідну транзакцію, придбавши криптовалюту самозахисної децентралізованої платформи даних.

3.2. Підхід до централізованої допоміжної серверної частини

Окрім написання самої децентралізованої програми, що буде підтримувати виконання делегованих транзакцій, важливо забезпечити автоматизовану роботу самих делегатів, що будуть проводити транзакції в децентралізованій платформі даних. Таким чином, постає задача створення автоматизованої серверної частини, що виконувала б наступні функції:

- Підтримка проведення делегованих транзакцій, в тому числі забезпечення відправки транзакції в мережу і її збереження в мережі (майнингу).

- Підготовка делегованих транзакцій: розрахунок доцільної комісії для сплати користувачем, визначення порядку транзакції у децентралізованій мережі, тощо.
- Валідація та перевірка клієнтських даних.
- Повідомлення результатів клієнту.

Була поставлена задача розробити таку допоміжну серверну частину, яка б виконувала всі вище зазначені функції, і однаково підходила для будь-яких систем децентралізованих платформ даних і децентралізованих програм. Дана технічна задача є нетривіальною, оскільки необхідно передбачити наступне:

- Захищеність централізованої системи від будь-яких атак: спам-атак, дублювання транзакцій, зламу, race condition, тощо.
- Послідовне виконання транзакцій при паралельних запитах (для деяких децентралізованих платформ даних).
- Визначення та керування параметрами транзакцій (необхідно брати завжди вигідну комісію за транзакції, перераховувати її відносно ціни токену на біржі, тощо).
- Обслуговування багатьох клієнтів одночасно.
- Універсальність допоміжної серверної системи, тобто можливість її використання для будь-яких децентралізованих програм з підтримкою делегування транзакцій.

Розглянемо необхідну автоматичну роботу, яку має виконувати допоміжна серверна система на прикладі децентралізованої платформи даних Ethereum. Взаємодія компонентів системи наведена на рисунку 3.4.

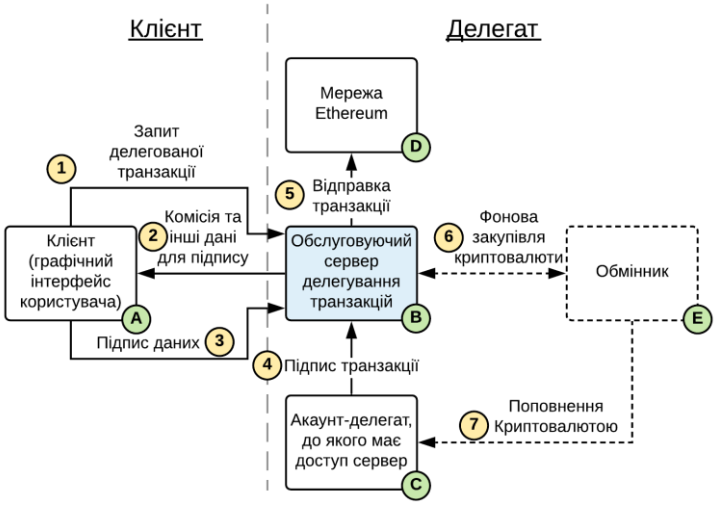


Рисунок 3.4 — схема взаємодії компонентів системи делегування транзакцій з допоміжною серверною частиною. Цифрами позначено порядок взаємодії компонентів системи.

Допоміжна система для делегування транзакцій включає в себе роботу п'яти компонентів, позначених латинськими літерами на рисунку 3.4. Серед них виділено наступні компоненти:

A. Клієнт (графічний інтерфейс користувача) — безпосередньо графічний інтерфейс, з яким взаємодіє користувач.
B. Обслуговуючий сервер делегування транзакцій, яких може бути декілька.
C. Акаунт-делегат або криптогаманець/система, що зберігає приватний ключ і підписує транзакції для їх подальшої відправки в децентралізовану мережу.
D. Децентралізована мережа, підтримувана самозахисною децентралізованою платформою даних.
E. Можливий обмінник для проведення автоматичних криптовалютообмінних операцій з боку обслуговуючого сервера делегування транзакцій для придбання криптовалюти, необхідної для сплати комісій в децентралізованій платформі даних.

Варто зазначити, що децентралізованим застосунком може бути передбачена інша конфігурація, наприклад, з відсутністю обмінника, якщо в цьому немає потреби (наприклад, при приватному делегуванні транзакцій, коли з користувачів комісія не стягується). Публічне делегування транзакцій відрізняється від приватного або обмеженного тим, що сервіс делегування транзакцій знаходиться в необмеженому публічному доступі, і за кожну транзакцію делегату (допоміжній серверній частині) необхідно стягувати комісію в токенах з користувача для того щоб не збанкрутіти, оскільки делегат постійно витрачає кошти на проведення транзакцій в децентралізованій мережі.

Дана комісія (або її частина) буде використана делегатом для обміну на криптовалюту, якою буде сплачуватися транзакція в децентралізованій мережі.

3.3. Взаємодія користувача з системою

Розглянемо порядок взаємодії користувача з системою і порядок взаємодії компонентів системи, а також обґрунтуємо чому саме такий порядок має використовуватися при публічному делегуванні транзакцій, так само як може використовуватися у приватному або обмеженому делегуванні транзакцій. Порядок взаємодії компонентів системи також продемонстровано на рисунку 3.4, а діаграма послідовності взаємодії користувача із делегатом, так само як і делегата з децентралізованою мережею, показана на рисунку 3.5.

Взаємодія користувача з делегатом передбачає наступні автоматизовані кроки:

1. Запит делегованої транзакції. Визначившись яку саме конкретну дію має на меті виконати користувач, клієнт підготовлює параметри транзакції. Необхідно отримати схвалення від допоміжного сервісу делегування транзакцій про те, що дана транзакція ним може бути проведена.

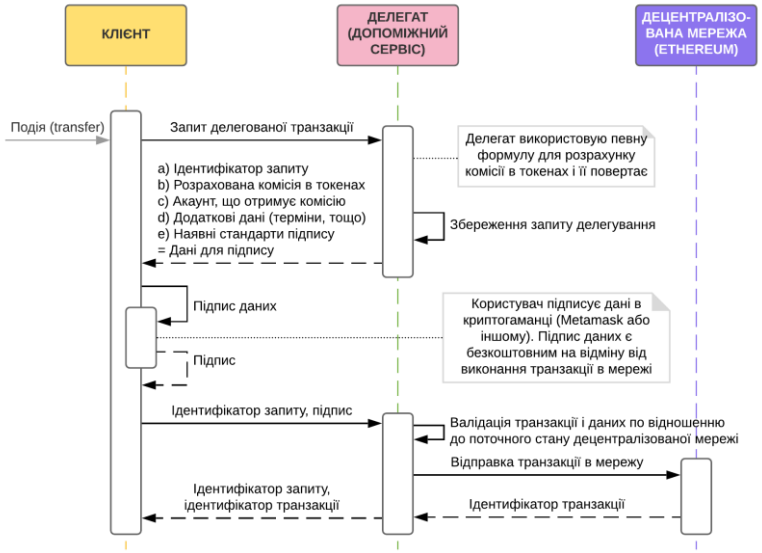


Рисунок 3.5 — діаграма послідовності взаємодії користувача з делегатом і децентралізованою мережею

2. Відповідь делегата. Використовуючи розроблений протокол для REST API, клієнт виконує такий запит, а делегат у відповіді повідомляє клієнта про:
    a. Можливість або неможливість виконати дану конкретну транзакцію.
    b. Можливу комісію, яку буде сплачена користувачем в токені (замість криптовалюти). При цьому делегат її самостійно розраховує, використовуючи підготовлений алгоритм розрахунку вартості комісії.
    c. Дані, які необхідно підписати користувачу, передбачені децентралізованою програмою. Варто зазначити, що якщо у децентралізованій програмі імплементовано декілька стандартів підпису, делегат повертає декілька варіантів підпису і клієнт має змогу вибрати з них той, який його найбільш задовольняє.
    d. Додаткова інформація (ідентифікатор делегованої транзакції, орієнтовний час проведення транзакції, тощо).
3. Підпис даних та підтвердження транзакції. Отримавши дані для підпису та мета-інформацію, клієнт приймає рішення про доцільність виконання транзакції і підписує дані (параметри делегованої функції) якщо його все влаштовує. Однак, наприклад, делегат може запросити занадто високу комісію за проведення делегованої транзакції, а тому рішення клієнта про її проведення залишається відкритим. У разі взаємної згоди делегата і клієнта, підписані дані відправляються делегату як підтвердження виконання делегованої транзакції.
4. Підпис транзакції. Делегат, в свою чергу, виконує підпис транзакції, яка відправляється в децентралізовану мережу. В даній транзакції зібрані:
    a. Параметри, які було надано користувачем для транзакції (з рисунку 3.3 — наприклад, для функції *transfer* це *to* та *value*).
    b. Додаткові параметри для децентралізованої програми, які було визначено делегатом (з рисунку 3.3 — fee, feeRecipient, sigId, deadline).
    c. Підпис всіх параметрів користувачем (з рисунку 3.3 — sigStd, sig).

d. Визначення порядкового номеру транзакції та актуальної комісії в мережі і підпис транзакції акаунтом делегата, який сплачує комісію децентралізованої мережі за дану транзакцію.

5. Відправка транзакції в децентралізовану мережу. Тільки після всіх вище перерахованих кроків транзакція вважається дійсною і буде прийнята мережею як правильна. Через певний короткий час транзакцію буде фіналізовано і делегат зможе повідомити клієнт про її успіх або помилку. При відправці підтверджуючого запиту до делегата, клієнт через розроблений протокол для REST API у відповідь отримує або помилку, або ідентифікатор транзакції в децентралізованій мережі.
6. Отримання результату. Під час очікування виконання делегованої транзакції, клієнт постійно запитує сервер про наявні зміни в статусі транзакції, і коли сервер відповідає позитивно — делегована транзакція вважається виконаною. Клієнт також може не залежати в даній ситуації від делегата, а самостійно перевіряти статус транзакції в децентралізованій мережі використовуючи ідентифікатор транзакції, який йому надав делегат у відповіді на підтвердження делегованої транзакції.

Таким чином, майже всю взаємодію користувача з системою при використанні делегування транзакцій вдається автоматизувати, за виключенням необхідного підпису користувача. Саме своїм електронно-цифровим підписом користувач засвідчує, що він погоджується з конкретними умовами делегування транзакції, які йому виставив делегат і водночас робить ці умови незмінними, оскільки делегат не може підробити електронно-цифровий підпис. Наданий користувачем підпис валідується в децентралізованій програмі та дозволяє виконання передбачених децентралізованою програмою дій.

3.4. Про застосування універсального підходу до делегування транзакцій в СППР

Системи підтримки прийняття рішень (СППР) — це системи, які беруть факти та дають рекомендації для осіб, які приймають рішення, спираючись на численні фактори. Ці рекомендації, як правило, використовуються лише в інформаційних цілях, однак вихід системи може вважатися основним висновком для багатьох складних операцій, з якими люди зазвичай не можуть впоратися самостійно (приклади: дослідження поведінки соціальних груп, складна мережа фактів та стосунків, стратегічне планування тощо) [17, 18].

Однією з ключових особливостей систем підтримки прийняття рішень є робота з експертами, під час якої експерти надають усі необхідні для системи дані [19], на базі яких в подальшому системою формулюються висновки і рекомендації. Підсистема, яка призначена для роботи з експертами, які здебільшого є звичайними користувачами, нерідко зберігає даних у незахищених або централізованих реєстрах, що підвищує ризик їх компрометації, підробки, несанкціонованого видалення, тощо.

Тому в першу чергу, важливо переконатися, що в обробку системою поступлять достовірні вхідні дані. Достовірність вхідних даних становить довіру до всієї системи підтримки прийняття рішень, так само як і до організації, що її підтримує. Тому одним із актуальних та необхідних векторів використання розробленої в статті системи підтримки делегування транзакцій, як і самозахисних децентралізованих платформ в цілому є підсистема роботи з експертними оцінками системи підтримки прийняття рішень.

Таким чином, експертам системи підтримки прийняття рішень буде запропоновано створення акаунта в самозахисній децентралізованій платформі даних замість реєстрації в системі через пару логін/пароль, що за складністю відрізняється лише встановленням додаткового програмного забезпечення у будь-якому актуальному інтернет-браузері. Подальша робота експертів з децентралізованою системою принципово не буде відрізнятись від централізованої, але дані, які будуть вводитись експертами тепер будуть захищені самою децентралізованою платформою даних. Це надає даним гарантію незмінності, або модифікації за строго визначеними правилами, попередньо прописаними у децентралізованій програмі. В подальшому, дані експертів, що тепер без виключень вважатимуться достовірними (окрім компрометації самих експертів), будуть завантажуватися з децентралізованої платформи даних в систему підтримки прийняття рішень для подальшої їх обробки та отримання рекомендацій системи.

Висновки

Існує багато підходів до делегування транзакцій в самозахисних децентралізованих платформах даних, але лише деякі з них задовольняють необхідним умовам безпеки і простоти використання. Розроблений універсальний підхід до делегування транзакцій, який комбінує в собі всі переваги розглянутих наявних підходів, має стандартизовану підтримуючу серверну і клієнтську частини, і в той же час не вимагає стандартизації децентралізованого застосунку, тобто може бути використаний для будь-яких існуючих або нових систем, що мають на меті спрощення роботи з децентралізованими застосунками шляхом впровадження системи делегування транзакцій.

Розроблений підхід та програмне забезпечення для підтримки делегування транзакцій в децентралізованих платформах даних сприятиме подальшій адаптації децентралізованих застосунків у всіх сферах життя та технології, оскільки даний підхід спрощує як роботу користувача з системою, так і необхідні зусилля розробників для впровадження системи делегування транзакцій в своїх проектах.

Робота універсального підходу до делегування транзакцій була протестована на проекті DreamTeam.gg, який має близько 2000 користувачів децентралізованого застосунку (кількість користувачів апроксимована виходячи з кількості власників токену проекта) та розгорнута за інтернет-адресою send-token.dreamteam.gg. Даний інтернет-сервіс є рішенням з відкритим початковим кодом, який може бути адаптований будь-яким іншим проектом на базі самозахисної децентралізованої платформи даних Ethereum.

Список літератури

1. Savchenko M., Tsyganok V., Andriichuk O., 2018. Decision Support Systems' Security Model Based on Decentralized Data Platforms / Selected Papers of the XVIII International Scientific and Practical Conference "Information Technologies and Security" (ITS 2018), pp. 199–208.
2. Mookherjee D., 2006. Decentralization, hierarchies, and incentives: A mechanism design perspective. Journal of Economic Literature, 44(2), pp.367-390.
3. Raval S., 2016. Decentralized Applications: Harnessing Bitcoin's Blockchain Technology. O'Reilly Media, Inc. pp.6-7.
4. Khovratovich D. and Law J., 2017, April. BIP32-Ed25519: Hierarchical Deterministic Keys over a Non-linear Keyspace. In 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) pp. 27-31.
5. Arapinis M., Gkaniatsou A., Karakostas D. and Kiayias A., 2019. A Formal Treatment of Hardware Wallets. IACR Cryptology ePrint Archive, p.34.
6. Pustišek M. and Kos A., 2018. Approaches to front-end iot application development for the ethereum blockchain. Procedia Computer Science, 129, pp.410-419.
7. Rezaeighaleh H. and Zou C.C., 2019. New Secure Approach to Backup Cryptocurrency Wallets. In submission to IEEE Global Communications Conference-Communication & Information Systems Security Symposium.
8. Cho S., Park S.Y. and Lee S.R., 2017. Blockchain consensus rule based dynamic blind voting for non-dependency transaction. International Journal of Grid and Distributed Computing, 10(12), pp.93-106.
9. Ouaddah A., Elkalam A.A. and Ouahman, A.A., 2017. Towards a novel privacy-preserving access control model based on blockchain technology in IoT. In Europe and MENA Cooperation Advances in Information and Communication Technologies. Springer, Cham. pp. 523-533.
10. Mulders M., A comparison between ERC20, ERC223, and ERC777 token standard [Electronic resource]/Michiel Mulders. Mode of access: https://www. cointelligence. com/content/comparison-erc20-erc223-new-ethereum-erc777-token-standard.–Title from the screen.
11. Chavan A.B. and Rajeswari K., 2019, July. Design and Development of Self-sovereign Identity Using Ethereum Blockchain. In International Conference on Sustainable Communication Networks and Application. Springer, Cham. pp. 523-531.
12. Ludovic Galabro. ERC865: Pay Transfers in Tokens Instead of Gas, in One Transaction #865. https://github.com/ethereum/EIPs/issues/865, [Last accessed Dec 21, 2019].
13. Wels S.J., 2019. Guaranteed-TX: The exploration of a guaranteed cross-shard transaction execution protocol for Ethereum 2.0 (Master's thesis, University of Twente).
14. Lee J.Y., 2019. A decentralized token economy: How blockchain and cryptocurrency can revolutionize business. Business Horizons, 62(6), pp.773-784.
15. Iansiti M. and Lakhani K.R., 2017. The truth about blockchain. Harvard Business Review, 95(1), pp.118-127.
16. Kim M., Hilton B., Burks Z. and Reyes J., 2018, November. Integrating Blockchain, Smart Contract-Tokens, and IoT to Design a Food Traceability Solution. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 335-340.
17. Saaty T. L., 2010. Principia Mathematica Decernendi - Mathematical principles of decision making - Generalization of the Analytic Network Process to neural firing and synthesis. Pittsburg: RWS Publications.
18. Tsyganok V., Kadenko S., Andriychuk O., Roik P., 2017. Usage of multicriteria decision-making support arsenal for strategic planning in environmental protection sphere / Journal of Multi-Criteria Decision Analysis. 24, pp.227-238.
19. Totsenko V.G., Tsyganok V.V., 1999. Method of paired comparisons using feedback with expert / Journal of Automation and Information Sciences. Volume 31, Issue 7-9, pp.86-96.

СПОСОБИ ПРОТИДІЇ ВИКОРИСТАННЮ СОЦІАЛЬНОЇ ІНЖЕНЕРІЇ

Оксана Цуркан [0000-0002-5524-8834], Ростислав Герасимов [0000-0002-4115-8344],
Ольга Крук [0000-0003-2994-6804]

Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова
Національної академії наук України, Київ-164, Україна
o.tsurkan24@gmail.com, gerasimov.rostislav@gmail.com, o.n.kruk@gmail.com

Анотація. Розглянуто особливості протидії використанню соціальної інженерії в кіберпросторі. Виокремлено типові способи її унеможливлення. Серед них приділено увагу тестуванню на проникнення. Цей спосіб автоматизованого виявлення уразливостей людини реалізується як інструментальний засіб SET. SET може використовуватися як окремий комплекс, так і може входити до Kali Linux. Окремо приділено увагу фреймворку протидії використанню соціальної інженерії. На відміну від SET, застосовується для підвищення обізнаності персоналу. Це досягається шляхом його навчання, удосконалення технологій і політик протидії. Крім цього розглянуто спосіб протидії використанню соціальної інженерії завдяки виокремленню об'єкта (персоналу, захисника) та суб'єкта (зловмисника) такого впливу. При цьому за основу береться припущення про відповідність дій зловмисника методам використання соціальної інженерії. Особливістю такого способу є можливість протидіяти соціоінженерному впливу вироблення протидій на кожну дію зловмисника. Таким чином, встановлено багатоаспектність способів протидії використанню соціальної інженерії. В кінцевому випадку це дозволить врахувати переваги та недоліки типових способів унеможливлення реалізації загроз соціоінженерного впливу.

Ключові слова: соціальна інженерія, використання соціальної інженерії, спосіб протидії, тестування на проникнення, обізнаність персоналу.

Використання соціальної інженерії в кіберпросторі полягає в маніпулятивному впливові на свідомість (підсвідомість) людини через властиві їй уразливості. Це одна з найбільш небезпечних загроз кібербезпеці [1], реалізація якої може призвести до значних збитків.

Протидія використанню соціальної інженерії здійснюється завдяки таким способам як, наприклад [2, 3], автоматизування виявлення уразливостей людини, підвищення обізнаності персоналу, виокремлення суб'єктів і об'єктів соціоінженерного впливу.

Протидія соціальній інженерії здійснюється завдяки автоматизуванню виявлення уразливостей людини (наприклад, працівника, клієнта). Для цього використовуються відповідні інструментальні засоби [2–3], наприклад, Social-Engineer Toolkit, SET; Social Engineering Defensive Framework, SEDF. Кожен зі зазначених засобів орієнтований на реалізацію загроз соціальній інженерії. Так, SET здебільшого використовується при проведенні тестування на проникнення за соціоінженерним підходом [3]. Розглядається як компонент Kali Linux.

Іншим прикладом інструментальних засобів є фреймворк протидії використанню соціальної інженерії (Social Engineering Defensive Framework, SEDF). Протидія досягається шляхом визначення впливу, оцінювання захисту, навчання працівників, удосконаленням існуючих технологій і політик [3].

Водночас може використовуватися два аспекти протидії: з боку суб'єкта (зловмисника), з боку об'єкта (захисника) соціоінженерного впливу [2]. Цей спосіб дозволяє протидіяти використанню соціальної інженерії завдяки розгляданню вірогідних дій з боку зловмисника. При цьому вважається, що послідовність його дій визначається виключно з огляду на методи використання соціальної інженерії. Такий спосіб дозволяє кожній дії протиставити протидію і, я наслідок, унеможливити реалізації загроз використання соціальної інженерії.

Таким чином, протидія використанню соціальної інженерії характеризується багатоаспектністю. По-перше, найбільш розповсюджений спосіб орієнтований на автоматизування виявлення уразливостей людини при тестуванні на проникнення. Тоді як, по-друге, не менш важливим є підвищення рівня обізнаності персоналу. І, по-третє, виокремлення суб'єкта та об'єкта соціоінженерного впливу для проставлення їх дій один одному.

В кінцевому випадку це дозволяє врахувати типові способи протидії використанню соціальної інженерії і, як наслідок, розробити відповідні моделі та методи з урахуванням їх переваг і недоліків.

Список використаної літератури

1. Мохор, В., Богданов, А., Килевой, А. Наставления по кибербезопасности (ISO/IEC 27032:2013). – ООО «Три-К», Киев, 2013.
2. Fathollahi-Fard Mohammad Amir, Hajiaghaei-Keshteli Mostafa, Tavakkoli-Moghaddam Reza. The Social Engineering Optimizer (SEO). Engineering Applications of Artificial Intelligence, 2018. – Vol. 72. – pp. 267-293. doi: 10.1016/j.engappai.2018.04.009
3. Mokhor V.V., Tsurkan O.V., Herasymov R.P., Tsurkan V.V. Information Security Assessment of Computer Systems by Socio-engineering Approach. Selected Papers of the XVII International Scientific and Practical Conference «Information Technologies and Security». Kyiv, 2017. - pp. 92-98. URL: http://ceur-ws.org/Vol-2067/paper13.pdf.

ПІДХІД ДО ПРОГНОЗУВАННЯ ДІЄВОСТІ ДЕРЖАВНОГО УПРАВЛІННЯ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ OSINT

Додонов О.Г.[1], Ланде Д.В.[1,2], Нестеренко О.В.[2], Березін Б.О.[1]

[1]*Інститут проблем реєстрації інформації НАН України*

[2]*Національний технічний університет України "КПІ ім. Ігоря Сікорського"*

1. Постановка проблеми

В період проведення реформ дієвість державного управління країною набуває особливого значення. Одним з важливих елементів державного управління є ефективність законопроектів, що приймаються. Ефективність приймаємих законопроектів та державного управління в цілому находить відображення у суспільній думці відносно законопроектів, показником якої можна вважати кількість новинних публікацій відповідної тематики в інформаційному просторі. На сьогодні, державне управління є однією з важливих сфер застосування технологій аналізу відкритих джерел OSINT. При цьому, прогнозування належить до числа невід'ємних складових системи OSINT [1,2].

В даній роботі розглянуто існуючи засоби прогнозування часових рядів кількості тематичних публікацій та запропановано підхід до прогнозування дієвості державного управління з використанням технологій OSINT.

2. Аналіз прогнозування дієвості державного управління з використанням існуючих засобів

Серед існуючих засобів прогнозування часових рядів, які можуть бути використані для оцінки дієвості державного управління, розглядалися відповідні засоби статистичних пакетів, бібліотек, що входять до складу сучасних мов програмування, а також відповідні онлайн сервіси. З цією метою, за допомогою систем моніторингу інформаційних ресурсів Інтернет було виконано збір даних за тематикою найбільш резонансних у суспільстві законопроектів та законів: про забезпечення функціонування української мови як державної, про розмитнення автомобілів на іноземній реєстрації, про Вищий антикорупційний суд, та деяких інших. Для отриманих часових рядів кількості публікацій відносно резонансних законопроектів було виконано прогнозування за допомогою відповідних бібліотек мови програмування *R* для статистичних задач та за допомогою засобів пакету Excel.

Серед популярних засобів прогнозування часових рядів відзначають моделі ETS, ARIMA, пакет Prophet та інші [3,4], реалізовані в різних мовах програмування та статистичних пакетах. В основі моделей ETS (Exponential Smoothing) лежить експоненціальне згладжування - метод прогнозування, при якому значення змінної за всі попередні періоди входять в прогноз, експоненціально втрачаючи свою вагу з часом. Це дозволяє моделі з достатнім ступенем гнучко реагувати на новітні зміни в даних, зберігаючи при цьому інформацію про історичну поведінку часового ряду.

Порівняно новий пакет Prophet (розробка 2017 р.), на відміну від раніше існуючих, повинен зменшити трудоємність налагодження параметрів при проведенні прогнозування. В його основі лежить процедура підгонки адитивних регресійних моделей з наступними чотирма основними компонентами: тренд (моделюється за допомогою кусочної лінійної регресії або кусочної логістичної кривої зростання); річна сезонність (моделюється як ряд Фур'є); тижнева сезонність (моделюється з використанням індикаторних змінних); "свята" (наприклад, офіційні святкові та вихідні дні). Оцінювання параметрів підгоняємої моделі виконується з використанням принципів байєсівської статистики. Пакет Prophet реалізовано для мов програмування *R* та Python.

В якості вихідних даних для прогнозування кількості публікацій по тематикам приймаємих законопроектів та законів, використано дані моніторингу ресурсів Інтернет з відповідними запитами

за період квітень-вересень 2019 р. (закон про забезпечення функціонування української мови як державної) та листопад 2018 р. – вересень 2019 р. (закон про розмитнення автомобілів на іноземній реєстрації). Для зменшення дисперсії вихідних даних було використано логарифмування. Далі, за допомогою методів ETS, Prophet, а також моделі Сорнетте, на основі даних моніторингу було отримано прогнозні значення логарифму кількості відповідних публікацій. На рис. 1 - 3 наведено вихідний часовий ряд кількості публікацій відносно закону про забезпечення функціонування української мови як державної. На осі абсцис графиків показано номери днів часового ряду, а на осі ординат – значення логарифму кількості публікацій.

На перші декілька днів часового ряду приходиться пік значення логарифму кількості публікацій, пов'язаний з голосуванням за прийняття закону (3-й день часового ряду). Спадаючий характер подальшої частини ряда пов'язан із зменшенням реакції на цю подію. Перша частина вихідного часового ряду (1 - 134 дні) використовується для навчання моделі прогнозування, друга частина (135 – 167 дні, приблизно 20% довжини ряду) для оцінки точності прогнозування. Отриманий прогноз в цій частині часового ряду порівнюється з вихідними даними (Таблиця 1).

На рис. 1 наведено графік часового ряду логарифму кількості публікацій відносно закону про забезпечення функціонування української мови як державної та графік відповідного прогнозованого часового ряду (пунктирна лінія), отриманого за допомогою метода ETS.

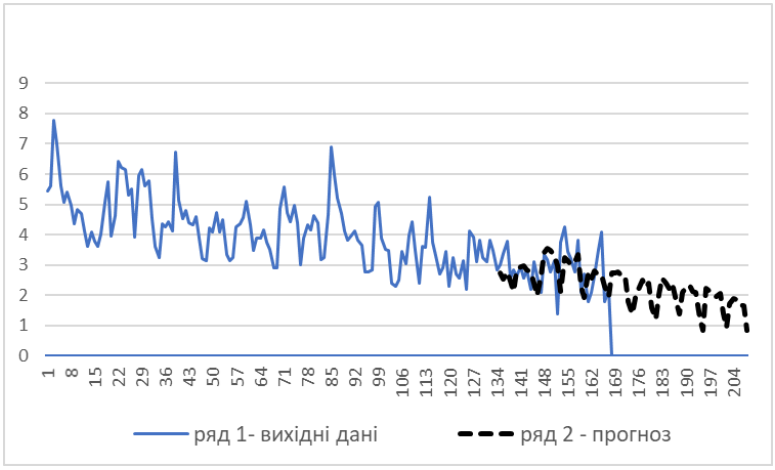


Рис. 1. Прогноз логарифму кількості публікацій відносно закону про забезпечення функціонування української мови як державної, отриманий за допомогою функції ETS. На осі абсцис показано номери днів часового ряду, а на осі ординат – значення логарифму кількості публікацій.

3. Підхід до прогнозування кількості публікацій за допомогою моделі Сорнетте

Серед публікацій, присвячених передбаченню фінансових криз відзначають роботи Д. Сорнетте [5]. Його метод заснований на аналізі закономірності руху ринкових цін на товарних і фондових ринках перед крахом. В роботі відзначається, що перед крахом ціна має степеневе зростання, ускладнене логоперіодичними коливаннями, які сходяться до нескінченності в критичній точці, де ймовірність краху досягає максимальної величини. Для прогнозування можливого краху пропонується аналізувати ринкові дані (часові ряди) з метою виявлення в них коливальних степеневих закономірностей. Для виявлення краху на фінансових ринках, степенева модель, що враховує лінійні логоперіодичні коливання, має слідуючий вигляд (1).

$$F(t) = A + B(t_c - t)^m \left[1 + C\cos\left(\omega \log\left(\frac{t_c - t}{T}\right) + \varphi\right)\right] \qquad (1)$$

У цій моделі t_c – критичний час (час краху на фінансовому ринку). Коефіцієнти моделі визначаються за допомогою процедури підбору. В роботі [6] розглянуто можливості передбачення криз на фінансовому ринку за допомогою методу Д. Сорнетте.

За допомогою використання моделі Сорнетте, на основі даних моніторингу було отримано прогнозні значення логарифму кількості відповідних публікацій. Для визначення коефіцієнтів моделі Сорнетте використовувались функції та пакети для підбору рішень у складі мови програмування *R* та пакету Excel (функція nls – розраховує параметри нелінійної регресійної моделі за методом найменших квадратів; функція solver для пошуку рішень нелінійних та лінійних задач та інші). Крім того, для визначення коефіцієнтів моделі Сорнетте було використано програму на мові Python [7]. На рис. 2 наведено графік часового ряду логарифму кількості публікацій відносно закону про

забезпечення функціонування української мови як державної та графік відповідного прогнозованого часового ряду (пунктирна лінія), отриманий за допомогою використання моделі Сорнетте.

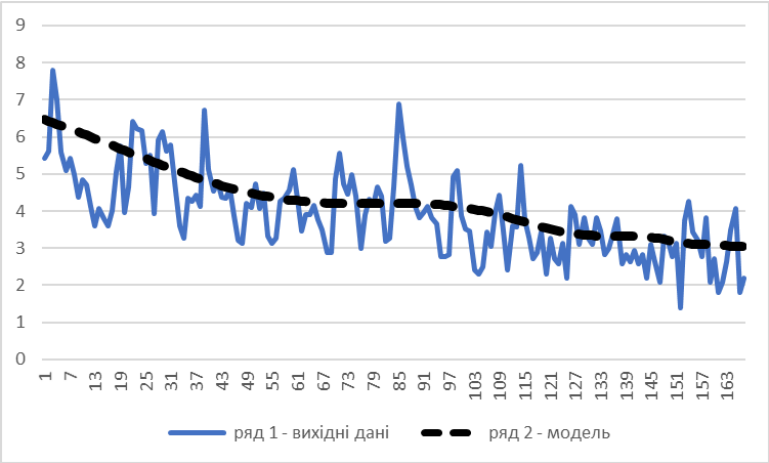


Рис. 2. Прогноз логарифму кількості публікацій відносно закону про забезпечення функціонування української мови як державної, отриманий за допомогою використання моделі Сорнетте. На осі абсцис показано номери днів часового ряду, а на осі ординат – значення логарифму кількості публікацій.

Результати прогнозування кількості публікацій за допомогою вище розглянутих моделей оцінювались на основі відомої метрики, в якій порівнюються відповідні значення вихідного та прогнозованого часових рядів (2).

$$\chi = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}(t) - y(t))^2} \qquad (2)$$

Для використання метрики у вихідних часових рядах було виділено 20% значень, які порівнювались із відповідними значеннями, отриманими шляхом прогнозування. Результати оцінки для різних засобів прогнозування та двох наборів даних про публікації наведено в Таблиці 1.

Таблиця 1. Оцінка точності прогнозування для різних засобів та вихідних даних.

Засоби, використані для прогнозування	Дані моніторингу про публікації:	
	відносно законів про забезпечення функціонування української мови як державної	відносно законів про розмитнення автомобілів на іноземній реєстрації
Функція ETS	0.83	1.39
Пакет Prophet	0.87	1.45
Модель Сорнетте	0.78	1.33

Для підвищення точності прогнозу з використанням моделі Сорнетте, пропонується враховувати сезонність аналізуємих явищ. У випадку кількості публікацій відносно законопроектів доцільно враховувати вихідні та святкові дні, коли активність публікацій знижується. З цією метою було розраховано сінусоїду, амплітуда та зсув якої підбиралися таким чином, щоб при накладанні її на модель, отриману у відповідності з виразом (1), значення метрики (2) зменшувалось. В розглянутих варіантах параметрів, зменшення метрики (2) складало від одного до кількох відсотків. На рис. 3 наведено графік часового ряду логарифму кількості публікацій відносно закону про забезпечення функціонування української мови як державної та графік відповідного прогнозованого часового ряду (пунктирна лінія), отриманий за допомогою використання моделі Сорнетте з врахуванням фактору сезонності -вихідних днів.

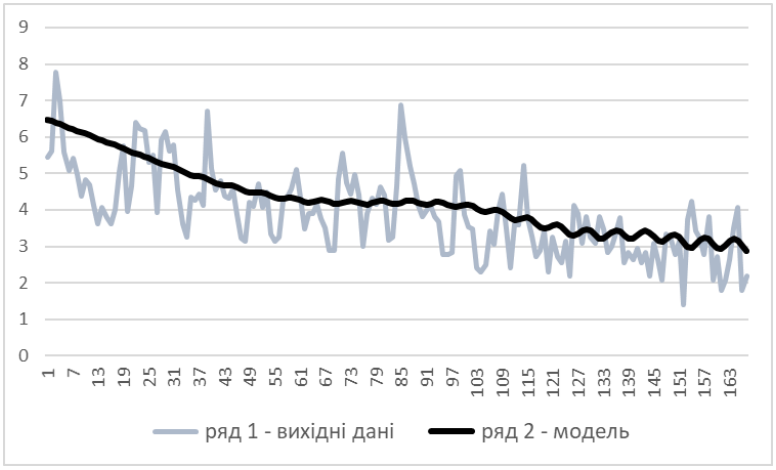


Рис. 3. Прогноз логарифму кількості публікацій відносно закону про забезпечення функціонування української мови як державної, отриманий за допомогою використання моделі Сорнетте з врахуванням сезонності. На осі абсцис показано номери днів часового ряду, а на осі ординат – значення логарифму кількості публікацій.

Отримані значення оцінки точності показують можливість впровадження запропанованого підходу до прогнозування кількості новинних публікацій (та відповідно дієвості державного управління) за допомогою моделі Сорнетте в розробках аналітичних систем для області OSINT [8].

4. Висновки

В роботі розглянуто прогнозування дієвості державного управління на основі оцінки часових рядів публікацій відносно приймаємих законопроектів та законів. На даних моніторингу публікацій в період прийняття та дії законопроектів показано можливості прогнозування часових рядів публікацій за допомогою існуючих програмних засобів. Запропановано підхід з використанням моделі Сорнетте для прогнозування часових рядів публікацій та показано особливості використання моделі. Отримані результати показують можливість впровадження запропанованого підходу до прогнозування дієвості державного управління в розробках аналітичних систем для області OSINT.

1. Распознавание информационных операций / А.Г. Додонов, Д.В. Ландэ, В.В. Цыганок, О.В. Андрейчук, С.В. Каденко, А.Н. Грайворонская – К.: Инжиниринг, 2017. – 282 с.

2. Додонов А.Г., Ландэ Д.В., Путятин В.Г. Применение OSINT в аналитической деятельности // Реєстрація, зберігання і обробка даних. Щорічна підсумкова наукова конференція 17-18 травня 2018 року: збірник / - Київ: ІПРІ НАН України, 2018. - С. 110-112.

3. Jain, G., & Mallick, B. A study of time series models ARIMA and ETS. Available at SSRN 2898968. I.J.Modern Education and Computer Science, 2017, 4, 57-63.

4. Yenidoğan, I., Çayir, A., Kozan, O., Dağ, T., & Arslan, Ç. Bitcoin Forecasting Using ARIMA and PROPHET. In 2018 3rd International Conference on Computer Science and Engineering (UBMK) (pp. 621-624). IEEE.

5. Сорнетте, Д. Как предсказывать крахи финансовых рынков. Критические события в сложных финансовых системах. Litres , 2017. - 394 с.

6. Уренцов, О. В. Проверка возможности предсказания кризисов на финансовом рынке с помощью метода Д. Сорнетте. Труды Института системного анализа Российской академии наук, 2008, 40: 174-191.

7. Nielsen J. Log-Periodic Power Law Singularity (LPPLS) Model for Bubble Detection https://boulderinvestment.tech/blog/2018/log-periodic-power-law-lppl-model-for-bubble-detection.

8. Aleksandr Dodonov, Dmitry Lande, Boris Berezin. Semantic Models at Task Monitoring Public Opinions // CEUR Workshop Proceedings (ceur-ws.org). Vol-2318 [http://ceur-ws.org/Vol-2318/paper1.pdf]

СОДЕРЖАНИЕ

Национальная академия наук Украины
Институт проблем регистрации информации

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И БЕЗОПАСНОСТЬ

Материалы XIX Международной научно-практической конференции

Выпуск 19

Підп. до друку 7.12.2019. Ум. друк. арк. 14,65. Обл.-вид. арк. 25,66.
Наклад 100 пр. Зам. № 15-200.

ТОВ "Інжиніринг"
ISBN 978-966-2344-72-1